



Universidade de Brasília
Faculdade de Economia, Administração, Contabilidade e
Ciência da Informação
Departamento de Ciência da Informação e Documentação

**A Integração entre Sistemas Legislativos,
Terminologia e *Web* Semântica na organização e
representação da informação legislativa**

Ailton Luiz Gonçalves Feitosa

Brasília, janeiro/2005

Ailton Luiz Gonçalves Feitosa

**A Integração entre Sistemas Legislativos,
Terminologia e *Web* Semântica na organização e
representação da informação legislativa**

Tese desenvolvida no Programa de Pós-Graduação em
Ciência da Informação, do Departamento de Ciência da
Informação e Documentação, da Universidade de
Brasília, como requisito para a obtenção do grau de
doutor.

Orientadora: Profa. Dra. Haruka Nakayama

Brasília, janeiro/2005



FOLHA DE APROVAÇÃO

Título: A Integração entre Sistemas Legislativos, Terminologia e Web Semântica na organização e representação da informação legislativa.

Autor: Ailton Luiz Gonçalves Feitosa

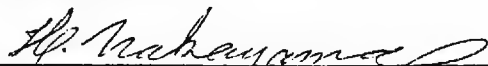
Área de concentração: Transferência da Informação

Linha de pesquisa: Arquitetura da Informação

Tese submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-graduação em Ciência da Informação do Departamento de Ciência da Informação e Documentação da Universidade de Brasília como requisito parcial para obtenção do título de **Doutor em Ciência da Informação**.

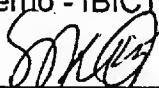
Tese aprovada em: 28/01/2005

Aprovado por:



Prof^a. Dr^a. Haruka Nakayama
Presidente - Orientador – (UnB/LET)

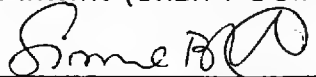
Prof. Dr. Hélio Kuramoto
Membro Externo - IBICT



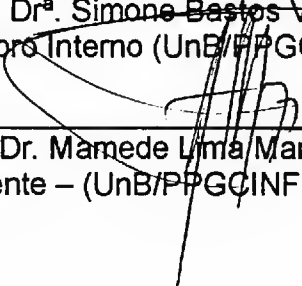
Prof^a. Dr^a. Maria Luisa Ortiz Alvarez
Membro Externo – (UnB/LET)



Prof. Dr. Murilo Bastos da Cunha
Membro Interno (UnB/PPGCINF)



Prof^a. Dr^a. Simone Bastos Vieira
Membro Interno (UnB/PPGCINF)



Prof. Dr. Mamede Lima Marques
Suplente – (UnB/PPGCINF)

A minha amada esposa Glaucione e a minhas amadas filhas Amanda e Lara que, em toda a sua grandeza espiritual, permitem-me partilhar do seu convívio nesta trajetória da minha existência e que, dia após dia, mostram-me que as maiores riquezas humanas encontram-se na simplicidade da vida em família.

A Natália, minha mãe, que desde minha tenra idade sempre incentivou meus estudos, e que em todos os momentos da minha vida agraciou-me com seu amor infinito.

A meus irmãos, que sempre partilharam comigo de muitos momentos importantes.

Agradecimentos

A Deus.

À minha orientadora, Profa. Dra. Haruka Nakayama, pela dedicação e seriedade com que me conduziu na realização deste trabalho e pelos ensinamentos durante todo o curso.

Ao mestre e amigo, Prof. Dr. Murilo Bastos da Cunha, pelas sempre pertinentes sugestões em todos os momentos da minha carreira acadêmica.

À Profa. Dra. Simone Bastos Vieira, por todos os ensinamentos, antes e durante a realização desta pesquisa.

Ao Prof. Dr. Helio Kuramoto, à Profa. Dra. Maria Luísa Ortiz Alvarez, ao Prof. Dr. Jaime Robredo, ao Prof. Dr. Mamede Lima-Marques, à Profa. Dra. Marisa Brasher Basílio Medeiros, à Profa. Dra. Maria Cristiane Barbosa Galvão e à Profa. Dra. Sely Maria de Souza Costa, que aceitaram participar da banca examinadora deste trabalho, contribuindo, dessa maneira, para que ele seja aprimorado.

Aos integrantes da Mesa Diretora da Câmara Legislativa do Distrito Federal e aos seus Assessores Especiais, por haverem acreditado que a realização deste trabalho traria contribuições para a democratização da informação legislativa.

À Profa. Dra. Suzana Mueller que, com suas orientações de sempre, ajudou-me a imprimir nova trajetória à minha vida acadêmica.

Ao Prof. Dr. Antônio Miranda, à Profa. Dra. Kira Tarapanoff e ao Prof. Dr. Emir Suaiden, pessoas que me brindaram com um pouco de sua sabedoria e que me introduziram nos caminhos e nos métodos de ensino e de pesquisa acadêmica em Ciência da Informação.

À Profa. Dra. Lígia Maria Café de Miranda, por todas as suas contribuições para que este trabalho se tornasse melhor do que o projeto de pesquisa inicial.

Às Professoras Dras. Hagar Espanha Gomes e Maria José Bocorny Finatto, pela importante colaboração e orientação virtual.

Ao amigo Eloi Yamaoka, companheiro das mais árduas discussões.

Ao amigo Cláudio José Heibel, por toda a ajuda com o levantamento de dados.

Aos amigos Roberto Miranda, Leonardo Lazarte e Ivette Kafure, pelas importantes trocas acadêmicas.

A Marcelo Dutra Vila Lima, Marisa Perrone Rocha e Wanderley Gonçalves de Freitas, técnicos envolvidos em atividades no Sistema de Normas Jurídicas do Distrito Federal (NJUR-DF).

A Fernando Ciciliati Júnior, analista responsável pelo Sistema de Apoio ao Processo Legislativo (SAPL), do Programa Interlegis.

A João Alberto de Oliveira Lima e Marcelo Nunes Gonçalves, técnicos envolvidos em atividades nos sistemas Normas Jurídicas (NJUR), Legislação Federal e Constituição Federal, do Senado Federal.

A Maria José Alves e Robson Marcelo Castelo Branco Barros, técnicos envolvidos em atividades no Sistema de Legislação Informatizada (LEGIN), da Câmara dos Deputados.

Aos amigos Elizabeth Fischer, Thomas Fischer e Cláudia Cnauvet, pela ajuda com os ~~trabalhos~~ estrangeiros.

A todos aqueles que embora não nominados aqui, contribuíram para que este trabalho fosse levado a termo.

Muito Obrigado.

“Com asa fechada ninguém voa.

Na falta de céu ninguém voa.

Vamos juntos essa lição aprender...”

(Ló Borges, com adaptações).

Resumo

A pesquisa investigou as interfaces teóricas e aplicadas das áreas de Terminologia e *Web Semântica*, com vistas à utilização em Sistemas de Informação Legislativa. O objetivo foi identificar fundamentos teóricos e metodológicos, bem como princípios e diretrizes referentes a essas áreas do conhecimento, cuja sistematização pudesse fornecer insumos para a elaboração de uma proposta metodológica de organização, representação e recuperação da informação legislativa. Para tanto, o estudo foi realizado nas seguintes etapas: a) levantamento dos fundamentos teóricos e metodológicos referentes a conceito, sistema de conceitos, controle terminológico, linguagem documentária e web semântica, tendo em vista suas aplicações em sistemas de informação legislativa; b) levantamento e análise das características de 29 Sistemas de Informação Legislativa brasileiros, compreendendo o Senado Federal, a Câmara dos Deputados, a Câmara Legislativa do Distrito Federal e as Assembléias Legislativas Estaduais; c) levantamento e análise das características dos sistemas NUJR, SAPL, NJUR-DF e LEGIN; d) elaboração de proposta metodológica para a organização, representação e recuperação de informações em sistemas legislativos. A proposta elaborada fundamenta-se: a) no losango semiótico, uma adaptação do modelo de triângulo semiótico de Ogden e Richards (1972), em face dos fundamentos teóricos e metodológicos da Terminologia e da Web Semântica; b) no Formato de Marcação de Terminologias (FORMATER), elaborado nesta pesquisa; c) no Formato de Marcação Legislativa (FORMALEGIS), elaborado nesta pesquisa. Conclui-se que a proposta metodológica apresentada é aplicável a Sistemas de Informação Legislativa e, feitas as adaptações necessárias, passível de extensão a outros sistemas de informação.

Palavras-chaves: informação legislativa, sistema legislativo, *web semântica*, terminologia, linguagem documentária, indexação, ontologia, XML, OWL, RDF.

Abstract

The present research investigated theoretical and applied interfaces in the areas of Terminology and Semantic Web aimed at using Legislative Information Systems. The purpose was to identify theoretical and methodological fundamentals, as well as principles and guidelines referring to these areas of knowledge, which systematization could provide inputs for the elaboration of a proposal that included a method of organization, representation and retrieval of the legislative information. In order to do so, the study was performed in the following steps: a) assessment of theoretical and methodological fundamentals regarding the concept, concept systems, terminologic control, documentary language and semantic web and their application in legislative information systems; b) assessment and analysis of the features of 29 Brazilian Legislative Information Systems, including the Senate, the House of Representatives, the Legislative Chamber of the Federal District and State Legislative Assemblies; c) assessment and analysis of the features of the following systems: NUJR, SAPL, NJUR-DF and LEGIN; d) elaboration of a methodological proposal for the organization, representation and retrieval of information in legislative systems. The proposal elaborated is based on: a) the semiotic losang, an adaptation from Ogden & Richards (1972) semiotic triangle model, due to theoretical and methodological fundamentals in the areas of Terminology and Semantic Web; b) the Terminology Markup Format (FORMATER), elaborated in this study; c) the Legislative Markup Format (FORMALEGIS), elaborated in this research. We conclude the methodological proposal presented is applicable to Legislative Information Systems and provided the needed adaptations, which can be extended to other information systems.

Keywords: legislative information, legislative system, semantic *web*, terminology, documentary language, indexing, ontology, XML, RDF, OWL.

Resumé

Dans ce travail de recherche, il s'agit de l'investigation des surfaces théoriques et appliquées dans les domaines de la Terminologie et du Web Sémantique comme but de l'utiliser en tant que système d'information législative. L'objectif fut d'identifier les fondements théoriques et méthodologiques tant que les principes et directives par rapport a ces domaines de connaissances. Leurs systématisation pourrait fournir des inputs pour élaborer la proposition qui inclut la méthode d'organisation, représentation et récupération de l'information législative. Pour y arriver, l'étude a été réalisée dans les étapes suivantes: a) rassemblement de fondement théoriques et méthodologiques par rapport aux concepts, système de concepts, contrôle terminologique, langage documentaire et web sémantique et leurs application dans les systèmes d'informations législative. b) le rassemblement et l'analyse des caractéristiques de 29 systèmes d'information législative brésiliennes, qui comprend le sénat fédéral, la chambre des députés la chambre législative du Distrito Federal et les assemblées législatives des états c) le rassemblement et l'analyse des caractéristiques des systèmes NUJR, SAPL, NJUR-DF et LEGIN. d) mise en place d'une proposition méthodologique pour l'organisation, la représentation et la récupération des informations dans les systèmes législatives. La proposition mise en place se base sur: a) le losange sémiotique, une adaptation du modèle triangulaire sémiotique de Ogden e Richards (1972) grace aux fondements théoriques et méthodologiques de la terminologie et du web sémantique. b) le format de marquage de terminologie (FORMATER) mise en place lors de cette recherche. c) le format de marquage législative (FORMALEGIS) mise en place lors de cette recherche et. Nous concluons que la proposition méthodologique présentée peut être mise en application aux systèmes d'information législative et, après avoir fait des adaptations nécessaires, s'étend aux autres systèmes d'information.

Mots clé: information législative, système législative, web sémantique, terminologie, langage documentaire, indexation, ontologie, XML, OWL, RDF.

Lista de ilustrações

ILUSTRAÇÃO 1: VOLUME DE NEGÓCIOS REALIZADOS NA <i>INTERNET</i> NO PERÍODO DE 1998-2003	23
ILUSTRAÇÃO 2: VISÃO GERAL DA METODOLOGIA UTILIZADA NA PESQUISA.....	36
ILUSTRAÇÃO 3: UNIVERSOS DA LÍNGUA GERAL E DA LÍNGUA DE ESPECIALIDADE.	55
ILUSTRAÇÃO 4: TERMINOLOGIA NO CONTEXTO DA LÍNGUA	56
ILUSTRAÇÃO 5: LÍNGUAS DE ESPECIALIDADE E LÍNGUA GERAL.....	62
ILUSTRAÇÃO 6: REPRESENTAÇÃO ESQUEMÁTICA DE UM SISTEMA DE CONCEITOS	65
ILUSTRAÇÃO 7: CLASSIFICAÇÃO DE COMPOSTOS QUÍMICOS EM GERAL E DE MINERAIS EM PARTICULAR	76
ILUSTRAÇÃO 8: DEFINIÇÃO DE COMPREENSÃO DE UM CONCEITO, COM BASE NAS SUAS CARACTERÍSTICAS	77
ILUSTRAÇÃO 9: COMPREENSÃO DE UM CONCEITO	77
ILUSTRAÇÃO 10: ESTRUTURA DE UMA DEFINIÇÃO SEGUNDO ARNTZ E PICT (1995)	87
ILUSTRAÇÃO 11: EXEMPLO DE UMA RELAÇÃO DIAGONAL	97
ILUSTRAÇÃO 12:RELAÇÕES CONCEITUAIS SEGUNDO DAHLBERG.....	99
ILUSTRAÇÃO 13: RELAÇÕES CONCEITUAIS SEGUNDO GORSKI	101
ILUSTRAÇÃO 14: RELAÇÕES CONCEITUAIS SEGUNDO CABRÉ (1993).....	103
ILUSTRAÇÃO 15: RELAÇÕES CONCEITUAIS SEGUNDO FELBER	106
ILUSTRAÇÃO 16: RELAÇÕES CONCEITUAIS SEGUNDO PICT & ARNTZ (1995).....	110
ILUSTRAÇÃO 17: RELAÇÕES CONCEITUAIS SEGUNDO SAGER (1990).....	113
ILUSTRAÇÃO 18: RELAÇÕES CONCEITUAIS, SEGUNDO A NORMA ISO 704.1.....	114
ILUSTRAÇÃO 19: RELAÇÕES CONCEITUAIS SEGUNDO A NORMA ISO 1087.....	116
ILUSTRAÇÃO 20: FORMATO TERMINOGRÁFICO SEGUNDO FAULSTICH	118
ILUSTRAÇÃO 21: FORMATO TERMINOGRÁFICO, SEGUNDO A NORMA NBR 13789.....	119
ILUSTRAÇÃO 22: FORMATO TERMINOGRÁFICO, SEGUNDO KRIEGER E FINATTO.....	122
ILUSTRAÇÃO 23: FORMATO TERMINOLÓGICO CONFORME DOCUMENTO DE REVISÃO DA NORMA ISO R 919	123
ILUSTRAÇÃO 24: FORMATO MARTIF DE INTERCÂMBIO DE TERMINOLOGIAS (SÁNCHEZ,1999).....	124
ILUSTRAÇÃO 25: FORMATO DE TERMINOLOGIAS ESTABELECIDO PELA NORMA ISO12620 (TTT.ORG, 2004).....	124
ILUSTRAÇÃO 26: COMBINAÇÃO DE DIAGRAMAS DE APRESENTAÇÃO DE TESAUROS	150
ILUSTRAÇÃO 27: RELAÇÕES CONCEITUAIS SEGUNDO AS NORMAS UNISIST E ISO 2788.....	151
ILUSTRAÇÃO 28: RELAÇÕES CONCEITUAIS, SEGUNDO AS DIRETRIZES DO IBICT (1984).....	156
ILUSTRAÇÃO 29: RELAÇÕES CONCEITUAIS SEGUNDO AITCHISON E GILCHRIST (1979, p. 50-55).....	157
ILUSTRAÇÃO 30: CLASSIFICAÇÃO DAS RELAÇÕES CONCEITUAIS SEGUNDO GOMES (1990A, p. 40-46).....	158
ILUSTRAÇÃO 31: RELAÇÕES CONCEITUAIS SEGUNDO LANCASTER (1987).....	160
ILUSTRAÇÃO 32: RELAÇÕES CONCEITUAIS SEGUNDO CINTRA (1994).....	162
ILUSTRAÇÃO 33: O CICLO DOCUMENTÁRIO - VISÃO SISTÊMICA.....	164
ILUSTRAÇÃO 34: ATIVIDADES DO CICLO DOCUMENTÁRIO	165
ILUSTRAÇÃO 35: UTILIZAÇÃO DE SERVIÇOS DE BUSCA NA <i>WEB</i> POR INTERNAUTAS AMERICANOS EM MAIO DE 2004.....	186
ILUSTRAÇÃO 36: PREDOMÍNIO DAS TECNOLOGIAS DE BUSCA NA <i>WEB</i> POR EMPRESAS EM MAIO DE 2004.	186
ILUSTRAÇÃO 37: RESULTADO DA BUSCA PELO TERMO LEGISLAÇÃO NO KARTOO.	189

ILUSTRAÇÃO 38: INICIATIVAS DE CONTROLE TERMINOLÓGICO NO SERVIÇO DE BUSCA SCIRUS.COM	190
ILUSTRAÇÃO 39 SINTAXE PARA A UTILIZAÇÃO DO OPERADOR SEMÂNTICO NO GOOGLE.....	192
ILUSTRAÇÃO 40: RESULTADOS DE UMA PESQUISA SEMÂNTICA NO GOOGLE.....	193
ILUSTRAÇÃO 41: <i>SITE</i> DO PADRÃO DUBLIN CORE NA INTERNET	200
ILUSTRAÇÃO 42: DISPOSIÇÃO DE <i>META TAGS</i> DUBLIN CORE EM UMA PÁGINA HTML.....	201
ILUSTRAÇÃO 43: PADRÕES DE META DADOS DA BANDA 3 DE DEMPSEY E HEERY (2000)	203
ILUSTRAÇÃO 44 - MODELO PROPOSTO PARA A <i>WORLD WIDE WEB</i> , SEGUNDO BERNERS-LEE.....	219
ILUSTRAÇÃO 45: GERAÇÕES TECNOLÓGICAS DA <i>WORLD WIDE WEB</i> – ADAPTADO DE HYVÖNEN (2002).	220
ILUSTRAÇÃO 46: O PROBLEMA DO SIGNIFICADO COM O USO DA CODIFICAÇÃO HTML	221
ILUSTRAÇÃO 47: O PROBLEMA DO SIGNIFICADO COM O USO DA CODIFICAÇÃO XML.....	221
ILUSTRAÇÃO 48: UTILIZAÇÃO DE ATRIBUTO COMO IDENTIFICADOR UNÍVOCO PARA UM REGISTRO.....	227
ILUSTRAÇÃO 49: UTILIZAÇÃO DE NAMESPACE PARA QUALIFICAR OS ELEMENTOS E ATRIBUTOS EM UM DOCUMENTO	228
ILUSTRAÇÃO 50: TRIPLA RDF (DACONTA, 2003, p. 88)	240
ILUSTRAÇÃO 51: ELEMENTOS DA LINGUAGEM OWL	251
ILUSTRAÇÃO 52: CONJUNTO DE ELEMENTOS DA LINGUAGEM OWL.....	255
ILUSTRAÇÃO 53: ESTRUTURA ORGANIZACIONAL DA CLDF.....	262
ILUSTRAÇÃO 54: FLUXOGRAMA DO PROCESSO LEGISLATIVO NA CLDF.	264
ILUSTRAÇÃO 55: LEVANTAMENTO DO USO DE SISTEMAS OPERACIONAIS PELAS 29 INSTITUIÇÕES LEGISLATIVAS	289
ILUSTRAÇÃO 56: LEVANTAMENTO DO USO DE APLICATIVOS DE SERVIÇOS PARA <i>WEB</i> PELAS 29 INSTITUIÇÕES LEGISLATIVAS	290
ILUSTRAÇÃO 57: MERCADO DOS SERVIDORES WEB MAIS POPULARES EM TODOS OS DOMÍNIOS AGO/95 – DEZ/04	290
ILUSTRAÇÃO 58: USO DE METADESCRITORES PARA A INDEXAÇÃO POR MECANISMOS DE BUSCA DA WEB.....	291
ILUSTRAÇÃO 59: DISPONIBILIDADE DE ACERVO LEGISLATIVO	294
ILUSTRAÇÃO 60: DISPONIBILIDADE DE MÓDULO DE PESQUISA	294
ILUSTRAÇÃO 61: USO DE OPERADORES DE BUSCA	295
ILUSTRAÇÃO 62: TIPOS DE CAMPOS DE BUSCA MAIS FREQUENTES	296
ILUSTRAÇÃO 63: TIPOS DE NORMAS MAIS FREQUENTES	297
ILUSTRAÇÃO 64: FORMATO DOS REGISTROS DOCUMENTÁRIOS – CAMPOS MAIS FREQUENTES.....	299
ILUSTRAÇÃO 65: DISPONIBILIDADE DE MÓDULO DE AJUDA AO USUÁRIO	300
ILUSTRAÇÃO 66: LINGUAGEM DE PROGRAMAÇÃO	302
ILUSTRAÇÃO 67: TELA INICIAL DO SISTEMA LEGISLAÇÃO FEDERAL.....	304
ILUSTRAÇÃO 68: FORMULÁRIO DE PESQUISA AVANÇADA.....	305
ILUSTRAÇÃO 69: FORMATO INTERMEDIÁRIO PARA CONSULTA A UM REGISTRO	305
ILUSTRAÇÃO 70: TELA DE PESQUISA BÁSICA DO SISTEMA SICON	307
ILUSTRAÇÃO 71: TELA DE PESQUISA AVANÇADA DO SICON.....	307
ILUSTRAÇÃO 72: ESTRUTURA DE PROJETO DE LEI ELABORADA PELA COMUNIDADE INTERLEGIS PARA SERIALIZAÇÃO COM XML.....	309
ILUSTRAÇÃO 73: TELA INICIAL DO SISTEMA SAPL	310
ILUSTRAÇÃO 74: FORMULÁRIOS DE CONSULTA A NORMAS JURÍDICAS.....	311

ILUSTRAÇÃO 75: FORMATO INTERMEDIÁRIO DE SAÍDA PARA A ESCOLHA DO REGISTRO PELO USUÁRIO.....	311
ILUSTRAÇÃO 76: FORMATO DETALHADO DE SAÍDA PARA O EXAME DO REGISTRO PELO USUÁRIO	312
ILUSTRAÇÃO 77: SISTEMA DE INFORMAÇÕES LEGISLATIVAS DA CLDF	313
ILUSTRAÇÃO 78: TELA DE CONSULTA TEXTUAL DO SISTEMA DE INFORMAÇÕES	314
ILUSTRAÇÃO 79: TELA DE CONSULTA TEXTUAL NO SISTEMA DE INFORMAÇÕES DA CLDF	315
ILUSTRAÇÃO 80: FORMATO DE SAÍDA DE UM REGISTRO	315
ILUSTRAÇÃO 81: DETALHAMENTO DA OCORRÊNCIA DE UM TERMO NO TEXTO PESQUISADO	316
ILUSTRAÇÃO 82: SISTEMA LOGIN - TELA DE CADASTRO DE NORMAS.....	318
ILUSTRAÇÃO 83: SISTEMA LOGIN – TELA DE PESQUISA ESTRUTURADA	319
ILUSTRAÇÃO 84: REPRESENTAÇÃO DOS CONCEITOS SEGUNDO OGDEN E RICHARDS (1972, P. 32).....	328
ILUSTRAÇÃO 85: LOSANGO SEMIÓTICO PARA REPRESENTAÇÃO DE CONCEITOS NA <i>WEB</i> SEMÂNTICA.....	329
ILUSTRAÇÃO 86: REPRESENTAÇÃO DA WEB, CONFORME BERNERS-LEE.....	331
ILUSTRAÇÃO 87: REPRESENTAÇÃO GRÁFICA DA WEB INTRODUZINDO-SE CONCEITOS, TERMOS, OBJETOS E LOCALIZAÇÕES.....	332
ILUSTRAÇÃO 88: DELIMITAÇÃO DO CAMPO SEMÂNTICO DE CONCEITOS EM UM SISTEMA PARA SUAS DEFINIÇÕES	333
ILUSTRAÇÃO 89: CONJUNTO DE ELEMENTOS DO FORMATER	336
ILUSTRAÇÃO 90: ELEMENTOS DA NORMA	347
ILUSTRAÇÃO 91: SUBELEMENTOS DO REGISTRO ELETRÔNICO	348
ILUSTRAÇÃO 92: SUBELEMENTOS DA PARTE PRELIMINAR.....	352
ILUSTRAÇÃO 93: SUBELEMENTOS DA PARTE NORMATIVA.....	353
ILUSTRAÇÃO 94: SUBELEMENTOS DA PARTE FINAL.....	354
ILUSTRAÇÃO 95: SUBELEMENTOS DE REMISSIVAS	354
ILUSTRAÇÃO 96: SUBELEMENTOS DO PROJETO.....	355
ILUSTRAÇÃO 97: SUBELEMENTOS DE DADOS DE PUBLICAÇÃO.....	355
ILUSTRAÇÃO 98: EXEMPLO DE BUSCA COM APOIO DE BANCO DE DADOS TERMINOLÓGICO	358
ILUSTRAÇÃO 99: SISTEMA DE ORGANIZAÇÃO E REPRESENTAÇÃO DE INFORMAÇÕES BASEADO EM BANCO TERMINOLÓGICO	359

Lista de tabelas

TABELA 1: QUANTIDADE DE USUÁRIOS CONECTADOS À <i>INTERNET</i> EM TODO O MUNDO	22
TABELA 2: POSIÇÃO DOS PAÍSES POR NÚMERO DE SERVIDORES DE <i>INTERNET</i>	23
TABELA 3: QUANTIDADE DE DOMÍNIOS REGISTRADOS NO BRASIL ATÉ AGOSTO DE 2003	24
TABELA 4: DISPONIBILIDADE DE LEIS NA <i>INTERNET</i> , POR CASAS LEGISLATIVAS BRASILEIRAS.....	26
TABELA 5: PERFIS DOS ENTREVISTADOS EM CADA UM DOS SIL	34
TABELA 6: PRINCÍPIOS ADMITIDOS SEGUNDO AS DIFERENTES “ESCOLAS DA TERMINOLOGIA”.....	53
TABELA 7: NÍVEIS DE ABSTRAÇÃO NO CONTEXTO DA LÍNGUA.....	63
TABELA 8: COMPARAÇÃO DE CONCEITOS PELA ABORDAGEM LÓGICO-FORMAL.....	97
TABELA 9: CLASSIFICAÇÃO DE CATEGORIA E SEUS CONSTITUINTES.....	98
TABELA 10: SIMBOLOGIA DE RELAÇÕES CONCEITUAIS SEGUNDO CABRÉ (1993).	102
TABELA 11: SIMBOLOGIA DA REPRESENTAÇÃO DE RELAÇÕES DE ACORDO COM FELBER.....	104
TABELA 12: ABREVIATURAS PARA CATEGORIAS GRAMÁTICAS	120
TABELA 13: FUNÇÕES DAS LINGUAGENS	127
TABELA 14: SEMELHANÇAS E DIFERENÇAS ENTRE AS LINGUAGENS NATURAIS E AS LINGUAGENS DOCUMENTÁRIAS	132
TABELA 15: SIMBOLOGIA DA REPRESENTAÇÃO DE RELAÇÕES DE ACORDO COM A NORMAS ISO 2788:1974, UNISIST (1975) E IBICT (1984).....	153
TABELA 16: TIPOS DE INDEXAÇÃO	170
TABELA 17: TIPOS DE ÍNDICES	174
TABELA 18: PRINCIPAIS META TAGS UTILIZADAS PARA A DEFINIÇÃO DE POLÍTICAS DE INDEXAÇÃO POR MECANISMOS DE BUSCA.....	194
TABELA 19: USO DO ELEMENTO LINK PARA OTIMIZAR A INDEXAÇÃO PELOS MECANISMOS DE BUSCA	195
TABELA 20: EXEMPLO DE REGISTRO DE META DADOS	198
TABELA 21: TIPOLOGIA DE FORMATOS DE META DADOS	202
TABELA 22: CARACTERÍSTICAS DOS FORMATOS DE META DADOS.....	203
TABELA 23: DEFINIÇÕES DE ONTOLOGIA.....	211
TABELA 24: ESTRUTURA DE UM DOCUMENTO ESCRITO EM XML	226
TABELA 25: REPRESENTAÇÃO DE DADOS POR MEIO DE ELEMENTOS FILHOS OU DE ATRIBUTOS	227
TABELA 26: USO DE NOMES IGUAIS EM ELEMENTOS QUE REPRESENTAM INFORMAÇÕES DIFERENTES	228
TABELA 27: ADIÇÃO DE PREFIXOS PARA DESAMBIGÜIZAR OS ELEMENTOS DO CÓDIGO XML.....	228
TABELA 28: DECLARAÇÕES DE ELEMENTOS EM UMA DTD	230
TABELA 29: DECLARAÇÃO DE ATRIBUTOS EM UMA DTD.....	230
TABELA 30: USO DE UM XMLSCHEMA PARA A ESTRUTURAÇÃO SINTÁTICA DE UM DOCUMENTO XML	232
TABELA 31: EXEMPLO DE ESTRUTURA DE UM XMLSCHEMA	233
TABELA 32: REFERÊNCIA A UM XMLSCHEMA EM UM ARQUIVO XML.....	233
TABELA 33: ELEMENTOS DEFINIDOS PARA UM XMLSCHEMA	234
TABELA 34: LINGUAGENS DA <i>WEB</i> SEMÂNTICA E SUAS FUNÇÕES	235
TABELA 35: ESTRUTURA DE UM ARQUIVO RDF.....	242

TABELA 36: INCLUSÃO DE NAMESPACES RELATIVOS A ELEMENTOS DUBLIN CORE EM UM ARQUIVO RDF	243
TABELA 37: EXEMPLO DE APLICAÇÃO DA ARQUITETURA RDF PARA UM ACERVO LEGISLATIVO	244
TABELA 38: CONTÊINER PARA DESCREVER RECURSOS COM VALORES ALTERNATIVOS	246
TABELA 39: CONTÊINER PARA DESCREVER RECURSOS COM VALORES NÃO ORDENADOS	246
TABELA 40: CONTÊINER PARA DESCREVER RECURSOS COM VALORES SEQUENCIAIS	246
TABELA 41: CLASSES RDF	247
TABELA 42: PROPRIEDADES RDF	247
TABELA 43: ATRIBUTOS RDF	248
TABELA 44: EXEMPLO DE APLICAÇÃO DE RDFSHEMA	248
TABELA 45: EXEMPLO DE APLICAÇÃO DE RDFSHEMA (SIMPLIFICADO)	249
TABELA 46: ESTRUTURA BÁSICA DE UM DOCUMENTO OWL (SINTAXE XML)	250
TABELA 47: NAMESPACES E PREFIXOS OWL	250
TABELA 48: ARGUMENTOS EM DEFESA DO UNICAMERALISMO OU DO BICAMERALISMO	258
TABELA 49: ESTRUTURA DOS PROJETOS DE ATOS LEGISLATIVOS	265
TABELA 50: SISTEMATIZAÇÃO DOS FUNDAMENTOS TEÓRICOS RELATIVOS A CONCEITO	267
TABELA 51: SISTEMATIZAÇÃO DOS FUNDAMENTOS TEÓRICOS RELATIVOS A CARACTERÍSTICAS	269
TABELA 52: SISTEMATIZAÇÃO DOS FUNDAMENTOS TEÓRICOS RELATIVOS A RELAÇÕES CONCEITUAIS (NÍVEL DE ANÁLISE E DETALHAMENTO 1)	274
TABELA 53: SISTEMATIZAÇÃO DOS FUNDAMENTOS TEÓRICOS SOBRE RELAÇÕES CONCEITUAIS (NÍVEL DE ANÁLISE E DETALHAMENTO 2)	276
TABELA 54: SISTEMATIZAÇÃO DOS FUNDAMENTOS TEÓRICOS RELATIVOS A RELAÇÕES CONCEITUAIS EM TESAuros (NÍVEL DE ANÁLISE E DETALHAMENTO 1)	280
TABELA 55: SISTEMATIZAÇÃO DOS FUNDAMENTOS TEÓRICOS RELATIVOS A RELAÇÕES CONCEITUAIS EM TESAuros (NÍVEL DE ANÁLISE E DETALHAMENTO 2)	282
TABELA 56: EVOLUÇÃO DO USO DE SERVIDORES WEB NOV/DEZ 2004	290
TABELA 57: FONTES DO DIREITO, SEGUNDO A CDDIR	297
TABELA 58: PROPOSTA DE REESTRUTURAÇÃO DA SUBCLASSE FONTES DO DIREITO DA CDDIR	298
TABELA 59: - REPRESENTAÇÃO DAS RELAÇÕES CONCEITUAIS NO VCBS E NO THES	303
TABELA 60: LINGUAGEM DE RECUPERAÇÃO NO SISTEMA LEGISLAÇÃO BRASILEIRA (QUE CONTÉM A BASE NJUR)	306
TABELA 61: DESCRIÇÃO DOS ELEMENTOS	309
TABELA 62: OPERADORES CONTEXTUAIS PREVISTOS PARA O NJUR-DF	317
TABELA 63: OPERADORES DE COMPARAÇÃO	317
TABELA 64: CARACTERÍSTICAS DE IDENTIFICAÇÃO DO ACERVO -	320
TABELA 65: CARACTERÍSTICAS DE IDENTIFICAÇÃO DO SISTEMA	321
TABELA 66: CARACTERÍSTICAS TECNOLÓGICAS DO SISTEMA	322
TABELA 67: TIPOS DE NORMAS LEGISLATIVAS DISPONÍVEIS NOS SISTEMAS ESTUDADOS	324
TABELA 68: FORMATO DO REGISTRO DOCUMENTÁRIO	325
TABELA 69: ELEMENTO COLEÇÃO	337
TABELA 70: ELEMENTO TÍTULO	337

TABELA 71: ELEMENTO TÍTULO EQUIVALENTE	337
TABELA 72: ELEMENTO AUTOR DA COLEÇÃO	338
TABELA 73: ELEMENTO PALAVRA-CHAVE	338
TABELA 74: ELEMENTO PALAVRA-CHAVE EQUIVALENTE	338
TABELA 75: ELEMENTO DESCRIÇÃO	338
TABELA 76: ELEMENTO DESCRIÇÃO EQUIVALENTE	339
TABELA 77: ELEMENTO INSTITUIÇÃO	339
TABELA 78: ELEMENTO DATA DE ATUALIZAÇÃO	339
TABELA 79: ELEMENTO FORMATO	339
TABELA 80: ELEMENTO ÍDIOMA	340
TABELA 81: ELEMENTO COPYRIGHT	340
TABELA 82: ELEMENTO CDU	340
TABELA 83: ELEMENTO CDD	341
TABELA 84: ELEMENTO REGISTRO	341
TABELA 85: ELEMENTO IDENTIFICADOR DO REGISTRO	341
TABELA 86: ELEMENTO DATA DO REGISTRO	342
TABELA 87: ELEMENTO AUTOR DO REGISTRO	342
TABELA 88: ELEMENTO NOTA	342
TABELA 89: ELEMENTO TERMO	342
TABELA 90: REPRESENTAÇÃO DE DADOS TERMINOLÓGICOS UTILIZANDO FORMATER	344
TABELA 91: ELEMENTO REGISTRO ELETRÔNICO	348
TABELA 92: ELEMENTO TÍTULO	349
TABELA 93: ELEMENTO NOME POPULAR	349
TABELA 94: ELEMENTO DESCRIÇÃO	349
TABELA 95: ELEMENTO PALAVRA-CHAVE	349
TABELA 96: ELEMENTO RESPONSÁVEL	350
TABELA 97: ELEMENTO INSTITUIÇÃO	350
TABELA 98: ELEMENTO DATA DO REGISTRO	350
TABELA 99: ELEMENTO FORMATO	350
TABELA 100: ELEMENTO COPYRIGHT	351
TABELA 101: ELEMENTO CDU	351
TABELA 102: ELEMENTO CDD	351
TABELA 103: ELEMENTO IDENTIFICADOR DO REGISTRO	351
TABELA 104: ELEMENTO OBSERVAÇÃO	352
TABELA 105: ELEMENTO PERSONALIDADE	352
TABELA 106: ESTRUTURA DE UMA COLEÇÃO DE NORMAS UTILIZANDO FORMALEGIS	356
TABELA 107: ELEMENTO ENTRADA	390
TABELA 108: ELEMENTO CATEGORIA GRAMATICAL	390
TABELA 109: ELEMENTO DEFINIÇÃO	390
TABELA 110: ELEMENTO ÍDIOMA	391

TABELA 111: ELEMENTO EQUIVALENTE.....	391
TABELA 112: ELEMENTO GÊNERO.....	391
TABELA 113: ELEMENTO VARIAÇÃO.....	391
TABELA 114: ELEMENTO FONTE.....	392
TABELA 115: ELEMENTO TERMO RELACIONADO.....	392
TABELA 116: ELEMENTO CONTEXTO.....	392
TABELA 117: ELEMENTO FONTE DA DEFINIÇÃO.....	392
TABELA 118: ELEMENTO ÁREA.....	393
TABELA 119: ELEMENTO URI.....	393
TABELA 120: ELEMENTO VARIANTE GRÁFICA.....	393
TABELA 121: ELEMENTO VARIANTE ORTOGRÁFICA.....	393
TABELA 122: ELEMENTO FORMA ABREVIADA.....	394
TABELA 123: ELEMENTO FORMA EXTENSA.....	394
TABELA 124: ELEMENTO REGISTRO TÉCNICO.....	394
TABELA 125: ELEMENTO REGISTRO VULGAR.....	394
TABELA 126: ELEMENTO REGISTRO GEOGRÁFICO.....	394
TABELA 127: ELEMENTO SINÔNIMO.....	395
TABELA 128: ELEMENTO ANTÔNIMO.....	395
TABELA 129: ELEMENTO HIPÔNIMO.....	395
TABELA 130: ELEMENTO HIPERÔNIMO.....	395
TABELA 131: ELEMENTO CONEXO.....	396
TABELA 132: ELEMENTO FONTE DO CONTEXTO.....	396
TABELA 133: ELEMENTO TEXTO DO CONTEXTO.....	396
TABELA 134: ELEMENTO AUTOR DA FONTE.....	396
TABELA 135: ELEMENTO TÍTULO DA FONTE.....	397
TABELA 136: ELEMENTO EDIÇÃO.....	397
TABELA 137: ELEMENTO LOCAL.....	397
TABELA 138: ELEMENTO EDITORA.....	397
TABELA 139: ELEMENTO DATA.....	398
TABELA 140: ELEMENTO PÁGINAS.....	398
TABELA 141: ELEMENTO PAGINAÇÃO.....	398
TABELA 142: ELEMENTO TÍTULO DO PERIÓDICO.....	398
TABELA 143: ELEMENTO VOLUME.....	399
TABELA 144: ELEMENTO FASCÍCULO.....	399
TABELA 145: ELEMENTO ISSN.....	399
TABELA 146: ELEMENTO ISBN.....	399
TABELA 147: ELEMENTO FÓRMULA.....	400
TABELA 148: ELEMENTO SÍMBOLO.....	400
TABELA 149: ELEMENTO SIGLA.....	400

Lista de abreviaturas e siglas

ABNT	Associação Brasileira de Normas Técnicas
ALCA	Área de Livre Comércio das Américas
BRASILTERM	Banco de Dados Terminológicos do Brasil
CDD	Classificação Decimal de Dewey
CDU	Classificação Decimal Universal
CE	Constituição Estadual
CEETT	Comissão de Estudo Especial Temporária de Terminologia
CI	Ciência da Informação
CITRAT	Centro Interdepartamental de Tradução e Terminologia
CLDF	Câmara Legislativa do Distrito Federal
CRG	<i>Classification Research Group</i>
DAML	<i>Darpa Markup Language</i>
DARPA	<i>Defense Advanced Research Projects Agency</i>
DC	Decreto
DF	Distrito Federal
DIN	Instituto Alemão para Normalização
DL	Decreto Legislativo
DTD	Definição de Tipo de Documento
EC	Emenda Constitucional
ELO	Emenda à Lei Orgânica
FGDC	<i>Federal Geographic Data Committee</i>
GDF	Governo do Distrito Federal
HTML	<i>Hypertext Markup Language</i>
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IDF	<i>Inverted Document Frequency</i>
INFOTERM	Centro Internacional de Documentação sobre Terminologia
ISO	<i>International Standardization Organization</i>
JSP	<i>Java ServerPages</i>
KIWC	<i>Keyword in Context</i>
LC	Lei Complementar
LD	Lei Delegada
LDAP	<i>Lightweight Directory Access Protocol</i>
LEXTERM	Centro de Estudos Lexicais e Terminológicos
LODF	Lei Orgânica do Distrito Federal
LOR	Lei Ordinária
MARC	<i>MAchine-Readable Cataloging</i>
MARTIF	<i>MAchine-Readable Interchange Format</i>
MERCOSUL	Mercado Comum do Sul
MIDI	<i>Musical Instruments Digital Interface</i>
NAFTA	Área de Livre Comércio da América do Norte
NISO	<i>National Information Standards Organization</i>
OIL	<i>Ontology Inference Layer</i>
OWL	<i>Web Ontology Language</i>
PDF	<i>Portable Document Format</i>
PDOT	Plano Diretor de Ordenamento Territorial
PT	Portaria
RDF	<i>Resource Description Framework</i>

RDFS *Resource Description Framework Schema*
RFC *Requests for Comments*
RI Regimento Interno
RINT Rede Internacional de Neologia e Terminologia
RITERM Rede Ibero-Americana de Terminologia
RS Resolução
SGML *Standardized Markup Language*
SIL Sistema de Informação Legislativa
TCDF Tribunal de Contas do Distrito Federal
TCT Teoria Comunicativa da Terminologia
TEI *Text Encoding Initiative*
TGT Teoria Geral da Terminologia
UE União Européia
URI *Universal Resource Identifier*
URL *Universal Resource Locator*
URN Nome Uniforme de Recurso
W3C *World Wide Web Consortium*
Web World Wide Web
WS Web Semântica
XHTML Linguagem de Marcação de Hipertexto Extensível
XML *Extensible Markup Language*
XML-S *Extensible Markup Language Schema*

Sumário

1	Introdução	21
1.1	Apresentação do tema	22
1.2	Definição do problema de pesquisa	30
1.3	Objetivos	31
1.3.1	Objetivo Geral	31
1.3.2	Objetivos Específicos	31
1.4	Metodologia	32
2	Revisão de Literatura	37
2.1	Terminologia	38
2.1.1	Histórico da Terminologia	39
2.1.2	Conceituação de Terminologia	46
2.1.3	Domínio da Terminologia	53
2.1.4	Campos de aplicação da Terminologia	57
2.1.5	Importância da Terminologia	60
2.1.6	Língua de Especialidade	61
2.1.7	Unidade terminológica	63
2.1.7.1	Conceito	66
2.1.7.1.1	Características	72
2.1.7.2	Termo	80
2.1.8	Definição	87
2.1.9	Sistema de conceitos e relações conceituais	94
2.1.9.1	Sistema de conceitos e relações conceituais segundo DAHLBERG	97
2.1.9.2	Sistema de conceitos e relações conceituais segundo GORSKI	99
2.1.9.3	Sistemas de Conceitos e Relações Conceituais segundo CABRÉ	101
2.1.9.3.1	Relações Lógicas	101
2.1.9.3.2	Relações Ontológicas	102
2.1.9.4	Sistemas de Conceitos e Relações Conceituais segundo FELBER	103
2.1.9.4.1	Relações Lógicas	104
2.1.9.4.2	Relações Ontológicas	105
2.1.9.5	Sistemas de Conceitos e Relações Conceituais segundo PICT & ARNTZ	106
2.1.9.5.1	Relações Hierárquicas	108
2.1.9.5.2	Relações não hierárquicas	109
2.1.9.6	Sistemas de Conceitos e Relações Conceituais segundo SAGER	111
2.1.9.6.1	Relações genéricas	111
2.1.9.6.2	Relações Partitivas	111
2.1.9.6.3	Relações Polivalentes	112
2.1.9.6.4	Relações Complexas	112
2.1.9.7	Relações Conceituais segundo a Normas ISO WD 704.1 e ISO 1087	113
2.1.10	Terminografia	116
2.2	Linguagens Documentárias	124
2.2.1	Conceituação	125
2.2.2	Controle terminológico	132
2.2.2.1	Tesauros	137
2.2.2.1.1	Macroestrutura, microestrutura e sistemas de remissivas	147
2.2.2.1.2	Relações entre descritores em Tesauros	150
2.3	Indexação e Recuperação da Informação	162
2.3.1	Documentos	162
2.3.2	Ciclo Documentário	163
2.3.3	Tratamento intelectual da informação	166
2.3.3.1	Indexação	167
2.3.3.2	Índices e Resumos	174
2.3.4	Recuperação da Informação	178
2.3.4.1	Modelos de Recuperação da Informação	180
2.3.5	Serviços de Indexação e Recuperação na <i>web</i>	181

2.3.5.1	Otimização de páginas web para mecanismos de busca	193
2.3.6	Padrões de meta dados na web	196
2.4	Web Semântica	204
2.4.1	Caracterização da web semântica	204
2.4.2	Ontologias	209
2.4.3	Infra-estrutura da Web Semântica	216
2.4.4	As gerações da web	220
2.4.4.1	A 1ª geração: HTML E XHTML	222
2.4.4.2	A 2ª geração: XML, DTD e XML Schema	223
2.4.4.2.1	XML	224
2.4.4.2.2	XML-Schema	225
2.4.4.3	A 3ª geração: padrões e linguagens de especificação de ontologias (RDF, RDF Schema, DAML+OIL, OWL)	234
2.4.4.3.1	RDF	238
2.4.4.3.2	RDFS	245
2.4.4.3.3	OWL	249
2.5	Informação Legislativa	255
2.5.1	O Poder Legislativo no Brasil	256
2.5.2	Poder Legislativo no Distrito Federal	259
2.5.2.1	Estrutura da Câmara Legislativa do DF	261
2.5.2.2	Processo Legislativo	261
2.6	Sistematização da Fundamentação Teórica e Metodológica	265
3	Levantamento e Análise dos Sistemas de Informação Legislativa (SIL)	287
3.1	Levantamento e análise dos SIL na Internet	288
3.1.1	Sistema operacional e aplicativo de serviços web	288
3.1.2	Tipo de indexação	291
3.1.3	Disponibilidade de acervo e de módulo de pesquisa	293
3.1.4	Linguagem de recuperação	295
3.1.5	Tipos de normas	296
3.1.6	Módulo de ajuda	300
3.1.7	Linguagem de Programação	301
3.2	Caracterização dos sistemas: NJUR, SAPL, NJUR-DF, LEGIN	302
3.2.1	Sistema de Normas Jurídicas do Senado Federal (NJUR)	302
3.2.2	Sistema de Apoio ao Processo Legislativo (SAPL)	308
3.2.3	Sistema de Normas Jurídicas do Distrito Federal (NJUR-DF)	312
3.2.4	Sistema Legislação Informatizada (LEGIN)	317
3.2.5	Comparação entre os sistemas	319
4	Proposta Metodológica de Organização e Representação de Informação Legislativa	327
4.1	Formato de Marcação de Terminologias (FORMATER)	333
4.1.1	Conjunto de elementos do FORMATER	335
4.1.1.1	Elementos identificadores da Coleção	337
4.1.1.2	Elementos identificadores do Registro Terminológico	341
4.1.1.2.1	Elementos identificadores do Termo	343
4.2	Formato de Marcação Legislativa (FORMALEGIS)	345
4.2.1	Conjunto de elementos do FORMALEGIS	346
5	Conclusões	360
5.1	Sugestões de pesquisa	366
	Bibliografia	368
	Apêndices	385
	Apêndice A – Sítios de Casas Legislativas do Brasil	386
	Apêndice B – Roteiro de entrevista semi-estruturada	387
	Apêndice C – Elementos de dados terminológicos do FORMATER	390
	Anexos	401
	Anexo A - Categorias de Dados Terminológicos ISO10620	402
	Anexo B – Perguntas de usuários do SRI/CLDF	405

Introdução

- 1.1 Apresentação do tema
- 1.2 Definição do problema de pesquisa
- 1.3 Objetivos
- 1.4 Metodologia

1.1 Apresentação do tema

O advento do serviço *World Wide Web* na *Internet* proporcionou grande crescimento na quantidade de informações, de usuários e de computadores ligados à rede. De acordo com dados das instituições de pesquisa *NUA Internet Surveys*¹ e *Nielsen NetRatings*², em todo o mundo, dados do ano de 2003 apontavam para cerca de 605,60 milhões de pessoas com acesso à *Internet* (tabela 1) e, no Brasil, em julho de 2003, cerca de 14,3 milhões. Segundo estimativas da NUA, somente nos Estados Unidos, até o final de 2003, foram realizados negócios entre empresas na rede, atingindo a soma de 1,3 trilhão de dólares (ilustração 1). Dados do Comitê Gestor da *Internet* no Brasil (2004), indicavam a existência, em janeiro de 2004, de mais de 223 milhões de servidores de *Internet* em todo o mundo, distribuídos entre trinta países. A tabela 2 apresenta as dez primeiras posições nessa classificação, destacando o Brasil em 8º lugar. Ao lado dessas estatísticas, o Serviço de Domínios para a *Internet* no Brasil – Registro.br (<http://registro.br>) contabilizava, em dezembro de 2003, mais de 509 mil³ domínios em todo o país (tabela 3). Em todo o mundo são quase 57 milhões de *sites* publicados na *web*. (SERVER WATCH, 2004).

Tabela 1: Quantidade de usuários conectados à *Internet* em todo o mundo

Region	Users
World Total	605.60 million
Africa	6.31 million
Asia/Pacific	187.24 million
Europe	190.91 million
Middle East	5.12 million
Canada & USA	182.67 million
Latin America	33.35 million

Fonte: NUA (2003, a)

A análise dos dados apresentados permite inferências sobre o enorme volume de informações disponíveis na *Internet*. Por certo, o crescimento da quantidade de informação eletrônica disponível por meio da *Internet* é uma realidade que pode ser percebida em todos os campos do conhecimento humano. O campo jurídico, um dos mais importantes para a sociedade, obviamente também tem sido bastante afetado por essa tecnologia – o diretório DireitoNet⁴, por exemplo, conta com um índice de mais de três mil *sites* jurídicos. De acordo com Paiva (2003), o avanço tecnológico tem penetrado decisivamente no universo jurídico não mais como uma

¹ http://www.nua.com/surveys/how_many_online/index.html

² http://www.nielsenratings.com/news.jsp?section=dat_to&country=br

³ Em dezembro de 2004, segundo dados da mesma instituição, esse valor aumentou para 705.463.

⁴ <http://www.direitonet.com.br>

perspectiva para o futuro, e sim como uma realidade imediata, trazendo uma espécie de revolução informática em aplicações que vão desde a simples consulta processual até a elaboração e o ajuizamento de petições iniciais. Blum (2001), outro jurista estudioso dos impactos da tecnologia sobre o campo jurídico, argumenta que, sendo as relações virtuais e seus efeitos uma realidade, há uma tendência à substituição gradativa do meio físico pelo virtual ou eletrônico, o que já ocorre e que já justifica a adequação, adaptação e interpretação das normas jurídicas em novo ambiente. O autor complementa seu pensamento, afirmando que é imprescindível o estudo, orientação e aplicação da *Internet* como ambiente de resultados legais sérios e com grande potencial de efeitos jurídicos.

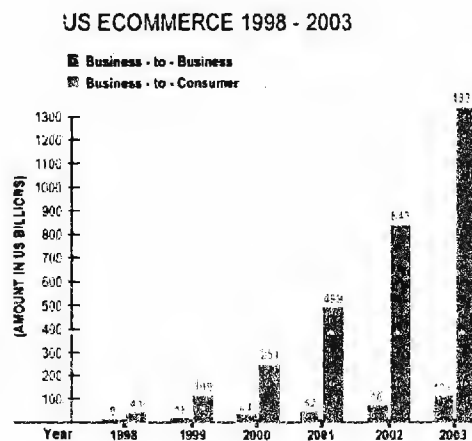


Ilustração 1: Volume de negócios realizados na *Internet* no período de 1998-2003.
 Fonte: NUA (2003, b).

Tabela 2: Posição dos países por número de servidores de *Internet*

Posição dos países por número de hosts		
	País	Janeiro/2004
1º	Estados Unidos*	162.195.368
2º	Japão (.jp)	12.962.065
3º	Itália (.it)	5.469.578
4º	Reino Unido (.uk)	3.715.752
5º	Alemanha (.de)	3.421.455
6º	Holanda (.nl)	3.419.182
7º	Canadá (.ca)	3.210.081
8º	Brasil (.br)	3.163.349
9º	Austrália (.au)	2.847.763
10º	Taiwan (.tw)	2.777.085

* (.edu, .us, .mil, .org, .gov e gTLDs)

Fonte: Comitê Gestor da Internet no Brasil (2004)

Tabela 3: Quantidade de domínios registrados no Brasil até agosto de 2003

Domínios Registrados por DPN - 22/08/2003 07:00:01					
DPN	QUANTIDADE	%	DPN	QUANTIDADE	%
Entidades			Profissionais Liberais		
AGR.BR	183	0.04	ADM.BR	665	0.13
AM.BR	50	0.01	ADV.BR	3129	0.61
ART.BR	1247	0.24	ARQ.BR	714	0.14
COM.BR	465.079	91.37	ATO.BR	60	0.01
COOP.BR	181	0.04	BIO.BR	111	0.02
ESP.BR	330	0.06	BMD.BR	2	0.00
ETC.BR	280	0.06	CIM.BR	226	0.04
FAR.BR	127	0.02	CNG.BR	8	0.00
FM.BR	104	0.02	CNT.BR	428	0.08
G12.BR	525	0.10	ECN.BR	87	0.02
GOV.BR	781	0.15	ENG.BR	1385	0.27
IMB.BR	291	0.06	ETI.BR	1301	0.26
IND.BR	3435	0.67	FND.BR	25	0.00
INF.BR	1509	0.30	FOT.BR	282	0.06
MIL.BR	14	0.00	FST.BR	49	0.01
NET.BR	368	0.07	GGF.BR	6	0.00
ORG.BR	13942	2.74	JOR.BR	276	0.05
PSI.BR	258	0.05	LEL.BR	87	0.02
REC.BR	78	0.02	MAT.BR	49	0.01
SRV.BR	1494	0.29	MED.BR	1416	0.28
TMP.BR	13	0.00	MUS.BR	383	0.08
TUR.BR	1292	0.25	NOT.BR	46	0.01
TV.BR	106	0.02	NTR.BR	28	0.01
491687	96.60		ODO.BR	498	0.10
			PPG.BR	428	0.08
Universidades			PRO.BR	1145	0.22
BR	1399	0.27	PSC.BR	246	0.05
EDU.BR	591	0.12	QSL.BR	29	0.01
1990	0.39		SLG.BR	12	0.00
Pessoas Físicas			TRD.BR	64	0.01
NOM.BR	1990	0.39	VET.BR	155	0.03
			ZLG.BR	3	0.00
			13343	2.62	
			Total	509.010	100.00

Fonte: Registro.br (2003)

Nesse sentido, com o objetivo de se acompanhar a evolução tecnológica – que por sua vez provoca uma mudança cultural na sociedade, fazendo com que o cidadão exija, cada vez mais, respostas exatas e velozes para os seus questionamentos –, Rover (2001) destaca algumas ações importantes quanto aos campos do direito e da tecnologia: a) empenho permanente dos juristas em implementar racionalidade ao sistema jurídico, de modo a tornar as informações acessíveis à sociedade, traduzíveis por sistemas de computação; b) democratizar e

popularizar o conhecimento das normas jurídicas, com o uso da tecnologia, de modo a permitir o acesso pelas grandes massas ao conhecimento jurídico; c) empenho de juristas e especialistas em tecnologia na simplificação do “mundo jurídico” por meio de sistemas inteligentes.

Como tem crescido o número de *sites* jurídicos, também tem aumentado a quantidade de instituições legislativas que já disponibilizam a matéria-prima do Poder Legislativo na *Internet*. Nessa linha de pensar, argumenta Borges (2000), a utilização de tecnologia de informação é um caminho natural para a implantação de qualquer processo modernizador nas instituições legislativas, na medida em que a matéria-prima e o produto gerado pelo legislador é informação, na forma de leis, pareceres, discursos. A autora ressalta ainda em seu pensamento a importância da informática para a área legislativa: o desenvolvimento tecnológico na área de informática parece trazer novas possibilidades de engajamento popular no processo político, diretamente e em larga escala. É fato que diversos setores começam a utilizar termos como *democracia eletrônica*, *teledemocracia* como alternativa ao regime representativo. Em se tratando de atitudes concretas do Estado, dados do Programa Interlegis⁵, vinculado ao Senado Federal, indicam que mais de mil casas legislativas no Brasil já têm presença na *Internet*. Até o final do ano de 2003 o Programa já estava presente em mais de 1.500 Câmaras Municipais, em todas as Assembléias Legislativas Estaduais, na Câmara Legislativa do Distrito Federal, no Tribunal de Contas da União (TCU), na Câmara dos Deputados e no Senado Federal. O número de parlamentares integrantes do Programa, incluindo senadores, deputados e vereadores, é de mais de 17 mil. O Programa prevê que, nos próximos dois anos, no total, serão atendidas 2.500 casas legislativas municipais (Interlegis, 2003b).

Como uma das estratégias para subsidiar as argumentações desta pesquisa, realizou-se levantamento na *Internet*, nos dias 30 e 31/08/03 e constatou-se que, somente três, em 22 casas legislativas brasileiras (incluindo Assembléias Legislativas Estaduais e a Câmara Legislativa do Distrito Federal), não possuíam ainda, no todo ou em parte, o seu acervo de leis na *Internet*. Foram analisados apenas 22 *sites*, pois os outros sistemas apresentaram problemas de conexão (ver tabela 4). Todas as demais dezenove Casas possuem versões eletrônicas da legislação já publicada em meio impresso ou de projetos que ainda estão em tramitação. É

⁵ O Interlegis é um programa desenvolvido pelo Congresso Nacional, em parceria com o Banco Interamericano de Desenvolvimento (BID), cujos objetivos são a modernização e a integração do Poder Legislativo nos níveis federal, estadual e municipal, bem como a promoção da maior transparência e interação desse Poder com a sociedade. O programa utiliza principalmente a *Internet* para realizar videoconferências, transmitir dados e transferir tecnologias, de modo a permitir a comunicação e a troca de experiências entre as Casas Legislativas e os legisladores, e entre o Poder Legislativo e a população. Todas essas iniciativas são realizadas para aumentar a participação da sociedade no processo legislativo. (INTERLEGIS, 2003a)

perceptível o movimento das instituições legislativas do país no sentido de democratizarem o acesso, por meio da *Internet*, às leis que produzem. Deve-se destacar, no entanto que as casas legislativas, entre os órgãos de governo, são aquelas que mais recebem recursos orçamentários e financeiros, além de possuírem um papel fundamental para a sociedade. Desse modo esperar-se-ia que houvesse maior investimento para a publicação de suas informações legislativas na web, como forma de democratização do acesso para a sociedade.

Tabela 4: Disponibilidade de leis na *Internet*, por casas legislativas brasileiras

Casa Legislativa no Estado	Acervo de Leis na <i>Internet</i>				
	Texto		Base de dados ou Mecanismo de busca		
	Integral	Estático	Formulário de pesquisa	Pesquisa textual	Busca Semântica
AC	Sim	Sim	Não	Não	Não
AL	Sim	Sim	Não	Não	Não
AP	Sim	Sim	Não	Não	Não
AM	Não avaliado*	Não avaliado*	Sim	Sim	Não avaliado*
BA	Sim	Sim	Não	Não	Não
CE	Não	Não	Não	Não	Não
DF	Sim	Sim	Sim	Sim	Não
ES	Não	Sim (ementas)	Sim	Sim	Não
GO	Sim	Sim	Não	Não	Não
MA	Não avaliado*	Não avaliado*	Não avaliado*	Não avaliado*	Não avaliado*
MT	Sim	Sim	Sim	Sim	Não
MS	Não	Sim (ementas)	Não	Não	Não
MG	Sim	Sim	Sim	Sim	Não
PA	Sim	Sim	Sim	Sim	Não
PB	Sim	Sim	Sim	Não avaliado*	Não avaliado*
PR	Não	Não	Não	Não	Não
PE	Sim	Sim	Sim	Sim	Não
PI	Sim	Sim	Não	Não	Não
RN	Não	Não	Não	Não	Não
RJ	Sim	Sim	Sim	Sim	Não
RO	Não avaliado*	Não avaliado*	Não avaliado*	Não avaliado*	Não avaliado*
RR	Sim	Sim	Sim	Sim	Não
RS	Sim	Sim	Sim	Sim	Não
SC	Sim	Sim	Sim	Sim	Não
SE	Sim	Sim	Sim	Sim	Não
SP	Sim	Sim	Sim	Sim	Não
TO	Não avaliado*	Não avaliado*	Não avaliado*	Não avaliado*	Não avaliado*

* Serviços não avaliados tendo em vista problemas de conexão com o servidor ou falhas nos sistemas

Vale lembrar, no entanto, que no ordenamento jurídico brasileiro há uma enorme produção de informações jurídicas e que a proliferação de atos normativos acaba provocando um

verdadeiro caos, vez que essa grande massa documental jurídica dificulta o trabalho de advogados, pesquisadores e dos próprios cidadãos (Passos, 2002). A autora argumenta que, desde 1969, Arnoldo Wald⁶ alertava para o problema, qualificando a produção de normas com o termo *inflação de diplomas* e destacando as dificuldades para que os especialistas da área encontrem, nessa diversidade de normas, as matérias necessárias ao apoio de suas decisões. Decorridos mais de trinta anos dessa afirmação, o termo permanece atual, como afirma Passos (2002). Aliado à proliferação de normas está o surgimento de uma grande quantidade de *sites* que pretendem disponibilizar normas jurídicas na *Internet* – o que aumenta ainda mais o volume de informações na rede –, cabendo-se enumerar os seguintes:

- SaraivaData [<http://www.saraivadata.com.br/>], *site* jurídico que possui livros, normas com texto integral e sistemas de remissivas;

- BuscaLaw [<http://www.buscalaw.com/>], pretende ser o maior portal jurídico da América Latina;

- Direito.com.br [<http://www.direito.com.br/>], oferece buscas em 121 tribunais do país, salas de *chat* e entrevistas virtuais com ministros;

De acordo com Hoeschl (2003), o desenvolvimento de sistemas inteligentes para a aplicação no universo jurídico é uma realidade. O autor cita vários estudos já publicados nessa área de pesquisa, entre os quais destaca-se o trabalho Bruninghaus e Ashley⁷, *Toward adding knowledge to learning algorithms for indexing legal cases*, no qual, os autores concluem que a integração ao sistema, de um vocabulário controlado e de informações lingüísticas, com um algoritmo de aprendizagem de máquina, pode ser importante aliada nos processos de recuperação da informação jurídica. Percebe-se, assim, a compreensão de especialistas na área de Inteligência Artificial, de que instrumentos originários da área de Ciência da Informação, como um vocabulário controlado, por exemplo, podem ser importantes ferramentas na construção de sistemas de informação mais inteligentes. Hoeschl prevê ainda o advento, “dentro de poucos anos”, de diversas soluções para as áreas de Tecnologia da Informação Jurídica, Governo Eletrônico e Sociedade da Informação, “que nos auxiliarão a resolver os problemas concretos da aplicação das leis e da organização social.” Como visto, a área jurídica, como todos os outros:

⁶ WALD, Arnoldo. A elaboração e revisão dos projetos de códigos. *Revista de Direito da Procuradoria Geral do Estado da Guanabara*, n. 21, p. 166-189, 1969.

⁷ BRUNINGHAUS, Stefanie; ASHLEY, Kevin D. *Toward adding knowledge to learning algorithms for indexing legal cases*. In: *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, p. 9-17, Oslo: Norway, June, 14-18, 1999. 220 p.

campos do conhecimento, tem acompanhado a tendência crescente de utilizar a *Internet* como meio de publicação.

Por outro lado, exatamente devido a esse volume de informações, tornam-se cada vez maiores os problemas de acesso e de recuperação de qualquer tipo de informação na rede. Alguns desses problemas podem ser caracterizados, como aponta Maedche (2000), por fatores como:

- crescimento acentuado das fontes de informação digitalizada;
- acesso, localização e resumo de informações tornam-se cada vez mais difíceis, tendo em vista a necessidade de se filtrar informações (a esses fatores contrapõe-se a indisponibilidade de ferramentas que possibilitem a filtragem);
- existência de um hiato entre o conceito de informação e as atuais formas de armazenamento – este último fator o autor defende, argumentando que as pessoas não podem compartilhar conhecimento via *web*, se ainda não é costume o uso em larga escala de formas padronizadas para o armazenamento de informações na *web*;

Baeza-Yattes e Ribeiro-Neto (1999) são outros estudiosos das disciplinas relacionadas com recuperação de informação, que apontam dificuldades na área, tais como:

- baixa efetividade ou elevado custo das técnicas de pré-processamento da linguagem natural, bem como de extração de texto semântico em grandes quantidades de dados – na maioria dos casos essas técnicas só são efetivas em textos bem estruturados com tesouros ou com outras informações contextuais;
- informações distribuídas em diferentes computadores, topologias, plataformas e larguras de banda diversas – esses fatores podem tornar o acesso bem mais complexo do que em ambientes de rede local, por exemplo;
- acentuada volatilidade dos dados – os autores estimavam, em 1999, que cerca de 40% da informação disponível na *Internet* já era modificada mensalmente, o que indica a ocorrência de baixos níveis de preservação da informação digital – preservação de originais;
- acentuada volatilidade de endereços e páginas – diversos *sites* e páginas mudam de endereço com elevada frequência;
- grande volume de informações – fator que torna cada vez mais complexa a tarefa de se recuperar informações na *web*;

- dados não estruturados e redundantes – muitos *sites*, além de não estarem estruturados rigorosamente segundo os padrões da *HyperText Markup Language* (HTML) – ou de outras linguagens de marcação –, não apresentam conteúdos originais; os autores estimavam que, naquela época, cerca de 30% das páginas da *web* já eram “duplicações”;

- baixa qualidade dos dados, seja por problemas de correção ortográfica, digitalização ou de correção conceitual.

Ao lado desses problemas, são destacadas, ainda, as dificuldades do usuário na recuperação da informação, seja pela falta de conhecimento ao elaborar questões que reflitam seus objetivos de busca, seja pela sua inabilidade para interpretar, classificar, priorizar ou filtrar as grandes quantidades de informação retornadas pelo sistema de busca, selecionando apenas aqueles documentos que lhe interessam. Quivev (2002), por exemplo, argumenta que, com o crescimento da *Internet* e das tecnologias de comunicação disponíveis, nunca foi tão fácil encontrar informação, como o é atualmente (2004). Tal facilidade é questionável, no entanto, já que freqüentemente as pessoas perdem horas examinando informações que podem lhes ser totalmente desnecessárias, excessivas ou redundantes. A autora lembra que a maioria dos mecanismos de busca trabalham com a localização de páginas para o texto específico digitado pelo usuário e simplesmente retornam milhares de fontes como ocorrências para o que foi pedido, sem nenhuma discriminação da informação; isso, obviamente, dificulta a análise do usuário que nunca pode ter certeza de que a informação exata que deseja está localizada entre as primeiras fontes retornadas. Assim, a *web* é, provavelmente, o mais rico repositório de informações na história humana, mas como a maioria dessa informação é desestruturada, não se pode saber o que ela contém e com que propósito, nem ao menos o que se pode esperar dela. A autora assinala a existência de alguns *sites* que armazenam informações de modo estruturado e que utilizam sistemas de busca, mas afirma serem apenas pequenas ilhas de ordem, no caótico mar de informações.

Do exposto nos parágrafos anteriores, é fácil depreender-se que a otimização das ferramentas para armazenamento, busca e recuperação de informações na *web* deve ser uma preocupação constante dos pesquisadores. Portanto, a falta de instrumentos que permitam a padronização de metodologias de estruturação de páginas e *sites* motivou diversas comunidades acadêmicas por todo o mundo ao estudo de alternativas de solução para esses problemas. Alguns exemplos desses projetos são: o *Dublin Core*, o TEI, a XML⁸ e, mais recentemente, a *Web*

⁸ *Dublin Core Metadata Initiative*. é uma organização dedicada a promover a ampla adoção de padrões de interoperabilidade de metadados e desenvolver vocabulários de metadados especializados para a descrição de

Semântica. *Web Semântica* e Ontologia de *Web* são denominações adotadas pelo *W3 Consortium*⁹, e que trazem iniciativas para a estruturação ou desenvolvimento de metodologias que viabilizem a organização semântica das informações disponíveis na *web* em todo o mundo.

O crescimento da informação compara-se a uma montanha que se torna cada vez mais alta e volumosa, o que torna necessária a busca de melhores métodos para significativamente se gerenciar a incrível quantidade de informação disponível. Portanto, na vanguarda dessa busca, está o conceito de *Web Semântica*, como um método de gravação de significados em páginas da *web*, unidas com um esquema para adicionar interpretação lógica aos documentos. A iniciativa busca oferecer instrumentos para se obter um sistema global, estruturado cuidadosamente, de modo a permitir que os usuários selecionem a informação desejada, em conformidade com as suas preferências individualizadas (QUIVEY, 2002).

1.2 Definição do problema de pesquisa

O já mencionado crescimento exponencial do número de publicações na *Internet*, bem como de usuários, tem fomentado uma considerável quantidade de estudos acerca do problema de recuperação da informação em mídia eletrônica. Diversos desses estudos abordam aspectos ergonômicos, funcionais ou sociológicos, na relação usuário-computador; outros se preocupam com aspectos técnicos como as linguagens e tecnologias de *software* e *hardware* empregadas; outros abordam metodologias utilizadas para o desenvolvimento e publicação de sistemas; outros, como o presente estudo, preocupam-se com aspectos da representação e da recuperação de informação na *web*.

Nesta última categoria de estudos, alguns se preocupam com o exame de algoritmos de busca; outros com a capacidade de cobertura de assuntos pelos sistemas de busca; outros com o comportamento dos usuários da *web* em relação aos sistemas de busca; outros comparam o desempenho de sistemas de busca utilizados. Este trabalho de pesquisa

recursos, de modo a habilitar a descoberta inteligente de informações na *web*. Mais informações estão disponíveis em [<http://www.dublincore.org>]. O *Text Encoding Initiative* (TEI) tem desenvolvido, com base na *Standard Generalized Markup Language* (SGML), uma codificação para um grande número de documentos em diversas extensões do conhecimento humano. A iniciativa TEI é um projeto de pesquisa internacional, patrocinado pela *Association for Computing in the Humanities* (ACH), pela *Association for Literary and Linguistic Computing* (ALLC), e pela *Association for Computational Linguistics* (ACL). Informações adicionais podem ser encontradas no site oficial do projeto: [<http://www.tei-c.org/>]. A *eXtensible Markup Language* (XML) é um formato de texto simples e muito flexível, derivado da SGML (ISO 8879) que tem desempenhado importante papel no intercâmbio de grande variedade de dados na *Internet* ou fora dela.

⁹ O *World Wide Web Consortium*, ou simplesmente W3C, foi criado em outubro de 1994 para liderar o desenvolvimento da *World Wide Web* em relação às suas tecnologias e interoperabilidade. O consórcio conta com cerca de quatrocentas organizações-membros, espalhadas por todo o mundo, e tem sido o referencial quanto às especificações normativas e metodológicas para praticamente todas as tecnologias envolvidas na *web*. Mais informações podem ser obtidas em [<http://www.w3.org/Consortium/>].

acompanhando as necessidades da sociedade nos campos jurídico e tecnológico, procurou trazer novas perspectivas no tocante ao tratamento e à publicação da informação legislativa na *Internet*, em benefício tanto das instituições legislativas, quanto da comunidade acadêmica, visto que aplica tecnologias e conceitos de *Web Semântica*, de Terminologia e de Indexação, para o processamento técnico de Informações Legislativas em ambiente *web*. O trabalho justifica-se, entre outras razões: a) pela sua relevância para a Ciência da Informação, uma vez que o núcleo temático aqui é o tratamento da informação; b) pela sua relevância para a sociedade, uma vez que procura resolver uma problemática de natureza social, a saber, a recuperação de informação legislativa na *web*; c) por serem oportunas as investigações acerca das contribuições da *Web Semântica* para a Ciência da Informação, visto que essas tecnologias são, atualmente (2004), ainda, muito recentes na área.

Em face dessa contextualização, formulou-se o seguinte problema de pesquisa:

É possível realizar controle conceitual na indexação e na recuperação de informações em Sistemas de Informação Legislativa - SIL, por meio da aplicação de instrumentos da Terminologia combinados com tecnologias da Web Semântica?

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo geral da pesquisa é:

Identificar como a Terminologia e a *Web Semântica* podem contribuir na indexação e na recuperação de informações em Sistemas de Informação Legislativa (SIL), no que se refere ao controle terminológico e conceitual, apresentando proposta metodológica de organização, representação e recuperação dessas informações.

1.3.2 Objetivos Específicos

Os objetivos específicos são:

– identificar e sistematizar os fundamentos teóricos e metodológicos, bem como os princípios e diretrizes referentes a conceito, a sistema de conceitos, a controle terminológico e conceitual, a linguagem documentária e a *web semântica*, no que se refere à sua aplicação na organização e na recuperação da informação em SIL;

– identificar e sistematizar as características técnicas dos SIL quanto a: a) linguagem de programação; b) formato do registro documentário; c) tipo de indexação; d)

linguagem de recuperação; e) módulo de pesquisa; f) tipos de normas; g) aplicativo de serviços *web*; h) sistema operacional;

– identificar e sistematizar as características de controle terminológico e conceitual na organização e na recuperação da informação nos SIL.

– elaborar uma proposta metodológica para a organização, a representação e a recuperação de normas legislativas que: a) utilize controle terminológico e conceitual na organização e na recuperação das informações; b) utilize tecnologias da *Web Semântica*.

1.4 Metodologia

Esta pesquisa, de natureza qualitativa, compreendeu três etapas principais: o estudo teórico, o levantamento de características de sistemas de informação legislativa e a elaboração de uma proposta de organização e representação de normas legislativas. O estudo teórico foi realizado com os objetivos de identificar e sistematizar a fundamentação teórica e metodológica referente a: língua de especialidade, conceito, sistema de conceitos, controle terminológico, linguagem documentária e web semântica, no que se refere às suas aplicações para a organização e recuperação de informações legislativas. O trabalho de levantamento bibliográfico para o estudo teórico foi realizado a partir da consulta a diversas fontes primárias, secundárias e terciárias, bem como de algumas consultas a especialistas em cada um dos assuntos pertinentes à temática desta pesquisa. Tais consultas foram realizadas por meio do correio eletrônico, com professores e especialistas brasileiros ou estrangeiros (apêndice A) ou, ainda, pelo envio de questões para fóruns de comunidades virtuais¹⁰. Os documentos levantados foram selecionados quanto à sua pertinência e relevância aos assuntos tratados nesta pesquisa. Os resultados do estudo teórico foram sistematizados em forma de diagramas implementados utilizando-se os aplicativos MindManager X5 e Microsoft Visio 2003¹¹.

O levantamento em Sistemas de Informação Legislativa foi realizado com o objetivo de identificar suas características quanto a aspectos técnicos de armazenamento, de processamento informatizado e de controle terminológico. A identificação de tais características foi realizada em duas etapas. A etapa 1 compreendeu o levantamento e a análise de informações

¹⁰ Foram consultadas as comunidades Letras do Brasil, Linguística – USP, Linguística – UFRGS e *Semantic Web*, no serviço de comunidades Orkut <www.orkut.com>, além dessas comunidades foi consultada a comunidade Websemântica no site Grupos.com.br <<http://www.grupos.com.br>>

¹¹ MindManager [<http://www.mindjet.com>] um aplicativo desenvolvido pela MindJet LLC e destina-se à construção de diagramas de brain storming para a modelagem do conhecimento; Visio 2003 [<http://office.microsoft.com/home/>] é um aplicativo destinado à construção de diagramas e modelos gráficos de naturezas diversas, desenvolvido pela Microsoft Corporation.

disponíveis nos *sites* institucionais das casas legislativas das esferas federal e estadual, bem como do Distrito Federal. A etapa 2 compreendeu o levantamento e a análise de informações coletadas por meio de entrevistas semi-estruturadas, realizadas com técnicos envolvidos em trabalhos nos sistemas legislativos cuja sede localiza-se no Distrito Federal.

Para viabilizar a etapa 1, foram utilizados vários mecanismos de busca na web, com a finalidade de identificar os endereços eletrônicos de diversos SIL (apêndice B). Esse levantamento abrangeu, exclusivamente, órgãos do Poder Legislativo, nas esferas de governo federal e estadual. Analisaram-se, assim, os SIL relativos aos seguintes órgãos: Senado Federal, Câmara dos Deputados, Câmara Legislativa do Distrito Federal, além das 26 Assembléias Legislativas Estaduais. Esta etapa de levantamento envolveu os seguintes procedimentos: a) visita ao sítio de cada uma das casas legislativas para levantamento de informações – foram considerados sistemas de informações legislativas, os serviços disponíveis em linha, por meio do sítio institucional, relativos a normas legislativas já publicadas; para efeito desta pesquisa, não foram considerados os acervos de proposições legislativas ainda não aprovadas, tendo em vista que os objetivos deste trabalho focalizam o levantamento de informações sobre normas legislativas vigentes; b) utilização do sistema de pesquisa e de identificação de servidores *web* disponível no sítio <<http://www.netcraft.com>>, para a avaliação dos sítios quanto às seguintes características: *aplicativo de serviços web* e *sistema operacional do servidor de web*; c) utilização do sistema SimSpider, disponível no sítio <http://www.searchengineworld.com/cgi-bin/sim_spider.cgi>, para o levantamento de informações relativas à organização do código HTML das páginas de entrada analisadas, com relação ao uso de *meta tags*¹² que podem auxiliar na indexação automática pelos mecanismos de busca na web, a saber: *meta keywords*, *meta description*, *title*, *author* e *robots*; d) interação com cada sistema legislativo, visando ao levantamento de características como: linguagem de programação; formato do registro documentário; tipo de indexação; linguagem de recuperação; módulo de pesquisa; tipos de normas.

Na etapa 2, como objeto de estudo, foram escolhidos os seguintes sistemas: o Sistema de Apoio ao Processo Legislativo (SAPL), do Programa Interlegis, que tem como objetivo promover a inclusão digital do maior número possível de casas legislativas no Brasil, em níveis estadual e municipal; o Sistema de Legislação Federal (NJUR), do Senado Federal, cujo objetivo é a disseminação de toda a legislação federal do país; o sistema Legislação Informatizada

¹² As *meta tags* são marcações introduzidas na linguagem HTML com o objetivo de fornecer metadados que descrevam o conteúdo de uma página da *web*, para facilitar o trabalho de indexação pelos mecanismos de busca. Como nos últimos anos houve grande incidência de spamming destinado à promoção de sítios nos mecanismos de busca, esses modificaram seus algoritmos, passando a indexar outras áreas da página.

LEGIN), da Câmara dos Deputados, cujo objetivo é a disseminação, pela *Internet*, de toda a legislação federal, desde o ano de 1946; e o Sistema de Normas Jurídicas do Distrito Federal (NJUR-DF), cujo objetivo é a disseminação de toda a legislação emanada da Câmara Legislativa do Distrito Federal. Os critérios de escolha desses sistemas foram, em primeiro lugar, o de importância dos sistemas, já que os sistemas LEGIN e NJUR contemplam todo o acervo legislativo brasileiro, em nível federal e que o NJUR-DF contempla todo o acervo emanado do Poder Legislativo local, cujas decisões influenciam diretamente a sociedade do Distrito Federal. Quanto ao SAPL, sua escolha deve-se ao fato de que possui importância nacional, uma vez que o propósito do Programa Interlegis é a sua disseminação por todo o país. O segundo critério de escolha dos sistemas proximidade geográfica, tendo em vista a realização da pesquisa com o menor custo possível.

Para cada sistema, foram entrevistados dois técnicos, sendo um analista de sistemas e outro especialista em informação legislativa. Quanto ao Sistema SAPL, apenas foi entrevistado o Analista de Sistemas responsável, uma vez que nenhuma instituição situada no Distrito Federal possui tal sistema em funcionamento.

Tabela 5: perfis dos entrevistados em cada um dos SIL

Sistema	Perfil dos entrevistados
NJUR – Senado Federal	um analista de sistemas e um bibliotecário
NJUR-DF – Câmara Legislativa do DF	um analista de sistemas e um bibliotecário
LEGIN – Câmara dos Deputados	um analista de sistemas e um advogado
SAPL – Programa Interlegis	um analista de sistemas

As entrevistas foram realizadas com o apoio de um formulário (apêndice C) contendo perguntas abertas e fechadas, cujo preenchimento foi realizado pelo próprio pesquisador. Previamente à aplicação, foram realizados dois pré-testes: um envolvendo uma bibliotecária e outro envolvendo um analista de sistemas, ambos no âmbito da Câmara Legislativa do Distrito Federal.

Para a realização do levantamento nos SIL foram consideradas as seguintes definições operacionais:

a) linguagem de programação – notação formal utilizada para descrever a execução de algoritmos em computador, composta por regras formais que especificam critérios de correção para o agrupamento de letras ou dígitos (sintaxe) e por regras formais que especificam critérios para o agrupamento de funções que executam operações (semântica) (SILVA; ASSIS, 1988). Para efeito desta pesquisa, foram consideradas as seguintes linguagens de

rogramação desenvolvidas por instituições, empresas ou grupos de especialistas: ASP, PHP, PERL, PYTHON, JSP e JAVA;

b) formato do registro documentário – conjunto de todos os elementos de informação que correspondem ao documento e que, descritos de forma precisa e detalhada, permitem a identificação do seu conteúdo ou de parte dele; (formulação baseada na definição de elementos essenciais da entrada, dada por Robredo e Cunha (1986, p. 102-103);

c) tipo de indexação – classificação do procedimento de indexação, quanto ao uso de computador (manual, automática ou semi-automática), bem como quanto ao uso de linguagem controlada ou não controlada (ROBREDO; CUNHA, 1986, p. 203);

d) linguagem de recuperação – caracterização do sistema de busca e recuperação com relação ao uso de operadores lógicos (e, ou, não) ou de operadores relacionais entre cadeias de caracteres (adjacência, proximidade, inclusão ou igualdade);

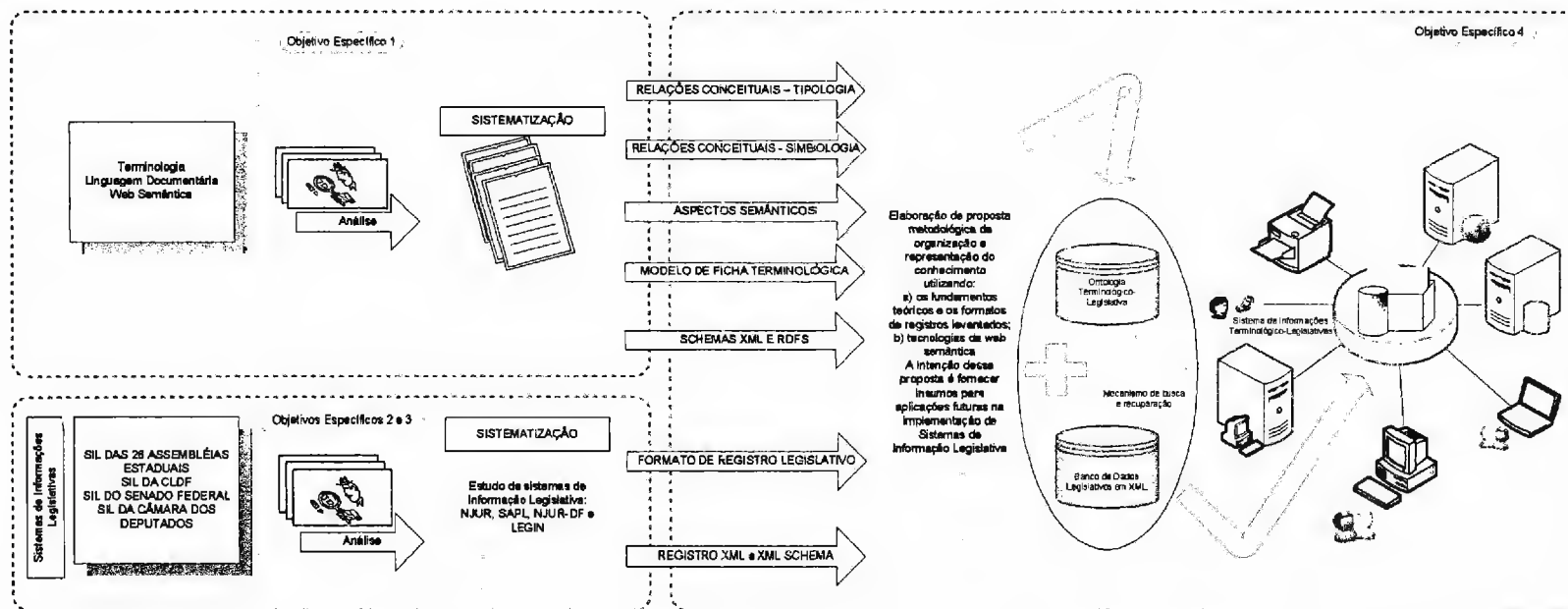
e) módulo de pesquisa – conjunto de formulários interativos que permitem ao usuário o envio, pela *Internet* ou *intranet*, de termos de consulta a um sistema informatizado;

f) tipo de norma – cada uma das diferentes espécies normativas emanadas do poder legislativo, de acordo com sua hierarquia e iniciativa;

g) aplicativo de serviços *web* – software que permite a interpretação de documentos de hipertexto, com a utilização do protocolo *Hyper Text Transfer Protocol* (HTTP), e cuja função é servir, via *web* ou *intranet*, a qualquer solicitação de arquivos que estejam hospedados no computador em que está instalado (KURNIAWAN, 2002; TANSLEY, 2002).

h) sistema operacional – *software* que informa ao computador como interagir com o usuário e como usar dispositivos como unidades de disco, monitor, teclado e outros periféricos (NORTON, 1996).

Após o estudo teórico e o levantamento nos SIL, os fundamentos teóricos e metodológicos levantados foram comparados com as características identificadas. Os resultados dessas análises culminaram na proposta metodológica de organização e representação de normas legislativas proposto no capítulo 5. A ilustração 2 apresenta uma síntese dos procedimentos adotados nesta pesquisa, bem como dos produtos obtidos em cada etapa de trabalho.



OE 1 Identificar e sistematizar os fundamentos teóricos e metodológicos referentes à teoria do conceito, aos sistemas de conceitos, ao controle terminológico e conceitual, às linguagens documentárias e à web semântica, no que se refere às suas aplicações na organização e na recuperação das informações em SIL

OE 2 Identificar e sistematizar as características técnicas dos SIL quanto a: a) linguagem de programação; b) formato do registro documental; c) tipo de indexação; d) linguagem de recuperação; e) módulo de pesquisa; f) tipos de normas; g) aplicativo de serviços web; h) sistema operacional

OE 3 Identificar e sistematizar as características de controle terminológico e conceitual na indexação e na recuperação da informação nos SIL

OE 4 Apresentar uma proposta de organização, representação e recuperação de informações legislativas

Revisão de Literatura

- 2.1 Terminologia
- 2.2 Linguagens documentárias
- 2.3 Indexação e Recuperação de Informação
- 2.4 Web Semântica
- 2.5 Informação legislativa
- 2.6 Sistematização da Fundamentação Teórica e Metodológica

O capítulo de revisão de literatura busca contextualizar o leitor quanto aos principais aspectos dos assuntos que permeiam o tema organização e representação da informação, trazendo um panorama sobre Terminologia, em seus aspectos históricos e fundamentos teóricos da disciplina; sobre Linguagem Documentária, sua tipologia e sua importância nos processos de tratamento da informação, sobretudo pelo uso de controle de vocabulário; sobre Recuperação da informação, quanto aos métodos e técnicas envolvidos nos processos de análise, armazenamento e recuperação de informação, enfatizando os procedimentos automatizados que se realizam na *web*; sobre *Web Semântica*, suas origens, infraestrutura tecnológica, objetivos e tecnologias envolvidas; sobre Informação Legislativa, abordando os contextos do Poder Legislativo.

Este capítulo apresenta também os resultados da compilação de estudos teóricos relativos a: Terminologia, Linguagem Documentária e *Web Semântica*, envolvendo os autores mais significativos para esta pesquisa, bem como normas ISO e ABNT e recomendações do W3C concernentes a tais assuntos. Ao final do capítulo, são apresentados os esquemas e diagramas que sistematizam as principais idéias dos autores, bem como os fundamentos teóricos e metodológicos, os princípios e as diretrizes definidos em normas pertinentes a cada um dos campos de investigação desta pesquisa, compreendendo conceito, sistema de conceitos, controle terminológico, linguagem documentária e *Web Semântica*.

2.1 Terminologia

A Terminologia é uma disciplina que estuda os conceitos e suas representações, abordando assuntos como natureza dos conceitos; relações conceituais; características dos conceitos; descrição dos conceitos; formação dos termos; padronização dos conceitos e termos; internacionalização dos conceitos. Esta seção trata dos aspectos teóricos da Terminologia, focando-se principalmente em língua de especialidade, unidade terminológica, conceito, denominação, definição, sistema de conceitos e relações conceituais. Foram estudados textos de autores mais representativos da área como: Helmut Felber, Pierre Auger, Maria Teresa Cabré, Juan Sager, Heribert Picht, Reiner Arntz e Alain Rey, entre outros. Além desses autores foram consultadas normas da ABNT ou ISO que tratam sobre Terminologia: ISO 1087, ISO WD 704.1, NBR 13789, NBR 13790 e NBR 13791.

2.1.1 Histórico da Terminologia

Embora a consolidação da Terminologia como disciplina seja relativamente recente, a prática terminológica remonta à idade média, período em que, como narra Campos (1992, p. 1), foram produzidas as primeiras obras relacionadas com a atividade terminológica. Nos séculos IX e X foram escritos glossários das ciências médicas com uma variação de quatro a dez idiomas. No século XII, Maimonides¹³ escreveu o Glossário de Nomes de Drogas, que consistia em 405 parágrafos curtos contendo nomes de drogas em árabe, grego, persa, sírio, berbere e espanhol. No século XVI Georgius Agricola, conhecido como o pai das ciências de mineração e metalurgia, desenvolveu uma obra considerada a primeira tentativa de se estabelecer uma terminologia internacional. O trabalho de Agricola teve grande importância pois o vocabulário técnico existente à época era muito confuso devido ao multilingüismo. Embora o trabalho tenha sido escrito em latim, uma língua que começava a desaparecer na época, foi a primeira tentativa de sistematização do vocabulário das ciências de mineração e metalurgia. O século XVIII foi marcado pelo surgimento de grande número de novos termos em todos os idiomas, devido aos avanços da Revolução Industrial. Notadamente nas áreas de ciências naturais e nas aplicações industriais como a mecânica, a metalurgia e a eletricidade, aos poucos começaram a surgir os glossários multilíngües. Nesse século destacam-se os trabalhos de Lavoisier e Berthollet – na química – e de Lineu – na botânica e na zoologia –, que originaram sistemas hierarquizados baseados em princípios que ainda hoje regem as nomenclaturas de suas respectivas disciplinas. Campos (1992, p. 5) relata ainda que o termo *terminologia* surgiu na Alemanha, para descrever a ciência dos termos técnicos e dos conjuntos de termos artificiais.

Como a Terminologia é uma disciplina jovem, somente a partir de meados do século XIX pode-se falar de um estudo de questões terminológicas de modo sistemático e em grande escala. Embora o estudo sistemático da Terminologia tenha origem nas ciências naturais,

¹³ Moses Maimonides nasceu em 1138, em Córdoba, Espanha e morreu em 1204, em Israel. Maimonides foi o mais importante médico judeu da Idade Média. Escreveu dez tratados médicos que mostram seu humanismo como médico e enfatizam a medicina preventiva. Nesses trabalhos pode-se ver que sua preocupação era tratar o paciente em duas dimensões: corpo e mente (ou psique e soma). Foi médico, filósofo, teólogo e astrônomo. Seus escritos médicos, como por exemplo as descrições sobre pneumonia, diabetes, asma e outras doenças, possuíam rigor metodológico comparável aos padrões da atualidade e demonstram que Maimonides possuía conhecimentos de idiomas como o grego e o persa. (ROSNER, Fred. *The Life of Moses Maimonides, a Prominent Medieval Physician*. In *Medical History*. Disponível em http://www.aecom.yu.edu/home/einsteinquarterly/PDFs/Volume_19/Number_3/19Rosner125.pdf. Acesso em 21/07/2004.)

isso não quer dizer que o trabalho sistemático terminológico seja aplicável somente a essas áreas. Pelo contrário, os princípios da Terminologia podem ser aplicados às mais diversas disciplinas, utilizando-se enfoques variáveis (ARNTZ; PICHT, 1995, p. 21). Como já exposto, a Terminologia foi utilizada primeiramente por cientistas ligados à química, à botânica e à zoologia (CABRÉ, 1993), mas no século XIX aparece a noção de língua técnica aplicável a diversos campos do saber (CAMPOS, 1992).

Inicialmente o crescimento da Terminologia foi motivado por questões que se limitaram à normalização de objetos, medidas e pesos; no entanto, logo os cientistas perceberam a necessidade de normalizar também os meios lingüísticos, com o objetivo de evitar problemas de comunicação. Datam do início do século XX obras como o Dicionário Técnico Ilustrado, do alemão Alfred Schlomann (1906 a 1910), escrito em alemão, inglês, francês, espanhol, italiano e russo, e abrangendo várias áreas da tecnologia e de suas aplicações industriais. Essa obra introduziu o método de ordenação sistemática, baseada no conceito. Nessa mesma época surgem, na União Soviética, diversos glossários bilíngües, tendo o russo como um dos idiomas (CAMPOS, 1992).

Em 1931, quando o engenheiro austríaco Eugene Wüster (1898 – 1977) apresentou a sua pesquisa de doutorado à Universidade de Viena, sua obra passou a ser vista como o marco inicial da escola de Terminologia Moderna (ARNTZ; PICHT, 1995). Wüster começa, a partir desse trabalho, a ser considerado o maior expoente da chamada escola de Viena. No entanto, de acordo com Rondeau (1984), foi o trabalho do russo D. S. Lotte que introduziu a Terminologia como disciplina científica. Para o autor, enquanto Lotte se preocupava com os aspectos teóricos e metodológicos, Wüster ocupava-se apenas do tratamento de dados terminológicos e somente no ano de 1979 sua teoria viria a se consolidar, quando seus escritos foram reunidos e publicados postumamente pelo seu discípulo Helmut Felber.

Com o progresso acelerado dos diversos ramos do conhecimento e com o rápido desenvolvimento da tecnologia, a Terminologia Moderna ganhou grande impulso, sendo preconizada principalmente no campo da engenharia. Assim, o fenômeno da língua técnica (língua de especialidade) não surgiu no seio da lingüística, mas das próprias necessidades lingüísticas das ciências e das técnicas, à medida que estas evoluíram (CAMPOS, 1992, p. 10). O maior desenvolvimento da Terminologia, tanto em sua vertente teórica como aplicada ocorreu, então, a partir dos anos 60 (CABRÉ, 1993), graças ao interesse de especialistas nos campos da ciência e da tecnologia e, principalmente, devido aos seguintes fatores:

- a) a partir do século XX, a sociedade experimenta um desenvolvimento científico sem precedentes, o que provoca o surgimento de grande quantidade de novos conceitos e de novos campos conceituais, que requerem novas denominações;
- b) o desenvolvimento tecnológico provoca o surgimento de novos campos de atuação econômica, bem como de novas formas de comunicação, e possui impacto direto sobre a área da informação, gerando a necessidade de renovação constante nas formas comunicativas, como as chamadas “indústrias da linguagem”;
- c) as relações internacionais, tanto políticas como culturais e econômicas, crescem de forma espetacular, passando-se de mercados restritos a mercados globalizados, caracterizados por empresas multinacionais. A produção dos bens de consumo exige um mercado cada vez mais extenso;
- d) a transferência de conhecimentos provoca o surgimento de mercados de intercâmbio científico, tecnológico, cultural e comercial; e, além disso, obriga a sociedade a planejar e desenvolver o multilingüismo;
- e) a explosão documental provoca o surgimento de grande massa de dados que demandam pela criação de sistemas de armazenamento e recuperação de informação que, por sua vez, acarretam a necessidade de padronização tanto dos sistemas automatizados de transferência, quanto dos conteúdos das grandes bases de dados;
- f) o desenvolvimento dos meios de comunicação de massa permite a difusão generalizada da Terminologia, com a conseguinte interação entre o léxico geral e o especializado; graças aos meios de comunicação, termos específicos da ciência e ou de áreas técnicas se popularizam;
- g) a intervenção governamental em matérias lingüísticas provoca a inclusão da Terminologia nos planos de normalização dos idiomas e a criação de organismos oficiais para geri-la;
- h) a concentração quase exclusiva da criação científica e tecnológica nos países de primeiro mundo provoca a transferência unidirecional de conhecimentos e novos produtos, o que acarreta a importação de grande quantidade de termos estrangeiros;

Embora, como já mencionado, diversos autores atribuam a Wüster a paternidade da Terminologia, ele próprio, em 1975, em sua palestra de abertura de um congresso do Centro Internacional de Documentação sobre Terminologia, mais conhecido pelo acrônimo INFOTERM, atribui a paternidade intelectual da teoria da Terminologia a quatro cientistas: o alemão Schloman, o primeiro a considerar o caráter sistemático dos termos de especialidade; o lingüista suíço Saussure, o primeiro a destacar a sistematicidade das línguas; o russo Drezen, um dos primeiros a destacarem a importância da normalização e pioneiro na fundação da *International Standardization Association* (ISA), órgão precedente da *International Organization for Standardization* (ISO); e o inglês Holmstrom, o primeiro a sugerir a criação de um organismo internacional a preocupar-se com a área de Terminologia; esta iniciativa culminou na criação, pela Unesco, do INFOTERM (CABRÉ, 1993, p. 27). Como mencionado, a Terminologia alcançou, então, o status de matéria autônoma e interdisciplinar com a obra póstuma *Teoria Geral da Terminologia*, de Wüster, publicada em 1979 pelo seu discípulo Helmut Felber (MARTINEZ, 2003).

Segundo Cabré (1993), em termos cronológicos, a evolução da Terminologia atual pode ser classificada em quatro períodos:

1930 a 1960 – origens da disciplina; nesta etapa surgem os primeiros textos teóricos, caracterizados pelo estabelecimento de métodos para o trabalho terminológico e preocupados com o caráter sistemático dos termos;

1960 a 1975 – estruturação; devido ao desenvolvimento da informática e das técnicas documentais, surgem os primeiros bancos de dados e inicia-se a organização internacional da Terminologia;

1975 a 1985 – eclosão; este período destaca a proliferação de projetos de planejamento lingüístico que incluem a Terminologia e destaca o papel que a Terminologia desempenha no processo de modernização de uma língua e da sociedade que a utiliza.

1985 a 1993 – ampliação; a informática converte-se em um dos elementos mais importantes para a mudança; por outro lado, os terminólogos dispõem de instrumentos e recursos de trabalho mais adaptados, sensíveis e mais eficazes no trato de suas necessidades; surge um novo mercado das indústrias da linguagem, no qual a Terminologia ocupa um lugar fundamental; consolida-se o modelo da Terminologia ligada ao planejamento da língua; consolida-se e amplia-se a cooperação internacional e criam-se redes internacionais.

Nesse último período foram muitas as iniciativas de consolidação da Terminologia como disciplina e como ciência. Em 1988 criou-se, no I Simpósio Ibero Americano de

Terminologia, realizado em Caracas, a Rede Ibero-Americana de Terminologia (RITERM)¹⁴, com o objetivo de se iniciar um trabalho de cooperação internacional na área de Terminologia; desde então essa rede tem sido responsável por grande parte do desenvolvimento da atividade terminológica, envolvendo os seguintes países: Argentina, Brasil, Chile, Colômbia, Cuba, Venezuela, México e Peru. Outro exemplo de projetos que congregam diversos países em torno dos objetos de estudo da Terminologia é a Rede Internacional de Neologia e Terminologia (RINT)¹⁵, uma organização intergovernamental francófone, orientada para o desenvolvimento terminológico e para a cooperação internacional em termos de organização lingüística. A RINT foi criada em 1986 e constitui um agrupamento de vinte países francófonos e tem como objetivos adaptar a língua francesa à expressão da modernidade científica e técnica; fornecer ferramentas de comunicação e expressão nos domínios das ciências e técnicas; estudar a criação terminológica nos países e comunidades de língua francesa; e favorecer, paralelamente ao desenvolvimento do idioma francês, o desenvolvimento das línguas nacionais nas regiões francófonas.

Para acrescentar uma atualização a esses quatro períodos referidos por Cabré (1993), ressalte-se que, a partir da década de 90, a popularização da *Internet* e a criação dos grandes blocos econômicos¹⁶ envolvendo diversos países – como a Área de Livre Comércio da América do Norte (Nafta), o Mercado Comum do Sul (Mercosul), a União Européia (UE) a Área de Livre Comércio das Américas (ALCA), entre outros –, deram significativo crescimento às atividades da Terminologia, especialmente as relacionadas com o desenvolvimento de instrumentos e obras de terminológicas como os dicionários eletrônicos disponíveis na *web* e as terminologias técnico-científicas, as terminologias jurídicas, entre outras obras que apóiam as comunicações entre as nações.

No Brasil, a Terminologia começou a desenvolver-se em meados da década de 80. Em 1986 foi criado o Grupo de Trabalho Lexicologia e Lexicografia da Associação Nacional de

¹⁴ A Rede Ibero-Americana de Terminologia é um organismo de intercâmbio e de trabalho na área da Terminologia e foi criada em 1988, com o objetivo de estabelecer um canal de cooperação entre seus membros, para consolidar as terminologias em espanhol e português, sugerindo atividades no âmbito terminológico, tais como: coleta e intercâmbio de informação terminológica; cooperação no planejamento e desenvolvimento dados terminológicos, lexicográficos e textuais; desenvolvimento da lingüística informática, informática aplicada à Terminologia; criação de um banco de dados terminológicos para a região ibero-americana, entre outras. Informações mais detalhadas estão disponíveis no *web site* da instituição: [<http://www.riterm.net/>]. Acesso em 31/07/2004.

¹⁵ Para maiores informações o leitor pode referir-se ao *web site* da organização, disponível em [<http://www.rint.org/>]. Acesso em 31/07/2004.

¹⁶ Para maiores informações sobre blocos econômicos, recomenda-se a leitura da seção Blocos Econômicos, no sítio <http://www.global21.com.br>. O Global21 é um portal brasileiro especializado em informações relativas a comércio exterior e relações econômicas internacionais e foi criado em maio de 2000, por uma equipe que envolve profissionais da *web*, jornalistas, especialistas em comércio exterior, especialistas em marketing e especialistas em relações internacionais. O portal publica diariamente notícias, entrevistas e informes setoriais e transmite newsletter eletrônica de segunda a sábado para cerca de 10.000 executivos (GLOBA21, 2004).

Pós-Graduação em Letras e Lingüística. Em 1988 realizou-se o III Encontro desse grupo de trabalho que, dando destaque à Terminologia como um campo de estudo, passou a denominar-se Lexicologia, Lexicografia e Terminologia. Em 1990 o IBICT sediou o II Simpósio Ibero Americano de Terminologia e o I Encontro Brasileiro de Terminologia Técnico Científica. Dois anos mais tarde resultou desse Simpósio uma comissão mista, composta pelo IBICT e pela ABNT, para a criação de uma Comissão de Estudo Especial Temporária de Terminologia (CEETT). Tal comissão reuniu especialistas ligados à Terminologia e apresentou os primeiros resultados de seus trabalhos nos anos seguintes, com as traduções e adaptações das normas ISO 639:1991, 10241:1992 e 860:1996. Seguindo as recomendações do Simpósio mencionado, em 1994 o IBICT iniciou a implantação do banco de dados terminológicos do Brasil, denominado *Brasilterm*. Tendo sido iniciado há dez anos, com os objetivos de difundir e de estimular o uso da Terminologia Científica e Técnica no território nacional, bem como de subsidiar a implementação do Banco de Dados Terminológico do Mercosul, até o momento o projeto não apresentou os seus primeiros resultados.

A partir dos anos oitenta, seguindo as tendências mundiais, diversas instituições brasileiras de ensino superior passaram a introduzir a Terminologia como disciplina em seus cursos de graduação e de pós-graduação (ALVES, 1998). Na Universidade de Brasília criou-se, no Departamento de Lingüística, Línguas Clássicas e Vernácula, do Instituto de Letras, a linha de pesquisa em Léxico e Terminologia. Tal linha de pesquisa tem como objetivo a formação de pessoal especializado em Terminologia e em Lexicografia Aplicadas. Com o impulso dessa linha de pesquisa, em 1990 foi criado o Centro de Estudos Lexicais e Terminológicos (Centro *Lexterm*), com os seguintes objetivos: desenvolver pesquisas em lexicologia, lexicografia e terminologia; realizar pesquisas sobre línguas neolatinas e lusófonas; difundir trabalhos de pesquisa; formar pesquisadores; e colaborar com outros centros, universidades e organizações não-universitárias (FAULSTICH, 1998a; 1998b)

Na Universidade de São Paulo, os estudos sobre terminologia iniciaram-se ainda em 1980, como parte de um curso de Especialização em Tradução, do Departamento de Letras Modernas. Em 1987 tais estudos ganharam feições de disciplina, com a denominação Terminologia Comparada. No mesmo período iniciaram-se as discussões que culminaram com a criação, em 1992, do Centro Interdepartamental de Tradução e Terminologia (CITRAT)¹⁷, que tinha como objetivos tornar-se um espaço colaborativo para a realização de projetos pertinentes à

¹⁷ Para maiores informações o leitor pode referir-se ao *web site* da organização, disponível em <http://www.fflch.usp.br/citrat/>. Acesso em 24/07/2004.

tradução e à terminologia. Desde então esse grupo vem desenvolvendo diversas atividades inerentes à produção de obras terminológicas, traduções, encontros de profissionais especializados na área, cursos de extensão e aperfeiçoamento, além de manter as atividades editoriais do periódico TradTerm¹⁸ e de outras obras como anais de congressos e manuais técnicos (AUBERT, 1998).

Projeto Termisul¹⁹ iniciou-se no Instituto de Letras da Universidade Federal do Rio Grande de Sul em 1991 e ganhou destaque com a constituição do Mercosul. Desde o seu início, Termisul atua no desenvolvimento de bancos de dados terminológicos, na formação acadêmica em terminologia e nas atividades de cooperação inter-institucional, voltada principalmente para a produção de instrumentos terminológicos. Além dessas linhas de atuação, são objetivos do projeto a pesquisa terminológica pura e aplicada, bem como a formação de tradutores (KRIEGER, 1998). Em termos de instituições universitárias, os principais centros de estudos terminológicos no Brasil ainda pertencem aos pólos de Brasília, São Paulo e Rio Grande do Sul, mas é possível observarem-se os esforços de outras instituições como a Universidade Federal de Pernambuco, a Universidade Estadual do Ceará, a Universidade Federal de Uberlândia, a Universidade Federal da Paraíba e a Universidade Estadual Paulista (ALVES, 1998).

A partir da década de 90 o Brasil iniciou sua participação ativa em diversas comissões e associações de cooperação internacional como a Subcomissão de Terminologia do Mercosul, a Rede Ibero-Americana de Terminologia (RITerm) e a Rede Panlatina de Terminologia (Realiter)²⁰ (ALVES, 1998). No âmbito normativo brasileiro, em 1997 a ABNT editou as normas 13.789, 13.790 e 13.791, com os respectivos objetivos de: a) estabelecer as regras para a elaboração e para a apresentação de normas de terminologia; b) fixar o enfoque metodológico para a harmonização dos conceitos, dos sistemas de conceitos, das definições, dos

¹⁸ Os sumários e resumos dos artigos publicados nesse periódico estão disponíveis em linha no *web site* da organização, na seção de publicações. [<http://www.fflch.usp.br/citrat/>]. Acesso em 24/07/2004.

¹⁹ Para maiores informações, o leitor deve referir-se ao *web site* da instituição, disponível em [<http://www.ufrgs.br/termisul/>]. Acesso em 24/07/2004.

²⁰ Rede Panlatina de Terminologia que engloba pessoas, instituições e organismos de países de línguas neolatinas que possuam projetos em Terminologia, com o objetivo de favorecer o desenvolvimento das línguas latinas (português, catalão, espanhol, francês, galego, italiano e romeno), por meio das seguintes ações: estabelecer princípios metodológicos comuns; promover pesquisas e criar instrumentos para favorecer o desenvolvimento das línguas latinas; acompanhar trabalhos terminológicos cooperativos e multilíngües; compartilhar materiais documentários de referência; favorecer a formação recíproca através do intercâmbio de formadores, especialistas e estudantes assim como de materiais didáticos. Informações mais detalhadas estão disponíveis no *web site* da instituição: [<http://www.realiter.net.br>].

termos e dos sistemas de termos em relação à elaboração de obras terminológicas; c) fixar códigos e simbologias para idiomas no âmbito da terminologia (ABNT, 1997a; 1997b; 1997c).

2.1.2 Conceituação de Terminologia

Cabré (1993) afirma que as autoridades em Terminologia dividem-se em dois pólos de pensamento principais: há aqueles que consideram-na como uma verdadeira disciplina científica, relacionada a outras diversas disciplinas, das quais toma emprestada uma série de fundamentos, sendo no entanto autônoma, pois tem sido capaz de efetuar uma reformulação e uma síntese original de princípios, chegando a constituir um campo original; outra corrente defende-a como técnica, uma prática ligada à simples resolução de problemas sociais e estreitamente vinculada a interesses de grupos políticos ou comerciais. Além desses pólos, deve-se ressaltar que existem conexões íntimas entre a Terminologia e outros campos do conhecimento que a precederam – como por exemplo a semântica, a lexicologia e a lexicografia (CAMPOS, 1992) –, que motivam um certo grau de indeterminação do seu objeto de pesquisa; basta ver que alguns autores consideram a Terminologia como parte da Linguística, outros a situam mais especificamente como uma prática corrente dentro da Linguística Aplicada, outros a definem como um conjunto de termos concernentes a uma especialidade.

Face à própria ambigüidade do termo Terminologia, Felber (1984) propõe três acepções para defini-lo:

- a) Ciência Terminológica – campo inter e transdisciplinar que estuda os conceitos e suas representações;
- b) terminologia – agregado de termos que representa o sistema de conceitos de um campo de assunto específico;
- c) terminologia – publicação na qual um sistema de conceitos de um campo de assunto específico é representado por termos;

Rondeau (1984, p. 18) também formula diferentes definições para o termo terminologia:

- a) conjunto dos termos de um domínio do conhecimento ou de uma disciplina;
- b) os métodos de coleta e de classificação dos termos, de criação neológica, de normalização dos termos, de difusão dos termos, utilizados por terminólogos e terminógrafos;

c) uma ciência cujo objeto é de ordem lingüística, mas que é essencialmente pluridisciplinar e participa ao mesmo tempo da lingüística, da lógica, da ontologia, da teoria da classificação e da informática.

Os pressupostos teóricos que definem a Terminologia variam de acordo com pelo menos quatro vertentes que ao longo dos anos acabaram sendo denominadas como as escolas de Viena, de Praga, de Moscou e de Québec.

A escola de Viena teve Wüster como o seu principal precursor e a sua principal vertente é a normativa. O trabalho de Wüster foi o mais significativo para essa escola e abordava assuntos como: natureza dos conceitos; relações conceituais; características dos conceitos; descrição dos conceitos; formação dos termos; padronização dos conceitos e termos; internacionalização dos conceitos. Outros autores significativos para essa escola são Ingetraut Dahlberg, Gernot Wersig, Helmut Felber, Christian Galinski, Pierre Budin, Heribert Picht e Reiner Arntz. Conforme relata Campos (1992), os princípios da escola de Viena são: o sistema de conceitos é fundamental em terminologia; o sistema de conceitos baseia-se na lógica e possui uma hierarquização estrutural na qual as relações dão-se no sentido vertical (extensão) ou no sentido horizontal (intensão); a correlação conceito-termo é tida como monorreferencial; o objetivo de uma definição é a delimitação do conceito; todo trabalho terminológico carece de documentação como um elemento indispensável; a padronização dos produtos terminológicos, realizada com o uso de métodos de terminografia, é uma parte essencial do trabalho da Terminologia.

Wüster considerava a Teoria Geral da Terminologia como um ramo distinto da ciência, caracterizada por ser uma área interdisciplinar entre a lógica, a lingüística, a ontologia as ciências da informação e as áreas de conhecimento específicas, que fornecem as línguas especializadas à língua geral. A TGT busca o estudo da natureza dos conceitos em relação a sua criação; características; relações com outros conceitos em um sistema de conceitos; descrição de conceitos; criação de termos e sua atribuição a conceitos (ENGEL, 1990). Para a TGT o domínio dos conceitos é considerado como sendo independente do domínio dos termos, isto é, os conceitos existem independentemente dos termos. Então, para a Terminologia, interessam apenas os termos relativos a conceitos, ou seja, as terminologias e seus significados, em um dado espaço temporal – o contexto atual, ou sincrônico –, não interessando as regras de flexão ou sintaxe.

A escola soviética, preconizada principalmente por Caplygin e Lotte contou com a colaboração de pesquisadores como Drezin, Vinokur, Kandelaki, Ahmanova, Danilenko, Kulebakin e Terpigorev. Os princípios defendidos por essa escola são: necessidade de uma ampla

pesquisa para a elaboração de uma teoria da terminologia; ênfase na concepção lingüística da terminologia; visão da terminologia como uma ciência aplicada, centrada na solução de problemas práticos como a padronização da língua russa e a criação de neologismos; construção do sistema de conceitos baseado na abordagem semasiológica, isto é, a partir dos termos de uma determinada área, busca-se estabelecer as definições dos conceitos (ênfase lingüística).

A escola de Praga, cujos principais expoentes são Trubetzkoy e Mathesius, iniciou seus trabalhos terminológicos nos anos trinta e recebeu grande influência do pensamento da lingüística funcional de Saussure. Os princípios dessa escola são: a terminologia insere-se num contexto de língua de especialidade; o termo é a menor unidade da linguagem funcional e é uma unidade léxica profissional, que insere-se em um estilo profissional funcional, inserido no contexto de uma língua nacional; o sistema terminológico compreende designações pertinentes a um sistema de conceitos; a padronização lingüística é vista sob o ponto de vista nacionalista; a padronização terminológica é vista sob o ponto de vista internacional.

A escola canadense iniciou seus trabalhos bem mais tarde, nos anos setenta. Seus trabalhos iniciais foram marcados pelo estudo de trabalhos relativos à terminologia sistemática e à lexicografia francesa, com o objetivo de se estabelecer princípios e métodos terminologia técnico-científica para o país sabidamente bilíngüe. Os expoentes da escola canadense são Rondeau, Dubuc, Boulanger e Auger. Os princípios dessa escola são: o conceito é fundamental para a terminologia; a unidade conceitual é que constitui o ponto de partida para a denominação; os sistemas hierarquizados de noções são considerados do ponto de vista onomasiológico; as relações entre conceito e denominação são regidas pela univocidade; o termo é essencialmente um signo lingüístico, que contém duas faces, conceito e denominação e que integra um sistema de conceitos pertinente a uma determinada língua de especialidade; a padronização terminológica é vista sob a perspectiva sociolingüística.

Resumindo as tendências, princípios e métodos das três primeiras escolas, Cabré (1993) afirma que os trabalhos que originaram as escolas de Viena, de Praga e a Soviética, influenciaram as seguintes correntes de pensamento:

- a) a escola de Viena considera a Terminologia como uma disciplina autônoma, de caráter interdisciplinar, a serviço das disciplinas técnico-científicas; esta orientação preocupa-se fundamentalmente com a natureza do conceito; as relações conceituais; a relação termo-conceito e a atribuição de termos aos conceitos;

- b) a escola de Praga considera a Terminologia como um subcomponente do léxico da língua, e as línguas de especialidade como subsistemas da língua geral;
- c) a escola de Moscou centra-se na filosofia e interessa-se fundamentalmente pela categorização lógica dos sistemas de conceitos e pela organização do conhecimento; preocupa-se prioritariamente com a relação que se estabelece entre o objeto da realidade e o conceito que ele representa;

Assim, a autora destaca que a palavra Terminologia pode designar pelo menos três conceitos, a saber: a) conjunto de princípios e de bases conceituais que regem o estudo dos termos – Terminologia vista como disciplina: é uma matéria de interseção que ocupa-se da designação dos conceitos das línguas e cujo objetivo primordial é a denominação dos conceitos; b) conjunto de diretrizes que utilizam a Terminologia – Terminologia vista como metodologia; c) conjunto de termos de uma determinada especialidade – Terminologia vista como instrumento (CABRÉ, 1993, p 82).

Segundo Martinez (2003), para todas as escolas tradicionais, o conceito é o ponto de partida da análise terminológica, seguindo, portanto, uma estratégia onomasiológica. A autora relembra, inclusive, que Wüster afirma que o conceito pode existir, independentemente da língua, e que a melhor forma de descrevê-lo é pela determinação de sua posição em um sistema de conceitos, por meio da formalização das relações lógicas e ontológicas. A Terminologia é, portanto, a representação de um sistema conceitual único e pré-existente. Os termos denominam o conceito, que é uniforme e que independe do momento histórico, da temporalidade, da sociedade, da ideologia; nessa concepção, as matérias ou a realidade se conceituam de maneira uniforme e universal.

Portanto, na visão da autora, o sistema conceitual é anterior e muito mais importante do que o sistema terminológico, de maneira que o “termo ideal” seja atribuído a determinado conceito em questão, do modo mais transparente e universal possível. O sistema conceitual é concebido independentemente do terminológico e, em consequência, os termos são independentes de contexto: o significado do termo é o conceito. Os termos, vistos dessa maneira, são meros rótulos dos diferentes elementos da realidade e, como tais, apenas importa a sua função denominativa, que de modo algum será alterada pelo contexto.

Os autores cuja visão é mais normativa, originária da Teoria Geral da Terminologia, admitem dois tipos de funções para a Terminologia vista como ciência: uma função descritiva e outra prescritiva. A função descritiva ocupa-se da documentação

terminográfica, registrando em dicionários os dados terminológicos de determinados conceitos (CAMPOS, 1992). A função prescritiva tem como objetivo a padronização, com vistas à eliminação de ambigüidades na comunicação. O controle terminológico é mais rigoroso e procura evitar a ocorrência de fenômenos como a polissemia e a sinonímia

Krieger (2000) relata que, a partir da década de 90, inicia-se uma profunda revisão crítica da Ciência Terminológica, em relação aos seus fundamentos teóricos e epistemológicos originais, justificada, em parte pela trajetória dos estudos das linguagens, em parte pela influência de novos paradigmas científicos, culturais e tecnológicos da atualidade. Para a autora, cujo pensamento é compartilhado por Martinez (2003), reconhecidamente o trabalho de Wüster culminou no que veio a ser conhecido como Teoria Geral da Terminologia (TGT), cujos fundamentos metodológicos são caracterizados pela sua natureza prescritiva e normativa, em detrimento da apreensão dos modos de funcionamento dos léxicos especializados. A autora cita Hoffmann, para esclarecer que a TGT foi um passo fundamental para o esclarecimento da essência das linguagens de especialidade. Ressalta, entretanto, que a teoria clássica não ampliou seu poder explicativo, consolidando-se somente aquelas orientações metodológicas necessárias à produção terminográfica. Isso, acrescenta, parece ter reduzido a atividade terminológica à escolha de conceitos e termos para a normalização dos termos (fixação de noções e denominações normatizadas), circunscrevendo os âmbitos da Terminologia ao universo da comunicação profissional e técnica Martinez (2003).

Não obstante, a partir da década de 90 a Terminologia começa a ser estudada como uma unidade da língua natural, incentivada pelo trabalho de pesquisadores de formação filosófica e voltada à tradução, bem como da área de inteligência artificial. Começa a consubstanciar-se, então, uma compreensão de que as Terminologias são “unidades lexicais”, e como tal, componentes naturais dos sistemas lingüísticos. Essa corrente de pensamento procura desfazer a idéia – equivocada, segundo Martinez (2003) – de que há uma nítida fronteira entre o léxico de especialidade e o léxico comum e de que a Terminologia é uma linguagem artificial.

Na mesma linha de argumentação, Krieger (2000) propõe uma reflexão de que a concepção clássica, impõe a idéia de que os termos técnico-científicos “são meras etiquetas designativas a serviço da consolidação e divulgação das ciências e das técnicas” em detrimento de sua significação lingüística que pode ser variável de acordo com o contexto discursivo e pragmático. E defende a Teoria Comunicativa da Terminologia (TCT) que, em contraposição à Teoria Geral da Terminologia, caracteriza-se pela sobredeterminação dos aspectos comunicativos das linguagens especializadas aos propósitos normativos da teoria wüsteriana.

O debate acerca da Terminologia situa-se, então, em dois pontos de vista: uma visão estática e normativa dos termos, expressão conceitual da escola vienense, que enfatiza princípios e métodos; outra lingüística, que entende o funcionamento das terminologias no contexto da sua pertinência aos sistemas lingüísticos. Essa dicotomia não representa a negação da contribuição da teoria clássica da Terminologia, cuja principal contribuição foi estabelecer os limites dessa ciência, mas pretende trazer à discussão uma nova teoria acerca da Terminologia.

Pode-se acrescentar a essa discussão o ponto de vista de Auger e Rousseau (1977), que classificam a Terminologia em três zonas de influência:

a) a orientação lingüístico-terminológica, representada sobretudo pelas escolas de Viena, Praga e Moscou, que busca atingir uma comunicação profissional precisa e eficaz, por meio da normalização conceitual e denominativa;

b) a orientação traducional, representada pelas organizações internacionais e pelos países plurilíngües, que buscam um ponto de referência para o trabalho de tradução, contribuindo para a qualidade do texto traduzido, com base na orientação para a fixação de equivalências terminológicas nas diferentes línguas;

c) a orientação do planejamento lingüístico preocupa-se com o planejamento atualizado e harmônico da língua, de modo a garantir a comunicação profissional; a Terminologia importada de países tecnologicamente dominantes é modificada, favorecendo a criação de termos na própria língua nativa.

Para Sager (1990), a Terminologia pode ser vista como um conjunto de práticas envolvidas na criação de termos, sua coleção e explicação, bem como sua representação em diversos meios eletrônicos ou impressos. Ressalta o autor, entretanto, que práticas bem estabelecidas não constituem uma disciplina, mas inegavelmente uma longa história de metodologias que requerem uma fundamentação teórica para justificar sua natureza distinta. Disciplinas estabelecem conhecimentos sobre coisas e, portanto, são justificadas pela sua própria natureza. Metodologias são apenas aplicáveis a um determinado fim; no caso da Terminologia, o como fazer coisas.

Para Barbosa (1995), enquanto objeto de estudo, uma terminologia é um conjunto de palavras (termos) técnicas ou científicas que constituem o vocabulário específico de uma ciência, de uma tecnologia, de um pesquisador ou grupo de pesquisadores, ou de uma área de conhecimento. Enquanto pesquisa fundamental, a Terminologia tem um objeto que contempla a produção de obras terminográficas, mas ultrapassa seus limites, porque envolve estudos como os

das relações de significações do signo terminológico. A obra terminológica visa à recuperação, compilação e armazenagem das denominações dos recortes científicos e tecnológicos. A Terminologia preocupa-se, assim, não com o significado de uma forma lingüística, mas com qual forma lingüística representa um conceito.

A tabela 6 resume e contrapõe os princípios admitidos por cada uma dessas correntes:

Tabela 6: princípios admitidos segundo as diferentes "escolas da terminologia:"

Escolas	autores	sistema de conceitos	relações conceituais
Viena	Dahlberg, Wersig, Felber, Galinski, Budin, Amtz e Picht	o sistema de conceitos é fundamental em terminologia; o sistema de conceitos baseia-se na lógica e possui uma hierarquização estrutural na qual as relações dão-se no sentido vertical (extensão) ou no sentido horizontal (intensão); a correlação conceito-termo é tida como monorreferencial; o objetivo de uma definição é a delimitação do conceito; todo trabalho terminológico carece de documentação como um elemento indispensável; a padronização dos produtos terminológicos, realizada com o uso de métodos de terminografia, é uma parte essencial do trabalho da Terminologia.	Terminologia como uma disciplina autônoma, de caráter interdisciplinar, a serviço das disciplinas técnico-científicas; esta orientação preocupa-se fundamentalmente com a natureza do conceito; as relações conceituais; a relação termo-conceito e a atribuição de termos aos conceitos;
Moscou	Caplygin e Lotte, Drezen, Vinokur, Kandelaki, Ahmanova, Danilenko, Kulebakin e Terpigorev	necessidade de uma ampla pesquisa para a elaboração de uma teoria da terminologia; ênfase na concepção lingüística da terminologia; visão da terminologia como uma ciência aplicada, centrada na solução de problemas práticos como a padronização da língua russa e a criação de neologismos; construção do sistema de conceitos baseado na abordagem semasiológica, isto é, a partir dos termos de uma determinada área, busca-se estabelecer as definições dos conceitos (ênfase lingüística)	centra-se na filosofia e interessa-se fundamentalmente pela categorização lógica dos sistemas de conceitos e pela organização do conhecimento; preocupa-se prioritariamente com a relação que se estabelece entre o objeto da realidade e o conceito que ele representa;
Praga	Trubetzkoy e Mathesius	a terminologia insere-se num contexto de língua de especialidade; o termo é a menor unidade da linguagem funcional e é uma unidade léxica profissional, que se insere em um estilo profissional-funcional, no contexto de uma língua nacional; o sistema terminológico compreende designações pertinentes a um sistema de conceitos; a padronização lingüística é vista sob o ponto de vista nacionalista; a padronização terminológica é vista sob o ponto de vista internacional	Terminologia como um subcomponente do léxico da língua, e as línguas de especialidade como subsistemas da língua geral;
Québec	Rondeau, Dubuc, Boulanger e Auger	o conceito é fundamental para a terminologia; a unidade conceitual é que constitui o ponto de partida para a denominação; os sistemas hierarquizados de noções são considerados do ponto de vista onomasiológico; o termo é essencialmente um signo lingüístico, que contém duas faces, conceito e denominação e que integra um sistema de conceitos pertinente a uma determinada língua de especialidade; a padronização terminológica é vista sob a perspectiva sociolingüística.	as relações entre conceito e denominação são regidas pela univocidade;

2.1.3 Domínio da Terminologia

De acordo com Coseriu (1979, p. 38), Saussure aponta pelo menos três definições para língua: a) a língua é uma realidade psíquica, compreendendo significados e imagens acústicas;

é um sistema gramatical que encontra sua existência virtual em cada mente ou, mais exatamente, nas mentes de um conjunto de indivíduos; estando presente na coletividade, a língua é um conjunto de hábitos lingüísticos que permitem a uma pessoa compreender e fazer-se compreender utilizando associações ratificadas pelo consenso coletivo e cujo conjunto constitui a língua; b) a língua é um produto ou instituição social; não está completa em nenhum indivíduo, mas apenas existe na coletividade; ao mesmo tempo que é um produto social da faculdade da linguagem, a língua é um conjunto das convenções necessárias adotadas pelo corpo social para permitir o exercício dessa faculdade nos indivíduos; c) a língua é um sistema de signos distintos correspondentes a idéias distintas; é um código, um sistema no qual de essencial existe apenas a união do sentido e da imagem acústica.

Analisando-se os três conceitos de Saussure, observa-se que correspondem, respectivamente, a três oposições: realidade psicofísica versus realidade psíquica; aspecto individual versus aspecto social; e conceito concreto ou abstrato versus realização ou sistema.

Jespersen (apud COSERIU, 1979, p. 19) advogando pela dimensão social da língua, argumenta que a língua é uma espécie de plural da fala, uma vez que é um fenômeno coletivo. O autor concebe a língua como um conceito comum que se extrai das linguagens individuais sendo, portanto, um conjunto de hábitos pelos quais os membros de uma coletividade podem comunicar-se. Pagliaro (apud COSERIU, 1979, p. 32) compartilha de tal pensamento e expõe o ponto de vista de que a língua pertence ao indivíduo e ao mesmo tempo à sua comunidade; a língua é uma objetivação concreta da atividade lingüística de uma comunidade, no tempo e no espaço; a língua é um sistema de elementos e relações que permitem as manifestações da fala e da criatividade individual.

De acordo com a norma DIN 2342 Parte I (1986:7), citada por Arntz e Picht (1995), o núcleo da língua do qual participam todos os membros de uma comunidade lingüística chama-se língua comum. Em um âmbito específico do conhecimento ou área profissional, a terminologia é o conjunto completo dos conceitos dessa área especializada e de suas definições, inserindo-se, assim, na linguagem especializada. Tal linguagem encontra-se na área da língua que aspira uma comunicação unívoca e livre de contradições em um domínio específico e cujo funcionamento encontra um suporte decisivo na terminologia estabelecida.

Os autores destacam também que a noção de linguagem especializada pode ser pensada considerando-se sempre dois pares opostos: léxico especializado x léxico não-especializado; compreensível para todos x incompreensível para todos. Assim, uma linguagem

especializada, embora possa perpassar o vocabulário comum, possui o seu grau de especialidade, determinado por fatores como: jargão, dialetismos e estrangeirismos (ARNTZ; PICHT, 1995).

De acordo com Picht (1984) a função principal da língua é a função comunicativa, que realiza-se nos limites de determinada comunidade. Cada comunidade, inserida num contexto social mais amplo, possui suas próprias características, o que motiva a ocorrência de diferenças na estrutura da linguagem, com relação a aspectos como fonética, léxico e semântica. Assim, as características, valores e costumes de certa comunidade, estratificada por camada social, idade, profissão ou gueto, determinam a especialização do modo como seus integrantes se comunicam e, conseqüentemente, da língua por eles utilizada. A essa língua, particularizada pelas características da comunidade em que se insere, o autor denomina *língua de especialidade*. Apesar de possuírem características próprias, intrínsecas à comunidade em que se originaram, as línguas de especialidade possuem algumas características comuns (PICHT, 1984, p. 161): a) possuem caráter monofuncional, uma vez que ocorrem em um limite social determinado e prestam-se apenas à função comunicativa; b) referem-se a um número restrito de usuários – apenas os integrantes de certa comunidade; c) não são imprescindíveis para a existência da sociedade em que se insere aquela comunidade lingüística.

O esquema apresentado na ilustração 3 mostra o contexto das línguas de especialidade em relação à língua geral:

Língua		
Língua Geral	Línguas de Especialidade	
	Profissional <i>(uso obrigatório apenas para a comunicação profissional)</i>	Com sentido restrito <i>(uso opcional, inclusive no grupo, embora possa determinar a aceitação por ele)</i>

Ilustração 3: Universos da Língua Geral e da Língua de Especialidade.
Fonte: adaptado de Picht (1984, p.162)

A análise da figura permite concluir que a existência das línguas de especialidade é motivada pela existência de uma comunidade lingüística. O autor restringe as línguas de especialidade àquelas de uso profissional, associadas a determinada carreira profissional; e àquelas com sentido restrito, associadas a determinado grupo ou gueto, o qual se diferencia, no contexto social mais amplo, pela adoção de usos ou costumes peculiares – por exemplo: grupos de adolescentes, grupos culturais, entre outros.

Assim, a língua profissional, que é o meio de comunicação das pessoas envolvidas em determinada carreira, geralmente é aprendida ao longo daquela profissão e é obrigatória para os membros da comunidade. A aprendizagem de uma língua com sentido restrito é facultativa; no entanto, a aceitação do indivíduo pelo grupo pode depender dessa aprendizagem.

São características das línguas de especialidade profissionais:

- a) possuem relação íntima com o grupo correspondente;
- b) sua existência depende do grupo;
- c) são obrigatórias para profissionais de uma área e para os mediadores de textos profissionais: tradutores, jornalistas especializados;
- d) têm função de comunicação;
- e) manifestam-se em formas estilísticas determinadas;
- f) distinguem-se da língua geral nos níveis morfológico, sintático e lexicológico;
- g) são línguas normalizadas;
- h) motivam o planejamento lingüístico;
- i) possuem diferenças da língua geral, com relação ao aspecto de ensino-aprendizagem.

Considerando as particularidades das línguas de especialidade, parece claro que a Terminologia tem papel importante nesse contexto. Barbosa (1995) pronuncia-se sobre o assunto destacando que o percurso da investigação científica da Terminologia começa no “recorte técnico-científico”, para chegar à denominação (termo); isto é, a Terminologia adota uma metodologia de dar nomes aos conceitos pertinentes a uma área específica do conhecimento. A ilustração 4 contextualiza a Terminologia em relação à língua geral e às línguas de especialidade:

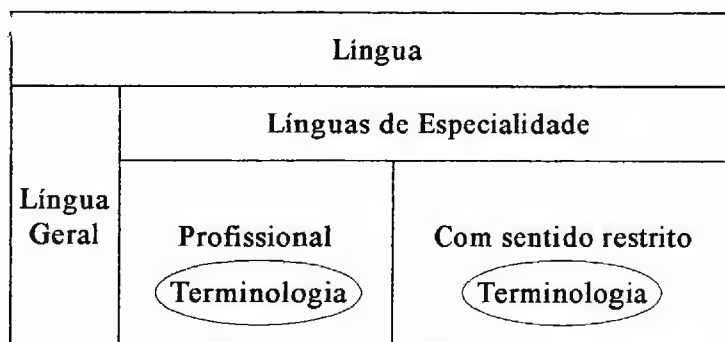


Ilustração 4: Terminologia no contexto da língua.

A ilustração mostra que a língua é o contexto mais amplo, dentro do qual distinguem-se a língua geral e as línguas de especialidade; nestas está o campo de trabalho da Terminologia, uma vez que essas línguas de especialidade são formadas por um conjunto específico de termos.

Nesse contexto, é perceptível que o vocabulário da língua geral acaba sendo mesclado por diversos elementos das diversas línguas de especialidade e que, como visto, as línguas de especialidade, em certa medida, derivam da língua comum, tornando-se sublínguas. Há níveis de linguagem tão especializados (como as fórmulas e equações matemáticas), que tornam-se praticamente incompreensíveis para um leitor comum. Assim a língua de especialidade depende da língua comum e a linguagem formalizada depende da linguagem natural especializada (ARNTZ; PICHT, 1995).

2.1.4 Campos de aplicação da Terminologia

Arntz e Picht (1995) argumentam que a Terminologia, em sua essência é uma ciência transdisciplinar, que tem interfaces com diversas áreas do conhecimento: o trabalho terminológico consiste na coleção e elaboração, tratamento e sistematização das terminologias, e realiza-se em âmbitos e com objetivos que podem divergir fortemente, segundo os requisitos de cada campo de atividade. Entre essas áreas podem ser relacionadas a lingüística, a filosofia, a informática, a ciência da informação, o direito, a engenharia do conhecimento, entre outras. Para os autores, como uma ciência de léxicos especializados, a Terminologia guarda estreita relação com a lingüística, cujo objeto de estudos é o léxico geral. No entanto essa ligação torna-se mais evidente nos campos da lingüística aplicada, que trata com línguas especializadas.

A Terminologia possui também estreita relação com a filosofia, uma vez que ambas as ciências possuem como objeto de estudo e investigação a essência dos conceitos, destacando diversas formas para a sua representação. O conceito, visto como unidade central do pensamento, é investigado sob a ótica da lógica, com preocupação especial nos campos de formação de conceitos, fusão de conceitos e de sua interrelação em sistemas de conceitos.

Em relação à Ciência da Informação (CI), a Terminologia guarda relações estreitas, posto que a CI tem preocupações com a elaboração, organização, armazenamento e recuperação do conhecimento; assim, para que possa desempenhar tal papel, a CI necessita de metodologias de organização e de representação do conhecimento. Os princípios terminológicos podem, então, ser aplicados à área de CI, de maneira a se otimizarem as operações de representação, organização, armazenamento e recuperação da informação. Na visão dos autores, existem três relações essenciais entre a Terminologia e a Documentação: a) a documentação é um

campo especializado, com Terminologia própria; b) a documentação utiliza a Terminologia; c) a documentação torna acessíveis os conhecimentos terminológicos, por meio da documentação terminológica.

Galinski e Krommer-Benz (1988) argumentam que a Terminologia, como campo da ciência e do conhecimento, afeta todos aqueles que trabalham com comunicação especializada. Assim, o trabalho terminológico é cada vez mais necessário para os diversos campos do saber, especialmente no que se refere a: a) organização do conhecimento (teoria geral da ciência, classificação dos conceitos para cada disciplina científica); b) transferência do conhecimento, habilidades especiais e tecnologias (educação e formação); c) tradução de textos científicos (tradução e interpretação); d) armazenamento e recuperação de informação especializada (linguagens de busca e recuperação, tesouros, índices, classificações).

No campo da Informática, a Terminologia traz sua colaboração a partir da elaboração dos conceitos para que a primeira, utilizando os elementos teóricos proporcionados pela Terminologia, elabore ferramentas que resultem em grande utilidade, sobretudo para a Terminologia Aplicada (Cabré, 1993). Assim, por um lado Terminologia utiliza a Informática para realizar seus trabalhos, abdicando de processos artesanais e obsoletos e, por outro lado, a Informática beneficia-se de elementos que permitem grandes progressos nas áreas de recuperação da informação, inteligência artificial e sistemas especialistas.

O maior vínculo entre a Informática e a Terminologia reside nos conceitos, unidades que constituem a base do conhecimento, uma vez que os sistemas conceituais descrevem a organização que cada disciplina efetua de seus conhecimentos. Os conceitos também estão na base da teoria dos termos e da inteligência artificial, que preocupa-se com o desenvolvimento de sistemas especialistas. Deste modo, ambas as teorias constituem a base para uma nova matéria: a engenharia do conhecimento, cujo tema central de estudo é a detecção, a representação, o tratamento, a transformação e a transferência do conhecimento. A engenharia do conhecimento tem como objetivo o desenvolvimento de bancos de conhecimento e sistemas especialistas, que utilizam conceitos da inteligência artificial e baseiam-se nas Terminologias correlatas (ARNTZ; PICHT, 1995).

Para Cabré (1993), na evolução da Informática orientada para a Terminologia, podem-se distinguir três etapas:

- a) Numa primeira fase, os computadores eram utilizados como instrumentos de armazenamento da informação, sendo inicialmente aplicados apenas à bibliografia e posteriormente à Terminologia. Nesta fase surgem os primeiros

bancos de dados documentais e bibliográficos; mais tarde, por volta dos anos 70, surgem os primeiros bancos de dados terminológicos, utilizados principalmente em sistemas de tradução;

- b) Na segunda fase, inicia-se o desenvolvimento de ferramentas que visam a facilitar a comunicação homem-máquina e máquina-máquina. Surgem as interfaces em linguagem natural, que modificam as condições de uso das grandes bases de dados pelos usuários, permitindo que os usuários realizem suas pesquisas de maneira autônoma. Melhora-se a estrutura dos sistemas de bancos de dados e dos sistemas de classificação conceitual; desenvolve-se a edição automatizada de dicionários e Terminologias, a partir de informações contidas em bases de dados;
- c) Na terceira fase, a partir das técnicas da inteligência artificial, desenvolvem-se os sistemas especialistas em questões terminológicas;

Ressalta a autora, que da concepção inicial de grandes bancos de dados como armazéns de informação destinados a pessoas, passa-se ao uso da informação pelo próprio computador, como elemento básico para a realização de operações mais complexas e mais aproximadas das habilidades e do conhecimento humano; assim, operações que antes exigiam a participação de um humano especialista em Terminologia, hoje podem ser realizadas automaticamente. Fausltich (1990) destaca que da junção da Terminologia com a Informática resultou um novo campo de investigação: a terminótica, cujo objeto de estudo é o tratamento automático do termo e requer que o pesquisador elabore (ou se utilize de) softwares para o gerenciamento de bancos de dados terminológicos; indexar termos com vistas ao levantamento de dados numéricos de frequência e disponibilidade de ocorrências nos textos; armazenar dados de obras terminográficas como glossários e dicionários técnicos especializados; editar as obras terminográficas em diferentes suportes: disquetes, mídia impressa, cd, entre outros.

Destacam-se algumas das aplicações da Informática à Terminologia: dicionários automatizados, verificadores ortográficos, programas que permitem criar bases de dados a partir de textos, editores que permitem redação semi-automatizada, programas de acesso a bases de dados terminológicas, entre outras. De modo análogo, a Terminologia também contribui para a evolução da Informática: computadores devem utilizar palavras ou termos para trabalhar e desenvolver todos os sistemas baseados em algum aspecto do processamento em linguagem natural. Assim, processos como tradução automática, verificação ortográfica, recuperação de informação, redação assistida, sistemas especialistas, requerem necessariamente a contribuição da

Terminologia, uma vez que necessitam de um tesauro que contenha as unidades conceituais que um computador deve compreender.

2.1.5 Importância da Terminologia

Para Felber (1984) Terminologias são importantes para a organização do conhecimento nas áreas de relacionamento dos conceitos (teoria geral da ciência e classificação conceitual de cada disciplina científica; para a transferência de conhecimento, habilidades e tecnologia, o que envolve educação, treinamento e leitura de textos técnicos; para a formulação e disseminação de informação técnica e científica; para a tradução de textos técnicos e científicos; para a elaboração de resumos de uma certa área de informação; bem como para o armazenamento e recuperação de informação técnico-científica (linguagens de indexação, tesouros e esquemas de classificação).

Conforme esse autor, o desenvolvimento da Terminologia como ciência, a aplicação de princípios terminológicos e a coordenação do trabalho terminológico internacional deverão ajudar para que se evite a desorganização da comunicação em cada área de conhecimento específico. Felber lembra que, em todos os idiomas, incontáveis novos conceitos, que surgem diariamente em todos os campos da atividade humana, devem ser expressos por termos existentes ou por novos termos, formados pela combinação de elementos das palavras, como raízes e afixos. Assim, para organizar o conhecimento e tornar possível a transferência de conhecimento, uma abordagem semântica e a aplicação de princípios terminológicos, bem como de terminografia auxiliada por computador, são absolutamente necessárias. O autor argumenta, ainda que, qualquer atividade relacionada com informação depende da disponibilidade de Terminologias confiáveis. Elas são necessárias quando a informação é formulada e quando é condensada ou recuperada. Os tesouros têm sido ferramentas indispensáveis para a recuperação da informação pois são flexíveis e podem incorporar novos desenvolvimentos científicos mais rapidamente do que os esquemas de classificação bibliográfica disponíveis. A construção de um tesauro inclui o trabalho de Terminologia Semântica e a explicitação das relações entre os conceitos.

De acordo com a ABNT (1997a), os princípios da Terminologia fornecem a base teórica para a construção dos sistemas de conceitos, uma vez que:

- a) ajudam a organizar os trabalhos de Terminologia de maneira eficiente;
- b) asseguram uniformidade e coerência de termos na área de especialidade e em áreas correlatas;

- c) coríbuen para a harmonização dos sistemas de conceitos e termos em diferentes línguas;
- d) fomentam o emprego eficiente da tecnologia da informação nos trabalhos de Terminologia.

Quanto maior o acervo de documentos, mais difícil a obtenção de documentos com precisão. O trabalho de recuperação pode ser facilitado por meio da indexação, destacando-se os termos (descritores) que descrevem da melhor maneira possível o conteúdo dos documentos que compõem o acervo. Obviamente os descritores somente cumprem a sua função se puderem seguir um padrão que, independentemente da época em que os documentos foram armazenados ou pesquisados, resultem na recuperação mais precisa possível. Esse nível de padronização pode ser conseguido por meio do uso de um tesauro, que é um conjunto de termos especializados, de um campo determinado, no qual especificam-se as relações conceituais e que assegura especialmente que um termo representa especificamente um conceito. Essa especialidade é alcançada por meio do uso de dois tipos de entrada: o termo preferido (descriptor) e o termo não preferido (não descriptor). Os não descritores guardam relação de sinonímia com os descritores. Assim, pode-se dizer que um tesauro é um instrumento de função terminológica, posto que constitui-se ferramenta para a realização de controle dos termos a serem utilizados na descrição de conteúdo (ARNTZ; PICHT, 1995).

2.1.6 Língua de Especialidade

Com o objetivo de garantir a comunicação, a língua utiliza palavras ou expressões como meio de representação dos conceitos. De acordo com a atividade humana especializada, os conceitos representados por uma palavra podem variar em relação a contextos mais amplos. Esse fenômeno introduz a necessidade de se definirem as fronteiras entre a língua comum (língua geral), pertencente à sociedade tomada como um todo, e a língua de especialidade (ou linguagem especializada). Os lingüistas e os terminólogos geralmente utilizam os termos *língua de especialidade*, *linguagem especial*, *linguagem especializada*, *linguagem especialista* ou *linguagem para propósitos especiais*, para designarem subsistemas léxicos cuja intenção é a comunicação não ambígua (WRIGHT; BUDIN 1997, p. 330). Pode-se, então, definir língua geral como o conjunto das palavras e expressões que, no contexto em que são empregadas, não se referem a uma atividade humana especializada (RONDEAU, 1984, p. 24). O autor apresenta em sua obra um diagrama esquemático – aqui reproduzido na ilustração 5 – que busca delimitar as fronteiras entre língua geral e língua de especialidade.

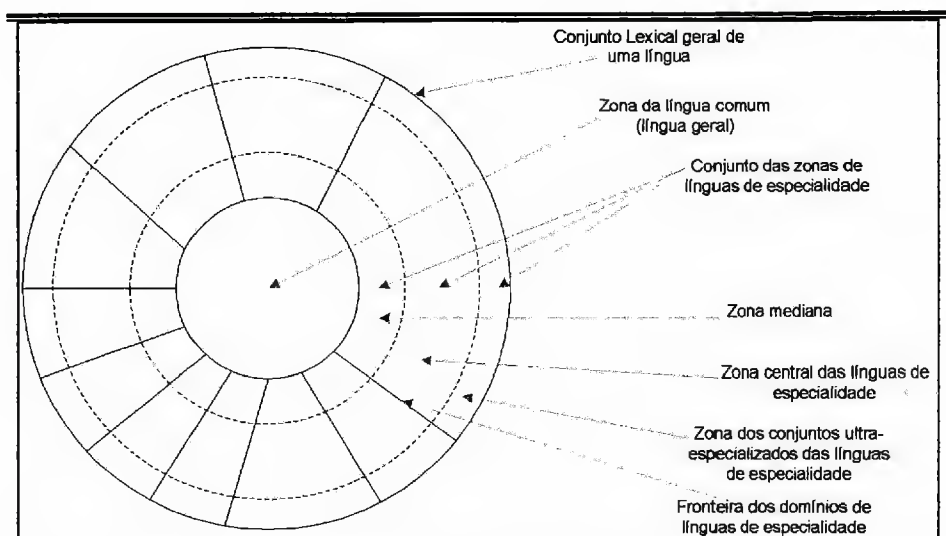


Ilustração 5: Línguas de Especialidade e Língua geral.

Fonte: (RONDEAU, 1984)

De acordo com o diagrama de Rondeau (1984, p.25), o conjunto das zonas de línguas de especialidade divide-se em três círculos concêntricos: uma zona mediana, a mais próxima da língua geral; uma zona central da língua de especialidade; e uma zona ultra-especializada, que congrega apenas os termos muito especializados da língua de especialidade. A zona mediana compreende os termos que representam conceitos que são utilizados em um domínio de especialidade, mas que também podem pertencer à língua geral com o mesmo sentido ou com um sentido aproximado. Em contraposição, o vocabulário ultra-especializado – terminologia ultra-especializada – compreende os termos que pertencem inicialmente apenas à área técnica ou científica e que originarão os neologismos. A zona central das línguas de especialidade compreende a terminologia principal de certa especialidade; essa terminologia é formada por termos que existem na língua geral, mas que, no domínio especializado, representam apenas conceitos pertinentes àquela língua de especialidade.

Segundo Wright (1997, p. 13) os textos em língua geral são produzidos quando se utiliza palavras pertinentes à linguagem cotidiana, combinados logicamente para se formar o discurso. Tais palavras são documentadas pelos dicionários lexicográficos. A gestão da terminologia, por outro lado, preocupa-se apenas com os termos, isto é, palavras atribuídas a conceitos utilizados em línguas de especialidade, que ocorrem em um dado campo ou domínio do conhecimento. As unidades terminológicas que são extraídas dessas línguas de especialidade geralmente representam um número limitado de partes do universo do discurso: nomes, verbos, adjetivos, advérbios, bem como expressões fraseológicas. Tais unidades são combinadas com conectivos utilizados na língua geral para produzir-se um discurso especializado coerente. Assim, o propósito das unidades terminológicas é o de delinear firmemente o espaço lingüístico e

conceitual ocupado por conceitos estritamente definidos e por determinadas unidades do conhecimento.

Para Auger e Rousseau (1977, p. 16) o campo semântico define-se como um conjunto de conceitos que podem ser agrupados sob um *conceito chave* ou *classe*. Tal conjunto estruturado de conceitos constitui o domínio, no sentido que adota a Terminologia. De acordo com Arntz e Picht (1995, p.37), a língua de especialidade, no domínio técnico-científico, divide-se em cinco níveis (de A a E), cujas características variam, conforme os critérios de avaliação: grau de abstração; representação lingüística; área ou âmbito e relações entre emissor e receptor. A tabela 7 representa os domínios das linguagens em relação às respectivas línguas. No nível mais baixo de abstração encontra-se a linguagem natural, no contexto da língua geral. Nesse nível há poucos termos, mas à medida que o grau de abstração aumenta, conseqüentemente aumenta o número de termos, até atingir o uso de símbolos.

Tabela 7: níveis de abstração no contexto da língua

Nível	Grau de abstração	Forma lingüística	Âmbito
A	Máximo	Símbolos artificiais para elementos e relações	Ciências teóricas
B	Muito alto	Símbolos artificiais para elementos; língua natural para relações	Ciências experimentais
C	Alto	Língua natural com grande porcentagem de termos técnicos e sintaxe determinada com rigor	Ciências aplicadas e tecnologia
D	Baixo	Língua natural com grande porcentagem de termos técnicos e sintaxe relativamente livre	Produção material
E	Muito baixo	Língua natural com poucos termos técnicos e sintaxe livre	Consumo

Fonte: (Arntz; PICT, 1995)

No universo da língua geral a originalidade de conteúdo e a unicidade de expressão são muito valorizadas. No universo da língua de especialidade, por outro lado, com o objetivo de se facilitar o compartilhamento global do conhecimento especializado, valorizam-se as funções cognitiva e referencial, em favor da consistência do conteúdo e da expressão. Em Terminologia, o princípio da consistência é mais importante do que o princípio da criatividade.

2.1.7 Unidade terminológica

Para Pavel e Nolet (2001, p. 18), a unidade terminológica é o mesmo nome ou designação de um conceito em um sistema de conceitos. Tal unidade pode ser uma expressão, um símbolo ou uma fórmula matemática ou química. Um termo ou unidade terminológica em

uma língua de especialidade distingue-se de uma palavra comum na língua geral pelo fato de que possui uma relação de monorreferencialidade com o conceito especializado que designa (monossemia).

Martinez (2003) define a unidade terminológica como “uma unidade de pensamento especializado, utilizada somente em situações de comunicação profissional...” A autora defende uma clara separação entre a língua natural e a linguagem especializada, argumentando que na primeira dão-se a polissemia e a sinonímia e que na segunda dão-se a univocidade e a monorreferencialidade dos termos. Assim, a motivação para desambiguar a comunicação científica e técnica, baseia-se na utilização da Terminologia como um instrumento de planejamento lingüístico e no estudo dos termos, não de forma isolada, mas nos textos da área de conhecimento a que se relacionam, e na sua correlação com outros termos, em contextos e situações reais.

As unidades terminológicas podem ser classificadas de acordo com diversos critérios, como por exemplo: a forma, a função, o significado e a procedência (CABRÉ, 1993).

De acordo com a forma, as unidades terminológicas podem ser classificadas:

- a) quanto ao número de morfemas – simples e complexas;
- b) quanto ao tipo de morfemas – derivadas (agregação de afixos a bases léxicas) e compostas (combinação de bases léxicas);
- c) quanto à estrutura sintática – combinação de palavras que seguem uma estrutura sintática (sintagmas terminológicos);
- d) unidades terminológicas de origem complexa – siglas (iniciais de palavras), abreviaturas (formas fixadas por consenso, que reproduzem o segmento inicial de uma palavra), acrônimos (palavras formadas por segmentos de outras) e formas abreviadas (uso de uma parte da palavra, normalmente para economizar o discurso).

Segundo Auger e Rousseau (1977), a atenção do terminólogo deve ser colocada sobre a estruturação dos conceitos relativos ao mundo real (campo conceitual) e tal estruturação é muitas vezes refletida nas atividades concretas próprias de um domínio do conhecimento. Desta sorte, toda unidade terminológica tem por função a denominação, isto é, o estabelecimento de relações entre uma forma lingüística (palavra ou termo) e uma realidade (abstrata ou concreta) pelo viés de um conceito. Toda unidade terminológica tem por função a denominação, isto é, o estabelecimento da relação entre uma forma lingüística (termo) e uma realidade (abstrata ou

concreta) enviesada por um conceito (AUGER; ROUSSEAU, 1977). Em outras palavras, unidade terminológica compreende a dicotomia conceito-termo, referindo-se a um objeto.

No contexto atual, a Terminologia fundamenta-se em um modelo tripartite, baseado em três pontos-chaves: o objeto, o conceito e o termo. A ilustração 6 representa esse modelo: um conceito é representado por um termo e corresponde a um saber sobre um objeto; um termo denomina um objeto e representa um conceito; um objeto é denominado por um termo e é conceituado por um conceito.

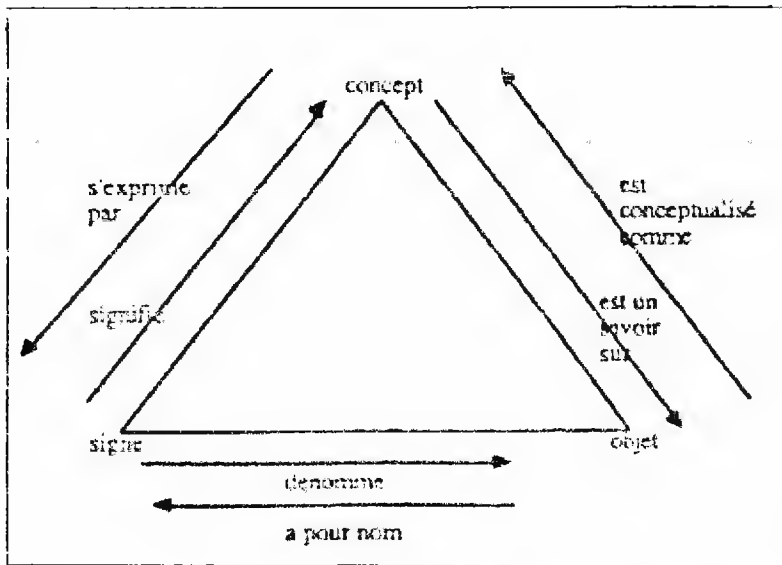


Ilustração 6: Representação esquemática de um sistema de conceitos
 Fonte: Van Campenhout, 1997

Segundo Felber (1984, p. 100), esse triângulo foi introduzido por Gomperz em 1908, sendo utilizado posteriormente por diversos outros autores: Dittrich, em 1913; Ogden e Richards, em 1923; Ullmann, em 1952; Knobloch, em 1956; e Baldinger, em 1959.

Com base nessa idéia, a norma ISO 1087 (1990) enuncia:

Conceito – unidade de pensamento constituída pela abstração a partir das propriedades comuns a um conjunto de objetos;

Objeto – elemento da realidade que pode ser conceituado;

Termo – designação, por meio de um símbolo, de um conceito definido em uma língua de especialidade.

2.1.7.1 Conceito

Para Mauritin (1958), no campo da lógica formal menor, o conceito é uma representação de uma coisa, de uma idéia sobre uma coisa; é o que o espírito produz ou exprime em si mesmo, e em que ele atinge ou apreende uma coisa. O conceito formal é uma imagem ou similitude espiritual da coisa no homem em si. Assim, o conceito formal faz o homem conhecer ou apreender alguma coisa, isto é, possui um certo objeto (objeto formal do conceito ou conceito objetivo) que é o objeto mesmo de uma simples apreensão: essência, natureza ou quiddidade (conjunto de condições que determinam um ser em particular) apresentada sob tal aspecto inteligível ou sob tal formalidade. O conceito objetivo é o que se conhece ou que se apreende da coisa pelos conceitos mentais. É a representação de um objeto pelo pensamento, por meio das suas características (FERREIRA, 1986). A idéia que se tem de um objeto só constitui um conceito quando se pode distinguir o que há nesse objeto de semelhante aos objetos que lhe são diferentes e o que há de diferente nele, em relação aos objetos que lhe são semelhantes (GORSKI, 1968).

Ainda nessa perspectiva, os conceitos são compostos exclusivamente das relações em que se dispõem os elementos que concorrem para a sua formação; isto é, os conceitos são representados por termos que, por seu turno, são derivados de suas próprias relações com outros termos. Assim, um conceito não pode ser concebido sem a existência de um sistema, isto é, o conceito se dispõe num conjunto conceptual de que participa integral e indissolúvelmente e dentro do qual encontra toda sua razão de ser. Um conceito não é concebível fora do sistema a que pertence. O conceito, como representação mental da realidade objetiva exterior ao pensamento, realiza essa representação não individualmente e numa correspondência bi-unívoca entre si mesmo e as diferentes feições daquela realidade; realiza-a por meio de relações com outros conceitos e nas quais tal conceituação, em conjunto (sistema), é estruturada. Nesse sistema, os termos, ou elementos constituintes de tais relações, possuem valor não como elementos em particular, mas unicamente pela sistemática em que eles se dispõem relativamente uns aos outros. Desse modo, é incalculável o número de relações que constituiriam ou representariam os diversos conceitos imagináveis pelo homem. (PRADO JR., 1968).

Para Gorski (1968), todo conceito é um pensamento acerca das propriedades do objeto; assim, o conceito pode ser pensado como referindo-se ao objeto, a uma propriedade do objeto ou à sua relação com outros objetos. Nos três casos, o conceito compreende um pensamento acerca de características. Qualquer conceito compreende um conjunto de características que lhe são peculiares e que, reunidas, distinguem-no de qualquer outro conceito. Tais características são definidas por alguns autores como notas ou propriedades do conceito

(GORSKI, 1968; MAURITAIN, 1958). A esse conjunto de características, denomina-se compreensão de um conceito. Essas características selecionadas denominam-se propriedades essenciais e cada uma delas tomada individualmente é necessária para distinguir o objeto, de tal modo que todas juntas devem ser suficientes para se realizar tal distinção. Para exemplificar, o conceito “cachorro” é definido por características, como animal, irracional, canino, possui quatro patas, entre outras. Tal conjunto de características distingue esse animal de qualquer outro animal. Por outro lado, algumas dessas características o tornam similar a certos outros animais. Tais características que podem ser comuns a certo número de conceitos formam o conjunto a que se denomina extensão de um conceito. A extensão significa a amplitude de um conceito em relação a outros conceitos aos quais se agrupa e se aplica em sua unidade. Os conceitos de compreensão e extensão são também denominados pelos termos conotação (ou, ainda, intensão) e denotação, respectivamente. Pode-se afirmar que as definições de intensão e de extensão são inversas: quanto maior a intensão, menor a extensão de um conceito, e reciprocamente (MAURITAIN, 1958).

Um conceito é uma unidade do pensamento, geralmente expressa por um termo, um símbolo literal ou por qualquer outro símbolo. Quando o conceito retrata um único objeto, é chamado de conceito individual e é geralmente representado na língua por um nome próprio ou por um símbolo; quando o conceito representa um conjunto de dois ou mais objetos, é chamado de conceito geral e a designação é representada por um termo (ISO 704, 1996). Assim, conceitos são representações mentais não apenas de seres ou coisas (que podem ser expressos por nomes), mas também de qualidades desses seres ou coisas; (tais qualidades podem ser expressas por meio de adjetivos ou de outros nomes); de ações (expressas por meio de verbos ou nomes); e, ainda, por localizações, situações ou relações (expressas por meio de advérbios, preposições, conjunções ou outros nomes) (ISO/1087, 1990).

Picht (1984) afirma que as diversas escolas têm definido “conceito” de diferentes maneiras, nem sempre convergentes. Mas o autor adota a definição na norma ISO/1087 que, no seu ponto de vista, não serve a todas as aplicações, mas apesar dos seus defeitos, serve como um instrumento de trabalho. Essa definição enuncia conceito como um elemento do pensamento, expresso em geral por um termo e por um símbolo literal ou outro diferente. O autor enfatiza no valor da norma o fato de que os conceitos podem ser a representação mental não somente de seres ou de coisas, mas, também, de qualidades, ações e, ainda, localizações, situações ou relações.

Do ponto de vista terminológico, a definição dos conceitos compreende um aspecto lingüístico e um aspecto filosófico. Pelo seu aspecto lingüístico, ela trata da semântica

lexical ou lexicologia, uma vez que os termos pertencem ao universo da língua como as outras unidades lexicais; trata também da semiologia, uma vez que os significados são mais importantes do que os significantes. Pelo aspecto filosófico, a Terminologia utiliza métodos da Lógica e da Ontologia (RONDEAU, 1984, p. 11). O autor argumenta que a Terminologia relaciona-se com a Ontologia, uma vez que o significado de um termo – conceito, para a Terminologia – corresponde a entidades concretas ou abstratas que são consideradas inicialmente pela sua existência para, em seguida, buscar a forma lingüística que as representa ou que deveria representar. Em outras palavras, destaca o autor, a Terminologia analisa os conceitos em uma direção contrária àquela dada pela lingüística: não parte de um nome para descobrir seus significados, mas sim de um significado para encontrar a melhor denominação que o represente. Com relação à Lógica, a Terminologia pode relacionar-se com essa disciplina de duas maneiras: a) os conjuntos de conceitos são reagrupados por domínios logicamente organizados; b) a terminologia analisa os conceitos com o apoio da compreensão (conjunto de características próprias de um objeto de um conceito) e pela extensão (conjunto das entidades que possuem as mesmas características). Rondeau advoga que a noção de conceito para a Terminologia equivale à dada pela Filosofia: uma representação abstrata composta por um conjunto de traços comuns essenciais a um grupo de entidades (objetos ou idéias) e obtida pela extração das características individuais de tais entidades (RONDEAU, 1984, p. 22). As características individuais podem ser classificadas em: forma, função e natureza.

A Teoria do Conceito, que é uma disciplina básica da Lógica, forma, juntamente com a Teoria da Denominação, os fundamentos da Teoria Geral da Terminologia (TGT). Para a TGT o conceito existe independentemente do termo e corresponde ao significado do que este é. Um termo é deliberadamente atribuído a um conceito após a devida consideração sobre se tal termo corresponde àquele conceito em questão. O termo, que é um símbolo lingüístico, já pode existir ou pode ser criado a partir das características que constituem integralmente o conceito em questão. A teoria do conceito é caracterizada pela investigação sobre a natureza dos conceitos, seus relacionamentos e sua representação em sistemas de conceitos (FELBER, 1984).

Conceitos são representações mentais de objetos individuais. Um conceito pode representar um objeto individual ou, por abstração, reunir um conjunto de objetos individuais que possuem qualidades (ou características) em comum. Dessa maneira um conceito, que existe apenas na mente humana, necessita de um símbolo lingüístico (termo, letra, símbolo gráfico) para ser expresso ou comunicado. O conceito é acima de tudo um elemento do pensamento. Os conceitos não apenas representam seres ou coisas, mas, em sentido mais amplo, podem representar qualidades (expressas por nomes), ações (expressas por verbos ou nomes) e até

localizações, situações ou relações (expressas por advérbios, preposições conjunções ou nomes). Os conceitos podem surgir da combinação de outros conceitos, mesmo que não haja correlação com a realidade. A agregação das características de um conceito constitui a sua intensão. O conjunto de todos os conceitos subordinados (específicos) no mesmo nível de abstração ou de todos os objetos individuais que pertencem ao conceito em questão constitui a extensão do conceito dado. No caso da agregação de objetos individuais, tal extensão é denominada classe (FELBER, 1984).

Para Sager (1990) os principais objetos da terminologia são os termos e são percebidos como símbolos que representam os conceitos. Os conceitos devem ser criados e devem vir a existir antes que os termos venham a ser formados para representá-los. A denominação de um conceito deve ser o primeiro passo na consolidação de um conceito como uma entidade socialmente aceita. A formação de um conceito compreende o agrupamento e a ordenação de objetos materiais e imateriais que o homem pode perceber, sentir ou imaginar em categorias abstratas. Assim, os conceitos são construtos mentais, isto é, abstrações que podem ser usadas na classificação de objetos individuais do mundo interior ou exterior. Tais objetos, sejam eles campos do conhecimento ou atividade humana, coisas, suas propriedades, qualidades, fenômenos, ou outros tipos de objetos, são representados por meio de conceitos.

Sager (1990) define conceito como um construto do processo de cognição humana que auxilia a classificação de objetos por meio de uma abstração sistemática ou arbitrária. Assim, o conceito é uma unidade do pensamento e é utilizado para estruturar o conhecimento e a percepção do mundo circundante. O conceito necessita ser expresso ou explicitado, uma vez que corresponde a uma referência a atividades mentais, seja como um resultado ou como um processo. Com o objetivo de ser aplicado no campo da Terminologia, é necessário que se considere o conceito como uma primitiva axiomática, expressa por meio de um termo ou sentença.

Para esse autor, no processo de formação dos conceitos, normalmente agrupam-se os dados da percepção e experiência humanas, de acordo com elementos comuns, que são usualmente denominados características. A soma das características utilizadas para a descrição de um conceito é chamada de intensão, isto é, um agrupamento único de características, que difere de um conceito semelhante por pelo menos uma de tais características. Quanto mais características particulares um conceito possui, mais específico ele é – sua intensão é mais específica; quanto menos características particulares, mais amplo é o conceito e poderá referir-se a diferentes conjuntos de objetos que se enquadrem na sua extensão. Assim como os objetos

materiais são organizados em classes para a formação de conceitos, os próprios conceitos são organizados em classes tais como entidades físicas, métodos científicos de análise e descrição, propriedades dos objetos, entidades abstratas ou concretas, produtos manufaturados, processos industriais, entre outras. Além de serem expressas como propriedades de um conceito, as características podem também ser expressas como relações de um dado conceito com outros conceitos.

O conceito desempenha um papel fundamental para a Teoria da Terminologia, sobretudo nos seguintes âmbitos: na normalização, para tornar claro o sentido em que se emprega o conceito; na informação e documentação, especificamente nos processos de tratamento e recuperação; na lexicografia especializada, na estruturação de dicionários. O conceito é uma unidade do pensamento que envolve as características comuns atribuídas aos objetos. Os conceitos não estão vinculados a uma dada língua, mas recebem a influência das transformações sócio-culturais em cada momento. O conceito é um elemento do pensamento que se refere àquilo que os homens detectam de comum em uma pluralidade de objetos e que utilizam como meio de organização do pensamento (concepção) e, conseqüentemente, como instrumento de comunicação. Em outras palavras: os conceitos são unidades do pensamento que servem para o reconhecimento dos objetos, para a comunicação sobre os objetos e para a organização mental de tais objetos (ARNTZ;PICHT, 1995).

Um conceito pode representar apenas um objeto ou um conjunto de objetos que possuem determinadas características. Tais objetos podem ser materiais (coisas animadas ou inanimadas que ocupam lugar no espaço e no tempo) ou mentais (ações, qualidades, entre outras). O conceito é uma unidade do conhecimento que sintetiza em uma forma lingüística os predicados necessários e verificáveis acerca de determinado objeto. Partindo-se de objetos individuais e gerais chega-se a conceitos individuais e gerais, respectivamente, estabelecendo-se predicados verdadeiros em relação a tais objetos; tais predicados são também chamados de elementos conceituais, elementos do conhecimento ou de características (ARNTZ;PICHT, 1995).

Para esses autores, a totalidade das características que permitem uma síntese mental de objetos individuais e a delimitação mútua dos conceitos constitui a compreensão de um conceito (intensão). Assim, a compreensão de um conceito é o conjunto de todas as características que o constituem. Não é possível a modificação do número de características de um conceito em um dado momento sem, ao mesmo tempo, modificar-se o próprio conceito. Por esse motivo, o conjunto de características de um conceito permite a sua clara delimitação em

relação a outros conceitos contíguos, cuja compreensão pode ser maior, menor, ou divergente daquele.

Por extensão de um conceito entende-se a totalidade de objetos individuais que possuem todas as características de tal conceito. Uma vez que nem sempre é possível a enumeração de todos os objetos individuais que correspondem a um conceito, a extensão desse conceito é freqüentemente expressa por meio dos conceitos subordinados pertencentes ao mesmo nível hierárquico (ARNTZ; PICHT, 1995).

Nessa linha de pensamento, os autores consideram que a definição de extensão é falha e destacam que a norma DIN 2342, de 1986, introduz uma distinção entre extensão de conceito e classe de conceito. Para essa norma, classe é definida como a totalidade de objetos que correspondem a um conceito; e extensão é o conjunto de conceitos situados em um mesmo nível e subordinado a outro conceito. Argumentam, ainda, que tais definições são também instituídas pela norma ISO 1087 em sua revisão de 1988: classe é a totalidade de objetos aos quais um conceito se refere; extensão é a totalidade de conceitos específicos incluídos em um conceito genérico. Quanto maior for a extensão do conceito ou a classe de um conceito, menor será o número de características que possui e maior será o número de objetos que ele descrever. Ao se adicionarem características a um conceito diminui-se a sua extensão, surgindo conceitos mais específicos; isto é, diminui-se, também, o número de objetos correspondentes a esse conceito.

Quando se cria um novo objeto cria-se, em conseqüência, um novo conceito. Na maioria dos casos, os novos conceitos têm sua origem na evolução de outros já conhecidos e entre os quais existe uma relação de parentesco ou de semelhança. Tal relação pode ser de quatro tipos: determinação (enriquecimento da intensão), conjunção (fusão de intensões), disjunção (fusão de extensões) ou integração (combinação ou fusão de constituintes). Se a um conceito original é adicionado um outro conceito para que o primeiro seja definido com maior precisão, sua compreensão tornar-se-á mais específica, resultando um novo conceito. Dessa maneira, as características adicionais diferenciam o novo conceito resultante do conceito original. No entanto, será estabelecida uma relação de semelhança entre eles, pois o novo conceito incluirá todas as características do conceito original. A esse processo denomina-se determinação.

A conjunção de conceitos consiste na fusão das intensões dos conceitos originais, de modo que as características de cada conceito são combinadas em partes igualmente proporcionais. O conceito resultante é a soma dos conceitos originais. Caso haja uma inversão na ordem de colocação dos termos constituintes do novo conceito, não haverá mudança de sentido quanto ao conceito representado. O conceito de máquina coletadeira-debullhadora pode também

ser representado pelos termos máquina debulhadora-coletadeira. Na disjunção de conceitos, o valor dos conceitos constituintes é equivalente, mas, nesse caso, a formação do novo conceito se dá a partir da fusão das extensões dos conceitos originais, resultando um novo conceito superordenado comum aos conceitos constituintes. A integração de conceitos consiste na reunião dos objetos atribuídos aos conceitos originais em torno de um novo objeto. Por exemplo: cubo, raios e aro somados introduzem o conceito de roda. Enquanto a conjunção, a disjunção e a determinação correspondem a relações de abstração, a integração corresponde a relações ontológicas.

2.1.7.1.1 Características

Para Felber (1984) uma característica é um elemento de um conceito que serve para descrever ou identificar certa qualidade de um objeto individual. A característica é, também, por si só, um conceito. As características são utilizadas para comparar conceitos, classificar conceitos, para formular a definição de um conceito, bem como para formar os termos atribuídos aos conceitos. As características podem ser classificadas em características intrínsecas – aquelas inerentes ao objeto –, que designam forma, tamanho, material, cor, entre outras propriedades; características extrínsecas de propósito – aquelas que designam o propósito do objeto –, como aplicação, funcionamento, performance, localização e posicionamento na montagem; e características extrínsecas de origem – aquelas que designam a origem do objeto –, como método de manufatura, descobridor, descritor, inventor, produtor, país de origem, fornecedor.

Com relação à relevância, as características intrínsecas são mais importantes do que as extrínsecas, uma vez que são mais facilmente verificáveis do que estas. Entre as características extrínsecas, as de propósito são mais importantes, porque geralmente relatam a essência de um conceito; ao contrário das características de origem, que não estão conectadas à natureza do objeto.

As características podem ser simples – referem-se a uma única propriedade de um objeto individual; ou complexas – referem-se a uma combinação de duas ou mais propriedades do objeto em questão. As características são utilizadas como elementos de classificação para a colocação de um conceito em um sistema de conceitos. As características são ditas equivalentes quando podem ser substituídas uma pela outra sem haver expressiva modificação no sentido do termo em questão. As características não equivalentes podem ser classificadas em independentes – que seguem umas às outras em diferentes níveis de uma série vertical de conceitos e que podem ser combinadas entre si; e dependentes – requerem que a característica superordenada tenha precedência em relação a uma subordinada, em uma série vertical de conceitos.

Para Sager (1990) as características são os elementos conceituais ou elementos do conhecimento. Obviamente, as características também são conceitos, por isso têm uma importância especial tanto para a definição de um conceito, quanto para a determinação das relações de tal conceito com outros. As características representam as propriedades dos objetos e podem ser determinadas com base na observação, medição, atribuição, entre outros processos. As características são a base para a classificação dos conceitos e são necessárias para se diferenciar um conceito de outros em uma dada área de especialidade. As características têm suma importância para a Terminologia, especialmente quanto a fatores como:

a) determinação da intensão de um conceito – a análise terminológica compreende um inventário ou uma recompilação de todos os elementos do conhecimento (características) acerca de um conceito. A totalidade das características de um conceito determinadas em certo momento constitui a soma dos conhecimentos sobre tal conceito. A associação de uma ou mais características de um conceito resulta na formação de um novo conceito.

b) fundamentação para a formação de termos – tanto a formação de novos termos ou a avaliação de termos já existentes podem ser realizadas por meio da seleção das características que melhor definem o conceito relacionado.

c) estruturação de sistemas de conceitos – a estrutura de um sistema de conceitos depende fundamentalmente da influência das características, uma vez que estas podem ser utilizadas como critérios de classificação ou de organização de tais conceitos no sistema.

d) determinação de equivalências – conceitos que possuem as mesmas características possuem identidade conceitual; assim também os termos que representam tais conceitos serão correspondentes – sinônimos, tratando-se do mesmo idioma; ou equivalentes, tratando-se de diferentes idiomas. A análise terminológica necessita conhecer as características conceituais.

Para objetos materiais, a norma DIN 2330 estabelece três tipos de características: a) características de estado – relacionam-se a forma, material constituinte, cor, posição e tempo; b) características racionais – relacionam-se a procedência, modo, comparação, valoração, situação no espaço; c) características funcionais – referem-se a potência ou rendimento, utilidade. A norma ISO 704 (1996), por outro lado, determina a divisão das características em características intrínsecas (forma, tamanho, material constituinte, cor, posição) e características extrínsecas (origem, performance, localização, descobridor, inventor) (SAGER, 1990).

O autor discute que nem todas as características estabelecidas para um conceito possuem a mesma importância. Há as características essenciais, ou seja, aquelas que refletem as propriedades de um objeto em uma dada situação, sob um ponto de vista específico e em certo campo de especialização; e as características de ordem ou não essenciais, que são todas as outras características utilizadas para a definição ou determinação das relações entre conceitos. De acordo com o campo de especialidade, as características essenciais e não essenciais de um conceito podem variar.

De acordo com a norma ISO WD 704.1 (1996, p. 5) os conceitos retratam ou correspondem a um conjunto de objetos e são representados na língua por meio de designações ou de definições; podem ser organizados em níveis hierárquicos. Assim, os objetos observados, percebidos ou concebidos, são abstraídos por meio de conceitos – são conceituados. Objetos concretos são ditos objetos materiais e objetos abstratos são ditos imateriais. Mesmo objetos imaginários – como sereias ou unicórnios – são considerados objetos imateriais passíveis de conceituação, uma vez que a preocupação principal da Terminologia não está na determinação da existência ou não-existência de um objeto, mas no modo como tal objeto é conceituado e na forma como as pessoas lidam com tais objetos para fins de comunicação.

A Terminologia interessa-se pela comunicação, que se dá por meio de línguas especiais afetas a certas áreas do conhecimento – ou línguas de especialidade; por esse motivo, pode-se dizer que o conceito é não apenas uma unidade do pensamento, mas também uma unidade do conhecimento. As propriedades específicas de um objeto ou comuns a um conjunto de objetos são abstraídas como características, que podem ser combinadas para a criação de conceitos (ISO WD 704, 1996).

Quando um conceito retrata um único objeto, é chamado de conceito individual e, geralmente, pode ser representado por um nome próprio ou por um símbolo. Quando o conceito representa um conjunto de dois ou mais objetos é chamado conceito geral e sua designação passa a ser um termo. Os objetos são descritos a partir de suas propriedades – características – e tais características podem ser utilizadas na análise dos conceitos para a construção de sistemas de conceitos, para a formulação de definições e para a formação das designações – termos (ISO WD 704.1, 1996). As características podem ser agrupadas em tipos, de acordo com critérios como constituição, uso e propriedades físicas do objeto, entre outras. Assim, ter-se-iam classificações como: cor, composição, função, uso, origem, forma, localização, movimento, entre outras. Normalmente é requerido certo domínio do campo de conhecimento – ou da língua de especialidade – para se abstrair as características de um determinado conceito.

Ao conjunto de características agrupadas para a formação de um conceito dá-se o nome de intensão. Ao conjunto de objetos que possuem intensões similares dá-se o nome de extensão. Assim, a intensão de um conceito determina a sua extensão. Nem todas as características de um conceito são fundamentais para a sua formulação. A norma ISO WD 704.1 (1996) denomina como características essenciais de um conceito aquelas consideradas indispensáveis para a compreensão do conceito em um determinado campo do conhecimento e cuja ausência altera fundamentalmente tal conceito. As características não-essenciais são aquelas cuja ausência não altera fundamentalmente o conceito a que caracterizam. Características que diferenciam um conceito em relação a outros podem ser chamadas de características delimitadoras. A análise das características semelhantes ou diferentes entre os conceitos permite situar o sistema numa posição específica em um sistema de conceitos.

Segundo a norma ISO/R 860, os conceitos devem ser unificados em sistemas de conceitos passíveis de internacionalização. Tal unificação deve preocupar-se com os seguintes elementos: conceitos e sistemas de conceitos; descrição dos conceitos; forma externa dos termos, especialmente a forma gráfica; forma interna dos termos, isto é, a forma literal de termos complexos e de termos transferidos; e descrição dos termos.

O conceito é uma tríade formada por um referente, pelas características essenciais a ele pertinentes e por uma forma comunicável. É importante considerar que esta definição não trata de uma estratificação de três partes, uma vez que o conceito está inteiro no referente; inteiro nas características pelas quais é definido; e inteiro em sua denominação: o *termo*. O termo tem, pois, uma função de exprimir, de maneira econômica, para fins de comunicação, todas as características do referente (IBICT, 1984, p.4-5).

Dahlberg (1978) considera a existência de pelo menos dois tipos de características:

a) características essenciais ou necessárias; para GORSKI (1968), tais características podem variar, de acordo com a necessidade que se tem em relação ao objeto; então, dependendo do uso prático que o homem possa ter em relação a tal objeto, suas propriedades serão de uma natureza; variando essa necessidade, variará o conjunto de suas propriedades essenciais;

b) características acidentais – as características acidentais são todas aquelas que, embora qualifiquem o conceito, caso sejam removidas, não alterarão fundamentalmente o conceito.

As características essenciais definem conceitos gerais, ao passo que as características acidentais definem os conceitos específicos. Tais características são categorizadas por Dahlberg (1978) em: a) essenciais constitutivas – aquelas que produzem a essência; b) essenciais consecutivas – aquelas que produzem conseqüências, devido a sua essência; c) acidentais gerais – aquelas que, embora acidentais, possuem uma certa generalidade; d) acidentais individualizantes – aquelas que possuem uma natureza puramente individualizante.

A ilustração 7 apresenta uma sistematização para essas definições:

Objetos materiais	Minerais
Características essenciais	
Essenciais-constitutivas	
possuir uma substância	composição química
possuir uma estrutura	estrutura cristalina
Essenciais-consecutivas	
possuir propriedades físicas	refratar a luz
possuir propriedades químicas	dureza ou solubilidade
possuir propriedades elétricas	conductibilidade
Características acidentais	
Acidentais-genéricas	
possuir uma certa forma	morfologia externa
possuir certos defeitos	deformações
possuir uma certa cor	cor
Acidentais individualizantes	
possuir uma certa localização	local onde o mineral é encontrado
possuir um determinado período	idade ou era em que é encontrado

Ilustração 7: Classificação de compostos químicos em geral e de minerais em particular

A ação de analisar as características de um conceito é denominada análise conceitual e pode ser considerada como uma representação de fatos conhecidos acerca de um conceito formal. Assim, Dahlberg (1978) define conceito como uma unidade do conhecimento sintetizando as características de um item de estado de referência, por meio de um termo ou nome. Um conceito geral é um conceito que sintetiza apenas as características essenciais; um conceito individual é tido como um conceito que sintetiza tanto as características essenciais, quanto as acidentais.

Segundo Dahlberg (1978b, p. 10) os conceitos, na sua dimensão extensional, podem assumir o caráter de verdadeiras classes. Por meio de um ou mais conceitos, é sempre possível obter-se uma classe de objetos desde que as características dos conceitos de tais objetos estejam contidas nos conceitos de um sistema de conceitos. Inicialmente, a autora, em consonância com as normas DIN 2330 e ISO-R 704, propõe uma definição de conceito como uma unidade básica do pensamento; e propõe como conceitos científicos somente aqueles

plenamente descritíveis ou definíveis. Mais adiante, apresenta em seu texto uma definição mais elaborada, segundo a qual conceito é uma unidade de conhecimento obtida pela síntese dos predicados – características, ou seja, aspectos perceptíveis – relacionados com determinado objeto e que, por meio de sinais lingüísticos, pode ser comunicado. Assim, cada enunciado acerca de determinado objeto corresponde a uma unidade conhecimento a respeito desse mesmo objeto. Em síntese, as características dos conceitos são obtidas por meio de predicados enunciados; os conceitos possuem elementos constitutivos denominados características; um conjunto de características determina um conceito; os conceitos são unidades de conhecimento constituídas pelas características dos objetos associados a elementos lingüísticos.

Os conceitos possuem características distintas e podem ser classificados, segundo essas características, em conceitos individuais ou conceitos gerais (PICHT, 1984). Com base nas características de um conceito, é possível definir sua extensão ou compreensão. A compreensão de um conceito refere-se ao conjunto de todas as características que se referem àquele conceito. As ilustrações 8 e 9 caracterizam o conceito de compreensão:

Características	Automóvel	Caminhão	Caminhão contêiner
1 Veículo	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
2 Propulsão a motor	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3 Dispõe de um mecanismo de direção	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4 Projetado para transportar cargas		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5 Transporta cargas líquidas			<input checked="" type="checkbox"/>

Ilustração 8: Definição de compreensão de um conceito, com base nas suas características

A interpretação das figuras fornece a compreensão do conceito “caminhão”, definido pelas características 1 a 4. O conceito mais específico “caminhão contêiner” é definido pelas características 1 a 5.

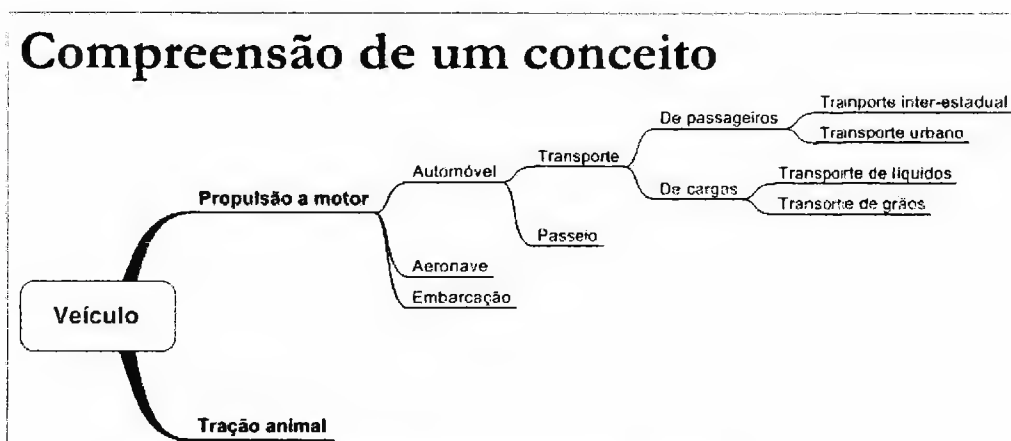


Ilustração 9: compreensão de um conceito

A extensão de um conceito pode ter duas acepções: a) extensão por semelhança – que corresponde a todas as espécies imagináveis de um conceito, consideradas em separado; este tipo de extensão pode ser exemplificado pelos conceitos mais específicos que compreendem um conceito mais amplo; por exemplo, a extensão “animal” compreende: mamíferos, anfíbios, aves e répteis; b) extensão por composição – que corresponde ao conjunto das partes que, em separado, constituem um todo; por exemplo, a extensão “água” compreende: hidrogênio e oxigênio.

De acordo com o grau de especialização do conceito, aumenta o número das suas características e diminui o número de objetos pertinentes a uma categoria em questão (Picht, 1984). Os conceitos e os termos evoluem de modo diferente de acordo com as peculiaridades da língua e da comunidade lingüística, em função de fatores históricos, geográficos, sociais e econômicos. Essas diferenças entre sistemas de conceitos paralelos e as pseudo-semelhanças entre os termos, prejudicam os processos de comunicação (ABNT, 1997a). Preocupando-se com esses processos de comunicação, portanto, o trabalho da Terminologia deve basear-se em conceitos e não apenas em termos, por duas razões principais:

- b) as diferenças ou semelhanças entre conceitos nem sempre aparecem claramente no nível dos termos;
- c) as similaridades no nível dos termos nas diferentes línguas não significam necessariamente que os conceitos aos quais esses termos se relacionam sejam idênticos (ABNT, 1997b).

Segundo Picht (1984), há basicamente quatro tipos de combinações de conceitos:

- a) determinação – um conceito determina outro, aumentando o número das suas características; exemplo: veículo + espaço => veículo espacial
- b) conjunção – dois conceitos são fundidos em um novo conceito; as características do novo conceito correspondem à soma das características dos formadores; exemplo: preto + branco => preto e branco;
- c) disjunção ou adição lógica – corresponde à soma da extensão de vários conceitos; neste caso, os conceitos constituintes convertem-se em conceitos subordinados a um conceito genérico; exemplo: garoto + garota => adolescente;
- d) integração – dois conceitos individuais são reunidos: império austríaco + império húngaro => império austro-húngaro

De acordo com a teoria vienense, os conceitos são compostos por um conjunto de características que são propriedades dos objetos conceituados e que permitem diferenciar ou comparar esses conceitos. Podem-se, assim, distinguir propriedades intrínsecas, que denotam características do tipo gênero-espécie; e extrínsecas, que denotam origem ou destino.

São essas características que servem para identificar os conceitos e determinar seu lugar em uma hierarquia espécie-gênero. Um sistema de conceitos depende dos tipos de características escolhidos para sua representação. Se as características forem do tipo dependentes, o sistema segue uma ordem determinada por essas características; se as características forem independentes, podem descrever qualquer sistema, uma vez que o sistema não necessita estar sub-determinado por elas (CABRÉ, 1993).

De acordo com Picht (1984), as características podem ser classificadas em:

- a) características de estado (ou próprias) – indicam propriedades de um objeto individual, envolvido por um conceito que pode ser compreendido pela própria análise do objeto; as principais características de estado são:

forma – oval, circular, largo, estreito, etc;

material – metálico, de madeira;

cor – preto, vermelho, azul;

posição – vertical, horizontal, suspenso;

tempo – velho, novo, moderno;

- b) características de relação – indicam a existência de relação entre conceitos em uma situação específica; as principais características de estado são:

origem – indicam a procedência dos objetos; inventor, país de origem, data de origem, fabricante, procedimento de fabricação;

medida de estado – transportável, maleável;

comparação – maior, menor;

valoração – caro, barato, insuficiente, suficiente;

classificação no espaço – situado na parte traseira, ante-posto, frontal

- c) características funcionais – mostram a combinação funcional dos objetos entre si; as principais são:

rendimento ou potência (capacidade do objeto) – velocidade de produção, velocidade máxima;

características de uso – para transporte, sala de cirurgia;

Quanto à função, o autor argumenta que as características podem ser utilizadas para: a) comprovar a compreensão de um conceito; b) para constatar mudanças na compreensão de um conceito – que podem até levar a um novo conceito: átomo indivisível; átomo divisível; c) para elaborar e formular definições, ou seja a verbalização de um conceito; d) para estruturar e ordenar conceitos em um sistema; e) para formar termos; f) para constatar a equivalência entre dois ou mais termos, em línguas diferentes; g) para constatar a sinonímia de dois ou mais termos em uma língua.

Gomes (1990a, p. 20) define característica como o elemento de um conceito que serve para descrever ou para identificar determinada qualidade de um objeto individual. É por meio das características que se comparam os conceitos, permitindo sua reunião em classes, subclasses, ou outro tipo de agrupamento não-hierárquico. É a partir das características que se definem os conceitos. As características de um conceito podem ser inerentes a um objeto individual, sendo denominadas características intrínsecas; ou podem ser relacionadas a propósitos ou origem de um objeto individual. As características de propósito identificam aplicação, uso, funcionamento, desempenho, lugar ou posicionamento de dado objeto em um conjunto. As características de origem referem-se ao método de manufatura ou nome do descobridor, inventor, produtor ou fornecedor. Compartilhando da definição de Dahlberg (1978), a autora admite a existência de características equivalentes – que podem ser substituídas entre si, sem haver alteração de significado para o termo – e características não-equivalentes. Na sua concepção as características não-equivalentes podem ser independentes – isto é, podem ser arbitrariamente combinadas entre si – ou dependentes – isto é, exigem que a característica superordenada preceda à subordinada, numa série vertical de conceitos.

2.1.7.2 Termo

Para Boutin-Quesnel (apud CAMPOS, 1992, p. 28), termo é toda unidade lingüística que designa uma noção de maneira unívoca dentro de uma área de atividade. Uma definição mais complexa apresenta o termo situado em um sistema de conceitos, relativo a um determinado campo do conhecimento: o termo é uma ou mais palavras cujo signo lingüístico está unido à noção correspondente dentro do sistema de noções de um dado domínio do conhecimento (KLIMOVICKIJ (1979, p. 32, apud CAMPOS, 1992, p. 30). Para Lotte (apud CAMPOS, 1992, p. 32), um termo deve possuir propriedades sistematizantes, cuja importância

umenta em função do acréscimo do número de noções compreendidas em cada nível de classificação de uma determinada ciência; não se pode restringir o papel do termo a uma única função de “denominação”. A terminologia científica não deve ser um simples conjunto de palavras, mas um sistema de palavras e de grupos de palavras ligadas entre si de maneira específica; nisto está a diferença entre a terminologia “tout court” e a terminologia científica.

Em suma, um termo é um signo lingüístico que representa um conceito identificado na estrutura conceitual de um campo específico do conhecimento (MEDEIROS, 1988). Um termo pertinente a um domínio específico do conhecimento difere de um não-termo por pelo menos três fatores: em primeiro lugar, o termo pertence ao léxico de uma língua de especialidade enquanto um não-termo pertence ao léxico da língua geral; em segundo lugar, cada termo representa um conceito especificamente, o que significa que para haver clareza na comunicação em um domínio específico, deve ser reduzida a ambigüidade do seu significado²¹; em terceiro lugar, o termo deve fazer parte de um sistema de classificação hierárquica.

Uma unidade terminológica, ou termo, é um símbolo convencional representando um conceito definido em um certo domínio do saber (CABRÉ, 1993). Para a autora, nesta concepção, inspirada diretamente no ensinamento de Wüster, falta acrescentar-se a importância de cada palavra, uma vez que um termo: a) é uma expressão, e esta um símbolo; b) resulta de uma convenção; c) possui importância pelo seu conteúdo; d) está associado a um conceito, que define-se explicitamente; e) refere-se à descrição de um conceito; tal descrição, por sua vez, reenvia a um domínio do saber no qual esse conceito está inserido.

Felber (1984) afirma que um termo é qualquer símbolo convencional utilizado para representar um conceito definido em um campo de assunto. Sager (1990) corrobora essa afirmação, argumentando que os termos são representações lingüísticas de conceitos e que, diferentemente da língua geral, em que a arbitrariedade do signo é aceita, as línguas de especialidade tentam sistematizar princípios de designação e nomeação de conceitos, de acordo com regras pré-estabelecidas ou princípios gerais.

Cabré (1993) ressalta que, como qualquer outra unidade significativa do sistema lingüístico, os termos fazem parte de um sistema estruturado, no qual ocupam um determinado nível e se relacionam, por um lado, com as demais unidades do mesmo nível e, por outro, com as unidades dos demais níveis, participando conjuntamente na construção do discurso.

²¹ Há autores, mais especificamente aqueles que pertencem à corrente da Teoria Geral da Terminologia, que defendem o ponto de vista da monorreferencialidade, isto é, a cada termo associa-se apenas um conceito indissociável; não se admitem sinônimos ou plurivalências de nenhum tipo e a padronização das terminologias é necessária para chegar-se à comunicação unívoca, sem ambigüidades (CAMPOS 1992, p. 411).

Em relação à formação de novos termos, Pavel e Nolet (2001, p. 20-21) admitem as seguintes formas de neologismos:

a) neologismo de significado ou neologismo semântico — origina-se de um termo já existente, mas cujo significado é alterado. Neologismos semânticos podem originar-se de processos de expansão, isto é, o termo recebe um significado estendido; de processos de metáfora, isto é, um termo já conhecido é empregado com valor metafórico — *backbone* em uma rede; de processos de conversão de categoria gramatical; de processos de adoção de um termo advindo de outra área do conhecimento – vírus, aplicado ao campo da computação.

b) neologismo morfológico – origina-se de diversos processos como derivação, composição, empréstimo, abreviação, junção.

As autoras lembram, ainda, que a aceitação de um neologismo depende de fatores como a brevidade, a facilidade de manipulação, facilidade de retenção, potencial para derivação, entre outros fatores.

O conceito de termo varia de acordo com os pressupostos teóricos a que se segue. A Lexicografia parte do significante para chegar ao seu significado, isto é, um signo lingüístico compreende um significado; a Terminologia parte do significado conceito para chegar ao significante (denominação ou termo). Em terminologia parte-se, portanto, de um conceito e busca-se uma denominação que o represente da maneira mais precisa; já em Lexicografia parte-se da denominação e buscam-se os conceitos que tal denominação possa representar.

Para Rondeau (1984, p. 19), o termo é essencialmente um signo lingüístico, isto é, uma unidade lingüística que compreende um significante e um significado. O autor adota a nomenclatura de *denominação* (termo) para o significante e de *noção* (conceito) para o significado. Uma característica do termo é o fato de que o seu significado é definido em relação ao um conjunto de conceitos pertencentes ao mesmo domínio. Em suma, um termo não pode ser considerado isoladamente; ele apresenta-se como uma representação de um objeto em um conjunto semântico que pode ser uma disciplina, uma ciência, um domínio de atividades, uma técnica, entre outras áreas, todas pertinentes a um domínio especializado. Rondeau destaca ainda uma outra característica do termo, que é o fato de que para um conceito dado, existe, teoricamente, apenas um termo correspondente. Tal característica compreende o postulado da univocidade, entre o significante (termo) e o significado (conceito), que é uma relação reflexiva (RONDEAU, 1984, p.19). O autor argumenta, no entanto, que o postulado da univocidade não se aplica a sempre a todos os termos, uma vez que, para um dado conceito, é possível encontrar-se, em uma mesma língua, diversas denominações diferentes, como é o caso dos sinônimos. Com

relação aos seus modos de formação, o termo é uma forma lingüística especializada representando um conceito definido no interior de um domínio particular, encontrando sua origem:

a) na especialização de uma palavra da língua geral;

b) na criação neológica a partir de raízes de línguas antigas ou a partir de diversas maneiras como: contração, justaposição, derivação, entre outras;

c) obtido pelo empregado em formas perifrásticas ou sintagmáticas mais ou menos complexas.

Segundo Rondeau (1984, p. 20) a homonímia não constitui um risco de ambigüidade, uma vez que todo termo pertence a um grupo semântico dado, de modo que sob o plano do discurso, um termo constitui um par denominação-conceito claramente identificado pelo contexto e que, sob o plano lógico, encontra lugar em uma estrutura de conceitos hierarquizada no interior de um domínio de conhecimento. Para exemplificar, o autor utiliza o termo *cromático*, que possui pelo menos três acepções: a) no campo da música, representa o conceito de uma sucessão de sons que distam entre si de um semi-tom; b) no campo da biologia, o termo *cromático* refere-se a cromossomos; c) no campo da óptica, denota os defeitos causados pela variação do índice de refração de uma lente em função do comprimento de onda originando os termos *aberrações cromáticas*. Tal característica da formação dos termos ilustra a necessidade, em terminologia, de se evitar conferir maior importância à forma lingüística do que ao conceito que essa forma representa.

Em resumo, o termo é um signo lingüístico composto por um significado (conceito) e por um significante (denominação). Então, segundo o Rondeau (1984, p. 21), a denominação é a forma lingüística externa do termo, o que implica dizer que a terminologia deve ocupar-se de estudar a forma lingüística em si mesma e a relação que tal forma guarda para um dado conceito em um certo campo do conhecimento. O autor defende o princípio da monorreferencialidade, isto é, para cada termo dado existe uma e somente uma denominação que corresponde a um certo conceito. Isto não significa que uma mesma forma lingüística não possa ser reutilizada, mas, ao contrário, significa que cada novo emprego de uma mesma forma lingüística externa associa-se a um conceito diferente.

Auger e Rousseau (1977, p. 31) definem termo como toda unidade lingüística que denomina um conceito de modo unívoco em um domínio do conhecimento. Destaca-se de tal definição que a principal função do termo é a denominação, isto é, o estabelecimento de uma

relação entre uma realidade concreta ou abstrata e um signo da língua. Tais termos podem ser classificados, segundo o autor, em unidades simples – formadas por apenas uma palavra – e unidades complexas – formadas por mais de uma palavra –, que não podem ser dissociadas sem que haja modificação do sentido dado pelo conjunto. Os autores definem os termos como monorreferenciais, isto é, cada termo exprime um único conceito como resultado da atividade de definição.

Para a ISO 1087, termo é uma designação de um conceito definido numa língua para propósitos específicos através de uma expressão lingüística. De acordo com a ISO/DIS 704 (apud CAMPOS, 1992, p. 32), um termo deve: ser lingüisticamente correto; ser preciso; ser conciso; permitir com facilidade a formação de derivados; ser padronizado somente quanto monorreferencial.

Os termos devem ser tão concisos quanto possível, sem, contudo, perderem a clareza do seu significado. Termos abreviados podem ser formados a partir de um termo simples, de um termo composto, ou de uma frase; podem reter sílabas ou letras isoladas do termo de origem, mas devem ser evitados se o termo completo for suficientemente conciso. Em relação à correspondência com a definição, o significado literal de um termo complexo deve refletir as características do conceito a ser designado ou, pelo menos, não devem contradizê-las. Termos complexos devem refletir a combinação dos conceitos-base (características) que os constituem. Não necessariamente um termo complexo inclui cada constituinte de uma combinação de conceitos mostrados em uma definição, mas somente aqueles necessários para distinguir o seu significado dos conceitos relacionados. Todo termo novo deve ser construído de tal maneira a permitir a formulação de derivações.

Termos que possuem mais de um significado – termos plurivalentes – podem ser homônimos ou polissêmicos em relação a outros termos; esse fenômeno faz com que aumente a ambigüidade em relação aos seus significados. Quando o fenômeno da ambigüidade ocorre, é necessário que se utilizem termos diferentes para a explicitação dos diferentes significados de um termo ambíguo.

Como definido na norma (ISO WD 704.1), um termo é uma designação que consiste de uma ou mais palavras e que representa um conceito geral numa língua especial. Quando um novo termo é criado para a designação de um conceito, tal fenômeno é denominado neologismo. No que tange ao aspecto semântico, as relações entre termos e conceitos podem ser classificadas como monossemita, mononímia, homonímia ou sinonímia. A monossemita e a mononímia reduzem a ambigüidade, enquanto a homonímia e a sinonímia a aumentam.

A monossímia é a relação entre termos e conceitos, na qual um termo representa apenas um conceito. A mononímia é a relação entre termos e conceitos em que um conceito possui apenas um termo que o designa em uma determinada língua. A homonímia é caracterizada por envolver a relação entre termos e conceitos, na qual os termos em uma determinada língua apresentam formas idênticas – sejam elas fonéticas ou gráficas –, mas que designam conceitos diferentes e não-relacionados. Termos que possuem fonia idêntica, mas com grafia distinta são ditos homófonos; termos de grafia idêntica, mas com pronúncia distinta são ditos homógrafos. Homônimos perfeitos possuem idênticas grafia e pronúncia. A sinonímia é a relação entre diferentes termos que designam o mesmo conceito em uma determinada língua – sinônimos. Se termos são atribuídos a conceitos cujas intensões são quase idênticas, são chamados de quase-sinônimos (ISO WD 704.1).

Em relação à sua composição, a norma (ISO WD 704.1) classifica como termos simples aqueles que contêm apenas um radical e como compostos os termos que contêm dois ou mais radicais. Em relação à formação de novos termos, a norma aponta três processos: criação de novas formas – por meio de derivação, composição ou abreviatura –; utilização de formas existentes; e empréstimo translingual.

O processo de derivação compreende a adição de um ou mais elementos afixos à raiz do termo. O processo de composição envolve a junção de termos ou elementos de termos já existente para a criação de uma nova forma que contém dois ou mais radicais, mas que designa apenas um conceito. No processo de abreviatura, novas formas reduzidas do termo original são utilizadas, com o objetivo de facilitar a sua utilização. O termo original é denominado termo completo ou por extenso.

A utilização de termos existentes para a criação de novos termos pode se dar pelos processos de conversão, terminologização, transferência semântica e empréstimo transdisciplinar. O processo de conversão compreende a alteração da categoria sintática, isto é, da função gramatical do termo de origem. A terminologização é o processo pelo qual uma palavra ou expressão da língua geral é transformada num termo que designa um conceito pertinente à língua especial. A transferência semântica compreende a utilização da extensão lógica de um termo para a designação de outro conceito. Os termos que designam um conceito concreto podem ser estendidos de modo a designarem um conceito abstrato ou uma parte estendida de tal conceito (ex.: tela: monitor; tela: informações no monitor).. O processo do empréstimo transdisciplinar, ou processo de empréstimo interno, compreende o empréstimo de um termo pertinente a uma área ou assunto e a sua atribuição a um novo conceito em outra área (ex.

memória: computador; memória: cérebro). O processo de empréstimo translingual compreende a introdução em uma língua de termos ou conceitos existentes em outra língua e pode ser direto ou por tradução literal. O empréstimo direto é caracterizado pela introdução do novo termo, que pode ter sua pronúncia ou ortografia alteradas (ex.: *lay out* – leiaute) e dá-se, preferencialmente, quando não há termo similar para a designação de tal conceito na língua de destino. O empréstimo por tradução literal ocorre pela tradução de um termo estrangeiro para a língua de destino (*heavy water* – água pesada).

De acordo com a norma (ISO WD 704.1) a formação de novos termos deve ser concordante com determinados princípios:

- é desejável que um termo seja atribuído a um único conceito;
- a criação de um termo deve ser antecedida de verificação da existência de um termo para o conceito em questão;
- o uso já consagrado deve ser respeitado, mesmo que haja deficiência na formação do termo;
- deve ser buscada a transparência, isto é, o quanto possível, o significado do termo deve ser visível na sua morfologia; isso pode ser conseguido utilizando-se, para a criação do termo, uma característica essencial que não se modifique a curto prazo;
- novos termos devem ser coerentes com os termos existentes no sistema de conceitos em questão;
- deve-se evitar a formação de termos que causem confusão, primando-se pela adequação aos padrões de significado conhecidos e estabelecidos pela comunidade lingüística (ex. deve-se preferir energia nuclear a energia atômica; e engenharia genética a manipulação genética);
- os termos devem ser concisos;
- os termos devem ser passíveis de derivação;
- os termos devem ser escritos em conformidade com as normas morfológicas, morfossintáticas e fonéticas da língua em questão;
- deve-se preferir a língua nativa para a formação de novos termos, evitando-se os empréstimos.

2.1.8 Definição

Uma definição consiste na determinação de um conceito com meios lingüísticos e serve para estabelecer uma relação o mais inequívoca possível entre conceitos e denominações. Além disso, delimita um conceito, colocando-o em relação a outros conceitos já conhecidos ou definidos. Uma definição é a equivalência entre algo que será definido (*definendum*) e como tal coisa será definida (*definiens*), com o objetivo de delimitar a compreensão do *definendum* em qualquer ato de comunicação. Uma definição pode ser representada graficamente como uma equação em que o *definendum* aparece do lado esquerdo e o *definiens* aparece do lado direito, separado por um sinal de dois-pontos ou por uma igualdade (ARNTZ;PICHT, 1995).

Com relação à sua estrutura, a ilustração 10 mostra como se forma uma definição:

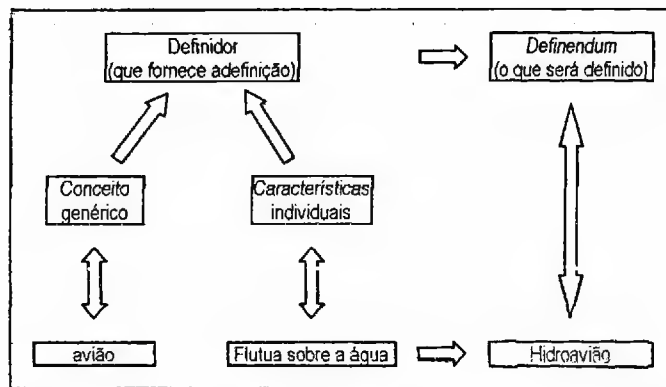


Ilustração 10: Estrutura de uma definição segundo Arntz e Picht (1996)

Para esses autores, os tipos de definições mais importantes para a Terminologia são as definições por compreensão e por extensão. Na definição por compreensão, parte-se de um conceito superordenado conhecido ou já definido e se indicam as características específicas que permitem distinguir o conceito em relação a outros conceitos da mesma série. As definições por compreensão utilizam a seguinte estrutura genérica:

definendum. definiens = conceito superordenado + características distintivas

As definições por extensão indicam a extensão de um conceito, seja enumerando-se todos os conceitos subordinados que encontram no mesmo nível de classificação; seja enumerando-se todos os objetos de uma classe; ou, ainda, indicando-se a regra pela qual se pode chegar à enumeração das características do conceito que está sendo definido.

De acordo com as normas DIN 2342 e 2330 (apud ARNTZ;PICHT, 1995), uma definição é a determinação de um conceito através de meios lingüísticos, e seu objetivo é estabelecer a melhor relação unívoca entre conceito e nome (termo); as definições representam a

delimitação de um conceito em relação a outro. Tais conceitos são colocados paralelamente de modo que um conceito já definido anteriormente (conhecido) é colocado ao lado de outro conceito ainda não definido (desconhecido), permitindo que se possa fazer a distinção entre ambos.

De acordo com a norma ISO WD 704.1 (p. 5), uma definição é um elemento utilizado para descrever um conceito. A fim de que tal descrição seja feita de maneira clara, deve-se determinar tanto a intensão quanto a extensão do conceito a ser definido. A definição de um conceito permitirá a determinação da sua posição em relação a outros conceitos em um sistema de conceitos. Conseqüentemente, a norma estatui que deve ser feita uma seleção das características a serem mencionadas na definição, com o objetivo de se delimitar o conceito em relação a outros conceitos. Nesse contexto, cada definição deve ser estabelecida em relação com as definições de outros conceitos do mesmo sistema, pois somente a comparação com outras definições poderá fornecer o significado preciso de um certo conceito. Todos os termos técnicos que ocorrerem em uma definição deverão ser definidos na mesma obra terminológica.

A norma ISO 704: 1968 classifica as definições em definições por extensão e definições por intensão. Nas definições por intensão, o conceito específico é definido a partir de características herdadas de um conceito mais amplo e de pelo menos mais uma característica que o diferencia de tal conceito. Este tipo de definição origina-se de relações genéricas, genérico-partitivas ou partitivas. As definições podem ser exaustivas, uma vez que novas espécies podem ser descobertas ou inventadas após uma definição por extensão ter sido estabelecida. Isto significa que as definições por extensão podem variar com o decurso do tempo. As definições por extensão podem ser muito úteis para finalidades práticas quanto utilizadas sozinhas ou como complemento de definições intensionais.

A definição genérica apresenta o conceito genérico superior ao conceito que está sendo definido, seguido das características essenciais que diferenciam o conceito em questão dos conceitos coordenados. Desse modo, as características que compõem a intensão do conceito superordenado são implicitamente assumidas na definição (ISO WD 704.1).

A definição genérico-partitiva apresenta o conceito genérico superior ao conceito que está sendo definido, seguido dos conceitos partitivos essenciais que compõem o conceito mais abrangente. As partes opcionais do conceito superior não são incluídas na definição; mas apenas as partes essenciais ou delimitadoras (ISO WD 704.1).

A definição partitiva descreve um conceito visto como parte de um determinado todo. Define-se o conceito abrangente em primeiro lugar, indicando-se a sua relação com os

conceitos partitivos. Normalmente uma definição partitiva é iniciada com formulações que indicam claramente que o conceito ora definido é parte de outro conceito mais abrangente, seguidas das características delimitadoras. Uma definição partitiva só é aceitável se o conceito em questão é parte essencial de um conceito mais abrangente e se sua extensão estiver completa (ISO WD 704.1).

Uma definição extensional consiste em substituir-se o enunciado de uma definição pela lista dos conceitos subordinados, designando, assim, os objetos que compõem a extensão do conceito. Segundo recomenda a norma (ISO WD 704.1), as definições extensionais só devem ser utilizadas se o número de conceitos a serem enumerados for limitado e se tais conceitos puderem ser esclarecidos por meio de definições intensionais, ou se forem bem conhecidos.

Uma definição é uma proposição lógica, composta de sujeito, verbo e predicado, utilizada para explicitar o que é um conceito. Nessa estrutura, o sujeito corresponde à designação – termo –, o verbo fica subentendido como a forma verbal “é” e o predicado constitui as características que formam o conceito. O enunciado pode ser apresentado em seguida a elementos tipográficos como dois pontos ou travessão, o iniciar-se em outra linha, destacando-se, assim, do termo designante. Ao elaborar-se uma definição para um determinado conceito, deve-se identificar as relações entre o conceito e seus conceitos relacionados e estabelecer-se, assim, um sistema de conceitos (ISO WD 704.1).

Alguns princípios devem ser seguidos na elaboração de definições:

- ao se substituir um termo pela sua definição, não deve haver perda de significado, isto é, deve ser possível verificar-se se a definição está correta;
- as definições devem ser elaboradas de modo sistêmico, sendo coordenadas entre si, de modo a possibilitarem a reconstrução do sistema de conceitos;
- uma definição deve ser claramente expressa, de modo a ser compreendida independentemente de outras;
- as definições devem ser tão breves quanto possível, contendo apenas as informações que tornam único o conceito definido ;
- uma definição deve incluir apenas um conceito, não incluindo definições ocultas para nenhuma das características;
- qualquer característica que requeira uma explicação deve ser definida separadamente, como um conceito;

– a extensão e as características refletidas numa definição devem ser adequadas ao sistema de conceitos, situando-se numa área específica de conhecimento;

A norma (ISO WD 704.1) relata alguns tipos de definições defeituosas, a saber: definições circulares, definições incompletas e definições por negação. A definição circular utiliza o termo, ou parte dele, como característica no enunciado da definição. Por exemplo: árvore perene – árvore com folhagem perene; uma reformulação adequada seria: árvore que conserva as folhas verdes durante o ano inteiro. A circularidade pode ocorrer também quando dois ou mais conceitos são definidos uns por meio dos outros (ISO WD 704.1). A definição incompleta caracteriza-se por omitir propriedades essenciais do conceito, podendo ser considerada abrangente demais ou restrita demais; em qualquer das situações, torna-se difícil a compreensão do conceito de forma unívoca. Uma definição por negação caracteriza-se por definir o que o conceito não é, quando deveria definir o que o conceito é.

Em terminologia, o papel da definição é descrever o conteúdo conceitual das unidades terminológicas, do modo mais apropriado possível. Para tal, ela deve ser precisa, clara, sucinta, mas deve evidenciar os traços significativos próprios do termo tratado (AUGER; ROUSSEAU, 1977, p.33). O autor argumenta que em Terminologia é preferível se optar por definições do tipo enciclopédico, do que por definições lingüísticas, uma vez que aquelas, sendo caracteristicamente descritivas, dão uma melhor imagem sobre as realidades; seu acento está mais nos aspectos concretos das realidades tratadas. Assim, a definição deve reunir todos os traços semânticos necessários para se chegar com precisão ao sentido do termo. Geralmente a definição indica a função de um objeto, a descrição de um procedimento, o funcionamento de um aparelho ou mecanismo. Portanto, sendo relativa a aspectos concretos, a definição deverá reunir todos os elementos necessários à melhor compreensão possível do sentido do termo pelos usuários ou especialistas no domínio tratado. A definição situa-se obrigatoriamente em um domínio específico de atividades ou conhecimentos e é monorreferencial, isto é, corresponde a somente uma realidade bem determinada. Os termos técnicos utilizados na formulação da definição devem ser conhecidos pelos usuários do domínio em questão, de modo que tornem-se facilmente acessíveis. Para Pavel e Nolet (2001, p. 23-24), uma definição terminológica é uma descrição concisa das características que delimitam um conceito, e apresenta-se no mesmo formato que as entradas em um dicionário lexicográfico. A definição deve dar o significado do termo, ao invés de tratar sobre sua utilização. Uma definição é a mais importante aplicação do princípio da monorreferencialidade e a principal motivação para o estabelecimento de uma correspondência textual – isto é, a apresentação de um contexto em que o termo assume determinado significado. Uma definição terminológica é uma breve declaração que fornece um claro entendimento do

significado de um termo especializado; ela inicia-se com uma entrada que identifica uma classe mais ampla à qual pertence dado conceito e, em seguida, especifica as características essenciais ou de delimitação que claramente distinguem o conceito em questão de outros conceitos relacionados naquela classe. As características de delimitação podem ser intrínsecas, tais como natureza, material de que se compõe, ou um tópico a que tal conceito se relaciona; ou extrínsecas, tais como a função, o modo de uso ou de operação, a origem, a destinação ou o referente.

Além das características essenciais, um conceito possui características não-essenciais como forma, inventor ou produtor, tempo e espaço. As definições podem ser enunciadas pelos seguintes métodos: definição por gênero e diferença; definição por função; definição operacional, listando partes ou passos de um processo; definição por paráfrase – utilização de um sinônimo (Pavel; Nolet, 2001, p. 25-26).

Segundo Bessé (1997, p. 68), escrever uma definição é uma operação que consiste na determinação de todas as características que unicamente identificam a intensão de um conceito. A definição terminológica é criada com referência a um objeto ou idéia denotado por um signo, e por isso é uma ação referencial, isto é, o *definendum* não é o termo em si, mas o objeto²² ou a idéia descrita e, mais precisamente, sua representação conceitual.

Ainda segundo Pavel e Nolet, alguns princípios são fundamentais para a formulação de definições: previsibilidade – a definição insere o conceito em um sistema de conceitos; simplicidade – a definição deve ser concisa, clara e não maior do que uma sentença; enunciado afirmativo – a definição estatui o que o conceito é e não o que ele não é; enunciado não circular – a definição de um conceito não deve utilizar termos cuja própria definição remeta novamente ao conceito que está sendo definido; ausência de tautologia – a definição não é uma paráfrase de um termo, mas sim uma descrição das características semânticas de um conceito. Para Bessé (1997, p. 70), a definição não deve ser exaustiva, mas apenas refletir o conhecimento em questão; deve permitir o estabelecimento de clara diferenciação entre objetos e conceitos; não deve ser redundante; não deve conter o termo que está sendo definido, mas sim a raiz do termo genérico, de modo a estabelecer uma relação entre termo a ser definido e seu hiperônimo imediato.

Em terminologia entende-se por contexto o ambiente lingüístico de um termo em um enunciado. Geralmente o contexto é representado por um segmento de frase ou por uma ou

²² A palavra objeto deve ser entendida em seu sentido amplo, isto é: qualquer coisa que possa ser conceituada, incluindo coisas materiais, fenômenos, propriedades, ações e estados associados a um dado domínio do conhecimento humano (DUBUC; LAURISTON, 1997, p. 80).

mais frases justapostas ao termo que está sendo definido. O papel do contexto em terminologia é esclarecer o sentido de um termo, por isso, deve colocar em evidência os traços semânticos do termo em questão. (AUGER; ROUSSEAU, 1977, p. 34)

Duquet-Picard (apud CAMPOS, 1992, p. 45) aponta cinco tipos de definições: a de sentido amplo, a específica, a genérica, a descritiva e a prescritiva. A definição de sentido amplo consiste na descrição de um conceito com a ajuda de conceitos conhecidos, sendo feita por intermédio de palavras. A definição específica ou de sentido restrito indica o conteúdo do conceito; ela parte de um conceito genérico conhecido e denomina as características restritas que são próprias do conceito a definir e que o distinguem dos outros conceitos numa mesma espécie. A definição genérica determina a extensão do conceito, isto é, consiste na enumeração de todas as especificações que se encontram no mesmo nível no interior do sistema de conceitos. A definição descritiva esclarece o significado de um termo pelo uso da língua. A definição prescritiva estabelece o que um termo deve significar. A autora apresenta ainda alguns princípios a serem seguidos para a elaboração de definições terminológicas:

- a definição de ser esvaziada de tudo que possa caracterizar o termo definido com uma palavra pertinente ao léxico da língua geral;
- a definição deve ser monorreferencial e unívoca, isto é, a um dado conceito corresponde apenas um termo e a esse termo corresponde apenas um conceito;
- a definição deve incluir a extensão do conceito e representar suas características extrínsecas e intrínsecas;
- a definição deve ser completa, sem englobar informações inúteis ou supérfluas;
- a definição deve respeitar o princípio da não circularidade, isto é, não se deve usar um sinônimo ou quase-sinônimo para definir o termo.

Quanto à forma, a autora sugere algumas regras a serem seguidas na formulação de definições:

- o termo definido não deve estar contido na definição;
- o termo deve ser definido em um domínio de referência a que ele pertence, utilizando-se termos próprios de tal especialidade;

- a definição deve permitir que se percebam as relações existentes entre o termo definido e o conjunto de termos da área;
- os termos utilizados para a definição devem ser definidos em outra parte da obra terminológica em questão;
- tanto em sua forma quanto em seu conteúdo, a definição deve ser adaptada às necessidades dos seus usuários, isto é, deve ser definido o público-alvo a que se destina a obra terminológica, bem como o domínio de especialidade.

A esses princípios Campos (1992) acrescenta:

- deve-se evitar a redação de definições modo negativo;
- deve-se evitar o uso de palavras de sentido vago ou figurado;
- a definição além de ser precisa, clara e sucinta, deve colar em evidência as características significativas do termo em estudo;
- a extensão da definição deve ser limitada aos elementos necessários à compreensão do sentido do termo pelos usuários da terminologia.

A definição de um conceito, de acordo com a norma alemã DIN 2330, é “...a fixação de um conceito, estabelecendo relações com outros conceitos (conhecidos ou já definidos) com a finalidade de delimitá-lo em relação a outros conceitos” (PICHT, 1984). Apoiando-se nessa premissa, o autor destaca três funções para as definições: a) fixar um conceito – normalmente mediante sua normalização; b) delimitar um conceito, isolando-o de outros conceitos; c) relacionar um conceito com outros, o que leva, conseqüentemente, à elaboração de um sistema de conceitos.

Quanto aos tipos, Picht (1984) classifica as definições em:

- a) específica – são enumeradas todas as características do conceito, considerando-se um contexto específico;
- b) genérica - baseia-se na extensão do conceito e pode ser obtida por meio da enumeração de todos os objetos que pertencem a um conceito mais amplo; pela enumeração dos conceitos subordinados (no mesmo nível) a um conceito mais amplo; pela indicação de uma regra, mediante a qual se chega a uma enumeração
- c) definição por exemplo – parte de exemplos para definir novos conceitos.

As definições são consideradas “defeituosas”, quando apresentam os seguintes problemas:

- a) definição incompleta – o elemento definido não apresenta características distintivas que permitam identificá-lo individualmente: “um parafuso é um elemento de união”; não estão definidas as características do conceito união;
- b) definição limitada demais – o elemento definido é restrito a um determinado contexto muito restrito, quando se poderia ampliar seu conceito a outros contextos: “frascos medicinais são frascos fabricadas de vidro e que servem para armazenar remédios”; nesta definição estariam excluídos os frascos plásticos;
- c) definição ampla demais – o elemento definido é inserido em um determinado contexto muito amplo: “um subúrbio é um distrito de uma cidade”; nesta definição não existe clareza quanto ao fato de que um subúrbio localiza-se numa região periférica da cidade;
- d) definição circular – o elemento é definido com base em um sinônimo e não nas suas características individuais ou elemento é utilizado como termo definidor e como termo a definir: “ISO é uma instituição que emite normas ISO” ou “Normas ISO são emitidas pela ISO”;

2.1.9 Sistema de conceitos e relações conceituais

A definição de conceito como uma síntese de seus elementos constitutivos (características) proporciona a compreensão de que dois ou mais conceitos podem compartilhar uma ou mais características, determinando algum tipo de relação entre eles. Outra possibilidade é a de que tais conceitos não possuam nenhuma característica em comum, o que poderia determinar algum tipo de relação de outra natureza entre eles. Diversos autores consideram diferentes tipos de relações entre os conceitos. Nesta seção serão exploradas algumas dessas principais abordagens.

Picht (1984) destaca que em Terminologia não se considera o conceito como uma grandeza isolada. Cabe avaliá-la e elaborá-la dentro do seu conjunto conceitual, que por sua vez está relacionado com um campo do saber, com uma disciplina ou com um ramo de atividade. É necessário, portanto, considerar-se o conceito inserido no seu sistema de conceitos. A norma DIN 2331 define sistema de conceito como um certo número de conceitos entre os quais ou existem ou foram estabelecidas relações formando-se, assim um todo coerente. Essa definição

coincide com o que postula a norma ISO 1087, segundo a qual um sistema de conceitos é um conjunto de noções vinculadas, unidas por relações lógicas ou ontológicas. Tal conjunto está constituído por séries horizontais e verticais de noções e compreende, no mínimo, uma dessas séries.

O autor compara os sistemas de conceitos a um edifício, em que os conceitos seriam os tijolos e as relações entre eles seriam a argamassa (PICHT, 1984). No entanto, para dar forma ao edifício é necessária uma planta, que pode fornecer informações como: o objetivo, o campo do saber com suas características e métodos específicos, os critérios de ordenação segundo os quais se deseja sistematizar os conceitos.

A gestão da terminologia é, em grande extensão, baseada em sistemas de conceitos, que indicam as relações que existem entre os conceitos de um campo de conhecimento específico. Por natureza, essas conexões podem ser hierárquicas ou não-hierárquicas. As relações hierárquicas mais comuns são as lógicas ou genéricas. Esse tipo de relação refere-se a conceitos que podem ser categorizados pela declaração “x é um tipo de y”. Relações hierárquicas pressupõem a existência de conceitos superordenados e de conceitos subordinados. Um conceito superordenado pode ser dividido em conceitos subordinados, de classificação mais baixa que, juntos, formam o conceito superordenado. Conceitos que existem no mesmo nível hierárquico são denominados conceitos coordenados. Outra maneira, mais utilizada pelos lingüistas, para a denominação desses tipos de conceitos compreende o uso dos termos *hiperônimos*, para designar os conceitos superordenados; *hipônimos*, para a designação dos conceitos subordinados; e *co-hipônimos*, para a designação dos conceitos coordenados. Especialistas em documentação tratam tais relações com o uso de outros termos. Assim, em um tesouro, por exemplo, o termo superordenado em uma relação hierárquica é chamado *termo amplo*, enquanto o termo subordinado é designado por *termo específico*; os termos que guardam entre si algum tipo de relação que não inclui subordinação ou superordenação são ditos *termos relacionados*, esta designação inclui também os conceitos coordenados, mas não apenas estes (WRIGHT; BUDIN, 1997, p. 337-338).

Outro tipo de relação descrita pelos sistemas de conceitos é a relação partitiva, a que os autores também denominam todo-parte ou meronímica e que descreve a condição em que “o objeto x é parte do objeto y”. Os autores argumentam que enquanto as relações lógicas referem-se a conceitos, as relações partitivas lidam com objetos. Embora haja autores que utilizam os termos *superordenado*, *subordinado* e *coordenado* para expressarem as relações hierárquicas no contexto partitivo, outros preferem o uso de conceito inclusivo, para a designação do todo e

conceito partitivo, para a designação da parte. São utilizados também os termos *termo amplo* e *termo específico* (WRIGHT; BUDIN, 1997, p. 337-338).

Relações seqüenciais, na visão desses autores, são aquelas que ocorrem quando os conceitos existem em uma seqüência temporal – estações do ano: primavera, verão, outono, inverno –; ou em suma seqüência espacial – organização das teclas de um piano ou de um teclado de computador. Em uma linha de montagem de fábrica, por exemplo, os processos podem ocorrer em seqüências temporais, mas também podem ser organizados sob determinado arranjo espacial. As relações seqüenciais são, em geral, não-hierárquicas.

Relações associativas, também chamadas de pragmáticas ou temáticas, são as mais difíceis de se descrever e são objeto de diversos tipos de opiniões. *Carrros* podem possuir relações associativas com diversos tipos de conceitos como *auto-estradas*, *garagens*, e *passageiros*; ferramentas estão associadas com profissões – martelo x carpinteiro; profissões estão associadas com locais – professores x escolas; juízes x cortes; atletas x ginásios. Relações associativas podem ser analisadas de acordo com uma grande diversidade de perspectivas, dependendo dos requisitos de dado campo conceitual (WRIGHT; BUDIN, 1997, p. 338).

Para Van Campenhoudt (1997), um dos principais méritos de Wüster foi ter demonstrado a importância de se levar em conta as ligações de sentido que unem os conceitos. O autor destaca que essas ligações são denominadas ligações nocionais ou relações nocionais (relações conceituais) e podem ser classificadas em lógicas; ontológicas; verticais; horizontais; e diagonais. A escola de Viena adotou uma distinção estabelecida por Wüster entre as relações lógicas e ontológicas. De acordo com essa distinção, uma relação lógica é genérica e abstrata; enquanto uma relação ontológica é partitiva e concreta. Uma é fundada nas semelhanças e outra na contigüidade, na justaposição, nas relações de tempo-espço e de causa-efeito. O tipo mais importante de relação ontológica é a relação partitiva, isto é, a relação entre o todo e suas partes.

Nesse quadro de uma relação conceitual fundamentada sobre uma subordinação lógica ou ontológica, a relação entre o conceito subordinado e o subordinante é denominada vertical. Se os conceitos situam-se no mesmo nível, a relação é classificada como horizontal e os conceitos são denominados coordenados. Wüster acrescenta um tipo de relação, a que denomina diagonal, característica de relações que não se enquadram nem como coordenadas nem como subordinadas. Para ilustrar, no domínio marítimo, por exemplo, pode-se dizer que os conceitos de *escuna de três* ou *quatro mastros* dependem, ambos, de um conceito genérico *embarcação*, o que estabelece uma *relação diagonal* entre esses conceitos (ilustração 11).

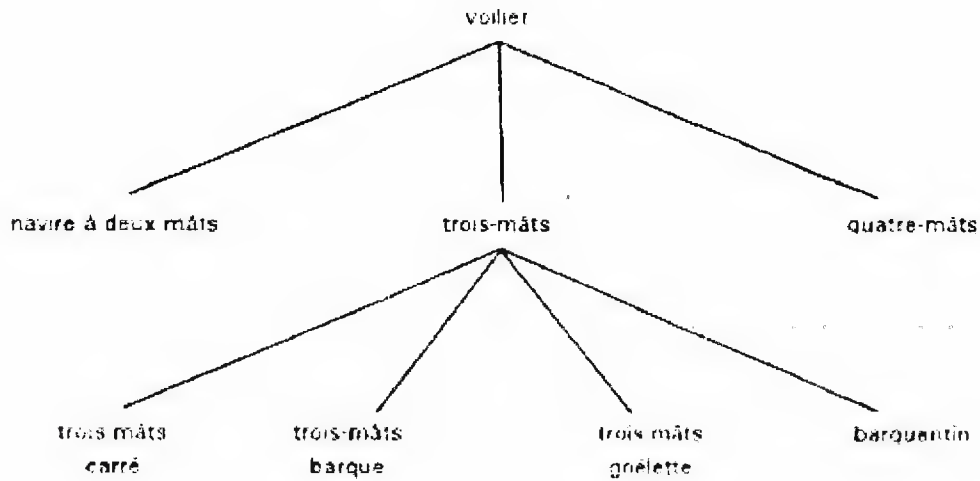


Ilustração 11: Exemplo de uma relação diagonal

2.1.9.1 Sistema de conceitos e relações conceituais segundo DAHLBERG

Para Dahlberg (1978) há dois tipos de abordagens para a identificação das relações conceituais: a abordagem quantitativa, na qual se observa as características de cada conceito, levando-se em conta apenas o aspecto lógico-formal; e a abordagem qualitativa, na qual tais características são analisadas levando-se em conta não apenas o aspecto lógico-formal, mas também o aspecto material.

Na abordagem quantitativa, os conceitos são considerados comparando-se logicamente as suas características, sob aspectos como identidade, inclusão, intersecção e disjunção. A tabela 8 explicita cada tipo de relação quantitativa.

Tabela 8: comparação de conceitos pela abordagem lógico-formal

relação	características do conceito A	características do conceito B
Identidade ($A = B$)	● ● ●	● ● ●
Inclusão ($A \subset B$)	● ●	● ● △
Intersecção ($A \cap B$)	● ● △	● △ △
disjunção ($A \notin B$)	● ● ●	△ △ △

Fonte: (DAHLBERG, 1978)

Na abordagem qualitativa, utiliza-se a categorização das características para, de acordo com três aspectos possíveis, realizar-se a comparação entre os conceitos. As relações são agrupadas da seguinte forma:

- a) relações formais ou categoriais;

b) relações material-paradigmáticas, que podem ser relações hierárquicas, relações partitivas ou relações de oposição;

c) relações funcional-sintagmáticas.

As relações formais ou categoriais estão diretamente ligadas à natureza do item em referência, que pode ser qualificado como exibido na tabela 9. O agrupamento realizado dessa maneira pode ser denominado de classificação por facetas.

Tabela 9: Classificação de categoria e seus constituintes

categoria	constituintes
entidades	fenômeno
	objeto imaterial geral objeto material
propriedades	quantidade
	qualidade
	comparação
atividades	operação
	estado
	processo
dimensões	período no tempo
	posição
	localização no espaço

Fonte: (Dahlberg, 1978)

As relações material-paradigmáticas podem ser hierárquicas, partitivas ou de oposição. As Relações hierárquicas ocorrem entre dois conceitos que possuem basicamente as mesmas características (idênticas), mas um desses conceitos possui uma ou mais características que o diferenciam do outro. Podem ocorrer relações hierárquicas do tipo gênero-espécie (abstração-especificação), que determinam um conceito amplo (genérico) e um conceito restrito (específico). Quando dois ou mais conceitos específicos possuem o mesmo conjunto de características contido em um conceito mais amplo, mas forem diferentes entre si por características que não determinam uma relação hierárquica, sua relação é definida como relação de contigüidade. As relações partitivas compreendem as características que designam o objeto formal e suas partes constituintes; desta maneira, o conceito de inteiro inclui suas características e também os conceitos de todas as partes que o integram. Há pelo menos quatro tipos de relações partitivas: partição em relação a um sistema natural (organismos vivos); partição em relação a um sistema artificial (objetos fabricados); partição em relação a uma organização humana (países, regiões, etc.) e partição em relação a um dado campo ou assunto (organização do conhecimento). As relações de oposição ocorrem quando as características dos conceitos comparados são ditas opostas. Podem ser de três tipos: - oposição contraditória: numérico – não-numérico; presença – ausência; - oposição contrária: amizade – inimizade; preto – branco; - positividade – indiferença – negatividade: favorável – neutro – desfavorável.

As relações funcional-sintagmáticas dependem das valências semânticas dos conceitos analisados e estão afetas à análise de tais conceitos em sua relação com processos, como por exemplo operações, usualmente expressas por verbos ou por derivações verbais que podem representar categorias como: instrumentalidade, condição co-ocorrência, lugar, causalidade, modalidade, resultado, temporalidade, finalidade, potencialidade, entre outras.

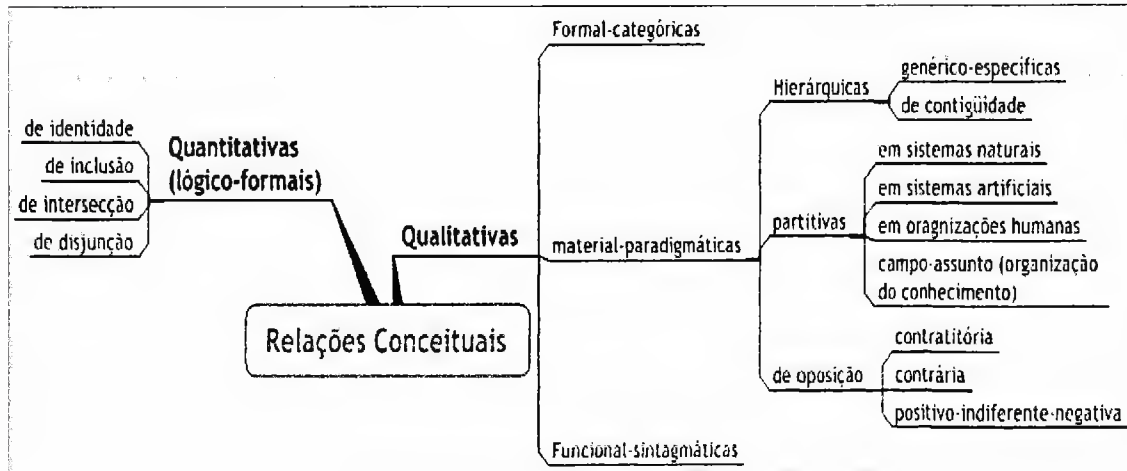


Ilustração 12: Relações conceituais segundo Dahlberg
Fonte: adaptado de Dahlberg (1978)

2.1.9.2 Sistema de conceitos e relações conceituais segundo GORSKI

Outro autor importante no campo da lógica formal menor é Gorski. Esse autor propõe a diferenciação dos conceitos baseando-se no seu conteúdo e na sua extensão. No seu ponto de vista, é impossível relacionar dois conceitos acerca de cujos conteúdos não se possa conceber alguma coisa em comum, uma vez que todos os objetos e fenômenos existem na realidade, não isoladamente, mas relacionados entre si de diversas maneiras. Conceitos que possuam traços semelhantes e que, portanto, pertencem a um mesmo campo de realidades, são denominados conceitos comparáveis. Conceitos em cujos objetos há mais características diferenciais do que comuns, uma vez que pertencem a setores ou campos de realidades diferentes, são ditos conceitos incomparáveis; por exemplo: “felicidade” e “raiz quadrada” (GORSKI, 1968).

Conceitos comparáveis podem ser categorizados em compatíveis, para denominar aqueles que, ainda que possuam conteúdos distintos, não excluem a possibilidade de coincidência ou de coordenação entre suas extensões, mesmo que isso se dê apenas parcialmente; e incompatíveis, para denominar os conceitos cujos conteúdos são muito diversos e cujas extensões não coincidem nem mesmo parcialmente.

Conceitos compatíveis são classificados em idênticos, mistos e conceitos com relação de subordinação. Conceitos idênticos - dois conceitos são considerados idênticos quando, mesmo possuindo conteúdo distinto, composto por características diferentes, possuem a mesma extensão. Conceitos mistos - são denominados mistos os conceitos que, embora possuam conteúdo diferente, possuem extensões parcialmente coincidentes. Isto significa que um objeto que pertence ao campo de um desses conceitos pertencerá ao mesmo tempo ao campo do outro conceito. Conceitos com relação de subordinação - existe uma relação de subordinação quando as características essenciais de um conceito constituem apenas uma parte das características essenciais de outro conceito, o qual possui, além de tais características, outras características que o diferenciam; a extensão do segundo conceito recai por completo no campo de primeiro conceito, como parte dele mesmo. São exemplos: “publicação periódica” e “revista”; “substância química” e “ácido”. Assim, o conceito mais amplo (de maior extensão) é denominado subordinante e o mais restrito (de menor extensão) é denominado subordinado. Quando a relação se dá entre dois conceitos universais, chama-se gênero (conceito genérico) o conceito subordinante e espécie o conceito subordinado (conceito específico). Um mesmo conceito considerado específico em uma dada relação pode ser um conceito genérico, quando considerado em relação a um terceiro conceito. Quando a relação de subordinação ocorre entre um conceito subordinante universal e um subordinado singular, o primeiro é denominado espécie e o segundo indivíduo.

Conceitos incompatíveis, aqueles cujas extensões não possuem nada em comum, são categorizados em conceitos contraditórios e contrários. Conceitos contraditórios - dois conceitos são ditos contraditórios quando em um deles se concebe uma certa característica do objeto que falta no outro conceito, de tal maneira que esta ausência constitui o caráter do objeto concebido pelo segundo conceito; um desses conceitos faz referência a um grupo de características essenciais; no outro, é a ausência desse mesmo grupo de características essenciais que constitui o caráter do objeto. São exemplos de conceitos contraditórios: “fumante” e “não-fumante”; “empregado” e “desempregado”. Conceitos contrários - dois conceitos são ditos contrários quando um deles não apenas nega o outro mas, além disso, possui características que se contrapõem às características do conceito negado; por exemplo: “saúde” e “doença”; “preto” e “branco”. Entre dois conceitos contrários pode haver conceitos intermediários de tal contraposição, constituindo uma graduação de variações (conceitos de transição). Por exemplo, entre os conceitos preto e branco existem diferentes graduações de cores que levam de um extremo a outro.

Dois conceitos são ditos coordenados quando pertencem a diferentes espécies de um mesmo gênero. Conceitos coordenados também podem ser categorizados como compatíveis ou incompatíveis.

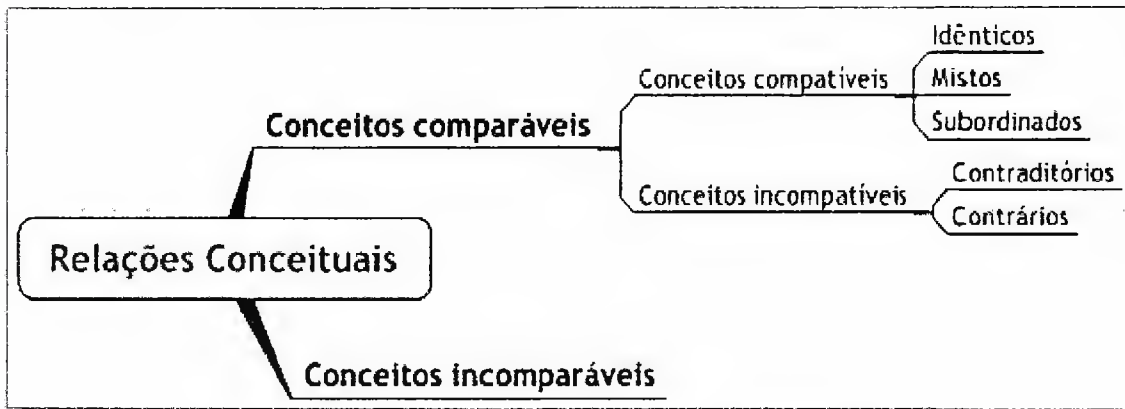


Ilustração 13: Relações conceituais segundo Gorski
Fonte: adaptado de Gorski (1968)

2.1.9.3 Sistemas de Conceitos e Relações Conceituais segundo CABRÉ

Cabré (1993) considera a existência de dois tipos de relações entre conceitos: as relações lógicas e as relações ontológicas.

2.1.9.3.1 Relações Lógicas

As relações lógicas são estabelecidas entre conceitos que apresentam semelhanças pelo fato de possuírem uma ou mais características em comum. A compreensão de tais semelhanças pode ser diferente, de acordo com o método de análise, que normalmente compreende a *comparação* ou a *combinação*.

A *comparação* entre conceitos pode fornecer dois tipos possíveis de inter-relação: a *hiponímia lógica* e a *coordenação lógica*. Nas relações de *hiponímia lógica*, um conceito mais amplo (hiperônimo) cede todas as suas características a um conceito mais específico (hipônimo). Este conceito específico possui, portanto, além de todas as características do conceito genérico, algumas que o particularizam em relação àquele. Uma relação de *coordenação lógica* (co-hiponímia) ocorre entre dois conceitos específicos de um mesmo conceito genérico. Neste caso, ambos compreendem as mesmas características do conceito genérico, mas cada um possui características particulares que os tornam distintos entre si.

A *combinação* entre conceitos introduz a noção de campo semântico hierárquico, que pode conter uma *série vertical* de conceitos – que mantêm uma relação de inclusão entre si – e uma *série horizontal* – na qual os conceitos são de mesmo nível e mantêm relações de oposição. Nesta abordagem, as *relações hierárquicas* mais importantes são as de *inclusão*, que compreendem as

relações genérico-específicas. Nas relações *gênero-espécie* os conceitos mais específicos fazem parte da extensão do conceito genérico que os inclui, isto é, herdam todas as características daquele conceito; além das características do conceito mais amplo, o conceito específico possui alguma que lhe é peculiar.

Em alguns casos a relação entre dois conceitos não é estabelecida a partir da combinação de suas características, mas da combinação de tais *conceitos simples* para formarem um terceiro conceito denominado *complexo*. Um *conceito complexo*, portanto, é formado pela combinação de dois ou mais conjuntos estruturados de características que individualmente correspondem a conceitos simples.

A combinação de conceitos simples, quando o conceito complexo resultante não compreende nem relações de hiponímia e nem de co-hiponímia, pode resultar relações de *determinação*, de *conjunção* ou de *disjunção*.

a) *determinação* – um dos conceitos simples é determinado pelo outro, isto é, tem sua intensão modificada, resultando um terceiro conceito complexo; exemplo: motor e diesel: motor a diesel.

b) *conjunção* – o terceiro conceito só existe se os dois outros conceitos de partida forem combinados e forma um conceito específico em relação aos dois conceitos de origem; exemplo: calça plástica (é um tipo de calça e um tipo de fralda ao mesmo tempo).

c) *disjunção* – o terceiro conceito é formado a partir dos traços diferenciadores de cada um dos conceitos geradores; portanto não é uma especificidade de nenhum deles. Por exemplo: indu e europeu: indoeuropeu.

Os diversos tipos de relações lógicas podem ser simbolizados graficamente, conforme a tabela 10

Tabela 10: Simbologia de relações conceituais segundo Cabré (1993).

Tipo de relação ou combinação conceitual	Símbolo	
relações (hiponímia, co-hiponímia, hierárquicas)	o mesmo sentido que	=
	sentido parecido a	≈
	genérico de	>
	específico de	<
combinações (conjunção, determinação, disjunção)	conjunção de conceitos	∧
	disjunção de conceitos	∨
	determinado por	+

Fonte: Cabré (1993)

2.1.9.3.2 Relações Ontológicas

As relações ontológicas interessam-se pela natureza dos seres da realidade e pelas interações que se estabelecem entre eles. Tais relações não se preocupam com a similaridade da intensão dos conceitos, mas pela proximidade situacional dos elementos na realidade, isto é, pelo contato entre eles. As relações ontológicas classificam-se em relações de *coordenação* e de *encadeamento*. As relações de *coordenação* descrevem o relacionamento entre um todo e suas partes e entre as distintas partes de um mesmo todo. Essas relações de *coordenação* baseiam-se na contigüidade dos objetos no espaço e são denominadas *partitivas*. As relações de *encadeamento* referem-se à sucessão dos objetos no tempo, isto é, a eventos de *causa-efeito*. Dito de outro modo, são relações *seqüenciais*.

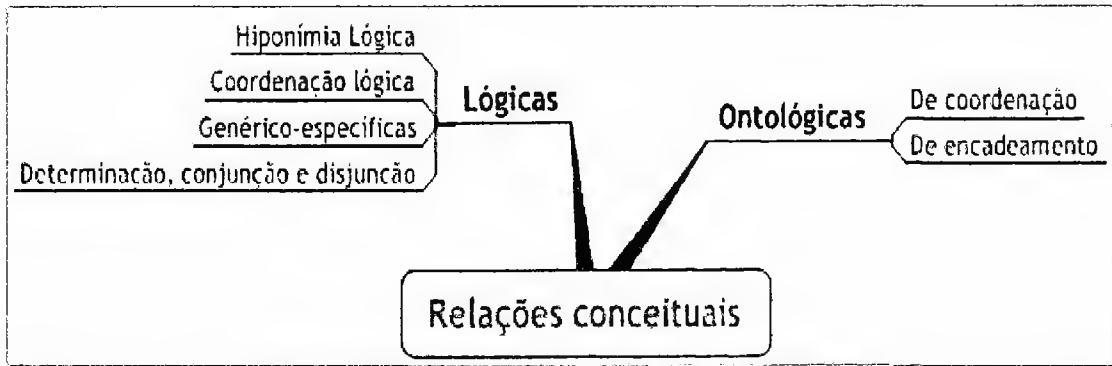


Ilustração 14: Relações conceituais segundo Cabré (1993)

2.1.9.4 Sistemas de Conceitos e Relações Conceituais segundo FELBER

Uma vez que os conceitos são formados por características, eles possuem relacionamentos com outros conceitos que possuam suas mesmas características, ou características similares. Conceitos podem ter também relações indiretas com outros conceitos, se os objetos individuais que representam forem vizinhos no espaço ou seguirem um ao outro no tempo. Tais relações ocorrem em uma subseção de um campo de assuntos somente acessível mentalmente; tal campo de assuntos estruturado é denominado sistema de conceitos. O autor admite os seguintes tipos de relações entre conceitos: *Lógicas*, *Ontológicas* e *Relações de Efeito*. A tabela 11 apresenta os símbolos internacionalmente utilizados para a representação dos diversos tipos de relações lógicas e ontológicas.

Tabela 11: simbologia da representação de relações de acordo com FELBER

Símbolos usados para exprimir relações lógicas (L) e ontológicas (O)				
Tipo	Subtipo	Simbologia		referencial teórico
		gráfica	Literal	
	Equivalência	=		ISO 1951
	significado diferente	≠		ISO 1951
	significado similar	≈		ISO 1951
L	intersecção lógica	x	IGR	ISO 1951
L	subordinação lógica (menor intensão) (maior intensão)	> ou <	BG ou NG	ISO 1951
O	Parte	>-		ISO 1951
O	Todo	-<		ISO 1951
L	conceito coordenado (relação lógica)		HGR	FELBER
O	coordenação partitiva	-	HPR	FELBER
L	relação diagonal lógica	/_ ou _\	DGR + ou DGR -	FELBER
L	conjunção lógica	^		FELBER
L	determinação lógica	+		FELBER
L	disjunção lógica	v		FELBER
O	subordinação partitiva	>- ou -<	BP ou NP	ISO 1951
O	intersecção partitiva	>-<	IPR	ISO 1951
O	relação diagonal partitiva	> ou <	DPR + ou DPR -	

Legenda das abreviaturas

BG	Broader Concept Generic
NG	Narrower Concept Generic
IGR	Intersecting Generic Relationship
HGR	Horizontal Generic Relationship
DGR	Diagonal Generic Relationship
BP	Broader Concept Partitive
NP	Narrower Concept Partitive
IPR	Intersecting Partitive Relationship
DPR	Diagonal Partitive Relationship

2.1.9.4.1 Relações Lógicas

As relações lógicas entre conceitos são relações de similaridade. Elas são também chamadas de relações *genéricas* ou *abstratas*. Conceitos que possuem relações de similaridade possuem certas características em comum, ou seja, sua certa parte de sua intensão é comum. As relações de comparação podem ser classificadas em relações de *subordinação*, *intersecção*, *coordenação* e relação *lógica diagonal*. Uma relação de *subordinação lógica* ocorre quando um conceito (espécie) possui todas as características de outro (gênero) e pelo menos uma característica adicional. Ocorre uma relação vertical entre o conceito *superordenado* e o seu *subordinado*. A relação de *intersecção lógica* ocorre quando um conceito possui apenas parte da intensão de outro, isto é, nem todas as características são as mesmas. A relação de *coordenação lógica* se dá entre dois conceitos subordinados a um mesmo conceito superordenado. Neste caso suas intensões são diferenciadas por pelo menos uma característica, mas possuem as mesmas características do conceito gênero.

Uma relação *diagonal lógica* ocorre em uma relação genérico-específica, quando dois conceitos não estão relacionados nem por coordenação e nem por subordinação.

Quando três ou mais conceitos são comparados levando-se em conta a sua similaridade, três tipos de relações poderão existir: relações de série *vertical lógica* – em que a subordinação lógica dos conceitos forma uma série vertical; relações de série *horizontal lógica* – em que a coordenação lógica dos conceitos forma uma série horizontal; e, por fim, relações de *ligação lógica* – em que dois dos três conceitos comparados são coordenados, enquanto o terceiro é superordenado em relação a esses dois. As relações de ligação lógica podem ser de:

a) *determinação* – ocorre quando um segundo conceito é integrado como característica do primeiro; deste modo a intensão do primeiro conceito é aumentada por pelo menos uma característica adicional; o conceito resultante é uma espécie do primeiro conceito;

b) *conjunção* – ocorre quando se dá a integração das intensões de dois conceitos membros; o conceito resultante será então uma espécie de cada um dos conceitos originais;

c) *disjunção* – ocorre quando se dá a integração das extensões de dois conceitos; o conceito formado passa a ser um superordenado em relação aos conceitos originais.

2.1.9.4.2 Relações Ontológicas

As relações ontológicas são relações indiretas entre os conceitos e são caracterizadas pela contigüidade ou justaposição de conceitos, no espaço ou no tempo, ou por conexões de causa e efeito. Entre as relações ontológicas, o tipo mais importante é a relação *partitiva*. As relações *partitivas* são categorizadas por esse autor em relações de *subordinação*, *interseção*, *coordenação*, *relação diagonal*, *séries verticais*, *séries horizontais* e *ligação partitiva*. Em uma relação de *subordinação partitiva*, se um objeto individual consiste das mesmas partes de outro objeto individual, mais pelo menos uma parte, então aquele objeto (todo) compreende este objeto como membro (parte). Uma relação de *interseção partitiva* ocorre entre dois objetos que possuem algumas partes em comum. Uma relação de *coordenação partitiva* ocorre entre dois objetos que representam partes de um mesmo todo. Em uma relação partitiva, quando dos objetos forem partes do mesmo inteiro, mas não estiverem relacionados nem por subordinação nem por coordenação, estarão em uma *relação diagonal*.

Quando três ou mais conceitos são comparados levando-se em conta as suas relações partitivas, haverá três possibilidades de relações: *série partitiva vertical* – em que a subordinação partitiva dos conceitos forma uma série vertical; *série partitiva horizontal* – em que a coordenação partitiva dos conceitos forma uma série horizontal; e a relação de *ligação partitiva* –

que ocorre a partir da integração de dois ou mais objetos em uma nova entidade. A ligação ontológica não combina dois ou mais conceitos, mas dois ou mais objetos individuais que pertencem a esses conceitos.

Diz-se que ocorre uma *relação de sucessão*, quando dois conceitos que encontram-se ordenados por *contigüidade*, no tempo. Relações *material-produto* referem-se aos diferentes estágios em um processo de produção, partindo da matéria-prima ao produto final. As *relações de efeito* envolvem fenômenos como *causalidade*, *instrumentação*, *descendência*. As relações de *descendência* o autor classifica como descendência *genealógica*, descendência *ontológica* e descendência *entre estágios de substâncias*.

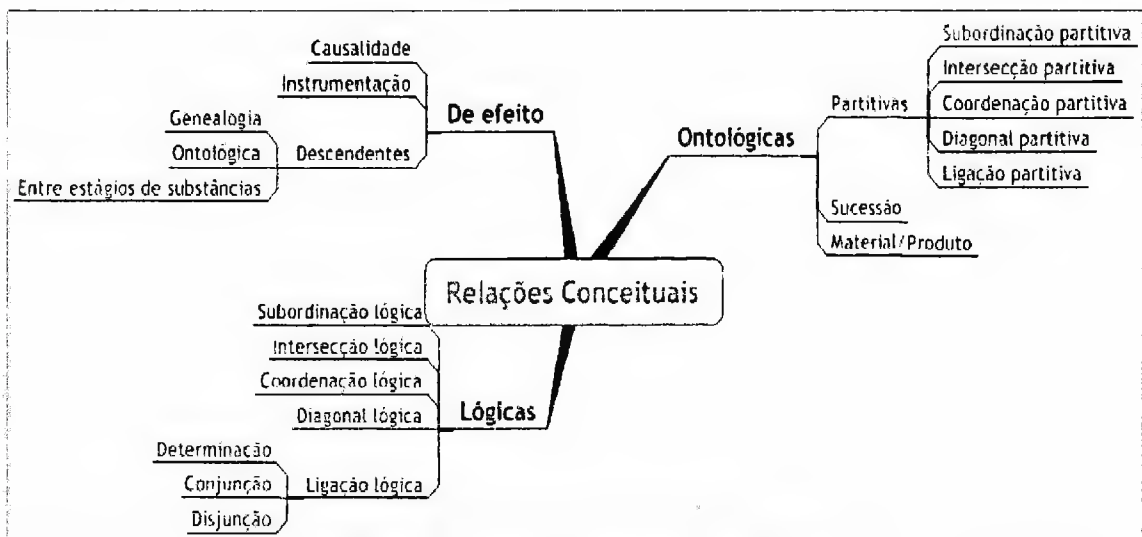


Ilustração 15: Relações conceituais segundo Felber
Fonte: Adaptado de Felber (1984).

2.1.9.5 Sistemas de Conceitos e Relações Conceituais segundo PICHT & ARNTZ

Os conceitos de um dado campo do conhecimento não devem ser estudados isoladamente, uma vez que possuem relações entre si e que ocorrem sistematicamente. Um sistema de conceitos é um conjunto de conceitos entre os quais existem relações, formando-se, assim, um todo coerente. Trata-se, por conseguinte, de um conjunto estruturado de conceitos, estabelecido de acordo com as relações existentes entre os mesmos, no qual cada conceito possui uma posição determinada no sistema.

Alguns princípios devem ser seguidos para a elaboração de um sistema de conceitos:

a) univocidade – a representação deve refletir de modo claro e inequívoco as relações e os critérios de classificação;

b) facilidade de compreensão – a representação deve ser compreensível para os usuários aos quais destina-se o sistema e deve moldar-se às características de tal perfil de usuários;

c) transparência – a representação deve ser amigável para o usuário; isto pode ser alcançado mediante a decomposição dos sistemas em conceitos complexos em sistemas parciais, por níveis hierárquicos;

d) possibilidade de aplicação – o sistema deve ser organizado de modo flexível e aberto à incorporação de modificações, quando necessário.

Um conjunto de conceitos de uma área especializada e entre os quais existe uma relação sistemática é denominado campo de conceitos ou grupo de conceitos que estão tematicamente relacionados. Os autores consideram a existência de *relações hierárquicas* e *não hierárquicas*.

Como já relatado, Picht (1984, p. 187) admite relações lógicas e relações ontológicas. Para o autor, as relações lógicas baseiam-se na semelhança das noções. Estas constituem um conjunto espécie-gênero. As relações ontológicas baseiam-se na contigüidade, isto é, no contato – no espaço ou no tempo – dos elementos que representam as noções. Os conjuntos ontológicos mais importantes são o conjunto todo-parte e o conjunto de evolução (por exemplo, o quadro genealógico de um animal ou o conjunto de formas sucessivas tomadas por um produto ou por uma língua no curso de sua história). O autor classifica as relações lógicas em: a) Identidade – as características dos conceitos são idênticas; b) Implicação – um conceito inclui as características do outro; c) Intersecção – um conceito possui parte das características do outro; d) Disjunção – os conceitos excluem um ao outro, ao mesmo tempo: exemplo garoto x garota => adolescente; e) Negação – relação de antonímia. Para o autor, de todas essas relações, as mais importantes são as de identidade e as de implicação, pois delas originam-se as relações de subordinação (gênero-espécie) e as relações de coordenação (relação entre espécies do mesmo gênero).

Como foi mencionado, na teoria vienense, um conceito compõe-se a partir de um conjunto de características que são propriedades dos objetos conceitualizados, permitindo diferenciar ou confrontar cada conceito. Um conceito específico é aquele que possui pelo menos uma característica de um conceito genérico ou abstrato. As relações entre conceitos genéricos e conceitos específicos são denominadas relações lógicas. Nas relações todo-parte um conceito é parte de outro. Alguns exemplos, desse tipo de relação são citados pelo autor: Bioquímica (refere-se à biologia e química relacionadas por uma intersecção partitiva); casal (ligação partitiva entre

um homem e uma mulher); muro (soma de suas partes: tijolos; ligação partitiva); madeira – armário (relação entre material-produto).

Diversos conceitos podem ser utilizados para a análise das relações coordenadas:

- Cadeia – caracteriza-se por uma ordenação seqüencial única. Ex. nascimento, infância, adolescência, idade adulta, velhice, morte.
- Hélice – a organização dos conceitos é helicoidal, as noções sucedem-se de maneira circular, de tal modo que, ao chegar-se à unidade máxima, volta-se para a unidade mínima, embora progredindo-se no tempo ou no espaço. Ex. janeiro, fevereiro, março.... dezembro, janeiro...
- Ciclo – são cadeias circulares sem progressão. Ex.: norte, sul, leste, oeste
- Escala – especifica uma relação espaço-temporal.
- Antonímia – os conceitos são opostos, em função de uma mesma característica diferenciadora; Ex. interno-externo; direto-indireto; horizontal-vertical, etc.

As relações conceituais não se limitam apenas às coordenadas e hierárquicas, uma vez que essas relações não podem dar conta de todos os campos do saber. Pierre Lerat²³ interessou-se pelas relações semânticas entre os conceitos, às quais denominou relações funcionais. Exemplos de verbos que representam relações funcionais: serve para, previne de, aparece quando, entre outras.

2.1.9.5.1 Relações Hierárquicas

Nesse tipo de relações, parte-se da combinação de conceitos originais para a formação de um novo conceito. As relações hierárquicas são categorizadas em relações de *abstração* (*lógicas* ou *genéricas*) e relações *ontológicas* (*partitivas* e *relações de integração*). As relações de *abstração*, por seu turno, podem ser classificadas em relações de *determinação*, relações de *conjunção* e relações de *disjunção*.

²³ LERAT, P. "L'hyponymie dans la structuration des terminologies", in *Langages*, nº 98, p. 79-86, 1990. ; LERAT, P. "La représentation des connaissances en terminologie assistée", in CLAS, A; BOUILLON, P. ed., T.A.-T.A.O. : recherches de pointe et applications immédiates. Troisièmes journées scientifiques du réseau thématique de recherche "Lexicologie, terminologie et traduction". Montréal, 30 septembre, 1er et 2 octobre 1993, Montréal, AUPELF-UREF et Beyrouth, F.M.A., p. 371-377, 1994.

Uma relação de *abstração* ocorre entre um conceito *superordenado* e o seu correspondente *subordinado*. Conceitos de mesmo nível de abstração relacionam-se por meio da *coordenação*. Os conceitos coordenados possuem o mesmo número de características do conceito superordenado, possuindo pelo menos mais uma que os distingue entre si. As relações de abstração podem ser representadas por meio da fórmula genérica:

Conceito *superordenado*: A

Conceito *subordinado*: A + C onde C = característica específica

Conceitos *coordenados*: (A + C₁) (A + C₂) (A + C₃) (A + C_n)

Entre o conceito superordenado e os conceitos subordinados ocorre uma *relação vertical*. Entre os conceitos coordenados ocorre uma *relação horizontal*. Um sistema é dito *mono hierárquico* quando se utiliza apenas um critério para a classificação dos conceitos subordinados de mesmo nível, em relação ao conceito superordenado. Quando se utiliza mais de um critério de classificação, o sistema é denominado *poliierárquico*. Os sistemas mono hierárquicos são mais transparentes, uma vez que apresentam menos possibilidades de classificação. Em um sistema mono hierárquico, os conceitos coordenados formam uma série monodimensional; já em um sistema poliierárquico, os conceitos coordenados formam uma série polidimensional.

As *relações ontológicas* baseiam-se na relação entre o todo e suas partes. Por esse motivo são denominadas relações todo-parte ou relações partitivas. O conceito de ordem superior é denominado inclusivo e o conceito subordinado é denominado conceito específico parcial. Como no caso das relações lógicas, as relações ontológicas possuem séries verticais e horizontais. A escolha da divisão do todo em suas partes constituintes recai sobre o objetivo da classificação e sobre a área de conhecimento específica. Assim, o mesmo conceito inclusivo pode ser dividido em partes diferentes, de acordo com os objetivos da divisão, resultando em classificações diferentes.

É possível se decompor um conceito superior em diversos níveis de detalhamento. A profundidade desses níveis dependerá dos objetivos da decomposição. Nas relações lógicas as características do conceito superordenado estão sempre presentes nos conceitos subordinados. Nas relações ontológicas, no entanto, esse fenômeno não ocorre. Nunca será possível, por exemplo, afirmar-se que um pistão (parte do motor) é um tipo de motor.

2.1.9.5.2 Relações não hierárquicas

As relações não hierárquicas são categorizadas em *relações sequenciais* e *relações pragmáticas*. As relações *sequenciais* são relações de contigüidade no tempo entre conceitos que

representam estados ou fases de dado transcurso, como por exemplo: processos de fabricação, processos ou procedimentos jurídicos, processos documentários, entre outros. Quando as fases ou subfases de um processo ocorrem de maneira seqüencial, diz-se que tais relações são consecutivas (série vertical); quando as fases ocorrem ao mesmo tempo, diz-se que tais relações são simultâneas (série horizontal). As relações seqüenciais podem variar de acordo com a área de conhecimento. A seguir são apresentados alguns exemplos de relações seqüenciais:

- a) *cronológicas* – em que há a coexistência ou sucessão no tempo;
- b) *causais* – em que há a ocorrência de causa e efeito;
- c) *genéticas* – em que se estabelece uma relação entre produtor e produto;
- d) *de produção* – em que se estabelece uma relação entre material e produto;
- e) *de transmissão* – em que se estabelece uma relação entre emissor e receptor;
- f) *instrumentais* – em que se estabelece uma relação entre instrumento e a utilização de tal instrumento;
- g) *funcionais* – em que se estabelece uma relação entre argumento e função;
- h) *de dependência* – em que se estabelece uma relação entre pupilo e tutor;
- i) *de obrigação jurídica* – em que se estabelece uma relação entre credor e devedor.

As relações *pragmáticas* baseiam-se na proximidade temática entre os conceitos, mas não podem ser classificadas como relações hierárquicas e nem como relações seqüenciais. No campo dos tesouros, tais relações são denominadas *relações associativas*.

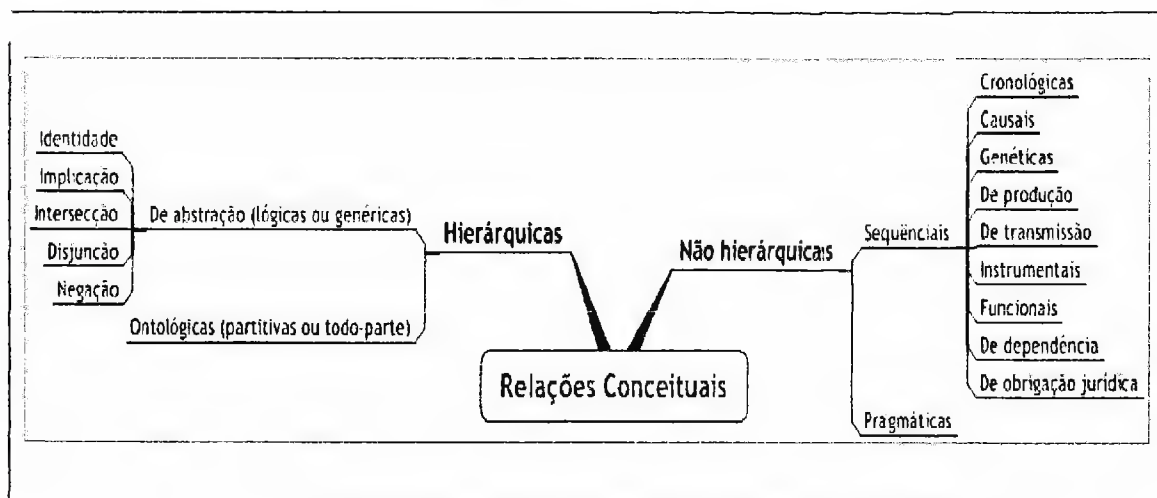


Ilustração 16: Relações conceituais segundo Picht & Arntz (1995)

2.1.9.6 Sistemas de Conceitos e Relações Conceituais segundo SAGER

Sager (1990) admite os seguintes tipos de relações entre conceitos: *genéricas*, *partitivas*, *polivalentes* e *complexas*.

2.1.9.6.1 Relações genéricas

Este tipo de relação estabelece uma ordem hierárquica, que identifica os conceitos como pertencentes à mesma categoria e na qual há um conceito amplo (genérico), denominado *superordenado* e um conceito mais restrito (específico), denominado *subordinado*. Os conceitos superordenados incluem os conceitos subordinados numa relação que não é reversível. Os conceitos subordinados possuem todas as características do conceito superordenado, bem como pelo menos mais uma característica distintiva. A relação genérica compreende tanto relações de hierarquia vertical, quanto horizontal e pode possuir diversas camadas, passíveis de representação por meio de um diagrama de árvore. Quanto mais baixo for o nível do conceito nessa estrutura, mais alto será o seu grau de especificidade e, conseqüentemente, mais restrita a sua intensão. O autor considera ainda a existência de relações *quase-genéricas*. Nesse tipo de relação, a atribuição de um tipo específico para um conceito é menos sólida, isto é, varia de acordo com a especialidade do assunto em que se inserem os conceitos; por exemplo: “algumas pessoas consideram que dente-de-leão é um vegetal; alguns vegetais são dentes-de-leão”.

Genericamente, as relações gênero-espécie podem ser expressas pela fórmula:

X é um *tipo* de A; ou A *possui um subtipo* X.

X, Y e Z são *tipos* de A; ou A *possui os conceitos específicos* X, Y e Z.

2.1.9.6.2 Relações Partitivas

Essas relações servem para indicar a conexão entre conceitos consistindo de mais de uma parte e suas partes constituintes. Sager sugere a fórmula para representar tais relações:

X é *parte* de A; ou A é *constituído por* X;

X, Y e Z são *partes* de A; ou A é *constituído por* X, Y e Z.

Algumas observações podem ser feitas em relação às relações partitivas:

- a) as partes são constituintes atômicos do todo;
- b) as partes são um conjunto finito;

- c) o inteiro constitui-se de vários grupos de partes numeradas ou não numeradas;
- d) a parte ou partes são constituintes opcionais do todo;
- e) as partes são constituintes e às vezes o próprio inteiro;
- f) em alguns casos, as partes são alternativas; dependendo da sua natureza, o inteiro possuirá a parte A ou a parte B, mas nunca ambas.

2.1.9.6.3 Relações Polivalentes

Alguns conceitos podem ocupar mais de uma posição em um sistema de conceitos, de acordo com suas relações em referência a outros conceitos do mesmo sistema. Nesses casos, diz-se que ocorre uma relação *poliierárquica (polivalente)*, isto é, um conceito é colocado em mais de uma hierarquia para um dado campo de assunto. Assim, sejam os conceitos B e C classificados como subtipos de A, haverá relacionamento poliierárquico se D for ao mesmo tempo subtipo de B e de C.

2.1.9.6.4 Relações Complexas

Quando não é possível a classificação direta de uma relação entre conceitos como sendo genérica ou partitiva, pode-se qualificá-la como uma relação do tipo *complexo*. Assim, podem-se criar subconjuntos de relações colocando-se os conceitos em classes conceituais de referência como, por exemplo, objetos, métodos, propriedades, qualidades, estados e processos. Dessa classificação resulta que o relacionamento entre dois conceitos pode se dar apenas em função de suas classes conceituais; por exemplo, a relação material-produto só poderá ocorrer entre duas entidades materiais. Surgem, dessa abordagem, padrões que mostram restrições na natureza das relações entre conceitos em virtude de suas categorias. Esse tipo de subdivisão de um conceito feito com base em características particulares é denominado classificação facetada, como uma analogia ao mesmo dispositivo utilizado na classificação de assuntos na área de Ciência da Informação.

As relações complexas podem ser, por exemplo:

- a) *relações causais* – causa e efeito;
- b) *relações materiais* – material e produto; material e propriedade; material e estado;
- c) *relações de processo* – processo e produto; processo e instrumento; processo e método; processo e paciente;
- d) *relações de objetos* – objeto e contra agente; objeto e contêiner; objeto e material; objeto e qualidade; objeto e operação; objeto e característica; objeto e forma; objeto e lugar;

e) *relações de fenômenos* – fenômeno e medida.

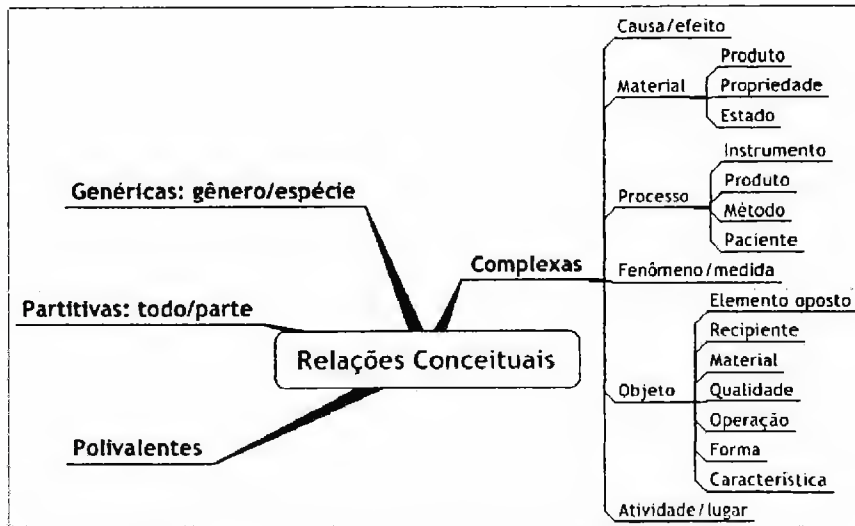


Ilustração 17: Relações conceituais segundo Sager (1990).

2.1.9.7 Relações Conceituais segundo a Normas ISO WD 704.1 e ISO 1087

A norma ISO WD 704.1 define sistema de conceito com um conjunto estruturado de conceitos, com base nas suas inter-relações. A posição de um conceito no sistema é definida pela sua intensão, ou seja, pelo conjunto das características que o definem. Um sistema de conceito destina-se a estruturar o conhecimento; tornar visíveis as relações entre os conceitos; fornecer uma base para a normalização de terminologia; auxiliar a análise comparativa de termos e conceitos em diferentes idiomas; e auxiliar a criação de definições. Essa norma classifica os sistemas de conceitos em: a) sistema genérico de conceitos – no qual todos os conceitos relacionam-se como genéricos ou específicos, em uma série vertical (relações genérico-específicas); b) sistema partitivo – no qual os conceitos relacionam-se em uma série vertical, por meio de relações partitivas; c) sistema associativo – no qual os conceitos relacionam-se com os outros por meio de associações; d) sistema misto – no qual os conceitos podem relacionar-se de acordo com diversos tipos de relações.

O desenvolvimento de um sistema de conceitos envolve operações como:

- seleção dos conceitos a serem tratados, de acordo com o campo conceitual;
- análise da intensão e da extensão de cada conceito;
- determinação das relações conceituais, que conferem a estrutura ao sistema;
- formulação das definições para os conceitos, com base em suas relações;
- atribuição de termos para cada conceito.

A norma ISO WD 704.1 classifica as relações entre os conceitos em relações *hierárquicas*, compreendendo as relações *genéricas* e *partitivas*, e em relações *associativas* (ilustração 41). Tais relações, como no caso de outros autores, são definidas com base na análise da intensão e da extensão de dois ou mais conceitos. Nas relações *hierárquicas* os conceitos são organizados em níveis, nos quais o conceito superordenado (subordinante) é subdividido em pelo menos um conceito subordinado (relação vertical). Aos conceitos subordinados de mesmo nível denomina-se coordenados (relação horizontal). Pode-se afirmar que uma relação entre dois conceitos é do tipo *genérico-específica* quando a intensão do conceito *subordinado* (conceito específico) inclui a intensão do conceito *superordenado* (conceito genérico) e pelo menos mais uma característica delimitadora. O critério de análise utilizado para a subdivisão de um conceito em conceitos subordinados é denominado dimensão; tal dimensão pode relacionar-se com diversos tipos de variáveis como: composição, função, área de utilização, entre outras. Quando é utilizada mais de uma dimensão para a análise o sistema de conceitos é denominado multidimensional. Pode-se afirmar que uma relação entre dois conceitos é do tipo *partitiva* quando o conceito *superordenado* (conceito abrangente) representa uma unidade inteira, enquanto os conceitos *subordinados* (conceitos partitivos) representam partes do todo. Uma ou mais partes do sistema podem ser essenciais ou não-essenciais. Para essa norma, pode-se dizer que há uma relação *associativa* entre conceitos quando for possível o estabelecimento de uma conexão temática entre tais conceitos ou quando se puder estabelecer uma dependência entre eles com base na sua proximidade no espaço ou no tempo. Relações *associativas* podem referir-se a: matéria-prima e produto; material e estado; matéria ou substância e propriedade; item concreto e forma; item concreto e material; ação e objeto; ação e lugar ou localização; ação e agente; causa e efeito; evento e tempo; entre outros fatores.

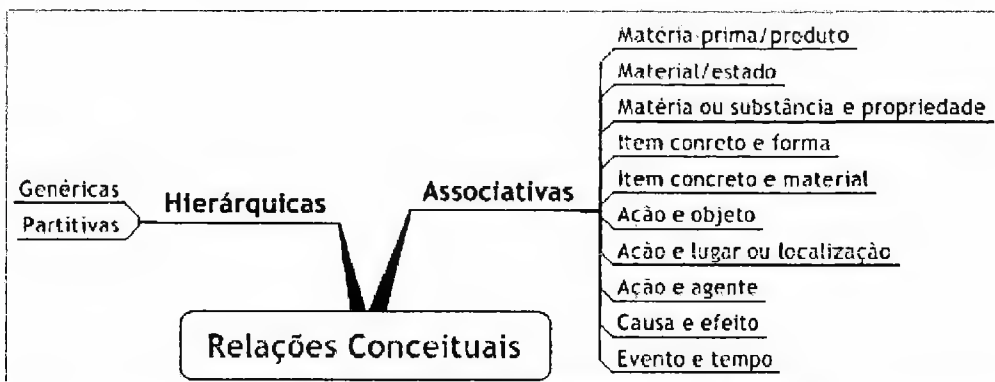


Ilustração 18: Relações conceituais, segundo a norma ISO 704.1

De acordo com a norma ISO 1087, os conceitos ocorrem sempre agrupados, formando um sistema de conceitos. Tal sistema é baseado em relações *lógicas* ou relações

ontológicas e é constituído por séries *verticais* ou *horizontais* de conceitos, ou pelo menos por uma dessas séries. Uma série corresponde a uma seqüência de conceitos relacionados, na qual cada conceito possui somente um predecessor imediato e um sucessor imediato. Desse modo, a série é um sistema linear e uma ou mais séries compreendem um sistema de conceitos. Em séries horizontais os conceitos são coordenados, uma vez que se encontram no mesmo grau de abstração ou divisão; neste caso, a diferenciação entre os conceitos se dá pela variação de uma característica. Em séries verticais os conceitos são subordinados uns aos outros, uma vez que se encontram em diferentes graus de abstração ou divisão. As relações *lógicas* estão baseadas na semelhança entre os conceitos e produzem tipicamente uma caracterização do tipo *genérico-específico*. As relações *ontológicas* são baseadas na contigüidade dos conceitos, isto é, no contato entre eles, que pode se dar no tempo ou no espaço. Os tipos de relações ontológicas são as *partitivas* (todo-parte) e as de *desenvolvimento* (por exemplo, uma árvore genealógica; uma árvore de produtos; uma árvore de uma linguagem).

As características que diferenciam os conceitos podem ser classificadas em características *intrínsecas*, que se referem a um objeto em si mesmo, não em relação a outros; em características *extrínsecas*, que pertencem a um objeto somente em suas relações com outro; e em características *equivalentes*, que referem-se à permutabilidade de características em uma dada intensão. Exemplos de características *intrínsecas* podem ser: forma, tamanho, material, cor, entre outras. As características extrínsecas são classificadas em características de *origem* e características de *propósito*. As primeiras indicam onde, por meio de quem ou como um objeto vem à existência, ao uso ou torna-se conhecido; são exemplos de características extrínsecas de origem de um objeto: descobridor, inventor, descritor, produtor, fornecedor, local de produção (cidade ou país), modo de produção, entre outras. As características de propósito indicam a finalidade a que dado objeto pode servir: modo de uso, campo de aplicação, posição em uma seqüência de montagem. As características *equivalentes* pressupõem uma substituição sem que haja mudança de sentido na definição do conceito e são causadas por uma conexão ontológica acidental, não por equivalência lógica. Como exemplo a norma utiliza o conceito triângulo equilátero (que é sinônimo de triângulo equiângulo): tanto a característica equilátero (que possui lados iguais) quanto a característica equiangular (que possui ângulos iguais) podem ser utilizadas para a compreensão do conceito de triângulo equilátero. A ilustração 19 apresenta os tipos de relações conceituais, de acordo com a norma ISO 1087.

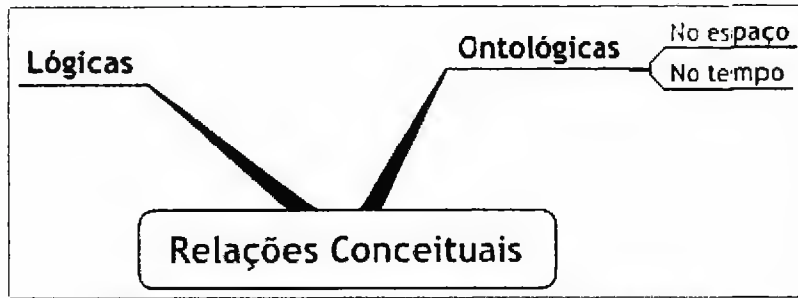


Ilustração 19: Relações conceituais segundo a norma ISO 1087
Fonte: (ISO1087)

2.1.10 Terminografia

Denomina-se Terminografia o processo de registro de dados terminológicos em suportes dos quais possam ser recuperados posteriormente em meios impressos e eletrônicos (ISO WD 704.1:1996). Segundo Nakayama (1990b), existem pelo menos dois tipos de obras terminográficas: as linguagens documentárias e os glossários. O glossário agrupa termos pertinentes a determinado domínio ou subdomínio do conhecimento, normalmente ordenados alfabeticamente e seguidos do seu enunciado. O enunciado é costumeiramente organizado em paradigmas: paradigma definicional, paradigma morfológico, paradigma gramatical, paradigma de sinonímia, paradigma etimológico, paradigma de pronúncia, paradigma de ilustração, paradigma de abonação, paradigma explicativo, paradigma referencial, e paradigma associativo. Cada verbete pode vir acompanhado de alguns desses paradigmas – não necessariamente todos –, cuja ordem também varia de acordo com a natureza da obra terminográfica. Conforme a sua estrutura, o glossário pode ter uma organização alfabética ou semântica – por grupos de campos semânticos ou nocionais. Com relação ao seu enunciado, os glossários podem ser simples – os termos possuem um enunciado simples, formado pelo mínimo de paradigmas possível – ou completos – os termos possuem em seu enunciado todos os paradigmas possíveis e necessários aos usuários.

A terminografia é a atividade de registro, processamento e apresentação de dados terminológicos, adquiridos em uma pesquisa terminológica (WRIGHT; BUDIN, 1997). A terminografia implica uma série de procedimentos e instrumentos, e entre eles a ficha terminológica, que normalmente são padronizados de acordo com os objetivos da atividade de registro. Uma ficha terminológica é um conjunto estruturado de informações sobre um termo (FAULSTICH, 1990). O arquivo de fichas terminográficas é a base da construção de um sistema de conceitos, uma vez que cada ficha contém não apenas as informações inerentes ao termo, mas também, com base na sua definição, as informações aos termos aos quais aquele se relaciona, como por exemplo: hiperônimos ou hipônimos; conceitos conexos e remissivas.

A ficha terminológica aplica-se como instrumento de entrada de dados em bancos terminológicos informatizados. Um banco terminológico é um banco de dados que contém informação sobre as unidades linguísticas (termos) que se usam especificamente em uma ou mais áreas do conhecimento, como unidades de denominação de conceitos, os quais organizam-se em conjuntos semântico-formais que designam um segmento de realidade em um domínio profissional. O banco terminológico é, portanto, um conjunto de dados formal e semanticamente organizados, de tal modo que sua recuperação possa ocorrer individualmente, por ficha terminológica, ou sistematicamente, por meio do cruzamento de informações organizadas em classes, categorias ou remissivas (FAUSLTICH, 1990). De acordo com Galinski e Nedobity (1988), a informação terminológica contém basicamente três tipos de dados diferentes: dados terminológicos (sobre termos: definição, sinônimos; definições; contexto e equivalentes em línguas estrangeiras; e informação adicional: número do conceito; símbolos de fonte e autoridade); dados bibliográficos (sobre literatura terminológica); e dados factuais (sobre a gestão do trabalho terminológico, especialistas, comitês, projetos, entre outras). Faulstich (1995), apresenta a seguinte estrutura para o registro terminográfico: a) número; b) entrada; c) categoria gramatical; d) gênero; e) sinônimo; f) variantes – gráfica (vários registros escritos ou orais em contextos diferentes) , lexical (os termos podem ser comutados, sem perda de sentido: software educativo x software educacional), morfossintática (o conceito não se altera pela alternância de elementos gramaticais: lombo-d’acém e lombinho-do-acém), socioprofissional (o conceito não se altera pela alternância de elementos gramaticais), topoletal (o conceito não se altera em função de mudança de região geográfica); g) área; h) subárea; i) definição ; j) contexto; k) remissivas – hiperônimo, hipônimo, conceito conexo; l) equivalentes – inglês, espanhol, francês; m) notas; n) autor da ficha; o) instituição; p) data (ilustração 13).

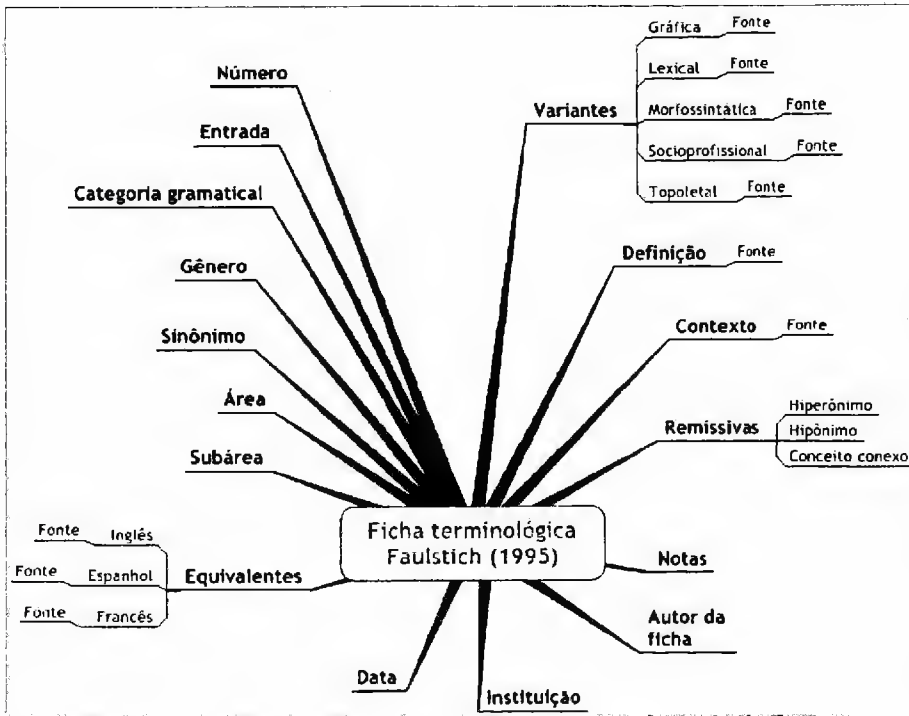


Ilustração 20: Formato terminográfico segundo Faulstich
 Fonte: adaptado de Faulstich (1995)

A grande vantagem da utilização de recursos informatizados para o gerenciamento dos dados terminológicos mantém suas bases na flexibilização dos formatos de saída. A manipulação informatizada de um banco de termos pode favorecer, por exemplo, a variação na organização da macroestrutura de uma obra terminológica, viabilizando o arranjo das entradas em ordem alfabética, ou por campo semântico e ordem alfabética ou, ainda, de acordo com suas relações conceituais como as de hiperonímia, as de hiponímia ou as associativas. Em relação à microestrutura da obra terminográfica, a fórmula de descrição da entrada pode fornecer diferentes maneiras para a recuperação de informações sobre um termo ou conjunto de termos, uma vez que tal fórmula normalmente corresponde a: *entrada = termo + enunciado lexicográfico*, onde *enunciado lexicográfico = Paradigma Informacional* (substantivo, sintagma) + *Paradigma Definição* (significado naquela língua de especialidade) + *Paradigma Pragmático* (abonações ou ilustrações) + *Paradigma lexical* (remissivas) (FAULSTICH, 1990). Cavalcanti (2000, p. 32) propõe microestruturas compostas pelos seguintes campos: entrada; variante; área; definição; fonte da definição; contexto; fonte do contexto; nota; fonte da nota; e remissiva.

Em termos normativos, no Brasil, segundo a norma NBR 13789:1997, o registro de dados terminológicos deve compreender pelo menos três categorias de dados (ilustração 14):

a) dados relativos ao termo: sinônimos; variantes ortográficas, morfológicas, sintáticas; forma abreviada; forma por extenso; símbolo; equivalente em língua estrangeira; grau de aceitabilidade – preferido, aceito, rejeitado, obsoleto, substituído –; informação gramatical, notas explicativas; antônimos;

b) dados relativos ao conceito: definição; contexto; outras representações – fórmula, figura, etc –; exemplos; notas;

c) dados administrativos: identificador do conceito; código do idioma; data de registro; identificação do responsável; fonte.

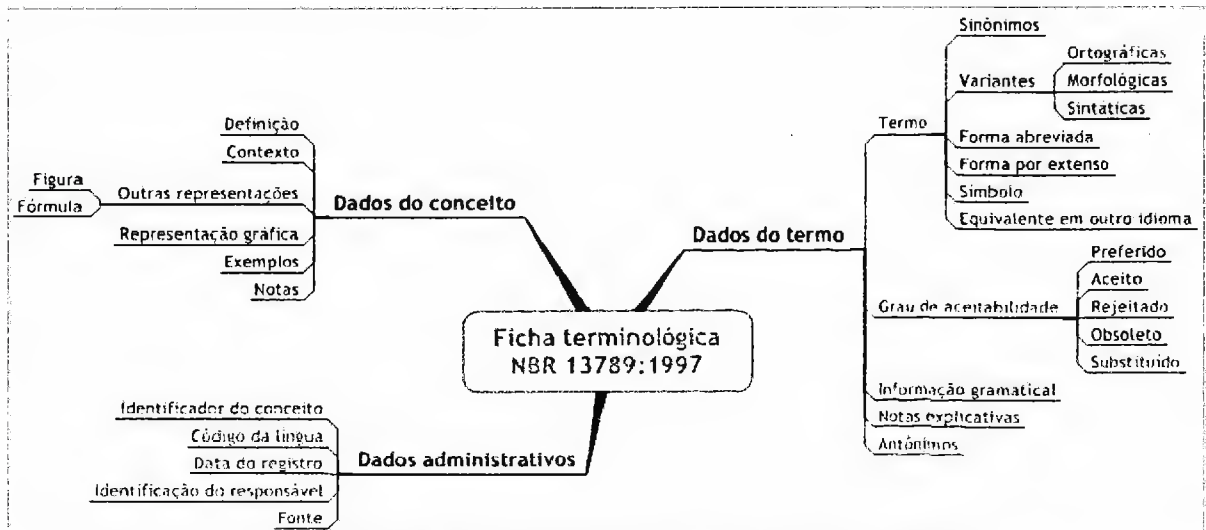


Ilustração 21: Formato terminográfico, segundo a norma NBR 13789
 Fonte: adaptado de Associação Brasileira de Normas Técnicas (1997a)

A utilização de um banco terminológico permite a recuperação de informações terminográficas baseadas em diversos critérios de busca: termo, forma abreviada, classificação, definição, fonte, equivalentes, termos relacionados, sinônimos, instituição, data de inserção no sistema.

Auger e Rousseau (1977, p. 67) sugerem os seguintes campos para a ficha terminográfica: entrada na língua original, entrada no segundo idioma, categoria gramatical e lexical, emprego específico, domínio associado, remissivas, fonte, definição, contexto. Em relação ao controle terminológico, os autores apresentam as seguintes recomendações:

Adjetivos – os adjetivos devem ser registrados no masculino e no singular;

Verbos – todos os verbos devem ser registrados no infinitivo;

Substantivos – a entrada dos substantivos é feita no singular, salvo de plural lexicalizado, que, neste caso, permanece na sua forma original;

Categorias gramaticais – são registradas por meio de abreviações, conforme a tabela 12.

Tabela 12: abreviaturas para categorias gramáticas

Abreviação	Categoria gramatical
adj	Adjetivo
f	Feminino
m	Masculino
n	Nome (substantivo)
pl	Plural
sing	Singular
v	Verbo

Fonte: (AUGER; ROUSSEAU, 1977)

Definições – as definições devem ser transcritas por inteiro, em minúsculas, salvo pela letra inicial que é registrada em maiúscula. Não se deve omitir a entrada, os domínios, os indicativos gramaticais e nem a noção fonética.

Contexto – deve ser transcrito em minúsculas, com exceção para a inicial maiúscula na primeira palavra da frase; se houver erro no texto original, deve-se transcrevê-lo seguido da expressão *sic*.

Remissivas – são assinaladas na ficha de informações em um campo destinado a esse fim. As remissivas podem ser abreviaturas, variações ortográficas secundárias de uma entrada principal, sinônimos, termos preteridos (*use*), termos preferidos (*ver*), termos equivalentes (*ver também*).

As formas abreviadas facilitam o uso de termos que possuam grande extensão e podem originar-se de diversas maneiras. Wright e Budin (1997) apresentam as seguintes:

– abreviação – criada pela omissão de partes de uma palavra ou de uma ou mais palavras, quando o termo for composto; em muitos casos, o uso apenas da primeira letra da palavra é suficiente para a sua compreensão. Ex.: p. = página; etc = *et cetera*;

– inicialismo – criado pelo uso da primeira letra (ou som) de cada palavra que compõe um termo. Ex.: UN = *United Nations* (Nações Unidas);

– forma curta – pode ser introduzida para termos extremamente longos e complexos, com o uso de poucas palavras, mas não utiliza abreviação ou inicialismo. Ex.: Grupo Intergovernamental de Vinte e Quatro em Assuntos Econômicos Internacionais (forma longa)/Grupo Vinte e quatro (forma curta);

– acrônimo – criado da mesma maneira que um inicialismo, mas pronunciado como se como uma única palavra. Ex.: laser = *light amplification by stimulated emission of radiation* (amplificação da luz por emissão de radiação estimulada);

– formas mistas – algumas formas podem funcionar como inicialismos ou acrônimos, dependendo das preferências regionais do falante. Ex.: ISO, NASA;

– mistura – é formada pelo corte e junção de palavras separadas. Como os acrônimos, é pronunciada como uma única palavra, mas pelo menos uma das partes que são juntadas é maior do que apenas uma letra. Ex.: quasar = *quase-stellar object* (objeto quase-estelar); ALGOL = *Algorithmic language*;

Os autores argumentam ainda que não há regras universais para a composição de termos abreviados, uma vez que cada idioma possui suas particularidades. Recomendam, no entanto, que a abreviação seja de fácil pronúncia, não seja idêntica a outra já conhecida e, quando trata-se de convenções internacionais, como as advindas do sistema internacional de unidades, por exemplo, não varie em número e não seja seguida de ponto: m = metro (s); l = litro (s).

Krieger e Finatto (2004) propõem como estrutura indispensável para a ficha terminológica (ilustração 22) os seguintes campos: entrada – a) termo (denominação) composto por uma ou mais palavras, sigla, abreviatura ou símbolo e que identifica o conceito que está sendo focalizado no registro; b) informações pertinentes ao conceito – área ou subárea temática; definição; remissivas (conceitos relacionados hiperônimos, hipônimos); notas relativas ao conceito; c) informações relativas à denominação – categoria gramatical, sinônimos, abreviaturas, formas truncadas, fraseologia, informações morfológicas; d) informações relativas a aspectos de uso – grau de confiabilidade, vigência, registro, variantes; e) equivalentes em outros idiomas; f) fontes – elenco das referências bibliográficas relativas à definição e ao contexto de aplicação do termo; g) dados de gestão da ficha – datas de entrada, revisão e rubrica do revisor; rubrica do redator.

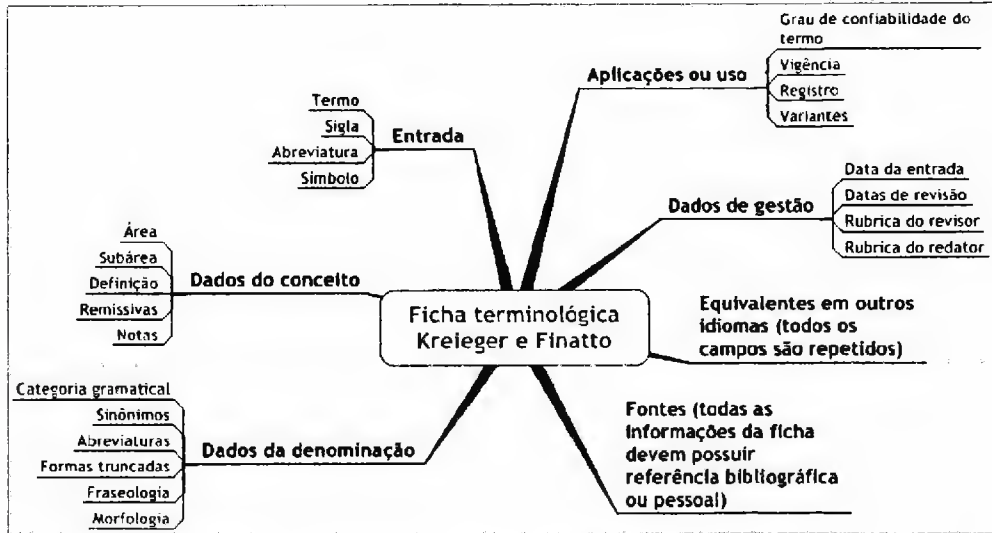


Ilustração 22: Formato terminográfico, segundo Krieger e Finatto
 Fonte: adaptado de KRIEGER e FINATTO (2004)

Felber e Picht (1984) classificam os bancos de dados terminológicos, segundo sua função em bancos do tipo dicionário – orientados à tradução – e bancos do tipo vocabulário – orientados para a normalização e para a investigação científica. De acordo com essa classificação, podem variar os campos de dados (ou entradas, como denomina a Terminologia) a serem incluídos no registro. Nos bancos de dados do tipo dicionário, os registros assemelham-se à microestrutura de um dicionário, sendo orientados ao termo. Nos bancos do tipo vocabulário, os registros são orientados ao conceito e às suas relações em um sistema de conceitos. Em sua obra, os autores apresentam inúmeros exemplos de formatos terminológicos, destacando-se o formato proposto no documento de revisão da norma ISO R 919 :1969 (ilustração 23).

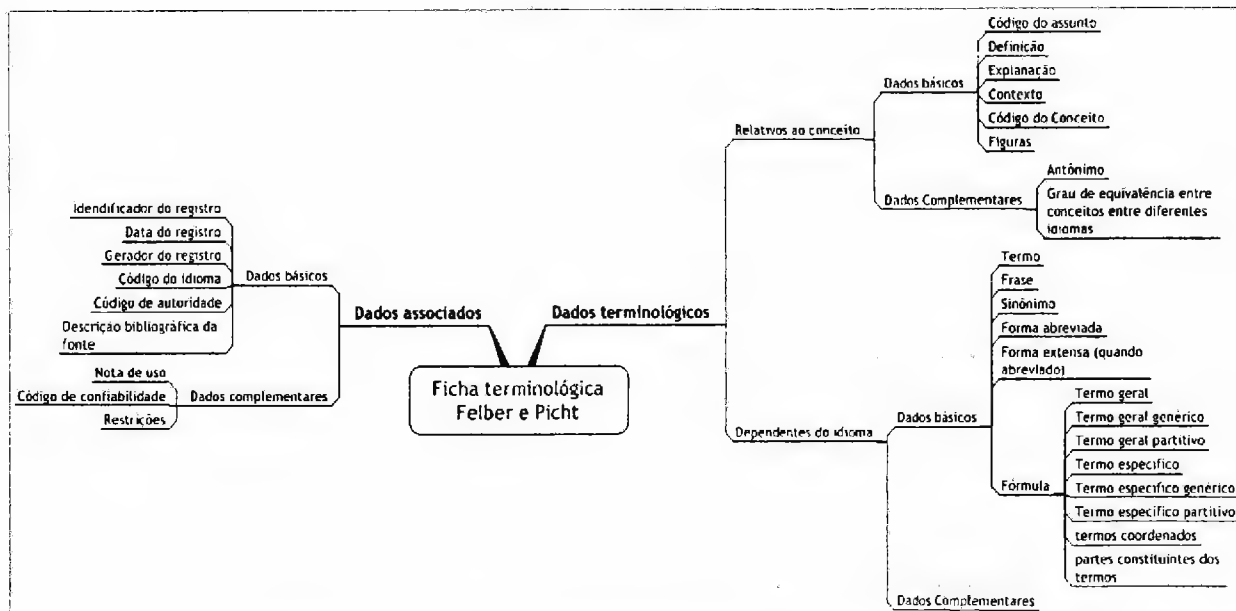


Ilustração 23: Formato terminológico conforme documento de revisão da norma ISO R 919
 Fonte: adaptado de FELBER e PICT (1984).

Além dos formatos citados, diversas iniciativas para a padronização terminológica foram realizadas. No âmbito da *Internet*, um padrão a ser destacado é o formato *Machine-Readable Terminology Interchange Format* (MARTIF), instituído pela norma ISO 12200 (ilustração 24). O formato baseia-se na linguagem SGML e utiliza uma DTD como instrumento de controle de semântica e de sintaxe. O objetivo do MARTIF é fornecer um formato que possa ser usado para o intercâmbio e para a disseminação de terminologias em meio eletrônico. Atualmente (2004), encontram-se em fase de desenvolvimento as especificações dadas pela norma ISO 12620: *Aplicações de computadores em Terminologia*, em duas partes: Parte 1: Modelo de descrição e procedimentos para manutenção de registros de categorias de dados para recursos de idiomas; Parte 2: Categorias de dados terminológicos. A ilustração 25 apresenta os elementos de primeiro nível para o formato sugerido pela norma ISO 12620. A lista completa dos elementos pode ser encontrada no anexo A.

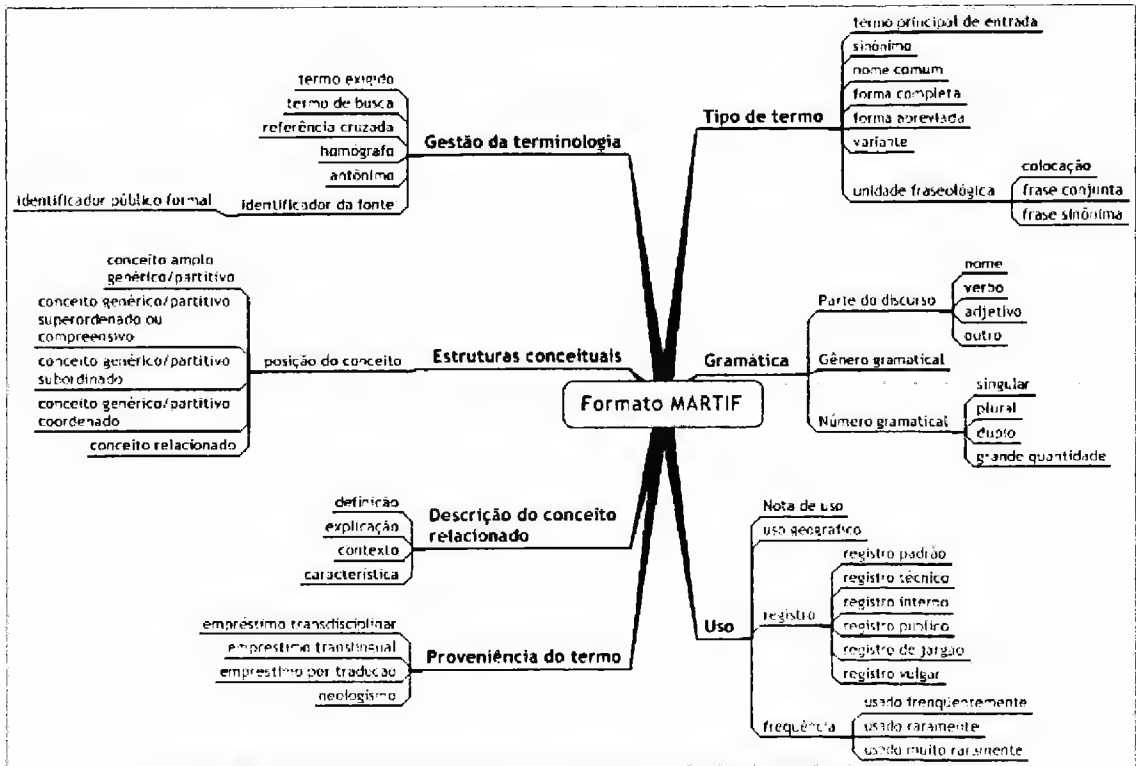


Ilustração 24: Formato MARTIF de intercâmbio de terminologias (SÁNCHEZ, 1999)

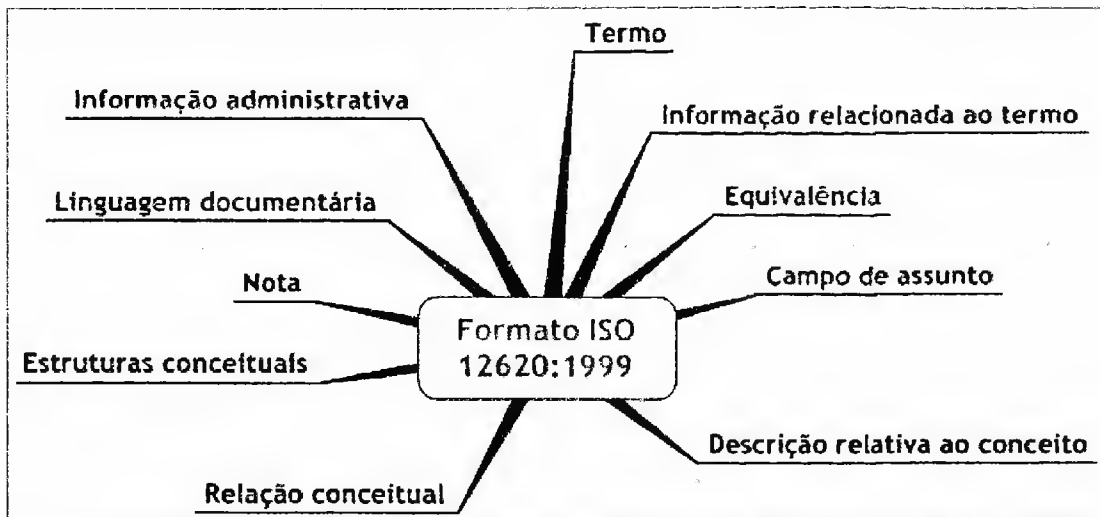


Ilustração 25: Formato de terminologias estabelecido pela norma ISO12620 (TTT.ORG, 2004)

2.2 Linguagens Documentárias

Esta seção trata dos aspectos teóricos de Linguagem Documentária, focando-se principalmente em aspectos semânticos, tipologia de tesouros, estruturação de tesouros e relações conceituais em tesouros. Foram estudados alguns textos de autores representativos da área: Aitchison & Gilchrist (1979), Gomes (1990), Cintra (1994), Lancaster (1993), Cavalcanti

(1978;1982), Além desses autores, foi consultada a norma ISO 2788:1986, que trata sobre a elaboração de tesouros.

2.2.1 Conceituação

As atividades de análise, organização e representação da informação podem valer-se de dois tipos de linguagens em seus processos: a linguagem natural e uma ou várias linguagens artificiais. A linguagem natural reúne sinais ou símbolos capazes de combinarem elementos perceptíveis ou sensíveis – por exemplo, gestos, olhares, fala, símbolos gráficos, literais ou numéricos – a um elemento inteligível – significado ou representação mental –, para constituírem uma relação que permite ao homem comunicar-se e expressar suas idéias para outros homens. As linguagens artificiais são elaboradas de acordo com regras previamente estabelecidas e com necessidades conhecidas, e têm como objetivo a padronização das significações, para que haja otimização da comunicação humana. Tais linguagens visam à uniformidade de armazenagem de informações, bem como à facilidade de recuperação, utilizando, para isso, o controle terminológico (CAVALCANTI, 1978). Em uma linguagem artificial, o vocabulário – a que nesta pesquisa pretere-se denominar como terminologia específica – compreende uma lista de termos apresentados alfabeticamente, seguidos por suas definições e deve se controlado em termos de semântica e de sintaxe.

A linguagem natural (LN) pode ser definida como a linguagem do discurso técnico-científico; no domínio da recuperação da informação, a expressão refere-se às palavras na acepção de “texto livre”. No entanto, a linguagem natural, ou linguagem falada, apresenta características que dificultam sua utilização para o tratamento da informação, uma vez que pressupõe nuances, associações de idéias, expressão de emoções e de valores. Assim, com o objetivo de facilitar as operações de busca e recuperação, as linguagens documentárias fazem uma condensação e uma simplificação da linguagem natural, de modo a reterem apenas uma parte do seu vocabulário e somente algumas regras gramaticais (LOPES, 1990; GUINCHAT, 1994).

Há diferentes definições para o termo linguagem, todas baseadas no conceito da língua como sistema de signos utilizados na comunicação humana. Martinet (apud, NOCETTI; FIGUEIREDO, 1978) a define como um sistema de signos utilizados para a comunicação entre dois seres vivos. O artigo de Nocetti & Figueiredo apresenta ainda outras definições de autores como Saussure, segundo o qual língua e fala são partes de um todo que é a linguagem; como Hughes, segundo o qual a linguagem é um sistema de símbolos vocais arbitrários utilizados para se transmitir o pensamento de uma pessoa a outra; e como Dubois, segundo o qual a linguagem é a capacidade específica da espécie humana de se comunicar com o uso de um sistema de signos

vocais (língua). Nocetti e Figueredo (1978) concluem, da análise de tais definições de língua e linguagem, sob a ótica daqueles autores, que essas são bastante similares, uma vez que consideram a linguagem como a capacidade humana para a criação e para a comunicação de mensagens; já a língua, é por eles definida como um sistema de signos vocais destinados às comunicações lingüísticas. Nocetti e Figueiredo resumem algumas características importantes da linguagem:

a) a linguagem é um sistema, isto é, um conjunto de regras em que cada item é o que é em função de sua colocação no sistema;

b) a linguagem é arbitrária, uma vez que inexiste uma necessidade intrínseca para que cada palavra tenha este ou aquele significado;

c) a linguagem é vocal, uma vez que é formada pelos sons produzidos pelos órgãos da fala dos seres humanos – dessa definição, percebe-se que os autores não pretenderam abordar outras formas de linguagens, como as linguagens visuais e linguagens baseadas em sinais²⁴.

d) a linguagem é simbolismo, uma vez que este é a base filosófica que torna possível a linguagem, isto é, para que haja comunicação são necessárias as habilidades de simbolizar e de distinguir símbolos.

e) veículo do pensamento, o que significa que é utilizada para expressá-lo.

Com relação às funções da linguagem, Nocetti & Figueiredo apresentam as abordagens de diferentes autores, que assim sintetizam: a linguagem serve para permitir a comunicação entre um emissor e um receptor, traduzindo realidades do universo interior ou exterior de um indivíduo, e permitindo-lhe que se expresse de modo mais ou menos elaborado. A tabela 13 apresenta as classificações de Bühler, Halliday, Martinet e Hutchins, de acordo com o artigo de Nocetti e Figueiredo (1978):

²⁴ Uma referência interessante nessa área é a Língua Brasileira de Sinais – LIBRAS, regulamentada e reconhecida como meio de comunicação e expressão, pela lei nº 10.436, de 2002. A Libras é definida naquele texto legal como “a forma de comunicação e expressão, em que o sistema lingüístico de natureza visual-motora, com estrutura gramatical própria, constituem um sistema lingüístico de transmissão de idéias e fatos, oriundos de comunidades de pessoas surdas do Brasil.”

Tabela 13: Funções das linguagens

Autor	Funções
BÜHLER	<ul style="list-style-type: none"> - representação do assunto; - exteriorização psíquica de um estado da alma; - apelo sobre o ouvinte, para despertar-lhe uma atitude.
HALLIDAY	<ul style="list-style-type: none"> - ideacional: serve para a manifestação de conteúdo, isto é, da experiência que o emissor possui acerca dos seus mundos interior e exterior; - interpessoal: estabelecimento e manutenção de relações sociais e interacionais; - textual: refere-se à comunicação por meio da escrita, baseada em suas características gramaticais e entonacionais.
MARTINET	<ul style="list-style-type: none"> - comunicação: uso do código para a transmissão de mensagens; - expressão: exposição de sentimentos, sem haver preocupação com as reações dos ouvintes; - suporte de pensamento: a linguagem dá suporte ao pensamento, pois esse é exercido no universo da língua; - estética: intencionalidade de se criar algo belo com a linguagem.
HUTCHINS	<ul style="list-style-type: none"> - descritiva ou informativa: comunicação da informação factual; - iniciativa, conotativa, imperativa ou manipuladora: provocação de determinados comportamentos nos receptores; - valorativa: indicação de juízos de valor; - prescritiva: orientação dos receptores para o desempenho de uma ação; - expressiva: expressão de sentimentos, experiências, disposições, necessidades, entre outras emoções; - evocativa: evocação de certos estados de espírito nos receptores; - fática: realização de contactos sociais; - metalingüística: discussão da linguagem em si mesma.

Fonte: (NOCETTI; FIGUEIREDO, 1978)

As linguagens documentárias, ou linguagens controladas, são também conhecidas na literatura como linguagens de indexação, línguas descritoras, codificações documentárias. Tais linguagens são definidas por Coyaud (apud NOCETTI; FIGUEIREDO, 1978, p. 29) como emanções das línguas naturais, que têm como objetivo facilitar a comunicação entre usuários de documentação e documentalistas, visando à recuperação. As linguagens documentárias são, assim, a ponte entre a linguagem dos documentos e a linguagem das consultas dos usuários. O autor argumenta, ainda, que as linguagens documentárias são meios de comunicação, mas que não constituem necessariamente uma língua. Em suma, a linguagem empregada para a descrição do conteúdo de um documento é chamada de linguagem documentária, linguagem de indexação, linguagem descritora ou linguagem de informação (PIEIDADE, 1977, p. 1).

Uma linguagem documentária é um sistema de representação do conteúdo dos documentos, que baseia-se na hierarquização, no controle e na padronização de uma lista de termos e que possui uma estrutura de apresentação definida, seja ela gráfica ou escrita. Para Cavalcanti (1978), uma linguagem documentária é definida como o conjunto de regras, símbolos e termos, previamente estabelecidos para formarem uma linguagem artificial destinada à indicação do conteúdo temático dos documentos. Segundo Nakayama (1990b), as linguagens documentárias são obras terminológicas baseadas em linguagem natural e utilizadas como instrumentos de tratamento – que envolve os processos de coleta, seleção, controle e normalização –, e recuperação da informação. A autora classifica as linguagens documentárias em quatro tipos: terminologia controlada; cabeçalhos de assunto; tabelas de classificação; tesouros. A terminologia controlada é uma lista de termos em ordem alfabética, coletados e selecionados conforme a necessidade do sistema de informação e pode ser utilizada para a elaboração de outros tipos de linguagens documentárias, especialmente os tesouros. O cabeçalho de assunto é uma obra terminológica que apresenta os termos em ordem alfabética e possui uma organização de remissivas, podendo aparecer, também, as relações hierárquicas e associativas entre os termos. As tabelas de classificação utilizam o sistema decimal e possuem códigos numéricos acompanhando os termos – estruturação alfa-numérica. O tesouro, a linguagem documentária mais importante no escopo desta pesquisa, apresenta os termos de um domínio ou subdomínio do conhecimento, organizados em ordem alfabética e sob um arranjo semântico que explicita as relações hierárquicas e associativas entre os termos, garantindo que esteja implícita a sua definição. Os tesouros terminológicos costumam trazer a definição, do conceito em referência, explicitada por meio de um enunciado.

Segundo Nakayama (1985), linguagem documentária, segundo sua finalidade, é um sistema de representação do conteúdo dos documentos e das perguntas, tendo como finalidade a recuperação de tais documentos. Quanto à estrutura, Nakayama (1985) lembra que uma linguagem documentária possui estrutura própria, é controlada, padronizada e geralmente hierarquizada.

Em algumas situações, como por exemplo os índices KWIC ou da indexação automática da *web*, é possível utilizar-se a linguagem natural como fonte dos termos de indexação. Tal procedimento facilita a entrada mas, por outro lado, dificulta a recuperação, devido aos fenômenos da polissemia e da homonímia. Para se utilizar um índice KWIC, às vezes é necessário o emprego de uma estratégia de busca que contenha tantos sinônimos quanto possível em relação a um certo termo de busca. O uso das Linguagens Documentárias, por outro lado, facilita a recuperação, uma vez que os termos são padronizados, sendo a entrada dificultada, uma vez que

é necessária a etapa de tradução de termos da linguagem natural para a linguagem documentária. Um dos grandes ganhos das linguagens documentárias é, portanto, a possibilidade de realização de controle terminológico, evitando problemas como a polissemia e a sinonímia (NOCETTI; FIGUEIREDO, 1978, p. 30-31).

Para Guinchat (1994), linguagem documentária é uma linguagem convencional utilizada para descrever o conteúdo dos documentos, com o objetivo de armazená-los e de recuperar as informações que eles contêm. Segundo Tálamo (1997), o termo “linguagem documentária” somente aparece explicitamente na literatura a partir dos anos 70. No entanto, de acordo com Cintra (1994), as Linguagens Documentárias passaram a desenvolver-se a partir das décadas de 50 e 60, tendo em vista necessidades de se melhorarem, devido à explosão documental registrada na época, os processos de armazenamento e recuperação da informação. Visando a esses objetivos, as linguagens documentárias foram desenvolvidas como instrumentos que pudessem traduzir o conteúdo dos documentos, utilizando técnicas de indexação, sistemas de símbolos, campos semânticos, campos associativos e categorias funcionais. Tálamo (1997), corroborando essa idéia, define linguagem documentária como “uma linguagem construída, oposta à natural, portanto, que tem como objetivo específico tratar a informação para fins de recuperação.”

Uma linguagem documentária é um conjunto de termos providos ou não de regras sintáticas, utilizado para representar conteúdos de documentos. Assim, uma linguagem documentária possui três elementos básicos: a) léxico - lista de elementos descritores, devidamente filtrados e depurados; b) rede paradigmática - compreende a organização lógico-semântica dos termos, que traduz relações essenciais entre eles; c) rede sintagmática - destina-se a expressar relações contingentes entre os descritores, relações essas que são válidas apenas no contexto em que aparecem (CINTRA, 1994). Desse modo, os elementos da linguagem são selecionados de universos determinados e seu sistema de relações é construído, sendo indispensável, para que se possa utilizar essa linguagem, a existência de regras explícitas, construídas a partir de uma visão particularizada da realidade.

A linguagem documentária é utilizada no processamento da informação, em dois momentos: no momento do tratamento intelectual do documento, ou seja, nas operações de entrada; e no momento de pesquisa, isto é, na recuperação da informação (GUINCHAT, 1994). Do mesmo modo que a linguagem natural, uma linguagem documentária é um sistema simbólico, que foi instituído com o objetivo de facilitar a comunicação. No entanto, sua função comunicativa é restrita a contextos documentários, isto é, devem tornar possível a comunicação

usuário-sistema de informação (CINTRA, 1994). As linguagens documentárias são utilizadas para representarem o conteúdo dos documentos e não os documentos em si; portanto, essa função de representação é eminentemente referencial, isto é, não se representa o texto individualmente, mas uma classe de assunto à qual ele pertença.

As linguagens documentárias mais conhecidas são o tesauro e os sistemas de classificação documentária; essas linguagens diferem entre si em função do maior ou menor grau com que as relações entre os termos, como encontradas na linguagem natural, se reproduzem em um campo específico do conhecimento. Os sistemas de classificação mais conhecidos são as classificações decimais: Classificação Decimal de Dewey (CDD) e Classificação Decimal Universal (CDU); a classificação facetada do *Classification Research Group* (CRG); e a *Colon Classification de Ranganathan*. Com base nessas classificações facetadas, surgiram os (tesauros) que possuíam como preocupação adicional, além da representação de conteúdo, o controle do vocabulário (CINTRA, 1994). Em relação aos tipos de tesouros existentes, Gomes (1990a, p. 16-17) apresenta as seguintes classificações: a) em relação à língua que abrangem: monolíngües ou multilíngües; b) em relação ao nível de especificidade dos termos: macrotesouros – os termos representam conceitos mais amplos; número de termos não muito extenso; grande número de remissivas – ou microtesouros – os termos representam conceitos mais específicos; pertencem a uma área mais restrita do conhecimento; c) em relação ao número de disciplinas que cobrem: multidisciplinares – os que cobrem diversas disciplinas ou monodisciplinares – voltados para apenas uma disciplina.

Segundo Lancaster (1993), os esquemas de classificação decimal (CDD e CDU), os cabeçalhos de assuntos e os tesouros são instrumentos adequados ao controle de vocabulário e possuem em comum as seguintes características: a) todos procuram apresentar os termos tanto alfabética como sistematicamente; b) nas classificações o arranjo alfabético é secundário; ocorre na forma de um índice que remete a um arranjo principal, o qual é hierárquico; c) no tesauro, o arranjo é alfabético, mas também é baseado em uma estrutura hierárquica implícita; d) a tradicional lista de cabeçalhos de assuntos é similar ao tesauro pelo fato de sua base ser alfabética, mas difere daquele instrumento, por incorporar uma estrutura hierárquica imperfeita e por não distinguir claramente as relações hierárquicas das associativas.

O autor argumenta, ainda, que um vocabulário controlado é uma lista de termos autorizados, que incorpora uma estrutura semântica. Essa definição equivale a dizer-se que, em geral, o indexador somente pode atribuir a um documento termos que constem de uma lista controlada. Sob essa perspectiva semântica, Lancaster (1993) afirma que o vocabulário

controlado destina-se a: a) realizar controle de sinônimos - optando por uma forma padronizada de entrada; b) controlar homógrafos (ver mais sobre isso na parte de controle terminológico; termos polissêmicos; ex. peru (país) peru (ave); c) reunir ou ligar termos cujos significados apresentem uma relação mais estreita entre si. O autor identifica dois tipos de relações: hierárquicas e não hierárquicas (ou associativas) exemplificando, o autor usa os termos: operárias - > mulheres (gênero - espécie); donas-de-casa -> mulheres (gênero - espécie); donas-de-casa -> emprego (associação não hierárquica).

Em termos estruturais, as linguagens documentárias procuram estabelecer uma integração entre a representação simbólica de uma área de conhecimento – o conjunto de termos que representam tal área de conhecimento –, e o domínio da linguagem natural conexas aos usuários. Tal representação simbólica, alcunhada como palavra-chave, descritor, cabeçalho de assunto, entre outras denominações é, portanto, o elemento fundamental da linguagem documentária, utilizado para a representação do conteúdo dos documentos. Sua utilização visa, primordialmente, o controle de fatores como a ambigüidade na comunicação, ocasionada pela sinonímia, pobreza informativa, redundância, entre outros motivadores. Assim, para Guinchat (1994), os elementos constitutivos de uma linguagem documentária são:

- descritores – termos simples ou compostos que descrevem a informação; são extraídos da linguagem natural e reduzidos a uma forma gramatical única e invariável (por exemplo: substantivo, singular);

- não-descritores – termos relacionados com o descritor por meio de uma remissiva; esses termos não devem ser utilizados nas operações de descrição do conteúdo informacional de um documento mas, em seu lugar, devem ser preferidos outros termos;

- relações entre descritores – as relações entre os termos podem ser hierárquicas, de equivalência ou de vizinhança;

- notações – as notações podem ser numéricas, alfanuméricas, alfabéticas, simbólicas ou sob a forma de sílabas e visam a representação dos descritores de forma abreviada;

- notas – destinam-se a restringir ou a orientar o uso dos descritores; podem ser notas gerais ou notas de escopo; por exemplo, o descritor “construção” pode ser acompanhado pela nota “utilizar apenas para a descrição dos elementos da construção civil, como muros, fundações e telhados”;

- representação gráfica – mostram graficamente as relações entre os termos;

Para Tálamo (1997), a compreensão dos aspectos sistêmicos de uma linguagem só é possível a partir da compreensão dos seus elementos constitutivos:

a) signo (termo) – uma unidade que representa um significado (conceito); a parte perceptível do signo é denominada significante e a parte ausente (ou intangível), significado; a relação entre signo e significado é denominada significação;

b) sintagma e paradigma – um termo possui relações paradigmáticas (categorizações) com outros termos (semelhança, inclusão, exclusão, implicação, pressuposição, entre outras); a combinação de diversos termos para integrarem o discurso é denominada relação sintagmática (combinação dos termos em cadeia linear);

A tabela 14 apresenta uma adaptação ao contraponto de semelhanças e divergências entre linguagens naturais e linguagens documentárias, proposto por Nocetti & Figueiredo (1978, p. 34-35).

Tabela 14: semelhanças e diferenças entre as linguagens naturais e as linguagens documentárias

Linguagem Natural	Linguagem documentária
Maior importância para a parte oral; as formas escritas são secundárias e geralmente derivam das formas vocais	o caráter vocal não é necessário; uma linguagem documentária normalmente é uma linguagem escrita.
Pertence a um domínio geral do conhecimento	pertence ao um domínio especializado do conhecimento.
está relacionada a milhares de pessoas	relaciona-se a um número reduzido de pessoas
são aceitos fenômenos lingüísticos, como a polissemia, a sinonímia e a ambigüidade	não admite polissemia e nem ambigüidade; há um esforço para a diminuição da sinonímia
sua aceitação pelas pessoas ocorre naturalmente	para que exista deve ser aceita pelos usuários
interessam os aspectos fonológicos, semânticos e sintáticos	interessam apenas os aspectos semânticos e sintáticos
são menos eficientes do que as linguagens documentárias quanto às operações de recuperação da informação	são mais eficientes nas operações de recuperação de informação
têm sua própria teoria	baseiam-se na teoria das ciências da informação e da lingüística
abriga diferentes classes gramaticais	não comporta pronomes, artigos, número, tempo e pessoas, adjetivos que ocorram isolados e nem advérbios

Fonte: (NOCETTI; FIGUEIREDO, 1978, p. 34-35)

2.2.2 Controle terminológico

Segundo Tálamo e outros (1990), os instrumentos de controle terminológico compreendem linguagens que possuem um número limitado e controlado de termos e podem ser utilizados tanto em processos que compreendem a análise, síntese e representação de discursos, quanto nos processos que compreendem a representação de buscas de informação. Disso, argumentam as autoras, depreende-se que a relação entre documentos e instrumentos de controle terminológico é determinada pela forma de representação. Um dos instrumentos mais

significativos para a realização do controle terminológico é o tesouro. Em primeiro lugar, porque sua construção apóia-se na organização, por meio da classificação, do campo conhecimento a ser representado; em segundo lugar, porque utiliza o princípio da garantia literária, isto é, os processos de coleta e de adoção dos termos para a representação daquele campo do conhecimento baseiam-se em um conjunto de documentos que contêm os discursos pertinentes a uma determinada língua de especialidade.

Segundo Aitchison e Gilchrist (1979, p. 28), em linguagens controladas, como as linguagens documentárias, é importante que as palavras no vocabulário sejam normalizadas, com o objetivo de se estabelecer a coincidência entre as palavras usadas na indexação e na busca.

Uma interpretação do texto de Lancaster (1987, p. 13), permite concluir que o controle terminológico destina-se a resolver problemas como sinonímia, polissemia ou ambigüidade, esta última provocada pela existência de relações de associação ou relações genéricas entre os termos. O autor sugere alguns meios para o estabelecimento do controle terminológico, a saber: a) termos sinônimos ou quase-sinônimos devem ser fundidos em um conceito, sendo um deles adotado como entrada, fazendo remissão ao outro; b) termos polissêmicos são acompanhados de qualificadores ou de outros recursos explicativos; c) termos associados entre si por meio de equivalência ou relações genéricas, são registrados com remissivas mútuas. Para o autor, o controle terminológico – a que ele denomina controle de vocabulário – tem como objetivos: a) promover a representação consistente de um campo de assuntos; b) facilitar a busca e recuperação por meio do agrupamento de termos cujos significados estão paradigmática²⁵ ou sintagmaticamente relacionados. Assim, o controle terminológico pretende fazer coincidir a linguagem dos autores, com a dos usuários e com a dos cientistas da informação.

Para GOMES (1990a, p. 15-16), a palavra é a menor unidade léxica, cujo significado é fornecido pelo contexto em que ela se insere. Tomada isoladamente, no entanto, uma palavra poderá possuir diversos significados, o que a torna inadequada para o uso em sistemas de armazenamento e recuperação de informações, tendo em vista fenômeno da ambigüidade. Dessa forma, deve-se escolher uma palavra, a que a autora prefere denominar como *termo*, que represente um único conceito. Dito de outro modo, o termo equivale a um conceito, mais uma designação. É comum os autores chamarem também os termos de descritores. O

²⁵ Segundo Houaiss (2004), uma relação paradigmática refere-se a “[...]uma série de unidades que possuem traço(s) em comum e que podem se substituir mutuamente num determinado ponto da cadeia da fala”. Paradigma designa, assim, o conjunto dos termos substituíveis entre si numa mesma posição da estrutura a que pertencem. Uma relação sintagmática ocorre entre “[...]unidades da língua que se encontram contíguas na cadeia da fala e que não podem se substituir mutuamente, pois têm funções diferentes (p.ex., em céu azul e eles chegaram, a relação entre céu e azul, e entre eles e chegaram)”. Relações paradigmáticas e sintagmáticas são opostas.

controle terminológico baseia-se, assim, na atribuição de apenas um termo representar um conceito, de tal sorte que cada conceito seja igualmente representado por apenas um termo. Uma característica importante de um tesouro é que sua cobertura refere-se a um domínio específico do conhecimento, não sendo recomendável a iniciativa de se desenvolver um tesouro geral. Esse fenômeno é também denominado *postulado da monorreferencialidade*, isto é, a denominação do conceito é um termo, o qual guarda com aquele conceito uma relação unívoca: para cada conceito existe apenas uma denominação e para cada denominação apenas é admitido um conceito²⁶ (GOMES, 1990a, p. 19).

Em relação ao processo de indexação, pode-se utilizar o controle terminológico sob dois pontos de vista: indexação pré-coordenada ou indexação pós-coordenada. Na indexação pré-coordenada combina-se os termos para se formar expressões compostas que são inseridas previamente na linguagem documentária; ou seja, no momento da armazenagem, insere-se os termos compostos na lista de termos aceitos para a representação de conteúdo de dado documento ou conjunto de documentos. A indexação pós-coordenada dá-se no momento da pesquisa; isto significa que os termos compostos são combinados ou correlacionados a partir da busca de termos simples na lista de termos autorizados, no momento da busca. Desses dois tipos de procedimentos resulta que a indexação pré-coordenada origina linguagens documentárias mais extensas, uma vez que os termos compostos são incluídos na lista previamente e, por outro lado, que a indexação pós-coordenada serve-se de linguagens documentárias com menor número de termos, posto que a combinação realiza-se no momento da pesquisa (CAVALCANTI, 1978, p. 21).

Para Cintra (1994), dizer que uma linguagem documentária possui controle terminológico equivale a afirmar: a) que os mecanismos de interpretação dos significados na linguagem documentária são próprios a ela mesma; b) a elaboração das linguagens documentárias orienta-se para a fixação de significados; c) as linguagens documentárias possuem um vocabulário próprio, cujos termos pertencem, preferencialmente, a linguagens especializadas.

Certos fenômenos lingüísticos, como a polissemia, a sinonímia, a homonímia, a neologia e a metonímia, podem influenciar sobremaneira o tratamento da informação, visto que, em geral, podem ser responsáveis pela ambigüidade na interpretação dos conceitos.

²⁶ Segundo a autora (1990a, p. 20) os termos não possuem conotação, fenômeno que ocorre apenas com as palavras da linguagem natural. Assim, a palavra "tênis", na linguagem natural tanto pode referir-se ao conceito de esporte quanto ao de calçado. Na linguagem documentária essas alternativas são mutuamente excludentes, isto é, o termo apenas pode designar um ou outro conceito. Quando houver necessidade, pode-se utilizar qualificadores, como recurso de diferenciação entre um conceito e outro.

Houaiss (2004) define a *polissemia*²⁷ como a “multiplicidade de sentidos de uma palavra ou locução” e afirma que a polissemia é um fenômeno comum nas línguas naturais, sendo raras as palavras que não a apresentam. Em geral, as causas da polissemia referem-se: a) aos usos figurados, por metáfora ou metonímia, por extensão de sentido ou analogia; ou b) ao empréstimo da significação da palavra em outra língua. A *homonímia*²⁸ ocorre entre palavras com origens diferentes que convergiram foneticamente, o que equivale a afirmar: é uma “relação entre formas lingüísticas que, com significados diferentes, têm a mesma forma gráfica e fônica ou apenas fônica”. A *metonímia*²⁹ é uma figura de discurso que “consiste no uso de uma palavra fora do seu contexto semântico normal, por ter uma significação que tenha relação objetiva, de contigüidade, material ou conceitual, com o conteúdo ou o referente ocasionalmente pensado”. O autor enumera alguns tipos de situações em que a metonímia pode ocorrer, por substituição de: a) matéria por objeto: *ouro* por *dinheiro*; b) pessoa por coisa; c) autor por obra: adora *Portinari* por *a obra de Portinari*; d) divindade: esfera de suas funções; e) proprietário por propriedade: vamos hoje ao *Venâncio* por *ao restaurante do Venâncio*; f) morador por morada; g) continente pelo conteúdo: bebeu *uma garrafa de aguardente* por *a aguardente de uma garrafa*; h) consequência pela causa: respeite *os meus cabelos brancos* por *a minha velhice*; i) a qualidade pelo qualificado: praticar a *caridade* por *atos de caridade*.

Na linguagem natural, a *sinonímia*³⁰ define-se como uma relação de sentido entre dois vocábulos diferentes, que têm significação muito próxima, permitindo que um seja escolhido em vez do outro, em alguns contextos, sem que haja alteração do sentido literal da sentença como um todo.

A linguagem pré-coordenada utiliza, na maioria dos casos, descritores compostos e que cobrem completamente a definição de um conceito.

“A linguagem pré-coordenada tem a vantagem de ser mais precisa e de facilitar a estratégia de busca, evitando falsas associações e relações incorretas. Sua desvantagem é o

²⁷ Ferreira (1999) define polissemia como “o ter uma palavra muitas significações”.

²⁸ Ferreira (1999), registra as seguintes definições para o termo homonímia: “Identidade: fonética entre formas de significado e origem completamente distintos, como entre *são*, presente do verbo 'ser', e *são*, 'santo'. Na escrita, palavras que têm a mesma pronúncia, e igual grafia (como *falácia*, 'que é falaz', e: *falácia*, 'fallatório') ou grafia diferente (como *lasso*, 'cansado', e *laço*, 'laçada’).”

²⁹ Ferreira (1999), define como metonímia o ato de se utilizar um termo designativo de um objeto para a definição de outro objeto que tenha com o primeiro “uma relação de causa e efeito (*trabalho*, por *obra*), de continente e conteúdo (*copo*, por *bebida*), lugar e produto (*porto*, por *vinho do Porto*), matéria e objeto (*bronze*, por *estatueta de bronze*), abstrato e concreto (*bandeira*, por *pátria*), autor e obra (um *Camões*, por um *livro de Camões*), a parte pelo todo (*asa*, por *avião*), etc.”

³⁰ Ferreira (1999), define como sinônimo qualquer “palavra ou locução que tem a mesma ou quase a mesma significação que outra”

aumento do custo da indexação devido ao crescimento do arquivo pela repetição de termos na entrada e uso de referências." (CARNEIRO, 1985)

A linguagem pós-coordenada permite maior revocação do que a pré-coordenada, mas não oferece alta precisão (CARNEIRO, 1985)

Para Tálamo (1997), até o advento do conceito de tesouro, o tratamento da informação limitava-se à organização do conhecimento registrado, segundo a classe de assunto a ele atribuída, respondendo apenas aos objetivos de padronização e deixando insolúveis os problemas relativos à comunicação. "Já através dos termos, do vocabulário, da linguagem documentária se do tesouro, reconhece-se a existência de um sistema que é simultaneamente um modo de organização e uma forma de comunicação da informação."

A organização básica dos tesouros obedece a critérios de relações hierárquicas. No entanto, há também as unidades que são relacionadas de forma não-hierárquica, por meio de associações ou proximidade. As relações hierárquicas representam associações mais estáveis entre os termos; as relações não-hierárquicas indicam a ligação entre termos que estão em campos semânticos distintos, porém próximos. O conjunto das relações que estruturam o tesouro é um elemento que permite, ao usuário ou indexador, escolher o termo mais apropriado para representar o conceito que deseja localizar ou armazenar. Assim, a estrutura permite, a partir de um termo conhecido pelo usuário, chegar-se a um ou mais outros termos que podem ser apropriados para a representação do conceito desejado (CINTRA, 1994).

Fenômeno comumente encontrado no domínio da linguagem natural, a ambigüidade pode ocasionar distorção de sentido na comunicação. No domínio das linguagens documentárias, a ambigüidade deve ser evitada – senão eliminada –, de maneira que se obtenha o maior controle possível na utilização de termos para representarem os conceitos. A ambigüidade pode decorrer de diversos fenômenos lingüísticos, como a polissemia, a homonímia, a sinonímia, entre outros. "Numa linguagem documentária, tanto a polissemia quanto a ambigüidade devem ser neutralizadas, para que seja garantida a monossema." Assim, na linguagem documentária, devem ser utilizados qualificadores que contextualizam o sentido do termo empregado. Por exemplo: planta (botânica) e planta (arquitetura); companhia (empresa) e companhia (pessoa). (CINTRA, 1994). O fenômeno da polissemia é caracterizado pelo fato de um termo possuir mais de um significado simultaneamente.

De acordo com Cintra (1994), a sinonímia é uma relação de equivalência entre dois ou mais termos, que podem possuir o mesmo significado. No domínio da linguagem natural, a existência de sinônimos absolutos é bastante controversa entre os linguistas, que

preferem adotar o conceito de quase-sinônimos ou para-sinônimos para os termos que remetem ao mesmo significado. Assim, para a autora, os quase-sinônimos podem ser classificados em:

- termos pertencentes a diferentes línguas de especialidade (de acordo com o fatores regionais, sociais, etários, entre outros): avião / aeroplano;
- termos pertencentes a diferentes estilos de registro: ácido clorídrico (química) / ácido muriático (construção civil);
- termos que guardam diferença emotiva ou valorativa entre si: país em desenvolvimento / país subdesenvolvido;
- termos que possuem ocorrência limitada à presença de outros: gilete / lâmina de barbear; “de barbear” é condicionado a lâminas;
- termos cujos significados são realmente muito próximos: belo / bonito; casa / residência.

O fenômeno da homonímia é caracterizado pelo fato de “uma mesma forma significante remeter a duas realidades vocabulares diversas com identidade fônica (homofonia) ou identidade gráfica (homografia)”. Por exemplo: o termo cadeira pode referir-se a uma disciplina acadêmica ou a um objeto utilizado como assento (CINTRA, 1994).

A hiponímia expressa a relação existente entre um termo mais específico (ou subordinado) e um termo mais geral (ou superordenado). Assim, a hiponímia opera com a noção de inclusão de termos em determinada classe, por exemplo: rosa → flor; vaca → animal. Essa relação de inclusão pode inclusive transcender mais de uma classe. Cintra (1994) argumenta que dados três termos, “x”, “y” e “z”, se x é hipônimo de y e y é hipônimo de z, então x é hipônimo de z. Por exemplo: vaca → mamífero → animal.

2.2.2.1 Tesouros

Segundo Gilchrist (1971), a palavra tesouro origina-se do grego e significa tesouro. O autor relata que o termo foi utilizado pela primeira vez em 1736, significando “um tesouro ou tesouraria de conhecimento, como um dicionário, enciclopédia ou similares”. De acordo com o dicionário Webster’s, relata, o termo tesouro também refere-se a “um livro de palavras ou de informação sobre um campo particular ou conjunto de conceitos, especificamente um dicionário de sinônimos.” No contexto da recuperação da informação, Vickery (apud GILCHRIST, 1971, p.

4), credita a Helen Brownson³¹ a primeira utilização da palavra tesouro no sentido como é utilizada hoje. A partir dessa data, o autor registra, ainda, quatro acepções com que a palavra começou a ser utilizada, a partir de 1960:

- tesouro visto como uma “família de conceitos” ou um “dicionário de famílias de conceitos”;

- tesouro visto como uma “rede” ou uma “treliça” em que palavras-texto estariam relacionadas a diversas palavras-chaves, com o objetivo de facilitar a busca;

- tesouro visto como uma “matriz de associação de palavras”, ou um índice de palavras-chaves, organizado alfabeticamente, no qual cada palavra-chave é seguida por referências cruzadas;

- tesouro visto como um instrumento de duas partes (concepção atual): a) uma lista alfabética de palavras-chaves, com a exibição de suas respectivas referências-cruzadas; b) uma lista de categorias e as palavras-chaves atribuídas a essas categorias;

Segundo Gomes (2004), tesouro é um conjunto de termos semântica e genericamente relacionados, que cobrem uma área específica do conhecimento e é utilizado para indexar ou recuperar informações. Os tesouros foram antecedidos pelas classificações gerais de assunto e, posteriormente, pelas classificações mais especializadas. No final da década de 50, com o aumento da atividade científica e técnica, que motivou o aumento do número de fontes de informação, com as primeiras tentativas de informatização de documentos, bem como com a publicação dos princípios de classificação de Ranganathan, seguida pelo trabalho do *Classification Research Group*³², foi possível idealizar-se o primeiro tesouro. Segundo Fujita (1990), o primeiro tesouro empregado para controle de vocabulário em sistemas de recuperação foi desenvolvido em 1959, pela empresa Du Pont. Nos anos que se seguiram, diversas outras iniciativas foram publicadas, entre elas o Cabeçalho de Assuntos Médicos e o Tesouro de Engenharia Química. Os tesouros foram concebidos com a função de se realizar controle terminológico na representação,

³¹ Segundo o autor, Helen Brownson falou pela primeira vez sobre a acepção da palavra tesouro, no domínio das linguagens documentárias, durante a Dorking Conference on Classification, em 1957. (CINTIRA, 1994)

³² O Classification Research Group (CRG) foi um grupo criado na Inglaterra, no ano de 1952, para atuar no âmbito de pesquisas relativas à classificação. O grupo foi um dos principais colaboradores para o estabelecimento de teorias e pesquisas na área de Classificação, na segunda metade do século XX. O trabalho teórico do CRG envolve o estudo da análise de facetas, de operadores relacionais e da teoria integrativa de níveis. O grupo opôs-se ao tradicional método de classificação baseado na abordagem “top-down”, segundo o qual as áreas do conhecimento são predeterminadas e em seguida quebradas em seus elementos constitutivos. Ao contrário, o CRG interessou-se primeiramente em colocar elementos individuais juntos, para somente após isso determinar as áreas do conhecimento a que pertenciam. O grupo lançou mão do olhar da teoria integrativa de níveis para fornecer os princípios básicos da abordagem “bottom-up” da classificação. Maiores detalhes sobre o grupo e sua abordagem estão disponíveis em [<http://www.lis.uiuc.edu/review/summer1995/spiteri.html>] (SPITERI, 1995).

por meio de linguagem sistemática, do conteúdo de documentos escritos em linguagem natural. Dito de outro modo, um tesauro é utilizado na etapa de indexação de documentos, mais precisamente na tradução, como instrumento de controle de vocabulário, com a intenção de indicar os relacionamentos semânticos entre as palavras (termos), de uniformizar a utilização dos termos sinônimos e homônimos, bem como, de assegurar a forma e a ordem em que tais termos são escritos. O tesauro é um instrumento dinâmico, uma vez que permite a inclusão, a exclusão ou a modificação de termos a qualquer tempo, se o domínio da área de conhecimento assim o exigir.

Cavalcanti (1978, p. 30) relata que a palavra *thesaurus* tem origem greco-latina e ocorre em dicionários mais antigos sob a forma *thesouro*. A origem da palavra remonta ao século XVI, quando foi publicada, em 1565, uma das primeiras obras a utilizarem o termo: *Thesaurus linguae romanae et britannica*, de autoria de Cooper. Quase trezentos anos mais tarde, em 1852, Roget publica o seu *Thesaurus of english words and phrases*. Até a primeira metade do século XX a palavra permanece esquecida até que, em 1958, é retomada na Conferência Internacional sobre Informação Científica, mais ligada à área de tradução automática. A partir de então, os especialistas em informação e indexadores passam a utilizá-la para a indicação de listas de descritores ou termos de indexação. A autora define tesauro como uma lista estruturada de termos associados, empregada para a descrição de um documento com a desejada especificidade e de modo a permitir aos pesquisadores a recuperação da informação desejada.

A norma ISO 2788 (1974, p. 121) considera a definição de tesauro de acordo com dois pontos de vista: a definição em termos de sua função e a definição em termos de sua estrutura. Em termos de função, a norma define tesauro como um dispositivo de controle terminológico utilizado na tradução da linguagem natural – de documentos, indexadores ou usuários – em uma linguagem mais restrita, a que a norma denomina como “linguagem sistêmica”. Essas mesmas diretrizes são também observadas no texto normativo do Unisist/Unesco (1975, p. 1-2). Do ponto de vista da estrutura, um tesauro é um vocabulário dinâmico e controlado que possui termos semântica e genericamente relacionados e que cobre um domínio específico do conhecimento. Quanto à realização do controle terminológico, a norma considera dois tipos de tesauros: a) os que realizam controle terminológico por meio de termos preferidos, isto é, somente um dos termos que denotam um conceito é permitido para a indexação e recuperação; b) os que permitem a utilização de todos os termos que denotam um conceito, mas que atribuem tais termos a uma notação ou representação não ambígua – como um código ou um número – para facilitar a atualização do tesauro, caso o termo mude de significado. Para Chaumier (1994, p. 60), o tesauro é um dicionário de palavras ou de expressões – termos –

normalizadas e preferenciais, organizado de modo conceitual, apresentando os termos agrupados por afinidade semântica e complementado pelas indicações das relações entre os conceitos. Acidentes da linguagem natural, como os fenômenos da sinonímia e da polissemia são descartados. Os termos preferenciais são padronizados – por exemplo, em termos de classe gramatical, gênero, número, entre outras variações gramaticais – permitindo que se evite o uso de diversas palavras para a representação de conceitos vizinhos, bem como o uso de termos polissêmicos.

Comparado com outras linguagens de documentação, como um dicionário, que fornece definições para dadas palavras ou termos, por exemplo, um tesouro fornece termos para expressar significados que estão implícitos nas relações existentes entre esses termos. Os termos pertinentes a um tesouro tanto podem denotar conceitos ou combinações de conceitos, como entidades individuais – nomes próprios. Embora um descritor possa ser constituído por uma ou mais palavras, ele deve conter a menor quantidade de palavras possível, desde que não haja prejuízo para o seu significado; além disso, as palavras que formam o termo devem ser apresentadas preferencialmente na sua ordem direta (ISO 2788:1974).

Normalmente os sistemas que se utilizam de tesouros para a indexação de documentos baseiam-se na premissa de que palavras possuem significados precisos, não havendo necessidade de extração de qualquer outro significado do contexto (FUJITA, 1990). Dessa premissa depreende-se que o mesmo termo deve ter sempre igual conceito tanto para o indexador quanto para o usuário, o que raramente ocorre. Sendo assim, o tesouro é uma ferramenta importante, porque propõe-se a minorar as distorções semânticas que ocorrem entre os processos da indexação, pelo profissional da informação, e o da recuperação, pelo usuário. Mas, ao mesmo tempo, um tesouro mal construído pode ser uma ferramenta limitadora, porque exige que usuário e indexador selecionem o mesmo termo para representarem um conceito particular do assunto. O rigorismo no controle terminológico acarretará uma baixa recuperação. Um bom tesouro deve possuir a flexibilidade necessária para indicar qual termo deve ser usado entre diversos sinônimos; distinguir diferentes homônimos; demonstrar os relacionamentos semânticos entre os termos e especificar a ordem e a forma correta das palavras.

A macro organização dos tesouros pressupõe: uma divisão hierárquica do conhecimento, comum também aos sistemas tradicionais de classificação bibliográfica, que tem sua base na lógica aristotélica e no conceito de indução de Bacon (TÁLAMO et al, 1990). Os tesouros podem oferecer, em parte ou no todo, a lista de categorias de conceitos compreendidos em um certo domínio do conhecimento e permitem estabelecer os relações hierárquicas e

analógicas que unem os termos de um mesmo campo semântico (AUGER; ROUSSEAU, 1977, p. 19). As subdivisões em classes podem, também, obedecer ao princípio de divisão dicotômica de Porfírio, segundo o qual, do objeto de estudos podem ser derivadas apenas duas classes: objeto e não-objeto (X e não-X). Nos tesouros, esse princípio pode ser utilizado para a obtenção das relações hierárquicas partitivas e genérico específicas. É falha, no entanto, a aplicação de tal princípio para a derivação de outros tipos de relações, uma vez que a divisão dicotômica baseia-se na análise da natureza do objeto considerando apenas uma de suas propriedades por vez. Isso significa, por exemplo, que não é possível analisar-se a natureza corpórea do homem – orgânica ou inorgânica – em contraposição à sua longevidade – mortal ou imortal. Para as relações não hierárquicas, os tesouros podem lançar mão dos pressupostos das teorias de classificação mais modernas, como a classificação facetada (TÁLAMO et al, 1990).

De acordo com a área de conhecimento a que se referem, segundo Dahlberg (1978b, p. 17), os tesouros podem ser classificados em:

- a) tesouros que dizem respeito a objetos (taxonomias): minerais, plantas, animais, etc;
- b) tesouros que dizem respeito a uma disciplina (classificações facetadas);
- c) tesouros que dizem respeito a todas as disciplinas (classificações universais);

Como instrumento de controle terminológico, seria esperado que o tesouro trouxesse definições precisas das propriedades dos seus descritores. Na prática, no entanto, os termos estão interligados geralmente por três espécies de relações semânticas: hierárquicas, de equivalência e associativas. Tal organização permite realizar inferências acerca do conceito a que se refere cada termo (em relação à sua intensão e extensão) e, desse modo, a definição é obtida por meio da posição do termo na estrutura do documento, pelo grau e pelo tipo de circunvizinhança com outros termos. Dizendo-se de outra forma: a estrutura do tesouro possui natureza subjetiva; essa subjetividade será determinada pelo campo do conhecimento a que se refira a análise e, às vezes, atenuada pelo uso de notas explicativas ou pela notação utilizada para a classificação dos termos; o tesouro poderia melhor contribuir como instrumento de controle terminológico se as definições dos conceitos ocorressem de modo mais explícito (TÁLAMO et al, 1990).

Para Gomes (1990a, p. 25-27) a definição de um termo é muito importante para a elaboração do tesouro, uma vez que fornece as características do conceito que permitirão seu agrupamento no sistema de conceitos e a indicação das relações conceituais existentes. Somente a

definição do termo poderá dar a sua posição no sistema, uma vez que os termos não são definidos isoladamente, mas se definem uns em relação aos outros. A definição de um conceito pode se dar epistemologicamente – isto é, com base no nível abstrato, o que requer mais tempo – ou pragmaticamente – isto é, os termos são registrados de acordo com interpretações adequadas, o que requer menos tempo. A autora argumenta que apenas os conceitos básicos, ou seja, aqueles dos quais derivam os demais, devem ser epistemologicamente definidos, para que se possa ganhar tempo na elaboração do instrumento.

A formação de um termo para a representação de um conceito em um tesauro pode ocorrer a partir da combinação de duas ou ^{mais} palavras, o que pode ocorrer antes da sua entrada no sistema (pré-coordenação) ou no momento da busca no sistema (pós-coordenação). A formação de termos pós-coordenados é justificável quando os descritores simples e/ou o termo pré-combinado são usados raramente. A formação de termos pré-coordenados deve ser sempre usada quando: a) os significados dos termos simples, tomados isoladamente, diferirem do significado do termo pré-coordenado; b) os descritores simples forem utilizados em conexões hierárquicas diferentes dos descritores pré-coordenados; ou c) o descritor pré-coordenado for um nome próprio (ISO 2788:1974). Cavalcanti (1978) sugere a utilização da pré-coordenação quando os conceitos representados por cada um dos termos, tomados isoladamente, são diferentes do significado expresso pelo termo composto ou podem representar conceitos ambíguos. A autora exemplifica com os termos *padrão* e *vida*. Quando o termo *padrão* antecede *vida* (*padrão de vida*), o significado pode relacionar-se, por exemplo, com o nível de qualidade dos bens materiais ou dos aspectos culturais, de educação e de lazer que alguém consegue obter em função dos seus rendimentos pessoais; Quando o termo *vida* antecede o termo *padrão* (*vida padrão*), o significado pode referir-se ao tipo de vida de alguém, que serve de modelo para outras pessoas. Se a combinação dos termos simples não dá lugar a ambigüidade, utiliza-se a pós-coordenação. Por exemplo, a combinação dos termos *instituição* e *educação* transmitirá sempre a idéia de instituição educacional.

Para Aitchison e Gilchrist (1979, p. 46), a adoção de termos pré-coordenados pode ser muito vantajosa, pois assegura grande precisão aos resultados de uma busca, uma vez que tais termos são mais exatos e mais específicos. Por outro lado, pode ser desvantajosa, uma vez que aumenta o número de termos no vocabulário e pode acarretar falhas na revocação.

Com relação ao método de coleta dos termos, há dois métodos: o dedutivo e o indutivo. O método indutivo consiste em utilizar os termos encontrados nos textos para a construção do vocabulário. O método dedutivo, por sua vez, extrai todas as palavras naturais que

fazem parte de dado campo semântico, para o qual se criam listas, glossários, dicionários especializados, classificações, entre outros instrumentos de controle terminológico. Cavalcanti (1978, p. 32) indica três métodos que podem ser utilizados para a seleção dos termos que comporão o vocabulário básico do tesauro:

a) analítico – baseia-se no exame de certa quantidade de documentos, dos quais são extraídos os termos significativos, que serão incluídos em inventário dos conceitos; os termos sinônimos são incluídos mediante o uso de remissivas; o arranjo hierárquico dos termos é realizado posteriormente;

b) sintético – os termos são coletados em dicionários, índices de obras, glossários, entre outras obras, relativos a um campo semântico dado;

c) a priori – baseia-se em listas apresentadas por grupos de especialistas em um dado campo do conhecimento; de tal trabalho resulta um esquema preciso, mas que pode apresentar problemas tais como a ausência de termos importantes ou a presença de termos que nunca são utilizados.

A autora conclui que o melhor método seria a reunião dos três anteriores: análise da terminologia encontrada em documentos, análise comparativa dos termos colhidos em obras de referência (dicionários, glossários, entre outras), consulta a especialistas.

Em relação ao controle terminológico, alguns cuidados são sugeridos pela norma ISO 2788-1974: a) a grafia adotada para um termo deve ser a mais amplamente conhecida; b) quando um termo originário de uma língua estrangeira e a sua tradução coexistirem com o mesmo significado, ambos devem ser inseridos no tesauro, realizando-se remissão de um a outro; c) na transliteração de termos estrangeiros é recomendável evitar-se, o quanto possível, o uso de diacríticos; d) é preferível que os descritores sejam utilizados na forma nominal (substantivos) do que em outras classes gramaticais embora, em alguns casos, seja aceitável o uso de adjetivos; e) em relação ao número (singular ou plural), a escolha de um padrão deve ser feita com base nas características da língua de especialidade e, quando os significados dos termos forem diferentes no plural e no singular, ambos os termos devem ser incluídos no tesauro; f) abreviaturas e acrônimos devem ser evitados, uma vez que podem não ser suficientemente genéricos ou seu significado pode depender do contexto – termos abreviados e sua forma por extenso devem ser tratados como referências cruzadas; g) caracteres de pontuação ou caracteres especiais devem ser evitados, sendo permitido o uso de parênteses e hífen. A norma do Unisist (1975, p. 10) sugere que a omissão dos sinais de pontuação seja referenciada em uma nota explicativa. Cavalcanti (1978, p. 37) sugere os seguintes critérios para o registro das entradas em um tesauro, de modo a

aprimorar o controle terminológico: a) o termo deve aparecer, na maioria das vezes, sob a forma de substantivo; quando necessário, quando inevitável, e quando, na pós-coordenação, os termos compostos não tiverem significados diferentes, deve-se utilizar o substantivo adjetivado; b) em relação ao número, deve-se preferir o singular para termos indicativos de materiais, propriedades específicas, processos, nomes próprios, disciplinas; deve-se empregar o plural para termos genéricos que designam classes de objetos como estrelas, horas de trabalho, relações humanas, entre outros; c) somente devem ser utilizadas abreviaturas quando essas forem indicadas nas instruções do tesouro; d) somente devem ser utilizadas siglas quando essas forem conhecidas pelos usuários do sistema, em um domínio específico do conhecimento; e) em relação ao uso de sinais diacríticos e sinais de pontuação, o hífen só deve ser utilizado quando for parte integrante da palavra; os parênteses devem ser utilizados para incluírem palavras de diferenciação de termos homógrafos (Vênus – planeta; Vênus – mitologia); o ponto, a vírgula, os dois pontos, o apóstrofo e outros sinais de pontuação devem ser omitidos; f) elementos numéricos de um termo devem ser indicados em algarismos arábicos, mesmo que sejam números romanos ou ordinais; g) termos homógrafos devem ser diversificados pela adjetivação (regime jurídico; regime político).

Aitchison e Gilchrist (1979, p. 29-37) estabelecem os seguintes princípios que visam controlar os termos em um tesouro: a) os termos devem ser registrados na forma substantiva, sendo evitados os verbos; em frases substantivas, deve-se omitir a preposição: *indústria pesqueira*, em vez de *indústria da pesca*; b) os adjetivos devem ser evitados ou, quando ocorrerem, devem coordenar um substantivo; c) termos que designam processos e propriedades, ou que sejam substantivos que denotem massa, ocorrem no singular; d) termos que designam objetos e entidades, ou que sejam substantivos contáveis, ocorrem no plural; e) termos compostos devem ser registrados na ordem direta; f) a grafia correta de um termo deve ser estabelecida pela consulta a um dicionário ou enciclopédia; g) são admitidas variações quanto ao uso de letras maiúsculas ou minúsculas, devendo ser adotado um padrão; h) devem ser estabelecidas regras para a transliteração – letras gregas são traduzidas foneticamente: α = alfa, π = pi; i) deve ser evitado o uso de abreviaturas, a menos que a forma abreviada seja de mais fácil manipulação do que a forma extensa: PTFE em vez de *Politetrafluormetileno*; j) sinais de pontuação devem ser evitados, sendo aceitos os parênteses para definirem qualificadores e os hífen, quando fizerem parte do termo; k) termos homógrafos devem ser diferenciados mediante o uso de qualificadores; l) em tesouros tradicionais não são utilizadas definições completas, mas deve-se incluir notas explicativas e definições sucintas, quando necessário.

Em relação ao controle de ambigüidade, a norma ISO 2788:1974 propõe o uso de qualificadores que, colocados entre parênteses, devem ser aplicados a termos homógrafos, com o

objetivo de distingui-los entre si. Assim, o termo homógrafo e seu qualificador formam um termo composto. Além dos qualificadores, constituem-se elementos de controle de ambigüidade as notas de escopo, as definições e as traduções. Uma nota de escopo ou nota explicativa é uma breve explicação do sentido em que o termo é empregado. Tal nota é colocada entre parênteses e destina-se a restringir o uso do descritor ou a explicar termos abreviados ou acrônimos. Uma definição, colocada entre parêntesis após o termo, especifica o seu conteúdo conceitual, mas não é considerada parte de tal termo. Uma tradução compreende a apresentação de um termo equivalente em outro idioma, quando for necessária para a compreensão do conceito a que se refere tal termo. De acordo com a norma do Unisist (1975, p. 12), a indicação da fonte bibliográfica do descritor ou de sua definição, pode ser importante para a evolução do instrumento, sendo recomendável o seu registro. Gomes (1990a, p. 54) define essas notas explicativas como notas de aplicação, necessárias à restrição ou ampliação do conteúdo conceitual de dado termo. As notas podem incluir data de inclusão ou exclusão do termo no sistemas, explicitam a definição do conceito e são indicadas em letras minúsculas, diferentemente das utilizadas no termos de entrada.

Conforme Cavalcanti (1978, p. 34), em relação à sua estrutura, o tesauro apresenta uma lista de entradas individuais (descritores), ordenadas alfabeticamente e também segundo as suas relações recíprocas. Cada descritor é, na verdade, um termo selecionado para representar um conceito, de preferência sem nenhuma ambigüidade. A organização dos termos pode se dar por meio da subdivisão de áreas específicas do conhecimento, que, por sua vez, podem ser subdivididas em categorias, que se sub-dividem em grupos. Com relação à sua composição, os termos podem ser simples, formados por apenas uma palavra, ou compostos, formados por mais de uma palavra. É importante lembrar-se que termos compostos podem representar conceitos simples (ex.: cana-de-açúcar, cruz vermelha, fruta-pão) e que termos simples podem representar conceitos compostos (ex.: menina – representa um conceito composto por duas idéias: criança + sexo feminino) (CAVALCANTI, 1978, p. 21 e 37).

O estabelecimento da estrutura de um tesauro é dependente das relações conceituais existentes no sistema. Aspectos lingüísticos como a sinonímia, a polissemia e a forma do termo podem determinar diversos tipos de relações entre os termos. O fenômeno da sinonímia é responsável pela origem de relações de equivalência. São considerados termos equivalentes: a) termos de origem lingüística diferente – antídoto = contraveneno; b) nomes populares e nomes científicos – alergia = hipersensibilidade; c) substantivos comuns e nomes comerciais – gilettes = lâminas de barbear; formas variantes para conceitos emergentes – contêineres = *containers*; grafias diferentes, inclusive com variações de radical – quociente =

cociente; termos de uso corrente e termos antigos – deontologia = ética; termos de culturas diferentes – pesquisa (Brasil) = investigação (Portugal); abreviaturas e nomes por extenso – FID = Federação Internacional de Documentação. Termos polissêmicos, isto é, que expressam diferentes conceitos, dentro de uma mesma área do conhecimento, normalmente deve ser seguidos por um qualificador que explicita os seus diferentes significados ou contextos: *tradução* (processo) e *tradução* (documento). Em casos como esse, deve ser escolhida a definição mais útil aos propósitos do sistema de conceitos, o que definirá o posicionamento de tal termo no sistema. Em relação à forma, termos figurados ou metáforas podem ocorrer com frequência na linguagem científica. Por esse motivo, é também possível que sua utilização seja necessária na construção de uma linguagem documentária, de modo a garantir a explicitação ou definição de novos conceitos por meio de tais metáforas (GOMES, 1990a, p. 47-53). Um exemplo é a expressão *farelo de soja*: embora a palavra *farelo* signifique resíduo de farinha, pode ilustrar adequadamente um processo similar de trituração da soja.

Segundo as diretrizes do IBICT (1984, p. 7-16), o controle terminológico em tesouros pode ser realizado por meio da estruturação consistente, embasada no princípio da monorreferencialidade, isto é, para cada termo há a correspondência de apenas um conceito; tal princípio deve transparecer na estrutura do tesouro, seja por meio da sistematização dos termos segundo as relações conceituais do sistema, seja pela adição de notas de aplicação. Outras recomendações dessas diretrizes são: a) casos de sinonímia devem ser tratados pela escolha de um termo preferencial que permita a recuperação a partir de qualquer ponto de acesso empregado pelo usuário na busca; b) termos compostos, que podem gerar ambigüidade, devem ser tratados como uma unidade conceitual, mesmo que grafados sem hífen – por exemplo, *carro de combate* ou *cobra-cipó*; c) termos homógrafos devem ser seguidos por qualificadores entre parênteses, explicitando-se, assim o seu significado; d) abreviaturas e siglas devem ser evitadas, a menos que sejam amplamente utilizadas e conhecidas; em caso de uso, deve-se fazer remissão à forma extensa; e) substantivos que denotam objetos contáveis (discretos) devem ser expressos no plural; partes do corpo e nomes de órgãos serão grafados no plural, se existirem em número de dois ou mais – dedos, pés, olhos; f) substantivos que denotam objetos contínuos, como materiais ou substâncias, devem ser grafados no singular – neve, chuva, vidro; caso denotem uma classe com mais de um membro, no entanto, ocorrerão no plural – plásticos; inseticidas; g) conceitos abstratos, como sistemas de crença, propriedades, atividades e disciplinas, devem ser expressos no singular, quando referem-se a um membro de uma classe – cristianismo, dureza, migração, história – e no plural, quando referem-se a uma classe com mais de um membro – ciências sociais; i) quando o mesmo termo designa conceitos diferentes como, por exemplo, atividades,

técnicas, materiais, entre outros, pode-se fazer a distinção por meio do número, grafando-se um no plural e outro no singular, sucedidos por um qualificador – comunicação (atividade) e comunicações (técnicas); j) em relação à grafia, usa-se a mais aceita, ou recomendada pelos dicionários, fazendo-se remissão à menos aceita – câmbra = cãibra; k) empréstimos lingüísticos devem ser tratados com sua forma mais aceita: a tradução ou o termo original em outro idioma; fazendo-se remissão à forma preterida – basquetebol = *basketball*; surfe = *surf*; l) em situações de transliteração, deve-se evitar o uso de sinais diacríticos; m) o uso de gíria ou jargão oriundo de um grupo social ou profissional deve ser incluído no tesouro, caso não haja outro termo que o substitua – bóia-fria; caso contrário, deve-se preferir o descritor mais estabelecido, incorporando-se a gíria ou jargão como termo não autorizado – orelhão (use telefone público); n) quando o termo se referir a nome comum e nome comercial, deve-se selecionar o mais aceito, fazendo-se remissão ao preterido – lambreta (usado por motoneta); o) em caso de termos populares e termos científicos usados para a designação do mesmo conceito, prefere-se o mais usado na literatura – cevada (usado para *hordeum*); p) é recomendável o uso de notas de aplicação ou definições para se explicitar o sentido de termos ambíguos ou de neologismos ou para indicar modificações no sistema de conceitos: borrachas termoplásticas (exclui todos os poliuretanos após 1975); liga de cobre (liga metálica em que o cobre é o principal constituinte).

2.2.2.1.1 Macroestrutura, microestrutura e sistemas de remissivas

No caso de um dicionário, a macroestrutura é fornada pelos seguintes elementos: a delimitação da área de conhecimento, o prefácio, a apresentação, as referências bibliográficas (CAVALCANTI, 2000).

De acordo com a norma ISO 2788:1974, em relação à estrutura, os tesouros devem conter uma introdução, uma parte principal e partes auxiliares. Segundo as normas do Unisist (1975, p. 4), o arranjo pode ser alfabético, de modo similar ao utilizado nos índices, diferindo apenas no fato de que os tesouros apresentam apenas termos, sem os endereços ou localizadores característicos dos índices. A classificação dos termos é apresentada hierarquicamente e apresenta as relações conceituais, de modo a facilitar a indexação e a recuperação. São admitidos termos mais genéricos e termos mais específicos – a norma do Unisist os denomina descritores. São aceitos como descritores em um tesouro os termos que designam conceitos ou combinações de conceitos e os termos que designam entidades individuais como: nomes de projetos; nomenclaturas; números ou símbolos de identificação; nomes geográficos; marcas; pessoas; organizações; siglas.

Microestrutura de um dicionário é o conjunto de informações ordenadas que seguem a entrada; tal conjunto possui uma estrutura constante e aplicável a qualquer entrada, dentro da mesma obra (CAVALCANTI, 2000).

Em relação à apresentação e arranjo do tesouro, Cavalcanti (1978, p. 49) argumenta que são elementos essenciais: introdução clara e completa que defina sua estrutura, finalidade e áreas de conhecimento a que se refere; menção às normas seguidas para a elaboração, fontes e métodos utilizados para a recolha dos termos; explicação da forma de apresentação, do significado das abreviaturas, bem como das regras de classificação utilizadas. No que define como parte principal do tesouro, a autora sugere o seguinte conteúdo:

a) representações dos conceitos – código numérico e/ou notações; descritores; não descritores ou termos proibidos; indicador de relações de equivalência e referências cruzadas (remissivas);

b) informações suplementares – definições, notas explicativas e indicações das fontes;

c) relações entre conceitos – indicações dos conceitos superordenados e subordinados; indicações das relações associativas.

Cavalcanti (1978, p. 50) recomenda, também, a inclusão de dispositivos auxiliares, como o índice alfabético e o índice sistemático. O índice alfabético deve ser utilizado quando a parte principal tiver sido organizada de modo sistemático; quando a parte principal apresentar arranjo combinatório alfabético-sistemático; ou quando a parte principal incluir apenas os descritores, sem as remissivas de termos sinônimos e outros. O índice sistemático é recomendado quando a parte principal tiver sido organizada alfabeticamente ou quando apresentar arranjo alfabético sistemático. Os índices alfabéticos devem registrar os descritores e os não-descritores e é aconselhável utilizar também índices permutados. Os índices sistemáticos devem registrar todos os descritores de acordo com as relações hierárquicas constantes do tesouro. Gomes (1990a, p. 18) argumenta que os conceitos de um dado campo do conhecimento formam um sistema de conceitos ou parte dele, o que implica um necessário relacionamento entre tais conceitos. Dada a existência dessas relações, é possível a ordenação dos conceitos sistemicamente e é recomendável que, além da parte alfabética, os tesouros apresentem uma parte sistemática.

A norma Unisist (1975, p. 18) determina que os tesouros devem conter a seguinte estrutura: a) introdução – define sua finalidade, sua estrutura, áreas de conhecimento que abrange,

significado das abreviaturas e símbolos, bem como as regras adotadas para sua elaboração; b) parte principal – inclui uma apresentação sistemática e uma alfabética, incluindo informações completas sobre cada termo: representação (código e termo propriamente dito), informações complementares (definição, nota explicativa, fonte) e relações conceituais (explicitadas no arranjo sistemático, por meio das siglas adotadas para cada caso, podendo haver, no caso de relações hierárquicas, vários termos genéricos ou específicos para cada descritor dado, uma vez que as relações são polierárquicas); c) partes auxiliares – índices alfabéticos, listas sistemáticas e representações gráficas.

Para a norma Unisist (1975, p. 22), as relações conceituais numa representação gráfica podem ser indicadas por meio da inscrição dos termos em um sistema de coordenadas; pela utilização de círculos concêntricos, indicando os diferentes níveis hierárquicos; mediante o uso de setas bidirecionais para as relações associativas, setas unidirecionais para as relações hierárquicas, apontando para o descritor mais específico, bem como de colchetes e setas que chegam ou partem do termo preferido, nas relações de equivalência.

Gomes (1990a, p. 56-65) recomenda que o tesauro seja composto pelas seguintes partes: apresentação sistemática, parte alfabética, apresentação planigráfica e outros elementos da apresentação. A parte sistemática apresenta os termos organizados por categorias ou classes e procura evidenciar as relações existentes entre os termos no sistema de conceitos. É indispensável que o instrumento contenha essa parte, uma vez que ela: permite a localização do termo que melhor representa dado conceito, se a necessidade de se especificar anteriormente qual o termo mais adequado; facilita a compreensão das relações conceituais no sistema e indica o grau de intensidade dos termos; facilita a manutenção do tesauro, uma vez que as características que levam à formação das classes e subclasses são evidenciadas. A parte alfabética, mais tradicional, permite a localização rápida de um termo, mas apresenta a deficiência de não evidenciar visualmente as relações conceituais no sistema – as diferenciações são feitas por meio dos códigos atribuídos aos termos. A apresentação planigráfica traz os termos representados hierarquicamente em gráficos, sendo os mais comuns: diagrama em árvore; diagrama em cadeia; gráfico com característica de divisão; diagrama em chave; ou outras formas combinadas – ver ilustração 26. Outro elemento que pode ocorrer nos tesauros é o índice permutado.



Ilustração 26: Combinação de diagramas de apresentação de tesouros
 Fonte: (GOMES; CAMPOS; MOTTA, 2004)

A norma do Unisist (1975, p. 5) recomenda que as entradas de um tesouro sejam formadas pelo menor número possível de palavras, de preferência apenas uma, sendo aceitos descritores compostos, os quais devem ser registrados na ordem direta.

De acordo com as diretrizes do IBICT (1984, p.1-2), quanto às suas finalidades, um tesouro pode: a) realizar controle terminológico, servindo como instrumento de tradução da linguagem natural dos autores, usuários e indexadores, para uma linguagem mais controlada; b) uniformizar, por meio do uso de um padrão coincidente, os procedimentos de indexação de profissionais em uma instituição ou numa rede cooperativa; c) propiciar uma lista com o número de termos necessário à explicitação dos conceitos expostos pelos autores de uma área; d) auxiliar a busca, fornecendo termos de adequados para a estratégia de recuperação. Quanto à sua estrutura, o tesouro explicita as relações existentes entre os termos – sinonímicas, hierárquicas ou outras – apresentadas num arranjo sistemático, determinado pelos elementos classificatórios de certa área do conhecimento.

De acordo com a norma ISO 2788-1986, os termos e suas relações em um tesouro devem ser apresentados por meio de três maneiras:

- a) apresentação alfabética - descritores, notas e aplicação e relacionamentos em uma lista alfabética;
- b) apresentação sistemática do índice alfabético - lista alfabética e agrupamento por categorias;
- c) apresentação planigráfica com índice alfabético- representação gráfica;

2.2.2.1.2 Relações entre descritores em Tesouros

Consoante a norma ISO 2788:1974, uma função indispensável aos tesouros é a representação das inter-relações conceituais pela representação das inter-relações entre os termos.

De tal modo, a rede de relações de um termo com outros fornece um tipo de definição pela localização do termo em um espaço semântico. As relações conceituais podem ser de três tipos: relações de equivalência, relações hierárquicas e relações de associação (ilustração 27).

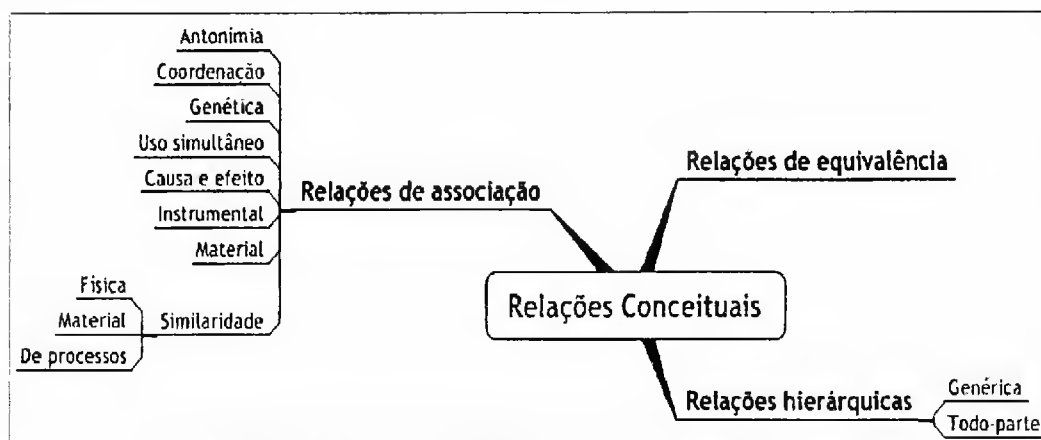


Ilustração 27: Relações conceituais segundo as normas Unisist e ISO 2788
 Fontes: (Unisist, 1975, p. 13-17; ISO 2788, 1974, p. 126-128)

As relações de equivalência ocorrem quando os termos são considerados similares, isto é, possuem quase o mesmo sentido. Assim, tais termos podem ser combinados em categorias de equivalência, sendo associados ao mesmo conceito de maneira que, no momento da busca, qualquer que seja o termo pesquisado, aquele conceito pertinente à categoria seja recuperado.

Para a norma do Unisist (1975, p. 13), as relações de equivalência são também denominadas relações preferenciais e utilizam sistemas de remissivas para orientar o emprego de termos sinônimos – que possuem o mesmo ou quase o mesmo sentido em dado campo do conhecimento – ou de quase-sinônimos – cujos significados podem diferir num vocabulário especializado, mas que são considerados como sinônimos para fins do sistema de documentação em questão. O sistema de remissivas pode ser utilizado para indicar um sinônimo que deve ser preferido; para remeter de um termo específico para um mais genérico que o represente; para indicar uma preferência ortográfica ou estender ou explicar uma abreviação; para recomendar o emprego de dois ou mais termos para expressar um conceito (fatoração semântica); para exprimir conceitos que possam ser considerados como sinônimos para fins de indexação ou recuperação; para remeter entradas indiretas para a ordem natural das palavras; para indicar a terminologia mais usual; para indicar termos técnicos no lugar do jargão; para indicar traduções (UNISIST, 1975, p. 15).

As relações hierárquicas expressam relações entre conceitos superordenados e conceitos subordinados e dividem-se em relações genéricas e relações todo-parte. Numa relação

genérica, o termo superordenado (genérico) denota uma classe de conceitos da qual o conceito subordinado (específico) é sempre um membro; o conceito específico difere do conceito genérico por pelo menos uma característica. Numa relação todo-parte, o termo superordenado (entidade) denota um objeto ou conceito do qual o termo subordinado (parte) é uma parte. Os conceitos parte são obtidos pela decomposição do todo. As classes de conceitos que enquadram-se nesse tipo de relações são: localizações geográficas; sistemas e órgãos do corpo de seres vivos; domínios do conhecimento. Em outros sistemas em que haja consenso nas opiniões dos membros de uma comunidade específica, as relações hierárquicas podem ser do tipo todo-parte, como por exemplo, para a caracterização mecanismos ou objetos e suas partes (ISO 2788:1974 e UNISIST, 1975, p. 15).

As relações associativas são geralmente empregadas para cobrirem outras relações entre conceitos que não as hierárquicas ou as de equivalência. As relações associativas podem ser utilizadas, por exemplo, para indicar: a) antonímia, isto é, a oposição de um conceito a outro; b) coordenação, isto é, os conceitos relacionam-se porque são derivados de um mesmo conceito superordenado e encontram-se no mesmo nível hierárquico; c) relação genética, isto é, um conceito é predecessor de outro; d) conceitos de uso concorrente; e) relações de causa e efeito; f) relações instrumentais; g) relações de similaridade entre tipos de material ou processo. Além desses tipos de relações associativas, a norma admite que poderá haver inúmeros outros, desde que sejam efetivos para a explicitação da relação entre os conceitos em questão (ISO 2788:1974 e UNISIST, 1975, p. 16-17).

Em termos de simbologia, a norma ISO 2788 propõe que as relações entre os conceitos sejam expressas de acordo com a tabela 15.

Tabela 15: simbologia da representação de relações de acordo com a normas ISO 2788:1974, Unisist (1975) E IBICT (1984)

Legenda das abreviaturas						
ISO 2788		UNISIST		IBICT		Descrição
BT	Broader Term	TG	Termo Genérico	TG	Termo Genérico	quando as relações genéricas e partitivas não podem ser distinguidas
NT	Narrower Term	TE	Termo Específico	TE	Termo Específico	quando as relações genéricas e partitivas não podem ser distinguidas
BTG	Broader Term Generic	TGG	Termo Geral Genérico	TGM	Termo Genérico Maior	quando as relações genéricas e partitivas podem ser distinguidas (termo genérico)
BTP	Broader Term Partitive	TGP	Termo Geral Partitivo	TGP	Termo Genérico Partitivo	quando as relações genéricas e partitivas podem ser distinguidas (termo entidade)
NTG	Narrower Term Generic	TEG	Termo Específico Genérico	—	—	quando as relações genéricas e partitivas podem ser distinguidas (termo específico)
NTP	Narrower Term Partitive	TEP	Termo Específico Partitivo	TEP	Termo Específico Partitivo	quando as relações genéricas e partitivas podem ser distinguidas (termo partitivo)
RT	Related Term	TC	Termo Correlato	TA	Termo Associado	relação associativa
USE	Use	US	Usar	USE	Use	relação de sinonímia, quase-sinonímia ou combinação de descritores
UF	Used For	UP	Usado Para	UP	Usado Por	relação de sinonímia, quase-sinonímia ou combinação de descritores
UFC	Used for Combination			—	—	combinação de descritores simples
				NA	Nota de aplicação	explicação concisa sobre o modo de emprego de um descritor
—	—	—	—	TO	Termo Oposto	conceito contrário, contraditório ou oposto

De acordo com as diretrizes do IBICT (1984, p. 21-35), há três tipos de relacionamentos entre os conceitos em um tesouro: relacionamentos semânticos – ou de equivalência –, relacionamentos lógicos e relacionamentos associativos. O relacionamento de equivalência ocorre entre dois termos que designam o mesmo conceito como, por exemplo: hidrovia = via navegável. A equivalência pode se dar por sinonímia ou por deslocamento genérico. São termos sinônimos aqueles cujos significados são considerados como sendo os mesmos devido aos seguintes fenômenos lingüísticos³³: a) origem lingüística diferente; b) nomes populares e nomes científicos; c) substantivos comuns e nomes comerciais; d) variantes para neologismos; e) grafias diferentes; f) termos de uso corrente e termos antigos; g) termos de culturas diferentes; h) termos abreviados e forma extensa; i) forma composta e forma fatorada – *mineração de carvão* = *mineração*; *carvão*. A equivalência por deslocamento genérico se dá quando o nome de uma classe e seus membros são tratados como equivalentes, sendo o termo mais geral escolhido como descritor: cereais (usado por arroz, cevada, trigo); arroz: use cereais; trigo: use cereais.

Um relacionamento lógico, segundo as diretrizes do IBICT (1984), pode ser hierárquico, lateral, ou de oposição. Um relacionamento hierárquico é o que mais distingue um tesouro sistemático de outras listas organizadas de termos, pois exprime os graus ou níveis de superordenação e de subordinação entre os conceitos. Tais relacionamentos podem ser classificados em: genérico, partitivo e polierárquico. O relacionamento genérico identifica a ligação entre uma categoria de conceitos e seus membros, sendo caracterizado pela comparação de características entre tais conceitos. Uma maneira de se verificar a validade desse relacionamento é assegurar-se de que os termos representam o mesmo tipo de conceito, denotando operações, coisas ou propriedades. Uma outra forma sugerida é a aplicação de um teste, a que as diretrizes do IBICT denominam “*todos-e-alguns*”, que consiste basicamente na seguinte verificação: dado os conceitos A e B, verifica-se que todo A é também B e que alguns B são também A. Por exemplo: todos os gafanhotos são insetos; alguns insetos são gafanhotos. Assim, A é uma espécie do gênero B.

O relacionamento partitivo abrange certo número de classes de termos, em que o nome da parte subentende o nome do todo. Os relacionamentos partitivos são basicamente de quatro tipos: sistemas e órgãos do corpo; localidades geográficas; disciplinas ou áreas do discurso; estruturas sociais hierarquizadas (IBICT, 1984).

³³ Ver também, na página 228 desta pesquisa, as definições de sinonímia por Gomães (1990a).

Um relacionamento polierárquico ocorre quando um conceito pode ser designado como membro de mais de uma classe simultaneamente. Por exemplo: *veículos de transporte* classificam-se em *rodoviários* e *de transporte de passageiros*; um *ônibus* pode enquadrar-se em duas classes: na de veículos rodoviários e na de veículos de passageiros (IBICT, 1984).

O relacionamento lateral refere-se a termos similares, com significados que se sobrepõem: *navios* e *barcos*. O relacionamento de oposição pode ser classificado em: oposição contraditória; oposição contrária; e relacionamento positivo/indiferente/negativo (IBICT, 1984).

O relacionamento associativo ocorre entre termos que não são equivalentes e nem formam uma hierarquia, mas que estão ligados de alguma maneira. Segundo as diretrizes do IBICT (1984, p. 32-35), os relacionamentos associativos ocorrem entre: a) uma disciplina ou campo de estudo e os objetos ou fenômenos estudados – entomologia e insetos; entre um processo ou operação e seu agente ou instrumento – iluminação e lâmpadas; entre uma ação ou processo e o produto ou resultado da ação: tecelagem e tecido; entre uma ação e seu paciente – extradição e criminosos; entre coisas e suas propriedades – venenos e toxicidade; entre pessoas ou coisas e suas origens – brasileiros e Brasil; entre conceitos ligados por dependência causal – doença e agente patogênico; entre uma coisa e seu contra-agente – inseto e inseticida; entre expressões sincategoremáticas³⁴ e os substantivos nelas incluídos – peixes e peixes fósseis.

A ilustração 28 apresenta a classificação das relações conceituais conforme as diretrizes do IBICT (1984).

³⁴ Uma palavra sincategoremática não possui significação própria e pode ser uma conjunção, uma preposição, um advérbio, entre outros elementos lingüísticos (HOUAISS, 2004). Assim, os termos sincategoremáticos são aqueles cuja significação plena só passa a existir quando tal termo for acompanhado por outra palavra, por exemplo, em determinado domínio do conhecimento, *bagre* é uma espécie de peixe, mas *peixe fóssil* não (IBICT, 1984).

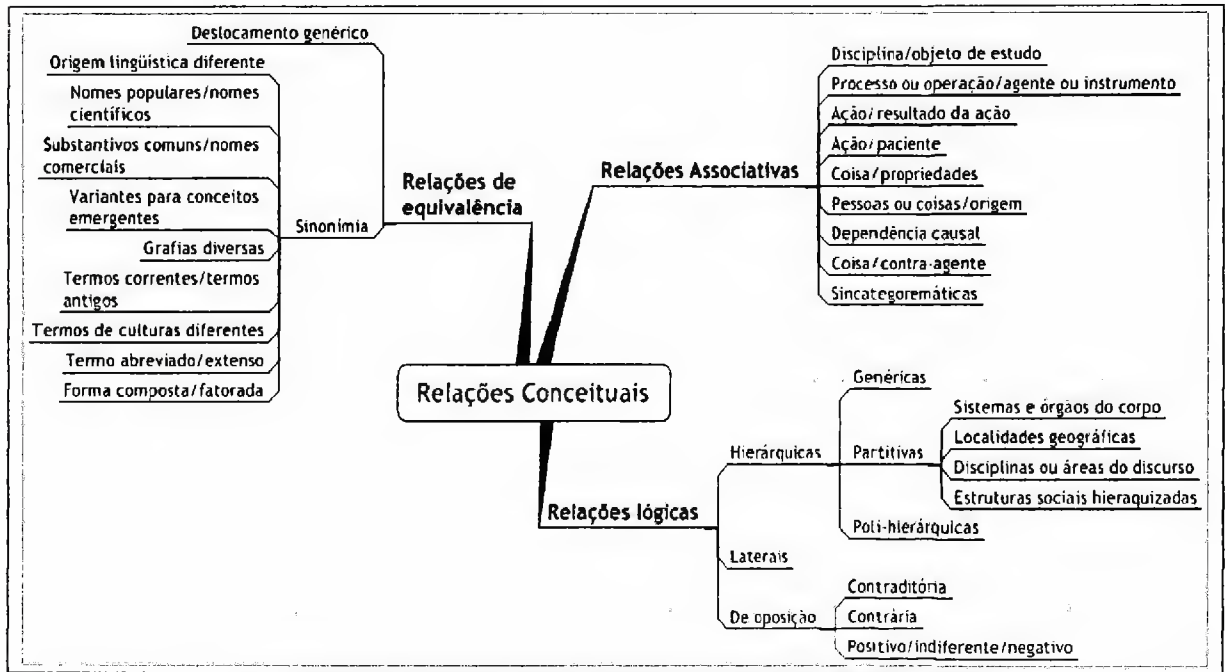


Ilustração 28: Relações conceituais, segundo as diretrizes do IBICT (1984).

Cavalcanti (1978, p. 43) destaca os três tipos principais de relações entre conceitos: equivalência, hierarquia e associação. Entre as relações de equivalência, a autora aponta:

- a) sinonímia – os termos possuem o mesmo ou quase o mesmo sentido;
- b) quase-sinonímia – os termos possuem significados diferentes num vocabulário especializado, mas são considerados sinônimos no sistema em questão.
- c) remissiva – um dos termos é considerado preferido na representação de um conceito; tais relações são utilizadas para indicar o termo preferencial (*bônus*: use *incentivo salarial*); para remeter de um termo específico para o termo mais genérico (cera vegetal: use cera); para indicar a preferência ortográfica ou explicar uma abreviatura ou sigla (*caraterística*: use *característica*); para aconselhar o uso de dois termos para a expressão de um conceito (*instituição educacional*: use *instituição e educação*).

As relações de equivalência podem ocorrer na forma de referências cruzadas, de tal maneira que sob o termo preferido haja informações sobre o preterido e vice-versa (CAVALCANTI, 1978). Quanto às relações hierárquicas, a autora admite os fenômenos de superordenação e de subordinação³⁵, que podem ser expressos pelas relações genéricas ou pelas

³⁵ Superordenação: estabelecimento da hierarquia sob os termos, com a indicação dos conceitos que fazem parte de um termo mais abrangente. Subordinação: indicação dos conceitos que fazem parte de um conceito mais amplo.

relações partitivas. Cavalcanti (1978) destaca, também, as relações de polierarquia, considerando-se hierarquia o encadeamento entre conceitos, baseado na ordem de precedência dos assuntos que esses conceitos representam. A mono-hierarquia ocorre quando o conceito representado pelo termo está subordinado apenas a um termo mais geral; a polierarquia ocorre quando o conceito representado pelo termo em questão está subordinado a mais de uma cadeia hierárquica, isto é, há um relação com mais de um conceito amplo ou genérico que se encontra em nível imediatamente superior. Quanto às relações associativas, a autora destaca as seguintes: a) relação de oposição – os termos representam conceitos opostos; b) relação genética – um objeto precede o outro em uma estrutura genealógica (pai – filho); c) relação de coordenação – os termos em questão derivam do mesmo termo genérico; d) relação de causa e efeito – os conceitos relacionam-se pelo fato de que a ocorrência de um deles determina a existência do outro; e) relação instrumental – um dos conceitos é utilizado como instrumento para obter-se o outro; f) relação material – um conceito origina o outro; g) similaridade – os conceitos pertencem a domínios similares do conhecimento ou representam objetos materiais ou imateriais que possuem similaridade em relação às suas características.

Aitchison e Gilchrist (1979, p. 50-3) admitem os seguintes tipos de relações: a) relações de equivalência, que incluem os sinônimos e os quase sinônimos; b) hierárquicas verdadeiras, que incluem as do tipo gênero/espécie e coisa/tipo; c) relação todo-parte, que os autores não consideram como relação hierárquica, a não ser quando se trata de localizações geográficas – França/Paris; d) relações polierárquicas, que indicam mais de uma hierarquia para dado termo; e) relações não hierárquicas, que compreendem vários tipos de relações como: coisa-propriedade; coisa-processo; coisa-coisa como atributo; propriedade-processo; propriedade-propriedade como atributo; processo-coisa; processo-propriedade (ilustração 29).

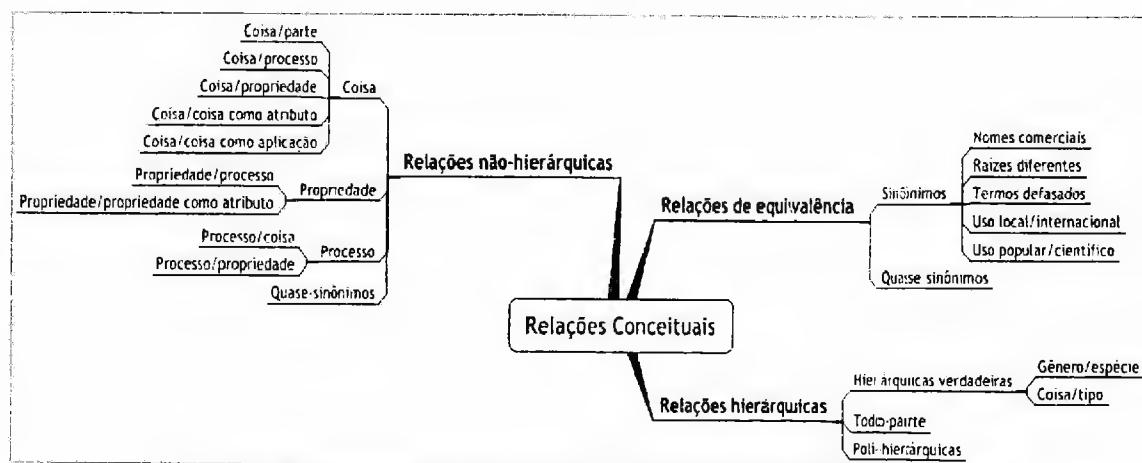


Ilustração 29: Relações conceituais segundo Aitchison e Gilchrist (1979, p. 50-56)

Para Gomes (1990a, p. 39) os tesouros mostram apenas relações dos tipos genérico-específicas, partitivas, associativas e de equivalência, embora tais categorias não sejam suficientes para a representação de todos os tipos possíveis de relacionamentos entre conceitos. Uma vez que os conceitos são definidos com base em suas características, sempre que houver características comuns entre dois ou mais conceitos, haverá relacionamento direto entre eles. Pode haver relacionamentos indiretos entre conceitos se os objetos individuais forem vizinhos no espaço ou sequenciais no tempo. Num sistema de conceitos, isto é, em uma área conceitual estruturada, cada conceito revela seu relacionamento com outros conceitos.

A ilustração 30 apresenta os diferentes tipos de relações entre conceitos que são encontrados, na concepção de Gomes (1990a, p. 40):

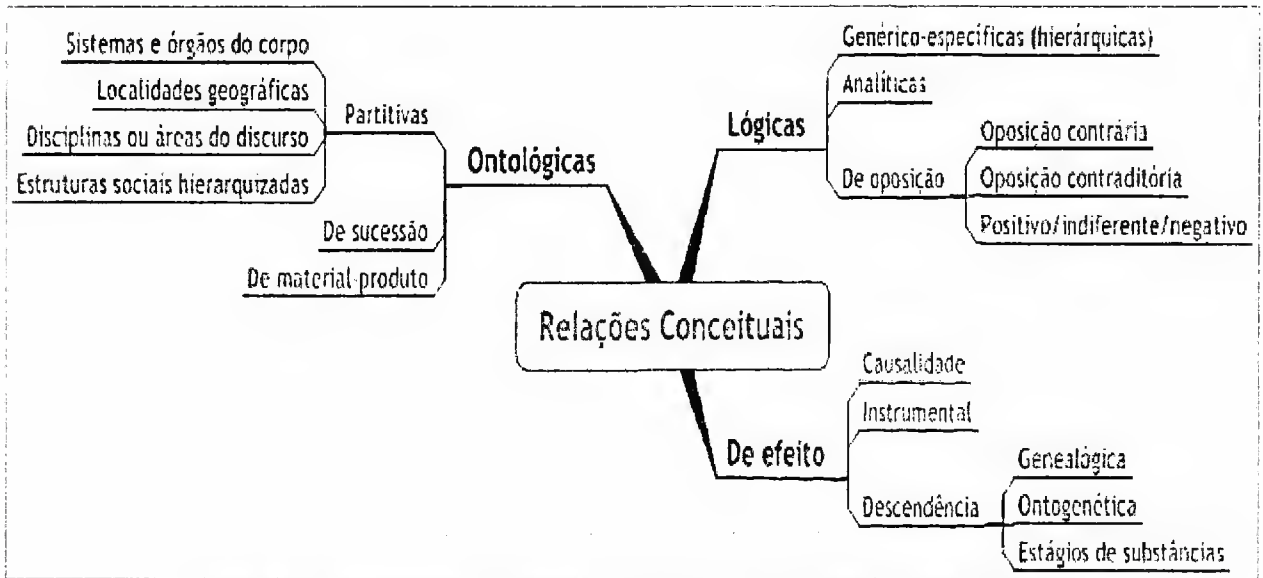


Ilustração 30: Classificação das relações conceituais segundo Gomes (1990a, p. 40-46)

As relações lógicas são relações de similaridade, isto é, são explicitadas por meio da verificação de características comuns entre os conceitos. Um relacionamento lógico genérico-específico permite a formação de classes de conceitos, sendo membros de uma mesma classe os conceitos que pertencem ao mesmo gênero. Tais conceitos membros, ou *específicos*, possuem todas as características do conceito genérico, mais uma que lhe é própria, específica. Como tal tipo estruturação forma uma hierarquia, essas relações também são ditas hierárquicas. A hierarquia é também vertical, pois liga os termos superordenados a termos subordinados. Um relacionamento lógico é dito analítico quando, entre os conceitos analisados, quando um deles for uma característica do outro e ambos não fizerem parte da mesma hierarquia. Um relacionamento lógico de oposição dá-se entre termos quase-sinônimos e pode ser classificado em três espécies:

a) oposição contraditória – numérico/não-numérico; b) oposição contrária (negação) –

amizade/inimizade; c) positivo/indiferente/negativo – muito valioso/valioso/pouco valioso (GOMES, 1990a, p. 40-43).

Relações ontológicas são relações que ocorrem de maneira indireta entre os conceitos e podem caracterizar-se pela contigüidade no tempo e no espaço ou pela conexão de causa-efeito. Um relacionamento ontológico partitivo ocorre quando se pode reconhecer em um conceito as partes que o compõem. As relações partitivas podem se dar em sistemas e órgãos do corpo – sistema nervoso/cérebro; localidades geográficas – América do Sul/Brasil; disciplinas ou áreas do discurso – ciência/física; estruturas sociais hierarquizadas – igreja católica/paróquia. Um relacionamento ontológico de sucessão é caracterizado pela contigüidade no tempo, podendo-se distinguir um conceito sucessor e seu antecessor para cada conceito no sistema. O relacionamento ontológico de material-produto mostra diferentes estágios de produção, desde a matéria-prima ao produto final (GOMES, 1990a, p. 43-45).

Os relacionamentos de efeito podem ser classificados em três tipos: de causalidade, de instrumentalidade e de descendência. O relacionamento de causalidade dá-se entre conceitos que designam a causa e seu efeito – pista molhada/acidente. O relacionamento de instrumentalidade dá-se entre o conceito relativo ao instrumento e a ação que tal instrumento provoca – broca de perfuração/broca. O relacionamento de descendência ocorre entre conceitos que possuem entre si relação genealógica, ontogenética – larva/borboleta – ou de estágios da substância – urânio I/urânio II³⁶ (GOMES, 1990a, p. 46).

Lancaster (1987, p. 26-29) categoriza as relações conceituais em hierárquicas e associativas. As relações hierárquicas abrangem as relações genéricas, partitivas e exemplares. Em relações genéricas, os termos específicos representam sempre uma espécie de coisa designada pelo seu termo genérico. Em relações partitivas, os termos subordinados não devem ser considerados como espécies, embora se convencie, em alguns casos, o tratamento de relações partitivas como se fossem genéricas. Tal é o caso de relações que designam partes do corpo, localidades geográficas, campos de estudo e estruturas sociais. Em relações exemplares, o autor define como termos exemplares aqueles que correspondem a algum nome próprio, seja de pessoas, grupos, prédios, entre outros. Assim, “*Catedral de Brasília*” refere-se a um exemplar de catedral, e não a um tipo de catedral.

³⁶ Com base nas definições do dicionário Houaiss (2004), pode-se afirmar que a genealogia refere-se ao estudo que tem por objeto estabelecer a origem de um indivíduo ou de uma família, estabelecendo um conjunto de antepassados segundo uma linha de filiação; ainda com base na mesma obra, define-se ontogenia como o desenvolvimento de um indivíduo desde a concepção até a idade adulta.

Os termos em um tesauro, geralmente estão associados a apenas um termo geral. No entanto, a depender da abrangência do tesauro, maiores as possibilidades de que a um termo estejam associados dois ou mais termos gerais. Por outro lado, quanto mais específico for o tesauro, mínimas serão as chances de esse fenômeno vir a ocorrer. Exemplificando: embora o termo *diamante* esteja associado aos conceitos de *pedra preciosa* e de *instrumento cortante*, em um tesauro relativo à metalurgia, presumivelmente o termo *diamante* estará associado ao conceito de *instrumento cortante*, e não ao de *pedra preciosa* (LANCASTER, 1987, p. 28).

As relações associativas ocorrem entre termos inter-relacionados em dado contexto que não envolva sua hierarquização. Dito de outro modo, esses termos relacionam-se inter-hierarquicamente, não intra-hierarquicamente. Termos associados são, assim, co-iguais na mesma hierarquia. A ilustração 31 resume a classificação de Lancaster para as relações conceituais.

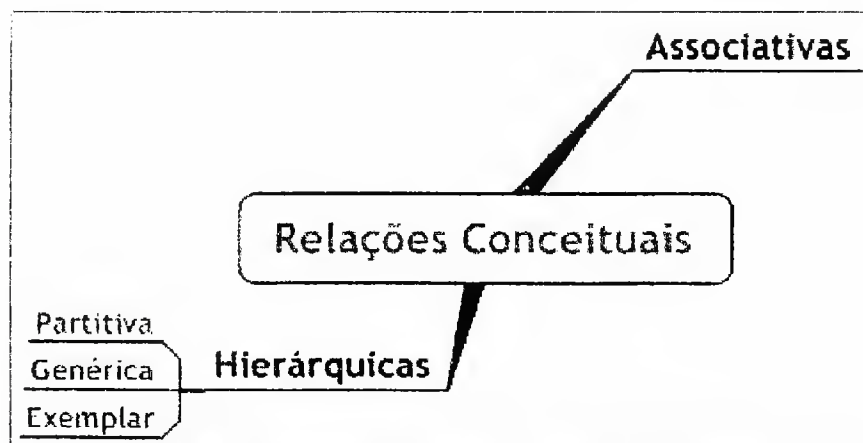


Ilustração 31: Relações conceituais segundo Lancaster (1987)

Cintra (1994) recorda que a qualquer campo do conhecimento associa-se um sistema estruturado de conceitos e que tal sistema define o universo vocabular daquele campo. Nessa direção, as linguagens documentárias encontram seu esteio nos sistemas de conceitos, uma vez que preocupam-se com o controle terminológico. Por certo, em termos de documentação, um sistema de conceitos pressupõe relações, que a autora classifica como hierárquicas e não-hierárquicas. As relações hierárquicas definem-se entre conceitos subordinados em um ou diversos níveis, de acordo com uma certa ordem. A ordem refere-se à posição hierárquica de um conceito em relação a outro. Assim, superordenação consiste na possibilidade de se subdividir um conceito em outros inferiores; analogamente, subordinação refere-se à possibilidade de agrupamento de diversos conceitos de mesmo nível em um conceito superior. Relações hierárquicas podem referir-se a dicotomias estabelecidas entre conceitos gerais e conceitos particulares, ou a dicotomias estabelecidas entre conceitos inteiros e suas partes. Assim, sugerem-

se três tipos de relações hierárquicas: genéricas, específicas e partitivas. Relações genéricas e relações específicas dão-se por meio da identidade parcial do conjunto de características dos conceitos superordenados e subordinados. O gênero, nesse ponto de vista, refere-se ao conceito superordenado, que comporta características comuns a todos os conceitos que lhe são subordinados. A espécie, por conseguinte, refere-se ao conceito subordinado, que comporta todas as características do conceito superordenado (gênero), e pelo menos mais uma característica distintiva. O conceito genérico representa, portanto, uma conjunção de características, ao passo que o conceito específico estabelece uma disjunção a partir da dada conjunção. Relações partitivas dão-se pela existência de uma dependência entre os conceitos subordinados e o superordenados, de tal modo que aqueles são partes integrantes deste. São quatro os tipos principais de relações partitivas, segundo sua classe: sistemas e órgãos do corpo; localidades geográficas; disciplinas ou ares de estudo; estruturas sociais hierarquizadas. Não há o conceito de parte sem que haja uma definição para o conceito de todo: *motor de automóvel* não pode ser definido anteriormente a *automóvel*. Partes que integram um mesmo objeto possuem entre si uma relação de coordenação (CINTRA, 1994).

Relações não-hierárquicas definem-se como todas aquelas em os conceitos não podem ser hierarquicamente organizados. Isto equivale a dizer que não é possível se pressupor a existência de ordem ou subordinação entre os conceitos. Desse raciocínio depreende-se que mesmo conceitos que participam de relações hierárquicas podem também participar de relações não-hierárquicas. É o caso, por exemplo, das partes de um objeto, que subordinam-se ao todo (relação hierárquica), mas que possuem entre si relações de coordenação (relação não-hierárquica). Noutro plano, relações não-hierárquicas podem ocorrer entre conceitos que não participam de uma hierarquia, mas que são interdependentes por causa da sua contigüidade no espaço ou no tempo. Por isso são também chamadas de relações seqüenciais. Relações seqüenciais podem se dar entre pares de conceitos como causa e efeito; produtor e produto; etapas de processos; conceitos opostos; conceitos contraditórios. Alguns autores denominam tais relações de associativas, mas Cintra (1994) rejeita esse conceito, argumentando que qualquer relação, de qualquer natureza é, em certo grau, associativa. A autora sugere uma detalhada classificação para os fenômenos que motivam relações não-hierárquicas entre termos que apresentam contigüidade espacial ou temporal. Tais relações são: a) atributiva – *economia/atividade econômica*; b) campo de estudos versus objeto estudado – *estética/beleza*; c) processo ou operação versus agente ou instrumento – *automação/computadores*; d); influência – *política monetária/inflação*; e) matéria-prima versus produto – *bauxita/alumínio*; f) coisa versus aplicação – *abastecimento de água/irrigação*; g) ação versus resultado da ação – *tecelagem/tecidos*; h) causalidade – *crescimento*

econômico/desenvolvimento econômico, i) efeito versus causa – *febre/infecção*; j) dependência causal – *doenças/agentes patogênicos*; k) atividade versus agente – *fumo/tabagismo*; l) atividade versus propriedade – *corte/usinabilidade*; m) atividades complementares – *compra/venda*; n) oposição – *vida/morte*; o) ação versus paciente – *pescado/pesca*; p) coisa ou atividade versus suas propriedades ou agentes – *venenos/toxicidade*; q) coisa versus seu contra-agente – *insetos/inseticidas*; r) atividade versus produto – *tear/tecido*; s) pessoas ou coisas versus suas origens – *brasileiros/Brasil*; t) associação implícita – *comércio internacional/balança comercial*; u) expressões sincategoremáticas versus substantivos nelas incluídos – *peixes fósseis/fósseis*; v) interfaceta – *nível de atividade econômica/política monetária*.

A ilustração 32 resume a classificação apresentada pela autora para os diferentes tipos de relações conceituais.

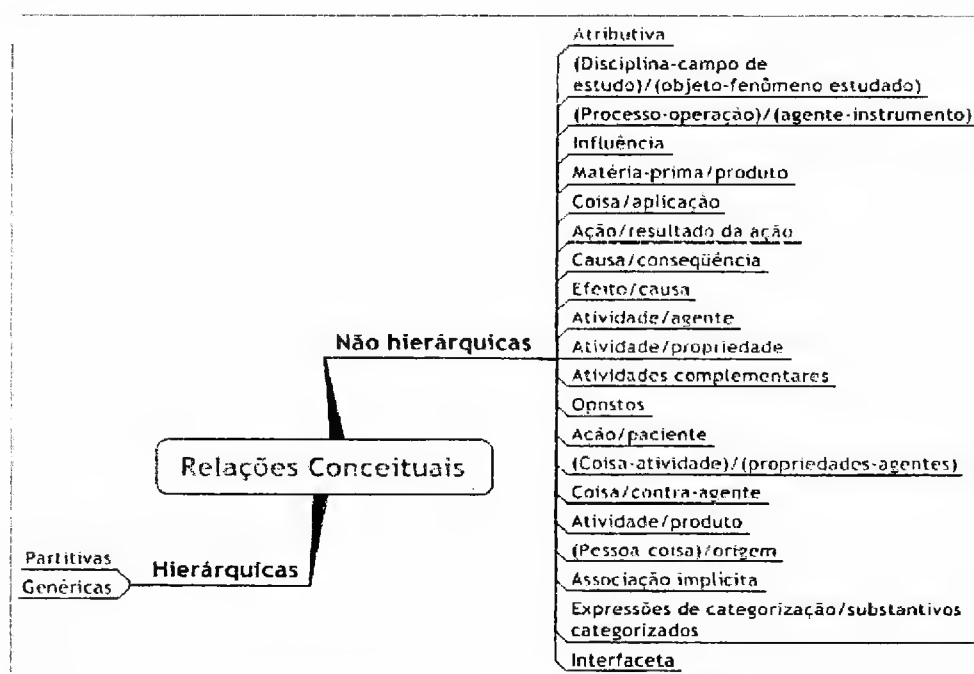


Ilustração 32: Relações conceituais segundo Cintra (1994).
Fonte: (CINTRA, 1994)

2.3 Indexação e Recuperação da Informação

2.3.1 Documentos

Um documento é um objeto que fornece um dado ou uma informação e pode ser diferenciado entre outros documentos, de acordo com suas características físicas ou intelectuais. As características físicas de um documento relacionam-se aos conceitos de material, natureza, tamanho, peso, forma de produção, suporte, entre outras. As características intelectuais relacionam-se aos conceitos de objetivo, conteúdo, assunto, tipo de autor, fonte, forma de difusão, originalidade, entre outras.

As características intelectuais de um documento permitem definir seu interesse, público alvo e valor. Entre as características intelectuais Guinchat (1994) destaca:

- objetivo – refere-se à razão pela qual o documento foi produzido: para servir como prova, para preparar outro documento, para expor idéias, para divulgar resultados de um trabalho, para ensino, entre outras;
- grau de elaboração – refere-se à autoria e à finalidade de um documento. Com base na noção de grau de elaboração é possível estabelecer-se uma classificação dos documentos em:
 - primários – documentos originais elaborados pelo autor;
 - secundários – documentos que descrevem documentos primários, como por exemplo as bibliografias, os catálogos e os resumos;
 - terciários – documentos elaborados a partir de documentos primários ou secundários e que reúnem, condensam e elaboram a informação original na forma mais adequada às necessidades de um usuário ou grupo de usuários.
- conteúdo – o conteúdo pode ser avaliado a partir da identificação do assunto, da forma de apresentação, da exaustividade da análise, do nível científico do texto, da novidade das informações, da representatividade das informações para um dado grupo de leitores, entre outras características.
- tipo – refere-se ao nível de relevância do documento; há documentos essenciais, isto é, que tratam de assuntos que possuem interesse direto para determinada comunidade de leitores e documentos marginais, ou seja, que possuem pouca ou nenhuma relevância para o usuário.

2.3.2 Ciclo Documentário

Em uma unidade de informação – que pode ser entendida como uma instituição ou sistema de informação – cada documento que entra recebe um tratamento visando a facilitar a recuperação das informações nele contidas. Esse tratamento relaciona-se a operações de seleção, avaliação, análise, tradução e recuperação de documentos capazes de responder a necessidades específicas, de acordo com o perfil do usuário e com seus objetivos. O conjunto de operações de

tratamento dos documentos é organizado num ciclo conhecido como *cadeia documentária* ou *ciclo documentário*. O ciclo documentário, como qualquer outro sistema, apresenta uma extremidade, na qual entram os documentos a serem tratados; e outra, na qual são apresentados os produtos documentários, que resultam desse processamento: referências bibliográficas, descrições de documentos, índices, instrumentos de pesquisa, documentos secundários e terciários. Os produtos documentários são utilizados em atividades de pesquisa e, novamente, transformados em documentos que realimentam o sistema (ilustração 33).

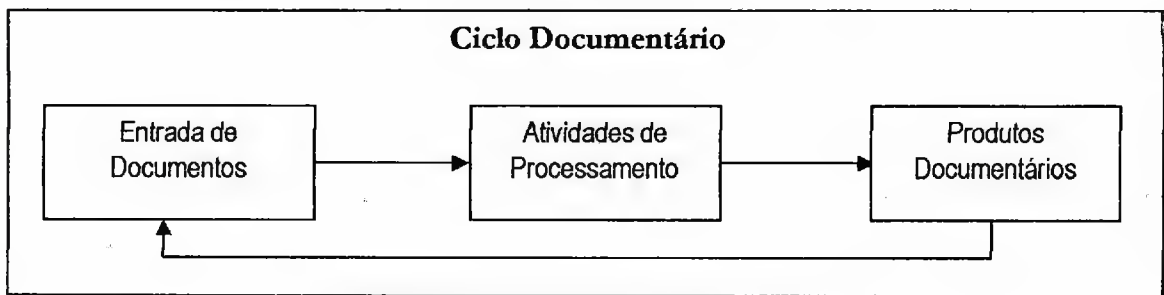


Ilustração 33: O Ciclo Documentário - visão sistêmica

As principais atividades do Ciclo documentário envolvem a coleta, o registro, o tratamento intelectual, a pesquisa e a difusão. A ilustração 34 representa um modelo esquemático para o ciclo documentário, obtido a partir da análise das definições de Guinchat (1994), de Lancaster (1993), Robredo e Cunha (1986) e de Nakayama (2001). A análise dessa ilustração permite identificar operações de entrada, de processamento técnico e de saída. As operações de entrada compreendem a seleção (que se dá pela localização e posterior escolha) e a aquisição (que pode ser realizada por meio de compra, permuta, doação, projetos, convênios, entre outras formas). As operações de processamento técnico, ou tratamento intelectual compreendem a catalogação, a classificação, a indexação e o resumo. As operações de saída correspondem àquelas que permitem tornar disponíveis as informações: armazenamento, disseminação (divulgação, distribuição, disseminação seletiva da informação), recuperação (acesso), ou alerta (forma de divulgação das novas aquisições, por meio de exposição ou impresso).

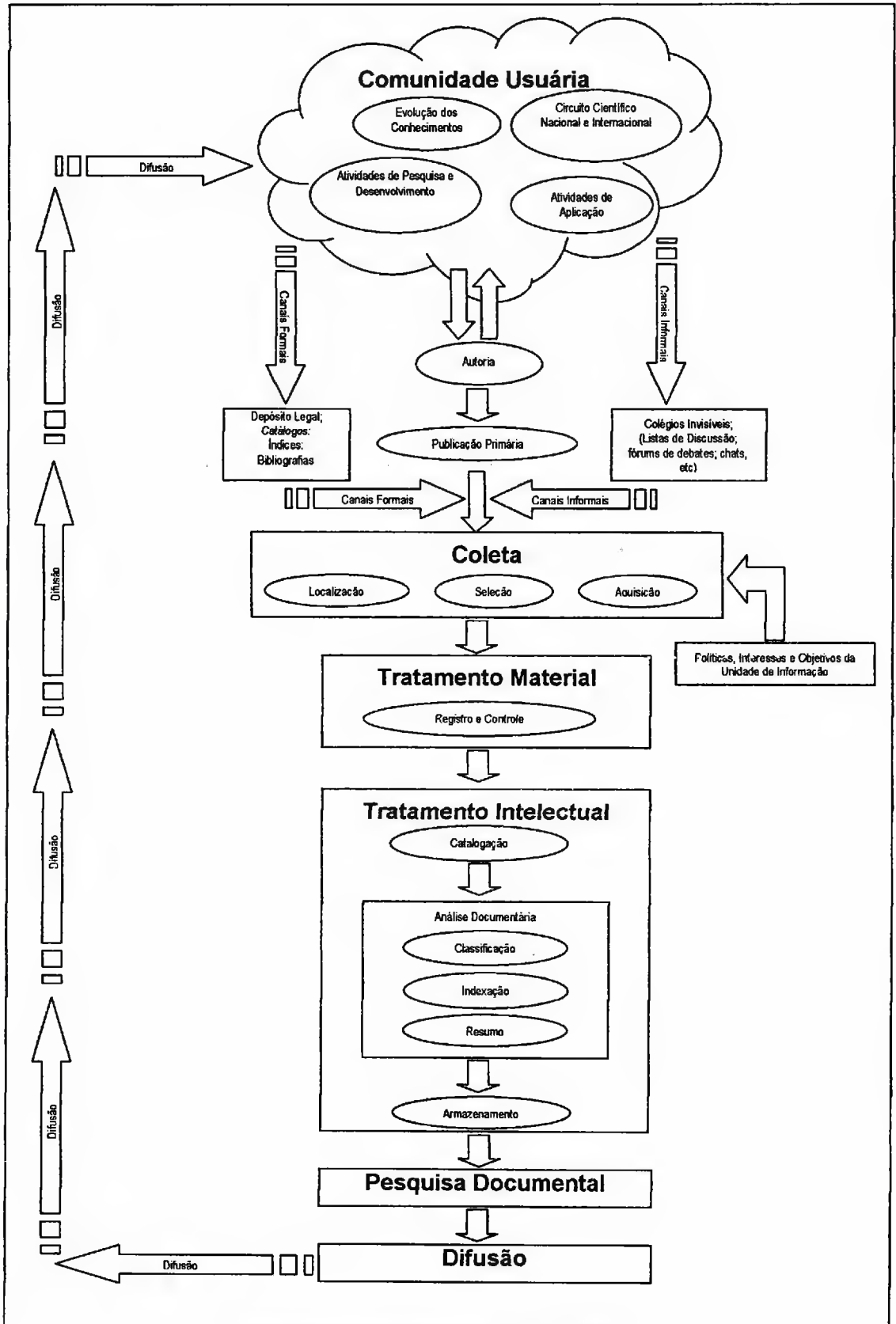


Ilustração 34: Atividades do Ciclo Documentário

2.3.3 Tratamento intelectual da informação

O tratamento intelectual da informação compreende operações como a catalogação e a análise temática da informação. A catalogação ou descrição bibliográfica é uma forma de referência que destina-se a fornecer uma descrição precisa do documento, identificando-o materialmente, de forma única e não ambígua, de modo a permitir sua identificação, localização e representação em catálogos ou em outros instrumentos que facilitem a localização física do documento. A análise temática da informação é relativa ao conteúdo informacional dos documentos e permite a identificação do tema ou assunto a que se refere o documento. Outros termos utilizados para definir essa operação são análise da informação, descrição de conteúdo, análise documentária, descrição de assunto, representação de conteúdo ou representação de assunto (NAKAYAMA, 2001).

A análise temática da informação ocorre no processamento técnico da informação, mais especificamente na classificação, na indexação, na disseminação, na recuperação e na busca. A análise temática compreende operações como classificação, indexação, resumo e análise. São características desejáveis na análise temática: pertinência, precisão³⁷, consistência, identificação e realização de modificações, julgamento, produto claro e de fácil acesso. A operação de classificação consiste principalmente na atribuição de um número ou outro símbolo para descrever o documento; A indexação consiste em atribuir um ou vários descritores, que podem ter ligações hierárquicas entre si. De certo modo, uma lista de termos pode ser vista como uma espécie de mini-resumo, uma vez que reúne os termos em um índice. Com relação à extensão do registro, a indexação pode ser seletiva (mais restrita) ou exaustiva (mais completa)

A operação de resumo consiste na condensação do documento em linguagem natural. O principal objetivo do resumo é indicar de que trata o documento ou sintetizar seu conteúdo. Conforme o nível de aprofundamento, o resumo pode ser classificado como resumo breve ou resumo detalhado (ampliado). Esse grau de aprofundamento é também denominado extensão do registro de um resumo. A maior vantagem do uso de resumos ampliados é que eles permitem melhorar a qualidade das informações recuperadas, uma vez que são representações mais claras sobre os documentos; quanto maior a extensão da representação, mais pontos de acesso ela proporciona, portanto mais fácil a recuperação da informação.

³⁷ A precisão procura exprimir se os termos no documento estão diretamente relacionados com os seus conceitos. A este conceito está associado o conceito de revocação (quantidade de documentos). A relação entre esses termos é inversamente proporcional: quanto maior a revocação, menor a precisão.

2.3.3.1 Indexação

A partir da realização das operações documentárias, um serviço de informações pode originar produtos com algum tipo de tratamento intelectual ou não. Produtos que recebem tratamento intelectual podem ser classificados, de acordo com o nível desse tratamento em simples, elaborados, analíticos, analítico-críticos ou complexos (CAVALCANTI, 1982). Uma das operações que compõem o ciclo de geração desses produtos documentários é a indexação, definida como "operação de representar o conteúdo dos documentos, qualquer que seja o método utilizado" (ROBREDO, 1982). Portanto, a indexação utiliza-se de instrumentos para o tratamento da informação, de modo a obterem-se termos que representem corretamente os conceitos contidos em determinado documento.

O principal propósito de um serviço de indexação é assegurar da forma mais eficiente e econômica possível, que qualquer documento ou informação seja fornecido ao usuário no momento preciso (CARNEIRO, 1985). Para a consecução desse objetivo a técnica de indexação necessita da utilização de instrumentos normativos (normas), bem como de instrumentos lingüísticos. Os instrumentos lingüísticos podem ser classificados em linguagem natural e linguagens documentárias.

Como discutido anteriormente, a linguagem natural é formada pela reunião de sinais utilizados e reconhecidos facilmente pelo homem; quando empregada na indexação denominada "livre": utiliza termos como se apresentam nos documentos. Como também já se discutiu, linguagem documentária compreende o conjunto de regras, símbolos e termos previamente estabelecidos, formando uma linguagem artificial para a indicação do conteúdo temático dos documentos; é, portanto, uma metalinguagem, que utiliza uma construção simbólica, concebida como instrumento para conversão de uma linguagem em outra.

A técnica da indexação necessita do uso de uma linguagem de indexação, que compreende uma especialização de uma linguagem documentária. A linguagem de indexação é uma "linguagem artificial utilizada para o registro ou indicação dos temas contidos nos documentos, dotada de vocabulário controlado e regida por uma sintaxe própria" (Cavalcanti, 1982). Um vocabulário controlado relaciona termos utilizados em sistemas de indexação, com vistas à uniformidade de armazenagem, bem como à facilidade de recuperação. A sintaxe compreende um conjunto de regras necessárias à tarefa de combinação dos elementos do vocabulário.

Embora fontes de informação contendo texto-completo tenham sido escassas por algum tempo, a partir de meados da década de 80 essas fontes começaram a proliferar-se. Uma

vez que os preços dos computadores começaram a cair e que sua capacidade de armazenamento começa a tornar-se cada vez maior, de modo que bases de dados de textos-completos correspondem a um segmento cada vez mais crescente no mercado. Certamente, a vantagem de se acessarem documentos completos, ao invés de apenas referências a esses documentos, é indiscutível. Entretanto, se por um lado o armazenamento de mais e mais documentos desse tipo é vantajoso, por outro traz uma série de dificuldades para a recuperação da informação. Os aspectos intelectuais da recuperação têm sido reduzidos à utilização de palavras-chaves, bem como de frases chaves, combinadas à utilização de operadores booleanos ou de proximidade, que são recursos menos efetivos em se tratando de recuperação em textos completos. Com base nessas argumentações, Warner (1994) procura discutir maneiras de resolver os problemas da recuperação em textos-completos, especialmente aqueles que não foram submetidos a processos de indexação manual ou resumo. As técnicas de indexação automáticas ou semi-automáticas normalmente estão baseadas em processos estatísticos, lingüísticos ou baseados na Inteligência Artificial. O trabalho de Warner (1994) discute o papel da análise lingüística na recuperação textual, enfatizando os problemas inerentes às bases de textos completos.

De acordo com Cavalcanti (1982), o processo de indexação compreende duas etapas: análise conceitual e tradução. No processo de indexação estão envolvidas as seguintes etapas:

- 1) seleção – nesta etapa o indexador realiza a escolha dos documentos a serem incluídos na sua base de dados, de acordo com critérios tais como assunto, tipo de documento, idioma, origem, data, entre outros;
- 2) descrição – nesta etapa, para cada documento, o profissional da informação realiza as seguintes operações:
 - a) descrição bibliográfica – refere-se à descrição física do documento, na qual o indexador identifica a autoria, título, fontes de publicação, entre outros elementos que descrevem bibliograficamente a obra;
 - b) resumo – refere-se a uma representação sintetizada do conteúdo do documento;
 - c) descrição de conteúdo – refere-se à identificação do assunto do documento; para isso o profissional pode lançar mão de vocabulários controlados ou de termos livres, retirados do próprio documento.

- 3) inclusão na base de dados – após a preparação dos índices e dos resumos, essas representações dos documentos são incluídas em bases de dados, de modo a permitir a posterior recuperação por parte do usuário; as bases de dados, como mencionado anteriormente, podem ser constituídas por índices impressos, catálogos de fichas ou arquivos em meio eletrônico;
- 4) recuperação na base de dados – para satisfazer suas necessidades de informação, o usuário elabora uma determinada *estratégia de busca* (um modo pessoal ou orientado de se procurar pela informação desejada na base de dados), partindo de um termo ou de uma combinação de termos a serem pesquisados na base de dados.

A Identificação da temática de um documento compreende o exercício de se obterem os termos adequados à representação dos conceitos contidos nesse documento. A essa tarefa está associado um processo detalhado por Nakayama (2001), que compreende as seguintes etapas: a) Leitura Técnica – é a etapa inicial da operação, na qual o indexador realiza a identificação do assunto de que trata o documento; b) Compreensão; c) Julgamento; d) Levantamento dos conceitos e/ou termos (usando técnicas como transcrição ou grifo); e) Transcrição em um formulário ou sistema (termos, frases, título, trechos); f) Seleção – controle de qualidade dos termos (seleção dos termos levantados); g) Tradução da Linguagem Natural para a linguagem de representação (linguagem documentária)

A indexação é o problema fundamental bem como o obstáculo mais dispendioso da recuperação da informação (LANCASTER 1993). Após a realização da análise conceitual, a segunda etapa da indexação de assuntos envolve a tradução dos conceitos levantados em um conjunto de termos de indexação. A tradução pode ser realizada por dois métodos:

- a) indexação por extração (ou indexação derivada) - as palavras utilizadas para a indexação de fato ocorrem no texto;
- b) indexação por atribuição - envolve a atribuição de termos a um documento a partir de uma fonte que não é o próprio documento; freqüentemente esse tipo de indexação faz uso de algum tipo de vocabulário controlado. (LANCASTER, 1993).

Com relação às técnicas, existem diversas formas para a realização da indexação. A tabela 16 apresenta os diversos tipos de indexação e suas características, de acordo com Cavalcanti (1982):

Tabela 16: Tipos de indexação

Tipo de indexação	Características
Associativa	indexação automática que indica a relação de vizinhança entre dois termos, sem que isso implique em associação funcional específica
Automática	efetuada por computador
Contextual	baseia-se na utilização das palavras significativas, apresentadas no seu contexto
Coordenada ou correlativa	indicação dos assuntos mediante emprego de termos de indexação de igual valor, que podem ser combinados no momento da indexação (pré-coordenação) ou da recuperação (pós-coordenação)
De citações	relação entre os documentos por intermédio das citações comuns que se fazem a esses documentos
De conceitos	Exige a análise do conteúdo temático dos documentos. Tipos de conceitos: entidades, atividades, abstratos, propriedades, heterogêneos.
Derivada	elaboração de índices baseada no uso de elementos que se encontram no próprio documento
Em cadeia	produção de índice alfabético dos termos, ou frases, correspondentes a partes de um índice de classificação
Hierárquica	indicação dos registros da informação sob as classes gerais, bem como sob as classes específicas que lhes são subordinadas
Na fonte	inclusão no documento, quando de sua editoração, dos termos de indexação necessários à recuperação da informação nele contida
Por palavra chave	utilização das palavras significativas de um texto para indicação dos assuntos nele tratados.
Por relações, ou indexação relacionada	representação estrutural de assuntos complexos, pela interposição das relações formalmente indicadas entre os termos.
Por truncamento	técnica automática que efetua a seleção de termos conforme critérios lingüísticos, fundamentados no uso da raiz das palavras.
Por unitermos	tipo de indexação coordenada que emprega palavras simples como termos de indexação.

Fonte: adaptado de Cavalcanti, 1982

A análise temática de um documento, como se constitui em processo intelectual, pode fornecer resultados bastante subjetivos. Portanto, no intuito de diminuir-se a subjetividade dessa análise, utilizam-se alguns critérios objetivos (CAVALCANTI, 1982) para a escolha dos termos de indexação, bem como para a redação de resumos: indicação do tema, baseando-se no título; indicação do tipo de estudo (experiência, levantamento, teoria, aplicação, etc); indicação dos problemas abordados; indicação das conclusões; indicação dos resultados que provam pontos afirmados; métodos de pesquisa empregados.

Existem mecanismos associados às linguagens de indexação que fornecem indicadores do seu desempenho quanto à técnica de representação dos documentos em relação ao seu conteúdo; ou seja, aos conceitos envolvidos nesses documentos. Esses indicadores são a revocação, definida como a relação entre o número de documentos relevantes acerca de um tema

e o número total de documentos do sistema; e a precisão ou relevância, que fornece indicações sobre a qualidade dos documentos recuperados pelo sistema em relação a determinado tema.

Entre os mecanismos de revocação, podem ser citados:

- a) Controle da sinonímia e da quase-sinonímia
- b) Controle de forma (número, forma gramatical, raízes)
- c) Relações estruturais
- d) Grupamentos e aglomerados
- e) Relações hierárquicas
- f) Conceituação específica de termos

Entre os mecanismos de precisão citam-se:

- a) Coordenação
- b) Elos
- c) Indicadores de função e relacionais
- d) ponderação

Para a obtenção dos produtos da indexação, na etapa de seleção dos termos representativos dos conceitos contidos no documento, dois aspectos devem ser levados em consideração:

- exaustividade – profundidade de análise de um documento, objetivando a indicação do assunto a ser especificado; (Foskett, apud Cavancanti, 1982)
- especificidade – grau de precisão de um sistema de indexação, quando aplicado ao assunto, ou aos assuntos de um documento.

Como afirma Lancaster (apud CARNEIRO, 1985): a exaustividade é definida como "uma medida de extensão em que todos os assuntos discutidos em um certo documento são reconhecidos na operação de indexação e traduzidos na linguagem do sistema". Assim, "um alto nível de exaustividade na indexação produz uma alta revocação e uma baixa precisão". Isso pode significar que, quanto mais exaustiva for a linguagem de indexação, maior será a quantidade de assuntos com os quais o indexador tenta relacionar o documento; portanto, menor será a precisão das respostas quando se busca um assunto mais específico.

A indexação de assuntos envolve duas etapas: análise conceitual e tradução. A análise conceitual implica decidir do que trata um documento, isto é, qual o seu assunto. Na etapa de análise de conteúdo o indexador deve responder a perguntas como: a) de que trata?; b) por que foi incorporado ao acervo?; c) quais de seus aspectos serão de interesse para os usuários? (LANCASTER, 1993).

A tarefa de análise temática de um documento deve preocupar-se, principalmente, com a identificação das idéias nele contidas, uma vez que a idéia é a unidade de pensamento que o identifica tematicamente. Assim, o entendimento sobre a categorização das idéias poderá facilitar significativamente esse processo. Para Cavalcanti (1982), as idéias podem ser categorizadas em:

- gerais: limitam as grandes fases do desenvolvimento do pensamento do autor
- principais: subdividem as idéias gerais, das quais constituem os argumentos-chave; a identificação das idéias principais pode se# dar pela pergunta: de que se trata?
- secundárias: são acrescentadas às idéias principais, para explicitá-las e ilustrá-las; a identificação das idéias secundárias pode se# dar pelas perguntas: o quê? por quê? quando? como? onde? quem? para quê?

A relação entre a lingüística e os sistemas de recuperação tem sido controvertida. Por um lado, pode-se dizer que a lingüística é um componente chave para a pesquisa e desenvolvimento; por outro, sistemas comerciais têm feito pouco uso dessa técnica no processamento automático. Há basicamente duas maneiras utilizadas para a recuperação de informação textual: a primeira baseia-se no uso de diversos operadores estruturais, tais como símbolos de truncamento, operadores de proximidade e operadores booleanos; a segunda utiliza métodos de ordenação estatística, de grande conjuntos de itens recuperados, de acordo com a relevância desses itens para a pesquisa realizada. Para Warner (1994), essas técnicas são e serão muito úteis; não obstante, do ponto de vista lingüístico são muito cruas e acarretam dois tipos de problemas:

- Utilizam o mínimo de estrutura gramatical encontrado no documento original;
- Fundamentam-se em termos chaves: assim há pouca exploração do significado dos elementos (conceito) que constituem o documento.

Enfatizando seu pensamento, autora argumenta, ainda, sobre a necessidade de sistemas baseados em métodos lingüísticos, que devem possuir as seguintes características:

- Para comunicar e receber informação, o sistema deve ter conhecimento lingüístico, isto é conhecimento de itens léxicos (palavras, frases), categorias gramaticais (verbos, nomes) e relações gramaticais (sujeito, objeto);
- Conhecimento extralingüístico – conhecimento de entidades, atributos, eventos, processos e relações entre os modelos de informação;
- Habilitar ao uso do conhecimento – isto é, o conhecimento de procedimentos para utilização do conhecimento lingüístico e extralingüístico com o objetivo de atingir uma meta em particular.

Do ponto de vista da recuperação da informação, os dois principais campos de análise da língua são a teoria lingüística e o processamento da linguagem natural. A teoria lingüística estuda a língua e linguagens com o único propósito de construir uma teoria para a sua estrutura e funções, em um dado ponto do tempo e sem preocupar-se com aplicações práticas; ou seja, o propósito da teoria lingüística é formular teorias adequadas sobre a estrutura individual das linguagens (gramáticas) e projetar uma teoria geral da lingüística que caracterize as propriedades estruturais de todas as linguagens humanas. O processamento da linguagem natural é uma área que estuda como a linguagem natural inserida em sistemas computadorizados pode ser manipulada e armazenada de modo a preservar certos aspectos do conteúdo original. Assim, o principal propósito do processamento da linguagem natural é produzir ferramentas úteis ou construir teorias de cognição dentro do campo computacional.

Warner (1994) defendeu a premissa de que a análise lingüística pode desempenhar um importante papel na pesquisa e desenvolvimento de técnicas de recuperação do textos-integrais. Isto pode ocorrer de duas formas: utilizando a análise lingüística como uma das diversas ferramentas na engenharia de um sistema de recuperação, em conjunto com outras técnicas, como por exemplo as técnicas de processamento estatístico; outra forma seria para a descrição de dados textuais contidos em sistemas de recuperação de informação. A autora lembra que, comparativamente, sabe-se pouco acerca dos padrões lingüísticos dos documentos em contraposição a outras técnicas; e finaliza sua abordagem enfatizando que quão longe a recuperação da informação pode ir com as técnicas lingüísticas permanece uma questão aberta e que, com ou sem a adição de um componente lingüístico, os objetivos de sistemas de recuperação totalmente automatizados e flexíveis podem nunca ser realizados (WARNER, 1994).

2.3.3.2 Índices e Resumos

A principal função da elaboração de índices e resumos, que são instrumentos utilizados para a representação do conteúdo de documentos primários, é facilitar a recuperação de informações relativas ao documento indexado ou resumido. Índices e resumos geralmente são incluídos em bases de dados que podem ser impressas, em fichas, ou em formato eletrônico (LANCASTER, 1993).

Comumente são produzidos, por meio do processo de indexação, termos de índices que podem ser utilizados para pesquisa em um arquivo em linha; esses termos podem ser armazenados em um arquivo invertido ou em outra estrutura de dados, de modo que pesquisas possam ser realizadas no índice, utilizando operadores booleanos, ou algoritmos baseados no peso dos termos, para obterem-se resultados ordenados de acordo com critérios estatísticos (HARMAN, 1994).

Os índices compreendem os produtos gerados a partir do trabalho de indexação. A tabela 17 apresenta alguns desses tipos de índices.

Tabela 17: Tipos de índices

Tipo de índice	Características
AKWIC (author and key-word in context)	Índice automático que relaciona autores e palavras-chaves
Alfabético	Termos em ordem alfabética
Analítico	inclui a referência bibliográfica e um resumo ou análise do documento
Assindético	não incorpora referências cruzadas
Coordenado ou correlativo	resultante da combinação dos descritores para a obtenção da indicação de assuntos específicos
Corrente	publicado periodicamente, com a finalidade de atualizar informações constantes de determinados documentos
Cumulativo	índice de publicações seriadas; reúne numa só lista as entradas relativas a vários fascículos ou volumes
De citações	associação das referências citadas por autor em um artigo A, com as referências feitas posteriormente ao próprio artigo A.
KLIC (key letter in context)	uso de truncamento na indexação automática, para a produção de índices
KWAC (key word and context)	índice obtido pela permutação das palavras de um título
KWIT (key word in title)	Espécie de índice KWIC baseado em permutação selecionada
KWOC (key word out of context)	as palavras significativas aparecem na devida ordem alfabética, seguidas pelo título completo, dentro porém, de um determinado espaço linear
KWOT (key word out of title)	as palavras significativas do título são extraídas e colocadas em ordem alfabética
KWUC (key word and UDC)	o arranjo deste índice é feito pelas classes gerais da CDU e subdivisões dessas classes em ordem alfabética das palavras-chaves dos títulos que constam de cada classe

Fonte: adaptado de Cavalcanti, 1982

Grandes quantidades de texto estão atualmente disponíveis para acesso em linha, incluindo tanto os textos criados para acesso eletrônico como publicações tradicionais.

Entretanto não é possível pesquisar informações nesses textos sem o auxílio da indexação automática. Baseado nesses pressupostos, o trabalho de Harman (1994) foi desenvolvido com dois objetivos principais:

- fornecer um tutorial sobre indexação de termos simples, discutindo questões críticas do processo de indexação, que devem ser resolvidas durante a indexação do texto completo, de modo a fornecer uma performance efetiva na recuperação;
- apresentar uma discussão sobre os avanços na indexação automática além da técnica termo-simples, utilizada na maioria dos sistemas de recuperação de informação.

Segundo Harman (1994), o ponto chave para a qualquer indexação é a escolha dos limites do registro que identifica uma unidade pesquisável. Um registro pode ser definido como um livro, um capítulo em um livro, uma seção ou um parágrafo. A decisão depende, em parte, da natureza do objeto a ser indexado. Por exemplo, quando se trata de uma pesquisa em registros bibliográficos, o registro é claramente definido como um dos registros do catálogo bibliográfico; quando se trata de uma aplicação em artigos de jornal, o registro poderá ser cada artigo. Dessa forma, a escolha do tamanho do registro, embora inicialmente pareça vaga, pode basear-se no tamanho do documento; para documentos muito grandes (de 100 páginas ou mais), a autora recomenda a divisão do registro por páginas ou por seções do documento.

A escolha do tamanho do registro não é tão importante para a exibição dos resultados; mas, principalmente, para a realização da pesquisa. Assim, um registro muito pequeno fornece pouco texto para os algoritmos de pesquisa, o que fornece resultados pobres; um registro muito grande, entretanto, pode diluir a importância das palavras encontradas e causar falsas ocorrências.

A segunda decisão chave para qualquer sistema de indexação é a escolha do que constitui um palavra e, por conseguinte, quais dessas palavras devem ser indexadas. Nos sistemas de indexação manual a escolha é facilmente feita por um indexador humano. Para a indexação automática, entretanto, é necessário definir que pontuação deve ser utilizada como separador entre as palavras e definir que palavras indexar. Essa separação normalmente é feita com todos os espaços em branco e todas as pontuações, embora haja exceções, dependendo da aplicação e do *software* utilizado. Definidos os limites das palavras, outro problema a ser resolvido é a indexação de números, uma vez que o conjunto de números únicos é infinito. Não indexar números, entretanto, pode acarretar problemas na recuperação, quando um número for imprescindível para a pesquisa, como é o caso de datas ou anos.

Harman (1994) aponta outro problema para a indexação, relativo aos caracteres individuais (a, b, c, etc); para a autora, se por um lado, indexar esses caracteres pode representar um aumento no tamanho do índice, por outro, não indexá-los pode representar perda de informação, quando um caractere for imprescindível à pesquisa (por exemplo: vitamina C). Como solução para definir que palavras devem ou não ser indexadas, a autora sugere os seguintes "compromissos", a serem estabelecidos com o usuário:

- a pontuação deve ser estudada e potenciais problemas identificados;
- deve-se estudar o uso de letras maiúsculas e minúsculas; uma técnica usual, relata a autora, é a conversão dos caracteres maiúsculos para minúsculos durante a indexação, o que pode acarretar problemas para a recuperação de nomes próprios;
- a indexação de números depende profundamente da aplicação; datas, rótulos de seções e números combinados com letras podem ser indexados; outros tipos de números não devem ser indexados;
- quanto à indexação de caracteres individuais, pode-se verificar o alfabeto e anotando-se as letras que possuem um significado particular para a aplicação; então essas letras podem ser indexadas.

No campo da indexação automática, diversas técnicas têm sido desenvolvidas, algumas com aplicações interessantes, tanto no âmbito de pequenas coleções, como para grandes acervos, como é o caso da técnica baseada no peso (ocorrência) dos termos. Citam-se ainda outras técnicas: indexação de palavras não significativas; uso de sufixos; e extensão da busca – esta última, segundo Harman (1994), muito bem sucedida em pequenas coleções.

Muitas técnicas de indexação automática trabalham com a utilização de listas de termos não significativos. Como a construção desse tipo de listas, entretanto, pode ser uma tarefa difícil do ponto de vista da escolha, muitos sistemas utilizam trabalhos desenvolvidos no passado por pesquisadores. Assim, palavras que ocorrem com frequência muito elevada nos textos são integradas a essas listas. Uma técnica mais apropriada é a produção de uma lista de termos não-significativos baseada no próprio corpus do texto a ser indexado, com base na frequência com que determinadas palavras ocorrem naquele texto.

Outros sistemas utilizam a sufixação ou derivação para substituir todas as palavras indexadas com suas raízes. Os algoritmos utilizados para realizar essa tarefa normalmente possuem três partes: uma especificação ou qualificação da terminação da palavra; uma lista de

exceções e uma ação necessária. A utilização de sistemas de sufixação ou derivação deve-se a dois fatores primordiais: a redução do tamanho do índice armazenado e o aumento da performance, devido ao uso de variantes das palavras.

Outra técnica conhecida é a da atribuição de pesos aos termos. Diversos sistemas comerciais de indexação fornecem a possibilidade de se ordenar os documentos com base no número de termos que correspondem à pesquisa do usuário e às ocorrências dos termos nos documentos. Essa técnica utiliza ordenações estatísticas complexas, envolvendo atribuições de pesos aos termos ou medidas similares mais acuradas e pode ser utilizada quando se desejar aumentar a precisão em uma busca com três ou mais termos. A técnica consiste em se atribuir pesos diferenciados para os termos, de modo que mesmo quando não contiverem um dos termos pesquisados, sejam retornados documentos que possuem os outros dois termos fornecidos na busca. Alguns sistemas de medida que podem ser utilizados por essa técnica baseiam-se no *Inverted Document Frequency* – IDF, que mede a escassez de um termo no texto; outros utilizam algum tipo de função de medida da frequência do termo no texto.

Um dos problemas encontrados nos sistemas de recuperação da informação é que documentos relevantes são perdidos porque não contêm os termos da busca. Um método de expansão de uma busca pode ser o uso de um tesauro. Esse tesauro pode ser utilizado como uma ferramenta de busca ou incorporado automaticamente ao sistema de algum modo. Embora haja pesquisas que se preocupem em desenvolver metodologias para a construção automática de tesouros, essa é uma tarefa difícil e extremamente dependente do domínio do conhecimento a que se refere o sistema de busca.

Grandes coleções de textos-completos necessitam não apenas de mecanismos de expansão de busca, mas também de dispositivos de precisão para melhorar sua acurácia. Um importante mecanismo de precisão é a atribuição de pesos aos termos, que permite melhorar a precisão na recuperação, pois os usuários não mais necessitam localizar informações aleatoriamente nos documentos, mas podem procurar em saídas ordenadas.

Harman (1994) considera que a indexação automática de termos simples possibilita pesquisas razoáveis em textos-completos; Entretanto, argumenta sobre a necessidade crítica de que pesquisas continuem sendo desenvolvidas com relação às técnicas avançadas, para que essas possam ser aceitas como técnicas de indexação automática pela comunidade de recuperação da informação como um todo.

2.3.4 Recuperação da Informação

De acordo com Lancaster (1993), para qualquer necessidade específica de informação, haverá sempre muito mais itens que não possuem relevância ou pertinência ao assunto pesquisado pelo consulente (itens inúteis) do que itens que possuam relevância ou pertinência às necessidades desse usuário (itens úteis). Portanto, a principal função de um sistema de recuperação de informações é permitir que o usuário localize o maior número possível de itens relevantes.

A relação entre o total de itens úteis para o usuário e o total de itens recuperados é denominada coeficiente de precisão. Essa relação pode ser representada pela fórmula

$$cp = niu/nt$$

onde

cp = coeficiente de precisão

niu = número de itens úteis

nt = número total de itens recuperados

A relação entre o total itens úteis encontrados pelo o usuário em uma determinada busca e a quantidade total de itens úteis sabidamente disponíveis na base de dados é denominada coeficiente de revocação. Essa relação pode ser representada pela fórmula

$$ir = niu/ntu$$

onde

ir = índice de revocação

niu = número de itens úteis

ntu = número total de itens úteis

O índice de revocação e o coeficiente de precisão são grandezas inversamente proporcionais. Portanto, quanto menor for o índice de revocação, mais preciso será o resultado da pesquisa, isto é, maior será a quantidade de itens úteis recuperados. É importante destacar-se, também, que resultados satisfatórios na recuperação dependem diretamente da qualidade com que a indexação foi realizada isto é, da política de indexação utilizada, das regras usadas para a redação do resumo, da qualidade do vocabulário controlado, da qualidade das estratégias de busca, entre outros fatores.

Segundo Piedade (1977), os dois fatores que mais influenciam a revocação e a precisão, são a exaustividade – isto é, a extensão ou número de conceitos escolhidos para se representar o conteúdo de determinado documento; e a especificidade – isto é, a exatidão com que os termos escolhidos descrevem o conteúdo do documento em questão. A escolha de muitos termos para a representação do conteúdo de uma base de documentos, isto é, alta exaustividade, poderá aumentar a revocação – mais itens serão recuperados –, mas alguns documentos recuperados tratarão apenas superficialmente sobre o assunto e, nesse caso, a precisão será baixa. Por outro lado a escolha de poucos termos, mas exatos, para a representação do conteúdo de uma base de documentos, isto é, alta especificidade, poderá aumentar a precisão – apenas itens que coincidam exatamente com o termo de busca serão recuperados –, sendo recuperado pequeno o número de itens que tratarão sobre o assunto em profundidade. Em resumo, o aumento da exaustividade aumenta a revocação e diminui a precisão. Por outro lado, o aumento da especificidade diminui a revocação e aumenta a precisão.

Em sistemas de indexação manual, o uso de mecanismos que permitam o aumento da revocação e da precisão deve aumentar, conseqüentemente, o custo do sistema; o mesmo ocorre com o nível de exaustividade. É normalmente antieconômico usar um vocabulário muito específico quando o tipo de documentos e as solicitações feitas ao sistema são mais genéricos. A carência de pessoal torna recomendável a opção por um sistema que exija pouca manutenção e menor esforço na indexação, principalmente quando se tem um volume muito grande de documentos a serem indexados.

O aumento do custo do sistema de indexação dá-se em função: a) do uso de mecanismos que melhoram os índices de revocação e de precisão; b) dos níveis de exaustividade; c) da quantidade de pessoas envolvidas na indexação do acervo (CARNEIRO, 1985).

Especificidade é a extensão em que o sistema nos permite ser precisos ao especificarmos o assunto de um documento que estejamos processando. Portanto, quanto mais específica for a indexação, mais precisas serão as respostas e menor será a revocação. (CARNEIRO, 1985). Tomando esse ponto de vista, a revocação, que é definida por Lancaster (1993, p. 4), como "a capacidade de recuperar documentos úteis" tem menor importância em um sistema de indexação automática. Então, o ideal é que se aumente o nível de precisão que o autor define como "a capacidade de evitar documentos inúteis", por meio do aumento da especificidade da indexação.

O conceito de revocação relaciona-se com a capacidade do sistema para recuperar um número desejável de documentos; já a precisão relaciona-se com a capacidade do sistema para

impedir a recuperação de documentos irrelevantes. Esses conceitos relacionam-se como grandezas inversamente proporcionais: quando se aumenta a precisão, a revocação tende a diminuir e vice-versa (Carneiro, 1985, p. 234-235).

Lancaster (1993) argumenta, ainda, que quanto maior for o acervo da base de dados, menos aceitável será uma baixa precisão, isto é, em bases de dados muito grandes, torna-se progressivamente mais difícil alcançar um nível de revocação aceitável, com um nível de precisão satisfatório.

2.3.4.1 Modelos de Recuperação da Informação

Os modelos de recuperação mais comuns são o booleano, o vetorial, o probabilístico e o de atribuição de pesos. De acordo com a definição do MOLE *Text Analysis Group* (1999), o método de recuperação booleana é o mais simples dos métodos de recuperação e, como diz o seu nome, baseia-se no uso de operadores booleanos. Assim, os termos em uma busca são ligados por meio dos conectores E, OU e NÃO. O método booleano é freqüentemente utilizado em mecanismos de busca na *Internet*, uma vez que é rápido e que pode ser utilizado em linha. No entanto, para que maior eficiência, esse método requer que o usuário tenha pelo menos algum conhecimento acerca do tópico de pesquisa desejado.

Atualmente a recuperação baseada no método booleano tem sido combinada com o método de navegação baseada no conteúdo, utilizando-se redes de conceitos, nas quais termos compartilhados de documentos obtidos previamente são utilizados para refinar e expandir a busca. Os operadores booleanos têm sido substituídos por operadores fuzzy³⁸ (MOLE, 1999a).

O modelo de vetor-espaco pode ser dividido em três estágios: a) indexação do documento, na qual os termos chaves são extraídos do documento; b) extensão dos termos indexados para melhorar a recuperação dos documentos relevantes para o usuário; c) classificação do documento com relação à busca, de acordo com uma medida de similaridade (MOLE, 1999).

A indexação probabilística é baseada no pressuposto de que existe algum nível de diferença na distribuição do conteúdo de termos-significativos e de termos não-significativos. Recentemente um método de indexação automática que utiliza um agrupamento serial de palavras tem sido introduzido. O valor de tal agrupamento é um indicador se o termo é ou não significativo. Obviamente muitas palavras em um documento não descrevem o seu conteúdo, como é o caso daquelas pertencentes às classes dos artigos, das conjunções e das preposições, por

³⁸ Um modelo baseado na teoria do conjunto fuzzy permite a interpretação de uma busca do usuário, com a utilização de um descritor lingüístico para cada termo (MOLE, 1999a).

exemplo. Assim, na indexação automática de um documento aquelas palavras não significativas são removidas do documento matriz; então o documento somente será indexado pelas palavras cujo conteúdo é significativo. Essa indexação pode ser baseada na frequência do termo. Isto é, os termos que têm alta ou baixa frequência em um documento são considerados não significativos. Desse modo, em geral, de 40 a 50% do número total de palavras em um documento são removidas, com uma lista de termos não significativos. (MOLE, 1999).

A ponderação de termos tem sido explicada pelo controle da exaustividade ou especificidade na busca, onde a exaustividade está relacionada com a revocação e a especificidade com a precisão (MOLE, 1999). A ponderação de termos para o modelo de vetor de espaço tem sido inteiramente baseada em estatísticas de termos simples. Segundo os autores, há três principais fatores de ponderação dos termos: a) fator de frequência dos termos; b) fator da frequência da coleção; c) fator da extensão da normalização. Esses fatores são combinados para, juntos, definirem o peso resultante para o termo.

2.3.5 Serviços de Indexação e Recuperação na web

Desde o início da *Internet* sempre houve preocupações de se disponibilizarem serviços que garantissem a recuperação dos documentos publicados. Entre as ferramentas mais antigas citam-se o Archie, o Veronica, o Jughead e o Gopher (CENDÓN, 2001). O advento da *World Wide Web* trouxe consigo o fenômeno da explosão exponencial da quantidade de documentos registrados na *Internet*, para o que foi necessária a implementação de outros serviços de organização e de recuperação de informações. Nessa área, alguns dos pioneiros da *web* foram o AltaVista, o Yahoo!, o Open Directory, a Virtual Libray, o Excite, o Lycos, entre outros serviços. Atualmente (2004) há centenas de serviços dessa natureza, categorizados, conforme as suas características tecnológicas, em diretórios de pesquisa e mecanismos de busca. Ambos os termos referem-se às ferramentas utilizadas para recuperação de informação na *Internet*, mas que funcionam de maneira diferente. Os mecanismos de busca criam seus índices automaticamente e percorrem continuamente a *Web* – utilizando *software* conhecido como *robot*, *crawler* ou *spider* –, visitando *sites* e indexando suas páginas. No momento da recuperação, as pesquisas são feitas utilizando-se essas informações colhidas. Os diretórios, por seu turno, constroem seus índices por meio de descrições de páginas fornecidas pelas pessoas no momento de submetê-las a um cadastro. As pesquisas baseiam-se nas informações fornecidas, que podem ou não refletir o conteúdo real das páginas. Assim, a indexação dos mecanismos de busca é feita de maneira diferente daquela dos diretórios e os fatores importantes para o julgamento da relevância das

páginas também diferem (BAX; CAMPOS, 2000). Algumas dessas diferenças e características são discutidas, a seguir, nesta seção.

Os diretórios foram a primeira solução proposta para a organização de conhecimento na *web* e surgiram com uma intenção de se coletar manualmente, ou por meio de indicações de usuários, a maior quantidade de informações possível, contando-se a grande variedade dos assuntos disponíveis na *Internet*.

Segundo Wall (2004), o primeiro mecanismo de buscas a ser disponibilizado na *Internet* foi o Archie, criado em 1990 pelo estudante Alan Emtage, na Universidade McGill em Montreal. A intenção original era a utilização do nome “*archives*”, o que não foi possível pelas restrições do sistema operacional Unix, que exigia nomes mais curtos. Basicamente, o Archie era composto por uma base de dados contendo nomes de arquivos na *web* e um sistema de busca para a localização e recuperação de nomes de arquivos coincidentes com uma pesquisa do usuário. Em 1991, Mark McCahill, da Universidade de Minnesota, desenvolveu o Gopher, como uma alternativa para o Archie e em 1993 um grupo de consultoria e pesquisas da Universidade de Nevada desenvolveu o Veronica, que tinha o mesmo propósito do Archie, mas que trabalhava com arquivos de texto plano. Na mesma época surgiu também o Jugheac.

Com o advento da *web*, Matthew Gray introduziu o seu sistema *World Wide Web Wanderer*, mais tarde conhecido como Wandex, que inicialmente tinha apenas a intenção de medir o crescimento da *web*, por meio da contagem dos servidores *web* ativos. Logo foi agregada ao sistema a capacidade de capturar URL's e ele fazia varreduras diárias, chegando a acessar a mesma página centenas de vezes no mesmo dia. Segundo Wall (2004), essa foi a primeira iniciativa do uso de tecnologias que originaram os termos *spiders* (aranhas), *crawlers* (rastejadores) ou *bots* (robôs), para designar qualquer tipo de mecanismos de coleta automática de dados na *Internet*. Para Willey (2004), robôs de computador são programas capazes de executar tarefas repetitivas em velocidades que seriam impossíveis para os humanos: no âmbito da *Internet* o autor define o termo para referir-se a programas que exploram a rede em busca de qualquer tipo de informação e que usualmente compilam uma grande base de dados para pesquisas futuras. Essa categoria de robôs é geralmente denominada *spider*³⁹.

Em outubro de 1993, Martijn Koster criou o sistema *Archie-Like Indexing of the Web* (Aliweb), que permitia aos usuários submeterem páginas para serem indexadas de acordo

³⁹ Willey (2004) utiliza também o conceito de *chatbot*, para referir-se a programas de computador que possuem capacidade de realizar algum tipo de diálogo interativo ou de prestar algum tipo de informação básica para o usuário de modo intuitivo e repetitivo.

com sua própria descrição. Em dezembro do mesmo ano, duas outras iniciativas foram implementadas: o JumpStation, que coletava informações do título e do cabeçalho das páginas; e o WWW Worm, que indexava títulos e URL's. Essas duas ferramentas possuíam o mesmo problema estrutural: ao invés de estabelecerem um sistema de classificação dos resultados, listavam-nos na ordem em que os encontravam.

Em fevereiro de 1993 estudantes da universidade de Stanford iniciaram as atividades do projeto Architext, que veio a tornar-se Excite em meados daquele ano; o *software* de busca, que previa a utilização de análise estatística de palavras para tornar a localização mais eficiente, foi liberado para a utilização em outros *sites*.

Wall (2004) relata que todos esses serviços possuíam deficiências, porque seus *spiders* não eram suficientemente inteligentes para compreenderem o significado dos links encontrados, resultando a necessidade de que o usuário tivesse plena certeza dos termos de busca a serem utilizados, sob pena de não conseguir localizar nenhuma informação relativa ao assunto em questão. Em janeiro de 1994 surge o diretório EInet Galaxy, que possuía características similares aos diretórios atuais e que contava também com recursos de Gopher e Telnet. Em seguida, em abril do mesmo ano, surge o Yahoo, criado por David Filo e Jerry Yang, como uma coleção de suas páginas favoritas. Filo e Yang logo tiveram que reorganizá-lo para tornar-se um diretório dotado de um mecanismo de busca local, uma vez que a quantidade de páginas referenciadas cresceu rapidamente. Diferenciaram-se também das outras iniciativas, introduzindo uma descrição para cada um dos links referenciados.

Em abril de 1994, Brian Pinkerton, da Universidade de Washington apresentou o WebCrawler que, segundo Wall (2004) e Willey (2004), foi o primeiro *crawler* (rastejador) com capacidade para indexar textos completos de páginas inteiras; antes dele, um usuário podia pesquisar apenas nas URL's ou e, descrições de páginas, fornecidas pelos seus autores. Logo o WebCrawler tornou-se tão popular que às vezes era impossível acessá-lo, devido à grande quantidade de usuários conectados – cerca de quinze mil acessos diários. Passado algum tempo, o provedor de serviços de *Internet* americano, American On-Line (AOL), adquiriu o WebCrawler e passou a rodá-lo em sua própria rede. Em 1997, o Excite comprou da AOL o WebCrawler e passou a oferecer sua base de dados para que aquela empresa mantivesse o seu serviço de busca, o NetFind. Em seguida ao lançamento do WebCrawler vieram o Lycos, o Infoseek e o OpenText.

O Lycos foi desenvolvido na Universidade Carnegie Mellon, sob a coordenação de Michael Mauldin e foi publicado com cerca de 54 mil documentos. Trazia como principais

características a recuperação categorizada por relevância; a localização baseada em prefixos ou na proximidade de palavras; e o tamanho do seu catálogo – após um mês do seu lançamento o Lycos já contava com cerca de 394 mil páginas indexadas; seis meses depois, com aproximadamente 1,5 milhão; um ano e meio depois, com mais de 60 milhões de páginas, mais do que qualquer outro mecanismo de busca da época.

Em 1995 o Alta Vista foi publicado, introduzindo novas características como a maior largura de banda para a época; técnicas de pesquisa avançada; técnicas de pesquisa em linguagem natural; permissão para que usuários adicionassem ou excluíssem suas URL's em 24 horas; pesquisa de links que apontavam para uma dada página; e dicas de pesquisa. Em 1996 vieram o diretório Looksmart e o mecanismo de busca HotBot, este último de propriedade da empresa Inktomi Corporation. Essa empresa foi fundada por Eric Brewer, professor assistente da Universidade da Califórnia em Berkeley, e pelo doutorando em Ciência da Computação Paul Gauthier. Segundo Willey (2004), o nome deve ser pronunciado “*ink-to-me*” e trata-se de uma apologia a uma aranha mitológica – *unktomi* ou *iktomi* – que vive nas planícies indígenas americanas e que é conhecida por trazer cultura para o povo.

Em 1997 foram lançados o Ask Jeeves e o Northern Light. Segundo Wall (2004), 1998 foi o último ano de lançamento de supermecanismos, como o Google, que introduziu a característica de classificação dos resultados com base nas ligações com outras páginas. Desde então, tornou-se tão popular que mesmo grandes serviços de busca como a AOL e o Yahoo passaram a receber resultados a partir de sua base de dados. No mesmo ano foram lançados o MSN, o DirectHit e o Open Directory.

Em 2000 foi lançado o Teoma, que utiliza classificação por *clusters* para organizar *sites*, de acordo com a popularidade específica do assunto. Em 2001, o Ask Jeeves comprou o Teoma, para substituir seu antigo Direct Hit. Nos últimos dois anos, na árdua corrida pelos melhores resultados financeiros, algumas empresas foram fundindo-se ou sendo vendidas. Tal é o caso do LookSmart, que adquiriu o WiseNut em março de 2002; e do Overture, que em 2003 adquiriu o AllTheWeb e o AltaVista. Em 2004, o Yahoo! adquiriu o Overture e deixou de utilizar o mecanismo de busca do Google, em favor de seu próprio mecanismo de busca. A empresa investiu também em um novo banco de dados que substituiu tanto a base do AltaVista, como a do AllTheWeb.

No tocante à busca na *web*, muitas iniciativas imbuídas da filosofia do *software* livre sempre foram implementadas; mas há que se destacar o Nutch.Org, que constitui-se um esforço para implementar um sistema de localização na *web* com código-fonte aberto e com estratégia de

desenvolvimento comunitário. Os argumentos da organização em favor da disponibilidade de um buscador com código fonte aberto enfatizam o risco de as atuais empresas tornarem-se monopólios. Certamente, apenas uma empresa, ou um pequeno grupo, controlando praticamente todo o sistema de localização e recuperação na *web* pode representar um sério risco à disseminação democrática da informação. Nesse sentido, o Nutch pretende ser uma alternativa transparente aos sistemas comerciais de localização na *web*, propiciando maior confiabilidade aos resultados gerados, quanto a não serem secretamente direcionados – se houver orientação, ela será pública. Outro argumento da instituição é o de que todos os principais sistemas de localização existentes têm fórmulas de classificação próprias e não têm interesse em explicar porque foi dada uma posição a um determinado resultado. Além disso, como são entidades que visam lucro, alguns sistemas de localização determinam o posicionamento dos resultados principalmente com base no pagamento (NUTCH, 2004).

Segundo texto de apresentação do seu *web site*, o Nutch espera permitir que qualquer pessoa possa facilmente colaborar para o desenvolvimento de um sistema de localização na *web*, com tecnologia de ponta conhecida em nível mundial, e com custos reduzidos. Tal sistema tem como objetivos: a localização mensal de bilhões de páginas; a manutenção de um índice atualizado dessas páginas; a realização de pelo menos mil pesquisas por segundo nesse índice; o provimento de resultados de alta qualidade; operação com o menor custo possível (NUTCH, 2004).

Relativamente ao panorama do uso de serviços de busca na *web* atual, Sullivan (2004a) apresenta um relatório acerca da popularidade dos serviços, obtida pela mensuração do volume de pesquisas. A ilustração 35 mostra a porcentagem de utilização de serviços de busca na *web*, por usuários norte-americanos, no mês de maio de 2004.

Os dados relativos ao Yahoo! contemplam também as pesquisas realizadas por meio do AltaVista, do Overture e do AllTheWeb, adquiridos recentemente por aquela companhia. Os dados relativos ao Google contemplam também o parceiro Go.com. Os dados relativos à AOL Search contemplam também o Netscape Search. Os dados relativos ao Excite referem-se também ao iWon, MyWay e My *Web* Search, todos de propriedade da rede Ask Jeeves Excite Network. Os dados relativos ao Ask correspondem ao Ask Jeeves e ao Teoma. O valor relativo ao InfoSpace engloba também seus outros serviços: o Dogpile e o WebCrawler. O HotBot está representado nos dados relativos ao Lycos, já que foi adquirido por aquela companhia.

Em relação ao mecanismo de busca, a ilustração 36 mostra a liderança das empresas Google e Yahoo! quanto à utilização de seus mecanismos de busca na *web*. Aproximadamente 95% do mercado americano, em maio de 2004, utilizava os mecanismos fornecidos por essas empresas.

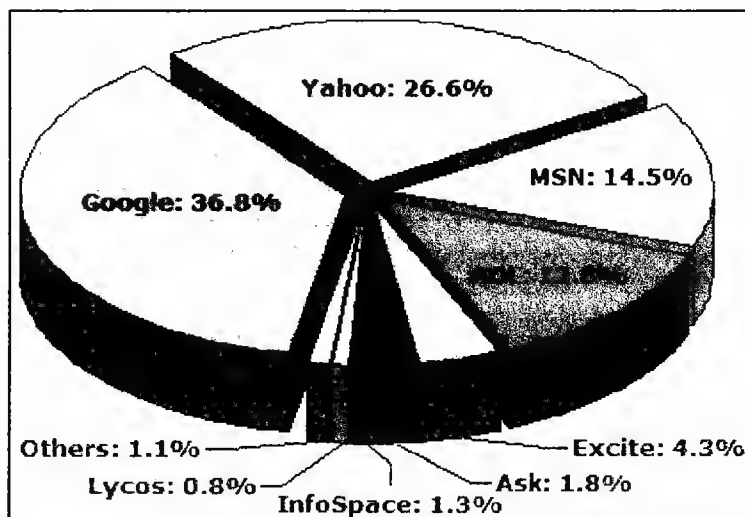


Ilustração 35: Utilização de serviços de busca na web por internautas americanos em maio de 2004.
Fonte: (SULLIVAN, 2004a)

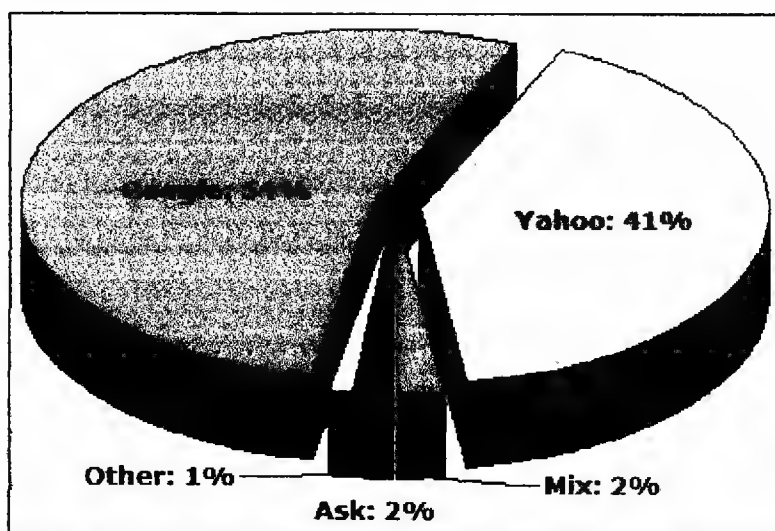


Ilustração 36: predomínio das tecnologias de busca na web por empresas em maio de 2004.
Fonte: (SULLIVAN, 2004a)

Os dados mostram o Google liderando a utilização, com 54% das buscas, e referem-se não apenas às buscas na base Google, mas também nos seguintes serviços que utilizam a tecnologia Google: a AOL, a Excite Network – iWon, MyWay e My *Web Search*. Os dados do Yahoo referem-se também à MSN Search. O rótulo Mix refere-se à combinação de buscas do Lycos – que acessa informações do LookSmart, do Google, do HotBot e do Yahoo,

além das suas próprias – e do InfoSpace, que executa meta busca em diversos mecanismos de busca.

Certamente, a mesma busca feita em diferentes mecanismos poderá trazer resultados completamente diversos, recuperando informações extremamente pobres em um sistema e verdadeiras minas de ouro em outro – o uso de operadores booleanos, por exemplo, pode representar uma grande diferença nos resultados de uma pesquisa. Como uma tentativa de solução para tal problema, surgiram as ferramentas de meta-busca. A primeira delas, segundo Willey (2004), foi o MetaCrawler, que fazia pesquisa no Lycos, no AltaVista, no Yahoo!, no Excite, no WebCrawler e no Infoseek simultaneamente. O MetaCrawler foi desenvolvido em 1995, por Eric Selburg, um estudante de mestrado na Universidade de Washington – a mesma origem do WebCrawler, alguns anos antes –, sob a orientação de Oren Etzioni. O MetaCrawler baseia-se na reformatação dos resultados obtidos em outros mecanismos, apresentado-os para o usuário final em uma página concisa. Além dessa iniciativa, outros metabuscadores que se destacam no mercado de busca na *Internet* são o Mamma, o Dogpile, o Vivíssimo, o Kartoo e o ProFusion.

Criado em 1996, como resultado de uma dissertação de mestrado, o Mamma foi um dos primeiros serviços de busca a introduzirem o conceito de metabusca na *Internet*. Devido à qualidade dos seus resultados, bem como aos benefícios da metabusca, o serviço cresceu rapidamente e, recentemente – fevereiro de 2004 –, recebeu uma menção honrosa na categoria de melhor serviço de metabusca, conferida pelos assinantes do *site* Search Engine Watch, organizado pelo consultor Danny Sullivan, especialista em mecanismos de busca⁴⁰.

O metabuscador Dogpile foi lançado em janeiro de 1996 e rapidamente tornou-se um dos serviços mais usados na *web*. No ano 2000 foi adquirido pela companhia InfoSpace. Atualmente (2004) o Dogpile permite a localização de informações principalmente no Google, no Yahoo, no Ask Jeeves, no About, no FindWhat e no LookSmart, entre outros. De acordo com o *site* do serviço, o uso do mascote de um cachorro como marca, é uma apologia à idéia de tempos antigos em que os cães recolhiam o jornal da manhã e entregavam a informação nas mãos dos seus donos (DOGPILE, 2004).

O Vivíssimo foi fundado por uma equipe de cientistas do Departamento de Ciência da Computação, na Universidade Carnegie Mellon, dando continuidade a uma pesquisa

⁴⁰ Para maiores detalhes, recomenda-se a leitura do seguinte artigo: SULLIVAN, Danny; SHERMAN, Chris. 4th Annual Search Engine Watch Awards. Publicado em 06/02/2004. Disponível em: <<http://searchenginewatch.com/awards/article.php/3309841#meta>>. Acesso em 04/10/2004.

originalmente subvencionada pela National Science Foundation. Em junho de 2000, após a universidade haver liberado os direitos da propriedade intelectual aos idealizadores do projeto, o grupo fundou uma organização comercial.

O sistema, que é uma combinação de conhecimentos das áreas de inteligência artificial, recuperação da informação, cognição e matemática, baseia-se, principalmente, na tecnologia de *clustering* que, por sua vez, fundamenta-se no agrupamento de documentos de acordo com suas descrições. Neste caso, as descrições são obtidas pela metabusca de URL's, títulos e resumos em outros serviços de busca direta, seguida do reagrupamento dos resultados por similaridade. Não há utilização de bases de dados, uma vez que o sistema não possui *crawler* ou *spider* e nem indexa os *sites* propriamente ditos.

A técnica de *clustering* difere de outras como a classificação, a construção de taxonomia, e a marcação de meta dados, uma vez que é totalmente automatizada, dispensando a intervenção humana. De acordo com informações obtidas no *site* do serviço, o maior desafio da técnica de *clustering* tem sido o de buscar e organizar rapidamente os grupos de conhecimento relativos à pesquisa do usuário. Tal iniciativa apóia-se no uso de um algoritmo heurístico que tem propiciado, segundo o *site*, bons resultados para páginas da *web* em geral; resumos de patentes; publicações acadêmicas e escolares; anais de encontros; entre outros tipos de informação (VIVISSIMO, 2004).

O KartOO é um serviço de metabusca de informação na *web* que trouxe uma maneira inovadora de apresentação dos resultados: trata-se da representação dos seus resultados na forma de mapas interativos. Os sítios encontrados são representados por círculos maiores ou menores, de acordo com a sua pertinência ao assunto. Entre as principais características de sua estratégia de recuperação, estão a capacidade de realizar buscas por palavras no endereço da página; a capacidade de realizar buscas avançadas; e a capacidade de realizar buscas em linguagem natural, inclusive com o uso de pontos de interrogação colocados no final da sentença (KARTOO, 2004). A ilustração 37 mostra os resultados de uma pesquisa pelo termo legislação utilizando-se esse serviço. Destaca-se como uma das principais características do serviço, o fato de que, mesmo que o termo seja digitado em português, o painel esquerdo da tela apresenta uma lista de termos relacionados, em idioma francês.

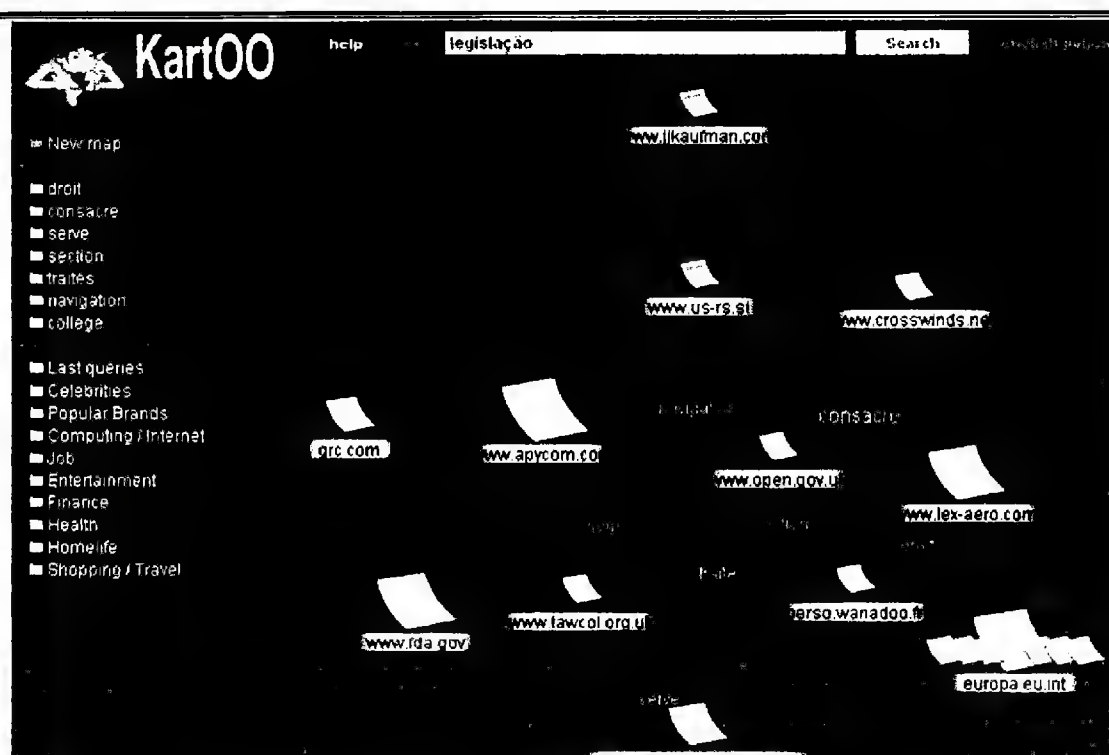


Ilustração 37: Resultado da busca pelo termo legislação no KartOO.

Fonte: (KARTOO, 2004)

O ProFusion organiza as fontes de informação em grupos ou categorias hierarquizados verticalmente, por assuntos. Os assuntos são organizados por colunas, sendo que a primeira, à esquerda, refere-se aos assuntos amplos e as demais, à direita, aos assuntos específicos de maior proeminência dentro de dado assunto amplo. O sistema permite ainda ao usuário a criação dos seus próprios grupos ou categorias de fontes de informação. Uma outra característica relevante é a página de Análise de Busca, que fornece ao usuário informações detalhadas sobre os resultados da busca como, por exemplo, a quantidade de documentos recuperados por fonte de informação; o número total de documentos recuperados; links aparentemente duplicados; links mortos; tempo que durou a busca, entre outras (PROFUSION, 2004).

Com o objetivo de aprimorarem os resultados na pesquisas realizadas na *web*, diversos serviços de busca têm utilizado recursos que procuram estabelecer maior rigor e riqueza ao controle terminológico. Um dos exemplos significativos é o serviço Scirus, especializado em pesquisas científicas. O sistema apresenta recursos tradicionais, como a área de busca por termos livres e a classificação dos resultados com base na ocorrência dos termos exatamente como foram digitados. Os resultados, portanto, exibem ocorrências literais do termo pesquisado nas páginas. Adicionalmente, no entanto, o serviço oferece um painel de termos sugeridos, que poderiam

guardar algum tipo de relação semântica com o termo pesquisado. A ilustração 38 apresenta os resultados de uma pesquisa pelo termo “criança”. Observe-se que, no painel central, são retornados resultados literais e, independentemente de o termo haver sido digitado em caixa alta ou baixa, serão exibidos resultados com todas as letras em minúsculas, todas as letras maiúsculas ou apenas com a inicial maiúscula. No painel direito, são apresentados termos alternativos, inclusive em inglês, sugerindo que algum tipo de controle terminológico está sendo realizado no sistema.

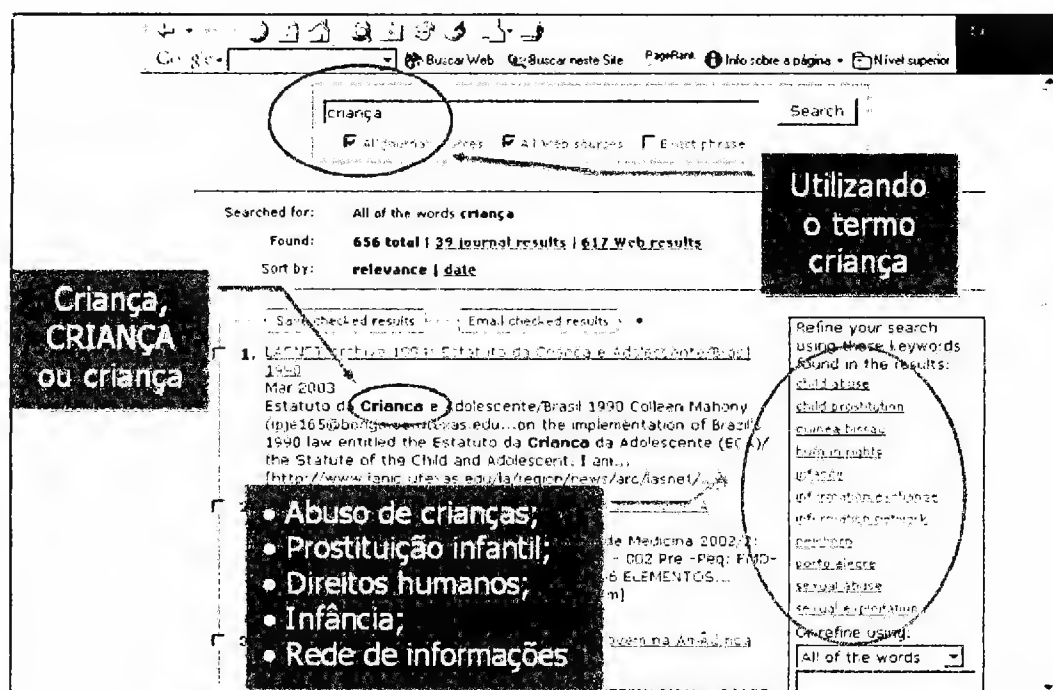


Ilustração 38: Iniciativas de controle terminológico no serviço de busca Scirus.com

O Scirus é um dos melhores serviços de busca, em se tratando de assuntos científicos – foi considerado o melhor serviço de busca especializada pelos usuários do *site* Search Engine Watch⁴¹. Utilizando tecnologia de ponta, o serviço busca aproximadamente 167 milhões de páginas que tratam exclusivamente sobre ciências, permitindo aos usuários a localização de artigos ou páginas científicas, escolares, técnicas e médicas; bem como de relatórios e manuscritos de artigos e jornais não catalogados por outros mecanismos de busca. O nome é uma homenagem a Scirus, um profeta da mitologia grega, e é justificado, segundo os idealizadores do projeto, pois assim como os visionários e profetas têm a capacidade de julgar os símbolos e sinais do que está por vir, também a ciência é uma disciplina visionária, que trabalha continuamente

⁴¹ Para maiores detalhes, recomenda-se a leitura do seguinte artigo: SULLIVAN, Danny. 2002 Search Engine Watch Awards. Publicado em 28/01/2003. Disponível em: <<http://searchenginewatch.com/awards/article.php/2155921#specialty>>. Acesso em 04/10/2004.

com novas idéias. O sistema, então, pretende ser uma ferramenta pró-ativa para dar suporte aos atuais visionários: os cientistas (SICRUS, 2004).

O fato de as páginas indexadas possuírem apenas conteúdo científico representa um diferencial muito interessante para o usuário, uma vez que os termos pesquisados apenas apontarão para conteúdos pertinentes à língua científica especializada, filtrando outras acepções pertinentes à língua geral; além disso, o sistema indexa os *sites* mais profundamente do que em apenas dois níveis, descobrindo maior quantidade de informações relevantes. Em relação aos delimitadores de pesquisa, o usuário poderá especificar sua busca por áreas, como saúde, física e ciências sociais; por nome de autor, periódico ou artigo; por intervalo de data; por natureza do acervo: conferências, resumos ou patentes; é possível também a personalização e a gravação das pesquisas.

Os resultados são classificados por relevância e por data. A relevância é obtida com base na localização e na frequência dos termos na página – classificação estática, correspondente à primeira metade do algoritmo de busca; bem como pelo cálculo do número de ligações para uma página, isto é, quanto mais ligações para uma página, maior sua relevância no contexto – classificação dinâmica, correspondente à segunda metade do algoritmo de busca. O peso final é dado pela soma dos valores estático e dinâmico. O sistema não utiliza meta marcações, pois são vulneráveis à intervenção dos autores das páginas em busca de melhores classificações nos sistemas de busca.

Em relação à natureza dos domínios, o sistema indexa: 58,5 milhões de *sites* .edu; 18 milhões de *sites* .org; 6,8 milhões de *sites* .ac.uk (acadêmicos no Reino Unido); *sites*, 18,6 milhões de *sites* .com; 5 milhões de *sites* .gov; mais de 45 milhões de *sites* especializados em ciência, tecnologia e medicina. Além de páginas da *web* o sistema indexa periódicos, incluindo: 14,6 milhões de citações na MEDLINE⁴²; 5,5 milhões de artigos com texto integral no ScienceDirect⁴³; 1,2 milhões de patentes a partir da USPTO⁴⁴; 261 mil e-prints no ArXiv.org⁴⁵;

⁴² MEDLINE – base de dados compilada pela Biblioteca Nacional de Medicina dos Estados Unidos. Para mais informações, recomenda-se uma visita ao sítio da instituição disponível em: <http://www.nlm.nih.gov/>.

⁴³ ScienceDirect – serviço implementado a partir de 1997 pela empresa de consultoria em informação Elsevier, a ScienceDirect autodenomina-se como a maior coleção eletrônica de bibliografias e textos integrais nas áreas de ciência, tecnologia e medicina. Mais detalhes poderão ser obtidos em: http://www.info.sciencedirect.com/licensing_options/index.shtml.

⁴⁴ USPTO – escritório de Marcas Registradas e Patentes do Estados Unidos. Para mais informações, recomenda-se uma visita ao sítio da instituição disponível em: <http://www.uspto.gov/>.

⁴⁵ ArXiv.org – iniciado em 1991, o serviço compreende um arquivo totalmente automatizado de artigos científicos eletrônicos. O acervo cobre principalmente a área de física e suas disciplinas correlatas: matemática, ciências não-lineares, linguística computacional e neurociência. Para mais informações, recomenda-se uma visita ao sítio da instituição disponível em: <http://www.arxiv.org/>.

5.352 artigos com texto integral na BioMed Central⁴⁶; 10.600 relatórios técnicos da NASA; e 7.672 artigos de texto integral do projeto Euclid⁴⁷ (SCIRUS, 2004).

Na mesma direção do aprimoramento das técnicas de busca na *Internet*, outro exemplo relevante é a introdução, em meados de 2003, do operador semântico no serviço Google. Como esse operador ainda não está disponível para o idioma português, as ilustrações 39 e 40 apresentam os resultados de uma pesquisa pelo termo “*Laws*”, utilizando o operador semântico “~”. Como se pode observar na ilustração n.º. 25, o sistema retorna resultados que contenham tanto sinônimos, quanto outros termos relacionados ao termo pesquisado, como por exemplo: *Law, Statutes, Code, Act, Rights, Rules, Legislation* e *Regulation*. Além desse recurso, o Google introduziu também o uso da busca por radical que, realizada automaticamente, retorna termos variantes de uma mesma palavra. Para exemplificar, uma busca pelos termos *run linux* retorna, também, *running linux*.

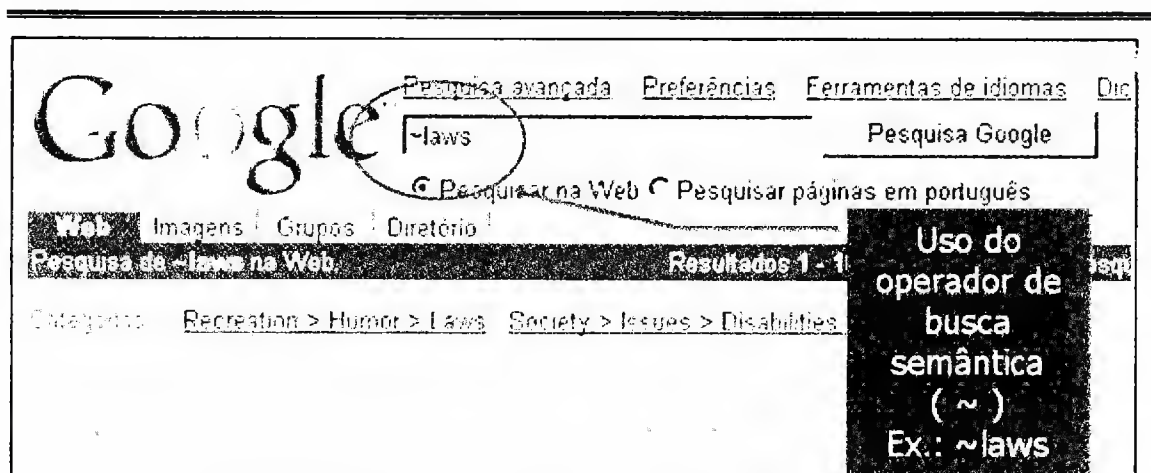


Ilustração 39 Sintaxe para a utilização do operador semântico no Google

⁴⁶ BioMed Central – editora independente que fornece acesso livre e imediato a documentos relativos a pesquisas médicas. Mais informações podem ser obtidas no sítio da instituição, disponível em: <http://www.biomedcentral.com/info/>.

⁴⁷ Projeto Euclid – Iniciativa da biblioteca da Universidade de Cornell, com o objetivo de promover a comunicação acadêmica nas áreas de matemática teoria e aplicada e de estatística. Para mais informações, recomenda-se uma visita ao sítio institucional, disponível em: <http://projecteuclid.org/Dienst/UI/1.0/Home>.

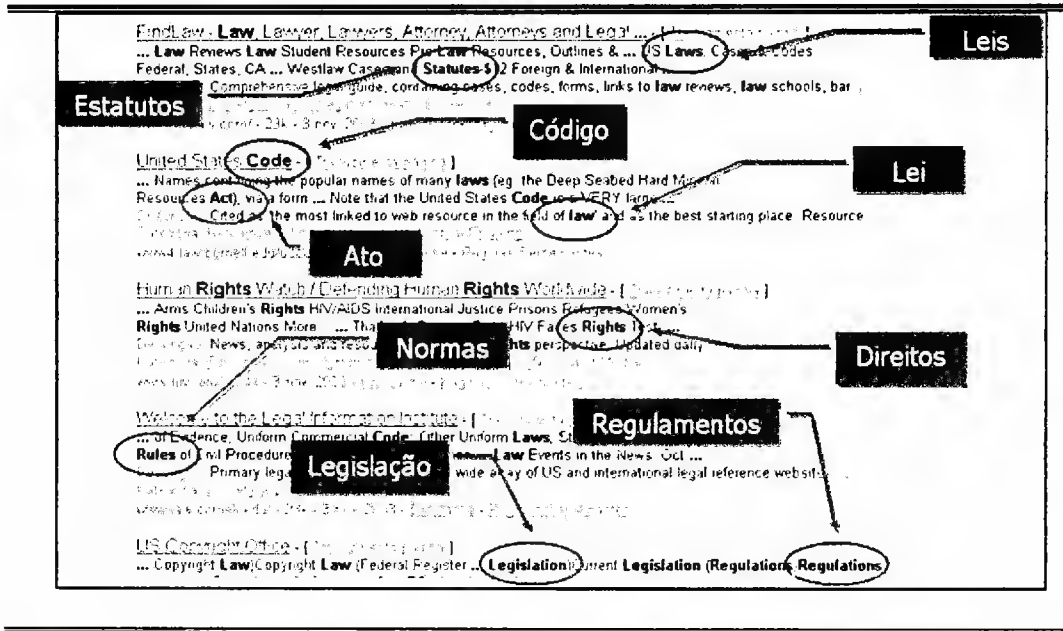


Ilustração 40: Resultados de uma pesquisa semântica no Google

No campo das inovações de serviços, em 2003 o Google lançou o *AdSense*, um programa de anúncios baseado em contexto, que permite com que pessoas consigam obter rendimentos apenas pela colocação de anúncios relevantes em suas páginas; nos últimos meses (2003-2004), o Google iniciou um investimento pesado na introdução dos citados elementos semânticos ao seu serviço de busca, embora muitos usuários continuem alheios a tais mudanças.

2.3.5.1 Otimização de páginas web para mecanismos de busca

Existem na web diversos recursos, como tutoriais ou guias, que abordam a otimização de páginas para a indexação por mecanismos de busca. Entre tais recursos, a maioria recomenda que se forneçam meta dados no código fonte de uma página web, para a descrição de informações. Tais meta dados podem ser: palavras-chaves; título; descrição do conteúdo; resumo; autor; informações de *copyright*; distribuição; idioma; entre outras. A gravação dessas *meta tags* no conteúdo da página, que pode ser realizada com o uso do elemento `<meta>`, que integra a biblioteca da linguagem HTML, foi muito utilizada para favorecer a indexação de páginas da web pelos mecanismos de busca. No entanto, devido aos abusos cometidos por *webmasters*, que desenvolveram técnicas de *spamming*, as empresas e organizações que desenvolvem mecanismos de busca passaram a utilizar outras metodologias de indexação. Desse modo, embora ainda possam se utilizar das *meta tags* para a descoberta de conteúdo de uma página, essa técnica passou a ter menor importância na indexação automática da web e seu uso tem sido mais bem fiscalizado com tecnologias *anti-spam* que, inclusive, apenas os *sites* cujo uso for considerado abusivo. Tal fato não é um indicativo de que se deva evitar o uso das *meta tags* mas, ao contrário, seu uso

responsável, respeitando as políticas *anti-spam*, pode favorecer a classificação de uma página web nos diversos mecanismos de busca – o próprio texto de ajuda do serviço Yahoo! Recomenda que se utilizem as tags *description* e *keywords* para a otimização dos resultados da busca (YAHOO!, 2004). A tabela 18 contém uma lista das principais *meta-tags* usadas para promover a classificação de uma página da web junto a um mecanismo de buscas, e foi adaptada de Bowman (2004) com complementos de SeoConsultants.com (2004a).

Tabela 18: principais meta tags utilizadas para a definição de políticas de indexação por mecanismos de busca

Nome da Tag	Sintaxe	Descrição
Robots	<meta name="robots" content="index, follow">	Indica ao robô, <i>spider</i> ou <i>crawler</i> se lhe será ou não permitida a indexação da página web
Abstract	<meta name="abstract" content="descrição ">	Similar à <i>meta tag Description</i> , pode ser utilizada para fornecer uma segunda descrição a respeito do conteúdo da página
Author	<meta name="author" content="webmaster da página ou responsável pela informação">	Fornece a indicação do autor da página
Copyright	<meta name="copyright" content="informação sobre o direito autoral">	Indica o detentor dos direitos autorais sobre o conteúdo da página
Distribution	<meta name="distribution" content="global ou local ou IU">	Define a audiência de um <i>site</i> (global, local, IU (uso interno – intranets))
Expires	<meta name="expires" content="indicação de data">	Indica se a página possui uma data de expiração
Language	<meta http-equiv="content-language" content="idioma">	Indica o código do idioma utilizado na página, de acordo com as convenções de código internacionais.
Refresh	<meta http-equiv="refresh" content="n" de segundos;URL=endereço">	Redireciona o visitante para outro endereço, em um intervalo de segundos. Essa tag é considerada <i>spam</i> e seu uso é apenas pelos mecanismos de busca.
Description	<meta name="description" content="descrição">	Indicação textual do conteúdo da página. Para evitar que o texto seja truncado pelos mecanismos de busca, ao apresentarem os resultados, é interessante que a tag <i>Description</i> possua apenas entre 140 e 200 caracteres. O truncamento varia de acordo com o mecanismo de busca e com o número de caracteres exibidos nos resultados da busca.
Keywords	<meta name="keywords" content="palavras-chaves que descrevem a página ">	Indica a lista de palavras-chaves que representam o conteúdo da página.

Fonte: adaptado de (BOWMAN, 2004)

Para que o uso das meta tags seja efetivo em termos de promoção de uma página da web junto aos mecanismos de busca, é importante que se evitem alguns erros comuns, que podem ser considerados como técnicas de *spamming*, como, por exemplo: não se deve escrever textos que possuam a mesma cor do fundo da página ou de uma célula de uma tabela; não é recomendável que se repita uma palavra muitas vezes na tag *keyword* – o mesmo vale para as tags *title* e *description* e esta técnica de *spamming* é conhecida como *keyword stuffing*; quaisquer palavras listadas no título da página devem ser também incluídas nas tags *keyword* e *description*, além de

ocorrerem em diferentes pontos da página, como texto comum (BOWMAN, 2004). A relevância da *meta tag keywords* atualmente (2004), é mínima; no entanto, caso seja utilizada, é importante que as palavras escolhidas estejam também grafadas no título e no primeiro parágrafo da página (SEOCONSULTANTS.COM, 2004a).

Tabela 19: Uso do elemento Link para otimizar a indexação pelos mecanismos de busca

Valor do atributo	Sintaxe	Descrição
start	<link rel="start" content="index, follow">	Indica o início de uma coleção de documentos
Alternate	<link rel="alternate" href="link-reference">	Designa uma versão substituta para um documento. Pode ser utilizada juntamente com o atributo lang, para especificar uma versão do documento em outro idioma; ou com o atributo media, para designar uma versão em outro tipo de suporte.
Start	<link rel="start" href="link-reference">	Refere-se ao primeiro documento de uma série; este link indica aos mecanismos de busca que o documento é considerado pelo autor como o ponto de partida de uma coleção.
Next	<link rel="next" href="link-reference">	Designa o próximo elemento em uma série linear.
Prev (ou previous)	<link rel="prev" href="link-reference">	Designa o elemento anterior em uma série linear.
Contents	<link rel="contents" href="link-reference">	Refere-se a um documento que serve como um sumário do site. Alguns mecanismos de busca podem interpretar esse valor como sinônimo de ToC (Table of Contents).
Index	<link rel="index" href="link-reference">	Refere-se a um documento que fornece um índice de assuntos para o documento corrente
Glossary	<link rel="glossary" href="link-reference">	Refere-se a um documento que fornece um glossário de termos para o documento corrente
Copyright	<link rel="copyright" href="link-reference">	Refere-se a um documento que contém uma declaração de copyright para o documento corrente
Chapter	<link rel="chapter" href="link-reference">	Refere-se a um documento que serve como capítulo para uma coleção de documentos
Section	<link rel="section" href="link-reference">	Refere-se a um documento que especifica uma seção para uma coleção de documentos
Subsection	<link rel="subsection" href="link-reference">	Refere-se a um documento que especifica uma subseção para uma coleção de documentos
Appendix	<link rel="appendix" href="link-reference">	Refere-se a um documento que especifica um apêndice para uma coleção de documentos
Help	<link rel="help" href="link-reference">	Refere-se a um documento que inclui informações de ajuda, como links para outras fontes, detalhamento de informações, entre outras.

Fonte: Adaptado de SeoConsultants.com (2004a).

Outro elemento que pode facilitar a indexação pelos mecanismos de busca é o elemento *link*. Esse elemento possui diversos atributos que podem fornecer uma indicação de ordenação do documento em relação a outros documentos em um *site*. Cria-se, assim, uma seqüência de páginas que pode ser útil à indexação automatizada. A tabela 14 fornece algumas indicações de uso do elemento link, com vistas à indexação.

Devido ao fato de que a *meta tag keywords* atualmente (2004) possui pouco peso na classificação dos resultados dos mecanismos de busca, Wall (2004a) sugere que as palavras-chaves escolhidas para a representação de conteúdo dos documentos de um *site* sejam colocadas em partes estratégicas, como: título; texto de links internos; cabeçalhos <h1>, sendo usado um por página; sub-cabeçalhos <h2> e <h3>; *meta description*; textos de links de barras de navegação; links externos, quando possível. O autor sugere, ainda, a utilização de negritos e listas com marcadores para o destaque de palavras importantes, além da descrição, com palavras-chaves, de todas as *tags* <alt> que ocorrerem na página. Uma outra técnica importante, lembra o autor, é a promoção de uma política de hiperlinks, isto é, o estabelecimento do maior número possível de hiperligações de outros *sites* para o *site* de que se pretende promover junto aos mecanismos de busca. Essa estratégia justifica-se porque os principais mecanismos de busca, a exemplo do Google, orientam suas buscas por meio da análise de hiperlinks.

2.3.6 Padrões de meta dados na web

A palavra meta dados origina-se do prefixo grego *metá*, que significa mudança, posterioridade, além, transcendência e reflexão crítica sobre alguma coisa. A palavra dado, no sentido informático, refere-se a informação em forma codificada (FERREIRA, 1986). Assim, juntando-se os termos, pode-se chegar a uma noção de transcendência sobre a informação, ou de uma possível descrição da informação que os dados representam. Nessa perspectiva, meta dado é todo dado físico, isto é, contido em algum software ou qualquer tipo de mídia, e que fornece informação sobre outra informação. De uma maneira mais simples, é possível se dizer, também, que meta dado é “*dado sobre dado*”. O conceito de meta dado está associado, ainda, a conhecimento – contido em pessoas ou mídias – dentro e fora de uma organização, incluindo informação sobre dados físicos, processos técnicos e de negócios, regras e restrições sobre os dados e estruturas de dados usados por uma corporação. Assim, falar sobre meta dados é, também, falar sobre conhecimento (MARCO, 2000).

Embora possa parecer que os conceitos de *meta dado* e de *repositório* de dados são conceitos recentes, eles têm suas origens no início dos anos 70. Nessa época, os primeiros repositórios de dados comerciais eram denominados dicionários de dados e tinham mais enfoque

nos dados, do que no conhecimento. A preocupação comum na época era, então, propiciar um repositório centralizado de informações sobre os dados em si – como suas definições, relações, origem, domínio, uso e formato –, com vistas a subsidiar o trabalho dos administradores de bancos de dados, cujas funções principais eram o planejamento, controle, avaliação, armazenamento e uso de tais dados. Dessa maneira, utilizando um dicionário de dados, o administrador de banco de dados estava apto a especificar requisitos, modelar dados corporativos e prestar suporte às bases de dados. A partir dos anos 90, o conceito de meta dados foi se alterando e, atualmente, o desafio com que se deparam os especialistas em informação é a habilidade para diferenciarem *repositórios de meta dados* de *dicionários de dados*. Os repositórios de meta dados realizam todas as funções de um dicionário de dados, mas possuem um escopo mais amplo, determinado, entre outros fatores, por necessidades que emergiram a partir dos anos 90: os sistemas tornaram-se mais flexíveis e não integrados; criaram-se os conceitos de *armazéns de dados* e *mercados de dados*; com o advento dos negócios realizados em linha, muitas necessidades de informação dos usuários ficaram ainda sem ser preenchidas; com a popularização dos negócios realizados em linha, aumentou, proporcionalmente, a necessidade de se implementar maior nível de segurança e de confiança aos dados envolvidos nas transações (MARCO, 2000).

No sentido do saber biblioteconômico, meta dado pode ser considerado como dado estruturado, que compartilha diversas características similares para a catalogação, e que descreve as características de um determinado recurso informacional. Portanto, um registro de meta dados consiste em um número pré-definido de elementos que representam atributos específicos de um objeto, sendo que a cada elemento pode estar associado um ou mais valores (TAYLOR, 2003). A cada elemento de um registro de meta dados estão associadas propriedades como nome, domínio, tipo, tamanho, formato, responsabilidade e, no caso de sistemas informatizados, origem. A propriedade *nome* identifica univocamente o elemento. O *domínio* especifica o conjunto de valores que o elemento de dado pode assumir; por exemplo, ao domínio do idioma de um documento poderiam estar associados os códigos relativos a esses idiomas. A propriedade *tipo* refere-se ao tipo de conteúdo que se poderá armazenar no elemento de dados, como por exemplo, valores textuais, numéricos, datas, entre outros. A propriedade *tamanho* refere-se ao número máximo de dígitos que poderão ser armazenados no elemento. O *formato* refere-se à forma com que a informação deverá ser representada, e define as posições de determinados símbolos como traços, pontos, barras, entre outros; nos sistemas informatizados, os formatos são também conhecidos como *máscaras*. A *origem* determina se os valores a serem armazenados no elemento de dados serão digitados pelo usuário, utilizando um formulário de

entrada de dados, ou se serão calculados ou gerados automaticamente pelo sistema. A tabela 20 representa um exemplo de um registro de meta dados:

Tabela 20: Exemplo de registro de meta dados

Nome do Elemento	Valor
Título	Catálogo da web
Criador	Dagnija McAuliffe
Editor	Biblioteca da Universidade de Queensland
Identificador	http://www.library.uq.edu.au/iad/mainmenu.html
Formato	Texto/html
Relação	Web site da biblioteca

Fonte: Adaptado de (TAYLOR, 2003)

O uso de meta dados apresenta muitos benefícios, como por exemplo: a) meta dados são estruturados e, desse modo, podem formar a base para o desenvolvimento de sistemas de busca mais avançados; b) meta dados podem ser convertidos para outros formatos, de modo que possam interoperar com diferentes protocolos de busca e recuperação; c) em relação a documentos publicados na *Internet*, pensando-se em recuperação automatizada, torna-se mais fácil a extração de conteúdo de uma base de dados de descrição de recursos, do que de arquivos textuais em formato HTML; d) meta dados administrativos podem ser utilizados para facilitar o gerenciamento do sistema de informação, uma vez que ajudam a avaliar quando os recursos devem ser revistos ou removidos da base de dados (DAY, 2000).

Na área de Ciência da Informação há diversos esquemas de meta dados, citando-se, entre as iniciativas mais populares, o Dublin Core, o AACR2 (*Anglo-American Cataloging Rules*) e o GILS (*Government Information Locator Service*). Obviamente, segundo as características da área de conhecimento em questão e do sistema de informações, os meta dados podem variar significativamente. Não obstante essa possibilidade de variação, para Taylor (2003), cada esquema deverá possuir pelo menos três características: número limitado de elementos; o nome de cada elemento; o significado de cada elemento. Em relação ao agrupamento dos elementos de meta dados de um recurso informacional, Day (2000) sugere uma classificação, segundo o uso, em: a) dados descritivos – são aqueles que contêm informação passível de utilização por sistemas de busca, como: título; título abreviado; título alternativo (em outro idioma, por exemplo); subtítulo; descrição; URI; autor; idioma; codificação do conjunto de caracteres; organização (de criação e de hospedagem do documento); formatos de arquivo (por exemplo, texto, imagem, áudio, vídeo), meio físico, propriedade intelectual ou *copyright*, disponibilidade (por exemplo, indicação se o acesso à informação é gratuito ou sujeito a registro); software requerido para acesso (navegadores específicos, software de MIDI); e público alvo; b) dados do assunto – são aqueles que descrevem o conteúdo do documento, como: palavras-chaves; código de classificação; sistema de

classificação; e termos do tesouro ou cabeçalho de assuntos; c) dados administrativos – são aqueles que facilitam a organização e a administração do sistema de informações, como, por exemplo: responsável pela manutenção do documento; data de adição do documento ao sistema; data da última modificação; data de expiração (por exemplo, de um anúncio de conferência); catalogador do documento; origem do registro; e proprietário dos direitos. Day (2000) considera que, embora essa lista de elementos não seja exaustiva, ela contém aqueles que, em sua avaliação, são fundamentais para a organização da informação.

A escolha de um formato de meta dados é uma das decisões mais importantes no desenvolvimento de sistemas de informação, e é essencial que tal formato permita a integração com o software que forma a base do sistema. É também importante que o formato contenha todos os campos considerados apropriados à descrição dos dados do sistema. Como já mencionado, o formato Dublin Core é, na atualidade (2004), um dos mais importantes meios de descrição de recursos na *Internet* e é um exemplo das iniciativas que se propõem a definir um *core* (núcleo) de meta dados, isto é, um conjunto de dados essenciais que possam ser utilizados como padrão para a descrição de recursos na *web*. Ao lado do formato Dublin Core, outra iniciativa que merece destaque no âmbito desta pesquisa é o *Resource Description Framework* (RDF). A arquitetura RDF é discutida em maior profundidade na seção Web Semântica, deste capítulo, e também no capítulo de Fundamentação Teórica e Metodológica.

O esquema Dublin Core caracteriza-se pela sua utilidade e flexibilidade na representação de dados. A semântica dos elementos foi projetada para ser claramente compreendida por um grande número de pessoas, sem a necessidade de treinamento, bem como para interagir com o maior número possível de tecnologias. Assim, a intenção da metodologia não é suplantando outros recursos usados para descrever documentos mas, ao contrário, complementá-los. Outra característica interessante do formato Dublin Core é que embora, por padrão, todos os elementos sejam opcionais, é possível que o administrador do sistema ou especialista em informação defina situações de obrigatoriedade, quando necessário (TAYLOR, 2003).

A *Dublin Core Metadata Initiative* (DCMI) (ilustração 27) iniciou-se em 1995, com um *workshop* realizado em Dublin, no estado de Ohio nos Estados Unidos, e congregou bibliotecários, pesquisadores especialistas em biblioteconomia e em bibliotecas digitais, provedores de conteúdos e especialistas em técnicas de marcação de textos. Como o evento, patrocinado pelo *Online Computer Library Center* (OCLC) e pelo *National Center for Supercomputing Applications* (NCSA), teve lugar em Dublin, o conjunto de elementos sugeridos recebeu o nome

de Dublin Core. O objetivo desse workshop era promover a melhoria dos padrões de descoberta da informação na web. A proposta original, inicialmente congregando interesses de um pequeno grupo, rapidamente atraiu interesse, em escala global, de uma grande variedade de fornecedores de informação nas áreas de artes, ciências, educação e negócios, além dos setores do governo. Devido à proliferação dos documentos eletrônicos em face da impossibilidade de os bibliotecários e especialistas em informação realizarem a catalogação de todos os documentos, pensou-se, inicialmente, em um conjunto de treze elementos que permitissem aos próprios autores e administradores de *sites* fazerem essa marcação, a partir da utilização de regras simples (ilustração 41). Assim, a tarefa tornar-se-ia factível, mesmo para aqueles que não dominassem os princípios de catalogação e classificação. Posteriormente, o número de elementos foi elevado para quinze – *título, criador, assunto, descrição, editor, colaborador*⁴⁸, *data, tipo, formato, identificador, fonte, idioma, relação, cobertura e direitos*. Embora o padrão tenha sido concebido originalmente para a descrição de documentos da web, o Dublin Core tem sido utilizado com outros tipos de materiais e aplicações (NISO Press, 2001; NISO Press, 2004).

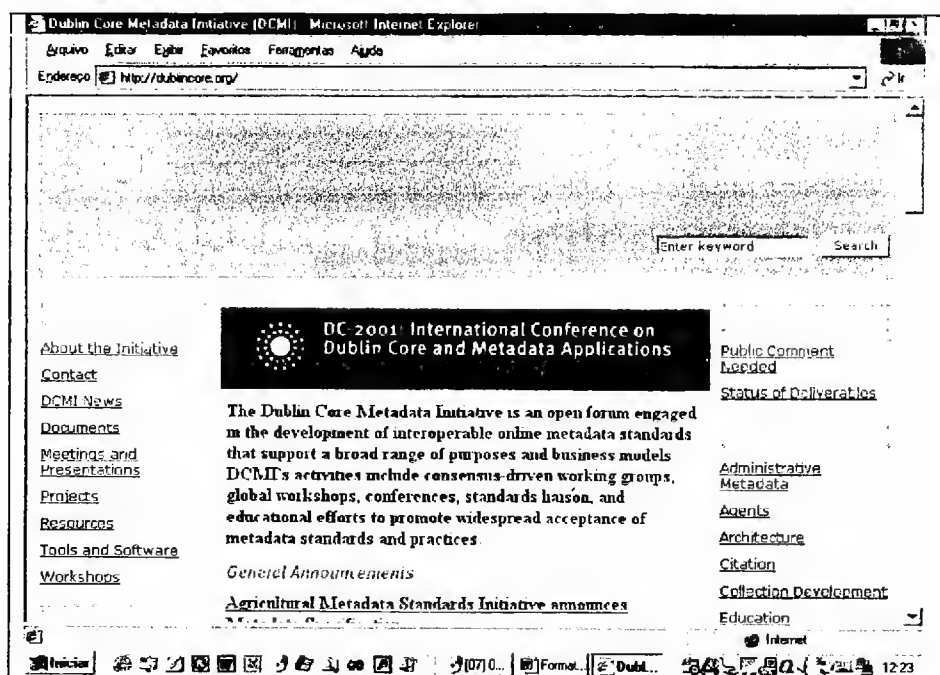


Ilustração 41: Site do padrão Dublin Core na Internet

Fonte: <<http://dublincore.org>>

⁴⁸ O elemento *contributor* foi traduzido aqui como *colaborador* – apenas por uma questão de estilo – embora haja registro do termo *contribuidor* no idioma português brasileiro na obra de Houaiss (2001).

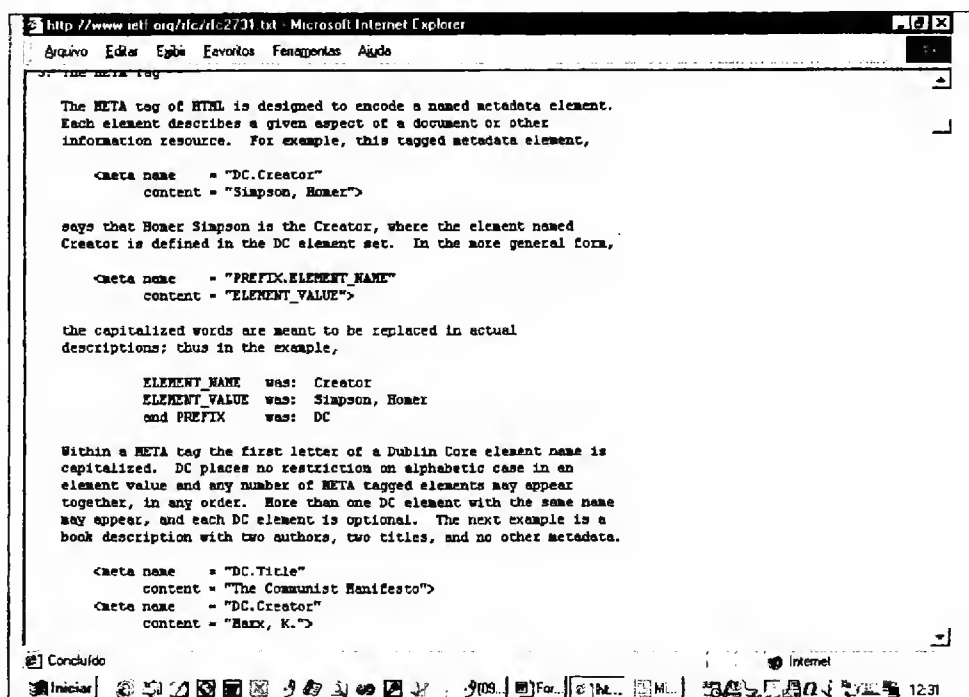


Ilustração 42: Disposição de meta tags Dublin Core em uma página HTML

Fonte: <http://www.ietf.org/rfc/rfc2731.txt>

Dempsey e Heery (2000) analisaram diversos formatos de meta dados e os sistematizaram, de acordo com sua complexidade, em bandas. A classificação estatui que a banda 1 compreende estruturas de organização mais simplificadas, aumentando gradativamente a complexidade até se atingir a banda 3. Assim, a banda 1 corresponde aos sistemas de indexação automática de texto integral, que é realizada pelos serviços de busca da *Internet*. Nessa banda, os meta dados são criados automaticamente, a partir da extração, realizada pelos *crawlers* ou *spiders*, de informações consideradas relevantes em um documento. Como abordado na seção que tratou dos mecanismos de busca, tais informações podem se referir algumas partes do documento, como título, parte inicial do texto ou a *metatags*. A desvantagem dos meta dados nessa banda, está no fato de que sua estrutura normalmente é oculta, uma vez que as arquiteturas dos sistemas são protegidas, devido a interesses comerciais. Como os serviços de indexação desse tipo geralmente utilizam os próprios documentos para a obtenção de informações e, também, pelo fato de serem muito amplos com relação à cobertura, tais serviços são muito mais efetivos na localização do que na descoberta. Se o usuário busca por um item conhecido, os resultados da busca provavelmente serão úteis, caso o tema seja mais amplo, é de se imaginar que a precisão seja mais baixa. Como a web apresenta um alto grau de desorganização, no que tange a estruturas de meta dados, e uma vez que as tecnologias incluídas nessa banda extraem informação com base exclusivamente nos documentos, terminam por operar em nível de cópia, isto é, não há como

reconhecer recursos duplicados ou estabelecer relações entre diferentes recursos, o conteúdo dos índices é apenas um extrato do conteúdo dos acervos indexados.

A banda 2 de Dempsey e Heery (2000) compreende serviços que se baseiam em padrões que traduzem os esforços da comunidade de Ciência da Informação para a busca em *sites* organizados por diretórios. Os formatos incluem recursos de descrição construídos manualmente e são um pouco mais complexos do que os formatos originados pela indexação automática. O formatos da banda 2 baseiam-se na criação de meta dados referentes a recursos selecionados; o padrão Dublin Core é o mais característico dessa banda.

Em geral, os formatos da banda 2 são criados de modo que possam ser aplicados por usuários não-especialistas na descrição de documentos de diferentes tipos e de diferentes domínios. As descrições dos documentos podem ser feitas apenas manualmente ou podem compreender o aperfeiçoamento manual de descrições realizadas automaticamente. Os formatos incluem elementos descritivos e outros atributos, e são projetados para permitirem buscas baseadas em campos, e não requerem conhecimentos aprofundados

Os formatos concernentes à banda 3 (ilustração 43) são mais complexos e requerem os conhecimentos de especialistas para serem criados e mantidos. Essa banda inclui o formato MARC, amplamente utilizado pela comunidade bibliotecária e nos catálogos de bibliotecas. A descrição dos registros é realizada intelectualmente e requer conhecimentos do não apenas sobre o formato, como também sobre as regras de classificação. Alguns formatos dessa banda são tão complexos, que podem ser utilizados não apenas para a localização e descoberta de informações em documentos, mas também para a organização de coleções inteiras. Esses formatos, como o TEI, por exemplo, propiciam níveis de análise de conteúdo e navegação entre categorias de objetos. Usualmente associados à atividade de pesquisa ou acadêmica, são expressivos o suficiente para envolverem uma ampla diversidade de relações, em diferentes níveis. A tabela 21 apresenta o esquema de classificação em bandas, de Dempsey e Heery (2000).

Tabela 21: Tipologia de formatos de meta dados

Banda 1 (índices de texto integral)	Banda 2 (formatos genéricos com estrutura simples)	Banda 3(i) (estruturas mais complexas, domínios específicos)	Banda 3 (ii) (integram uma arquitetura semântica mais ampla)
Formatos proprietários	Dublin Core IAFA/WHOIS++ RFC 1807	FGDC MARC GILS	TEI headers ICPSR EAD CIMI

Fonte: (DEMPESEY; HEERY, 2000)

Como já mencionado, a banda 3 agrupa padrões mais avançados de *meta dados*, baseados na linguagem SGML (*Standard Generalized Mark-up Language*) e suas derivadas (HTML,

XML e XHTML). As principais características dessa banda estão ligadas ao avanço dos algoritmos associados de descoberta da informação não mais de localização. Destacam-se como padrões dessa banda: TEI (*Text Encoding Initiative*), ICPSR (*Inter-University Consortium for Political and Social Research*), USMARC e EAD (*Encoded Archival Description*);

TEI tags	AACR2 (Chapter 9)	MARC tags
<titleStm> <title>	9.1B-E	245 a, b, a, p, h
	9.7B4	246
<titleStm> <author>	Main Entry See Chapter 21	100 a, b, c, d, e
<map> <name>	9.1F	500 700, 710, 711
<extent>	9.3B	256
<publicationStm> <publisher>	9.4B	260 b
<publicationStm> <pubPlace>	9.4B	260 a
<idno type>		089
<availability> <p>	9.7B9	500
<availability> <p>URL	Holdings	856 42
<date>	9.4B	Fixed field: Date1 260 c
<seriesStm> <p>	9.6B1	440
<notesStm> <note>	9.7D9	500
<bb:Full> <titleStm>	Title added	240
<title>	entry	700 X2 t 730 773 t 534 t
<bb:Full> <titleStm> <author>	9.1F, 9.7B6	245 c
<bb:Full> <editionStm> <p>	9.2B	534 b
<bb:Full> <extent>	9.3B	534 e
<bb:Full> <publicationStm>	9.4B	534 c
<publisher>		

Ilustração 43: Padrões de meta dados da banda 3 de Dempsey e Heery (2000)

Fonte: <http://www.libraries.psu.edu/tas/fca/ccda/tf-tei4.html>

A tabela 22 apresenta uma sistematização da classificação em bandas, de acordo com as características de cada uma.

Tabela 22: Características dos formatos de meta dados

Característica	Banda 1	Banda 2	Banda 3
Ambiente de uso	Serviços de indexação da internet global; serviços de indexação da web	Serviços de busca seletiva na internet; diretórios de serviços	Descrições de coleções acadêmicas e de outros repositórios importantes
Função	Localização	Localização; seleção	Localização, seleção, avaliação, análise, documentação
Criação	Gerados por robôs, <i>crawlers</i> ou <i>spiders</i>	Gerados por robôs, mais entrada manual	Necessária expertise intelectual de profissionais de informação
Estrutura	Não estruturados	Pares de atributos-valores	Campos e subcampos, qualificadores, marcação estruturada
Protocolos de busca associados	http com interface proporcionada por formulário	http com interface proporcionada por formulário; protocolos de serviço de diretórios (whois++, LDAP) com rotina de busca (Common Indexing Protocol)	Z39.50. Navegadores e buscadores SGML
Status	Proprietário	Padrões emergentes para Internet	Padronização de domínios específicos

Fonte: (DEMPSEY; HEERY, 2000)

2.4 Web Semântica

Esta subseção trata dos aspectos teóricos da *Web Semântica*, focando-se principalmente em ontologias, linguagem XML, esquemas XML, RDF, esquemas RDF e linguagem OWL. Foram estudados alguns textos de autores representativos da área: Berners-Lee (1989;1998,2003), Hyvonen (2002), Lassila (1997), Miller (2002), Daconta (2003), Brickley (2001), Hendler (2004), Miller (2004), Smith (2004), Horrocks (2004), Hori (2004), entre outros. Além desses autores foram consultadas recomendações normativas do W3 Consortium sobre tecnologias da *web* semântica: OWL, RDF, RDF-Schema, XML, XML-Schema. Foram consultadas as especificações semânticas do DARPA para DAML+OIL, DAML-R, DAML-L e DAML-S (DAML,2004)

2.4.1 Caracterização da web semântica

Segundo Nafria (2001) a *Internet*, como a havia imaginado Berners-Lee em 1989, não era exatamente o que é hoje, mas algo muito mais revolucionário e que ainda está por chegar: a *web* semântica. O autor lembra que a idéia da *web* semântica foi levantada no famoso artigo *The Semantic Web*⁴⁹ de Tim Berners-Lee, James Hendler e Ora Lassila, no qual os autores falam de uma *Internet* na qual os computadores não apenas são capazes de apresentar a informação contida nas páginas *web*, mas além disso, de compreendê-las.

Na prática, isto significa que as máquinas – computadores pessoais ou qualquer outro dispositivo conectado à *Internet* – poderão realizar, quase sem necessidade de intervenção humana, uma infinidade de tarefas que simplificam a vida das pessoas. Assim, a *web* semântica “(não) é uma *web* separada, mas uma extensão da atual, na qual a informação é dada com um significado bem definido, melhor habilitando computadores e pessoas a trabalharem em cooperação.” (BERNERS-LEE, 2001). Dessa maneira, a *web* tornar-se-á um ambiente com capacidade de processamento automático de informações, deixando de ser apenas um repositório de documentos para pessoas (NAFRÍA, 2001).

No entanto, para se chegar a essa nova *web*, são necessárias muitas iniciativas que não são fáceis de serem realizadas. Entre essas iniciativas está o desenvolvimento de uma linguagem padrão, que permita a construção de coleções de informações estruturadas, bem como de conjuntos de regras de inferências, de modo que os computadores possam utilizá-las para

⁴⁹ BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. *The Semantic Web*. A new form of *Web* content that is meaningful to computers will unleash a revolution of new possibilities. *American Scientific.com*. may, 2001. Disponível em:

<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso em 07/11/2003.

desenvolverem sua capacidade de localização automática de recursos (BERNERS-LEE, 2001; NAFRÍA, 2001). É necessário, portanto, adicionar-se lógica à *web*, por meio da utilização de uma linguagem que seja capaz de expressar dados e regras para inferências sobre tais dados, fornecendo também as regras para a exportação para a *web* de qualquer sistema de representação do conhecimento existente.

A concepção da *Web Semântica*, que almeja tais objetivos, é uma estrutura de representação de dados na *World Wide Web*, por meio de um esforço colaborativo liderado pelo W3C, com a participação de pesquisadores e de empresas de desenvolvimento de *software* de todo o mundo. A idéia da *Web Semântica* é uma evolução do grupo de trabalho *Metadata Activity*, vinculado ao W3C, cujos maiores objetivos eram o estudo e o desenvolvimento de um *framework* para a representação de informação compreensível por máquinas na *web*. O principal trabalho desenvolvido por esse grupo foi o *Resource Description Framework – RDF*, que tem como propósito fornecer um sistema de ontologia superficial para dar suporte à troca de conhecimentos na *web* (W3C, 2003a). As funções do *W3C Metadata Activity* foram absorvidas pelo *W3C Semantic Web Activity*.

O principal objetivo no desenvolvimento da *web semântica* é o compartilhamento de informações, tanto por máquinas, quanto por pessoas. A organização de sítios e páginas com tal escalabilidade que permita a completa integração entre os recursos, mesmo quando tiverem sido desenvolvidos em plataformas completamente diferentes – e isso refere-se não apenas a *hardware*, mas também a aplicativos –, habilitará os computadores a usarem as informações disponíveis não apenas com propósitos de exibição, senão também com possibilidades de automação, integração e reuso em diferentes sistemas (PALMER, 2001).

A *web* atual tem sido desenvolvida para o leitor humano (BERNERS-LEE, 2001). Uma máquina não pode entender muito do conteúdo da *web*, mas apenas oferecê-lo para que as pessoas o interpretem. Entretanto, a interpretação automática do conteúdo é vital para o desenvolvimento de aplicações inteligentes para a *Internet* (NAFRÍA, 2001). Para Hyvönen (2002), a *Web Semântica*, ou *Internet de significados* é uma visão da próxima geração da *web* que poderá ser utilizada não apenas por humanos, mas também por máquinas. Assim, o autor prevê que, com a ajuda dos padrões e ferramentas da *web semântica*, será possível a representação – de uma maneira compreensível para os computadores – dos significados dos conteúdos da *web*; isso facilitará a implementação de uma novas aplicações e serviços inteligentes.

Segundo Quivey (2002), a idéia da *Web Semântica* tem sido desenvolvida por pelo menos seis anos⁵⁰ e, embora com desenvolvimento lento, o projeto tem tido significativo desenvolvimento: diversas tecnologias estão sendo desenvolvidas com esse esforço. Pelo menos duas tecnologias são chave para o sucesso da *Web Semântica*: a *eXtensible Markup Language* (XML) e o *Resource Description Framework* (RDF) (QUIVEY, 2002; BERNERS-LEE, 2001), que são tecnologias promissoras e que atualmente têm encontrado grande aceitação na *web*. Então, a *Web Semântica* baseia-se principalmente em aplicações que utilizam XML como tecnologia de suporte à estruturação sintática e URIs para a denominação dos recursos (W3C, 2003).

Conforme lembra Nafria (2001) a linguagem XML é bem aceita pois permite aos usuários adicionarem uma estrutura arbitrária a seus documentos; entretanto, essa linguagem nada pode dizer a respeito do significado de tal estrutura. Por esse motivo, a tarefa de atribuir-se significado tem sido realizada com a linguagem (RDF) uma linguagem de marcação em que é possível descrever-se a informação para processamento de significado pelo computador. O terceiro elemento fundamental para construção dessa *web* do futuro são as ontologias. Conforme Berners-Lee (2001), na filosofia, uma ontologia é uma teoria sobre a natureza da existência, dos tipos de coisas que existem; então a ontologia, como disciplina, estuda tais teorias. O autor lembra, ainda, que pesquisadores nas áreas de Inteligência Artificial e de *Web* têm feito empréstimo desse termo, dando-lhe a significação de: um documento ou arquivo que formalmente define relações entre termos. Um exemplo típico de ontologia para a *web* possui uma taxonomia e um conjunto de regras de inferência (BERNERS-LEE, 2001).

Para Palmer (2001), a *web* semântica é uma malha de informação ligada de maneira que seja facilmente processável por máquinas, em uma escala global e pode ser pensada como sendo uma forma eficiente de representar dados na *web* ou como uma base de dados globalmente vinculada. Segundo Quivey (2002), com o advento da *Web Semântica*, as buscas não tornar-se-ão mais fáceis e mais rápidas porque os computadores poderão realizá-las com a mínima intervenção humana. No entanto, para que isso ocorra, é necessário que as páginas sejam escritas em uma linguagem clara, de modo que os computadores possam compreender seu significado. Para a autora, com a linguagem certa e com o conjunto semântico apropriado, nossos computadores estarão aptos a decifrar e fazer inferências sobre o material encontrado em uma página *web*. É importante destacar-se que o computador não “compreende” verdadeiramente qualquer tipo de informação em uma página *web*, mas pode, com o apoio das ontologias,

⁵⁰ Aproximadamente sete anos, contando-se o ano de 2003.

manipular os termos de uma maneira muito mais efetiva, permitindo que essa informação torne-se mais útil e significativa para um usuário humano (BERNERS-LEE, 2001).

Quivey (2002) destaca que o verdadeiro poder da *web* semântica será realizado quando forem criados programas que colem conteúdo de diversas fontes, processem essas informações e troquem os resultados com outros programas; a efetividade desses programas crescerá exponencialmente, à medida que mais e mais programas com a mesma capacidade estiverem disponíveis para interagirem. Ora Lassila (1997) também prevê que a *web* se transforme em algo compreensível para as máquinas, por meio de um mecanismo que permita uma descrição mais precisa das coisas na *web*; isso, afirma o autor, poderia elevar o status da *web* de legível-por-máquina para algo que poderíamos chamar de compreensível-por-máquina.

Em termos de aplicações, os exemplos a seguir mostram alguns benefícios da *web* semântica:

- se for corretamente desenvolvida a *web* semântica pode ajudar a evolução do conhecimento humano, facilitando a colaboração entre equipes de pesquisadores em localidades remotas;

- como os agentes melhorarão sua eficácia em razão dos conteúdos com significado semântico, as buscas serão mais precisas e retornarão melhores resultados, combinando informações em diferentes páginas;

- as escolas poderão integrar-se com sua comunidade: bibliotecas públicas, videotecas, museus, entre outras instituições;

Berners-Lee (2001), postula que o primeiro passo para que se consiga organizar a informação na *web*, de modo a torná-la inteligível para máquinas, é uma mudança no paradigma de como tal informação é armazenada, de modo que os dados sejam processados direta ou indiretamente por computadores. Para Daconta (2003), essa mudança de paradigma tem evoluído em quatro estágios:

- a) informações organizadas em forma de texto ou armazenadas bancos de dados (pré-XML) – nesse estágio inicial, em que a maioria das informações é de propriedade de um aplicativo ou serviço; assim, as “inteligências” concentram-se nas aplicações e não nos dados;

- b) informações organizadas no padrão XML em um domínio específico – nesse estágio, a informação registrada adquire independência dos aplicativos em um domínio específico do conhecimento; a informação agora é “inteligente” o suficiente para que possa ser compartilhada automaticamente de um aplicativo ou serviço para outro;

c) taxonomias e documentos com vocabulários mistos – nesse estágio, a informação pode ser composta por diversos domínios e cuidadosamente classificada em uma taxonomia hierárquica; assim, a descoberta dos dados é feita por meio da classificação; podem ser utilizados relacionamentos simples entre as categorias da taxonomia, de modo a permitir a correlação e a combinação de informações; a informação é “inteligente” o suficiente para ser descoberta e combinada com outras;

d) ontologias e regras – nesse estágio novos dados ou informações podem ser inferidos a partir de dados existentes por meio de regras lógicas pré-existentes; em sua essência, a informação é descrita com relacionamentos concretos e com sofisticados formalismos em que cálculos lógicos podem ser realizados a partir de uma “álgebra semântica”; isso permite a combinação e a recombinação de dados em um nível mais “atômico”, isto é, a informação é analisada a partir de unidades menores para um contexto mais amplo; nesse estágio os dados não existem individualmente, mas fazem parte de um sofisticado microcosmo; tal estruturação encontra aplicações em diversos âmbitos do conhecimento, como, por exemplo na tradução automática do conteúdo de um documento de certo domínio para outro. Nesse nível de estruturação os dados definem-se inteligentes, pois são independentes de aplicativos, são classificados, sua composição é flexível e são parte de um sistema de informação mais amplo (uma ontologia).

Finalizando essa abordagem, os autores relembram que os benefícios da *Web Semântica* não estão destinados apenas para a *World Wide Web* mas, ao contrário, representam também um conjunto de tecnologias que funcionarão igualmente bem em intranets e que resolverão um grande número de problemas existentes nas atuais arquiteturas de tecnologias de informação.

Em termos de tecnologias, a *web semântica* será construída com base nas potencialidades da linguagem XML, da arquitetura RDF e do esquema RDF-Schema, para a representação de dados por meio da descrição de esquemas de marcação personalizados e flexíveis. Além dessas tecnologias, como argumenta Powers (2003), é necessária a utilização de uma linguagem de ontologia que permita a descrição formal da semântica de classes e propriedades utilizadas nos documentos da *web*. Tal linguagem, lembra a autora, deve ir além da semântica básica fornecida pelo RDF-Schema.

Miler e Koivunen (2002, p. 27-33) consideram que os seguintes princípios fundamentam a *web semântica*:

1) tudo pode ser representado por meio de URI's – é possível referenciar pessoas, lugares e coisas do mundo físico na *web* semântica, com a utilização de diversos identificadores unívocos;

2) recursos e links podem ter tipos – a *web* atual consiste de links e recursos destinados à interpretação humana e, ainda, não é amplamente difundido o uso de meta dados para explicitar seu significado, bem como, suas relações com outros documentos, de modo que máquinas possam realizar inferências sobre o seu significado; tais relações podem ser, por exemplo: *depende de, é uma versão de, possui assunto, é autor de*, entre outras;

3) informação parcial é tolerada – a *web* atual sacrifica a integridade dos links, em função de sua escalabilidade; de tal modo, páginas podem estabelecer ligações com outras sem a necessidade de que haja uma reciprocidade nessa ligação; de maneira similar, na *web* semântica, pode-se declarar qualquer coisa sobre qualquer coisa, sem que haja necessidade da criação de ligações entre os recursos declarados;

4) a verdade absoluta não é necessária – nem toda informação na *web* atual é verdadeira, e a *web* semântica não prevê mudanças nessa orientação.

5) à evolução é suportada – a *web* semântica utiliza convenções descritivas que podem expandir-se como a compreensão humana se expande; tais convenções permitem a combinação do trabalho independente e compartilhado entre diversas comunidades, mesmo quando estas utilizam vocabulários diferentes.

6) o projeto é minimalista – um dos objetivos para garantir o desenvolvimento da *web* semântica é padronizar apenas o que for necessário, favorecendo a implementação de aplicações flexíveis.

2.4.2 Ontologias

Desde o início dos anos 90, as ontologias tornaram-se um tópico de pesquisa popular nas áreas de inteligência artificial, engenharia do conhecimento, processamento da linguagem natural e representação do conhecimento. A razão de tal popularidade é acreditar-se que, por seu intermédio, um entendimento comum e compartilhado sobre um determinado domínio de conhecimento poderá ser comunicado tanto entre pessoas como entre computadores.

Originário da Filosofia, o termo Ontologia, em sentido estrito, pode ser definido como estudo do ser e de suas propriedades fundamentais. Etimologicamente o termo provém do

grego: *ontos* (ser, ente) e *logos* (saber, doutrina). Filosoficamente, há dois aspectos pelos quais se pode conceber ontologia:

a) aspecto existencial – a ontologia é um saber sobre aquilo que é fundamental ou irreduzível, comum a todos os entes singulares;

b) aspecto essencial – busca determinar as leis, estruturas ou causas do ser em si.

Historicamente, a ontologia como disciplina foi cultivada desde o século ~~VIII~~ e, graças ao filósofo alemão Christian Woldd, o termo ganhou projeção. A partir do século XIX, três correntes tornaram-se importantes:

a) para Husserl – a ontologia é a ciência das essências e pode ser de dois tipos: formal – que é o fundamento de todas as ciências e interessa-se pelas essências e suas relações com outras essências; ou material – conjunto de ontologias setoriais que se preocupa com fatos;

b) para Hartmann – a ontologia presta-se ao reconhecimento daquilo que é metafisicamente insolúvel; essa corrente examinava os diversos momentos do ser, relacionando-os à existência ou à essência; à realidade ou à idealidade; à possibilidade, realidade, causalidade ou necessidade;

c) para Heidegger – a ontologia é uma metafísica da existência; é baseada na fenomenologia.

O dicionário Dictionary.com⁵¹ fornece as seguintes definições para o termo ontologia: a) uma área da ciência da metafísica que investiga e explicita a natureza, as propriedades essenciais e as relações entre os seres, como tais, ou os princípios e causas do ser; b) uma especificação explícita e formal de como representar objetos, conceitos e outras entidades existentes em uma área de interesse, bem como as relações entre tais objetos, conceitos e entidades; c) no campo da Ciência da Informação, a estruturação hierárquica do conhecimento sobre coisas, pela sua subcategorização, de acordo com suas qualidades essenciais (ou pelo menos relevantes ou cognitivas). Como informação complementar, o Dictionary.com anota que, para os sistemas de Inteligência Artificial, o que “existe” é algo que pode ser representado. Assim, quando o conhecimento de dado domínio é representado em uma linguagem declarativa, o conjunto de objetos que está sendo representado é denominado universo do discurso. Pode-se,

⁵¹ Dictionary.com é um serviço de busca produzido pela empresa Lexico Publishing Group, e baseia-se na reunião de definições advindas de diversos outros dicionários disponíveis na *web*. O serviço pode ser localizado em [http://www.dictionary.com] e inclui as obras: *The American Heritage Dictionary of the English Language*, da editora Houghton Mifflin Company; *Webster's Revised Unabridged Dictionary*, da editora MICRA; *WordNet*, editado pela Princeton University; *The Free On-line Dictionary of Computing*, de Denis Howe; entre outras.

portanto, descrever uma ontologia como um programa que define um conjunto de termos que podem ser representados. As definições associam os nomes das entidades em dado universo do discurso – por exemplo, classes, relações, funções ou outros objetos – com textos legíveis por humanos, que descrevem o que tais nomes significam, e com axiomas formais, que restringem a interpretação e o uso desses termos (Dictionary.com, 2004).

Para Hyvönen (2002, p. 112), ontologia é um campo da metafísica que especifica as categorias fundamentais da existência, isto é, classifica os conceitos e examina as distinções que sustentam cada fenômeno no mundo. A ontologia vista na acepção filosófica remonta aos trabalhos de Aristóteles – entre 384 e 322 a.C – nos quais o pensador referia-se à natureza do ser, investigando sua essência última e à razão de ser, por trás da natureza última como o ser é percebido. Não obstante a antecipação dos princípios, por Aristóteles, o termo “ontologia” foi utilizado pela primeira vez em 1613, por Rudolf Göckel, em sua obra *Lexicon Philosophicum* e, aparentemente de modo independente por Jacob Lorhard, com sua obra *Theatrum Philosophicum* (GUARINO; WELTY, 2001). Hyvönen (2002, p. 114) afirma que, na perspectiva lingüística, *ontologia* é vista como um método destinado a aprimorar a compreensão da linguagem pelos sistemas de processamento da linguagem natural. Dessa maneira, as teorias ontológicas tentam representar o conhecimento humano de maneira estruturada.

Do ponto de vista da Ciência da Computação, há diversas definições para as ontologias. Gómez-Pérez registra as visões de diversos autores para o termo. A tabela 23 apresenta algumas dessas visões:

Tabela 23: Definições de ontologia

Autor	Definição
NECHES (1991)	Uma ontologia define os termos básicos e as relações compreendendo o vocabulário de uma área de tópico, bem como as regras para a combinação de termos e as relações para definir as extensões do vocabulário; esta definição fornece também as linhas gerais para a construção de uma ontologia: identificar os termos básicos e as relações entre eles; identificar as regras para combiná-los; fornecer definições para tais termos e relações.
GRUBER (1993)	uma ontologia é uma especificação explícita de uma conceituação. Nesta definição: conceituação refere-se a um modelo abstrato de algum fenômeno, sendo identificados os conceitos relevantes desse fenômeno; explícita significa que o tipo de conceitos utilizados e as restrições a esse uso são explicitamente definidos; formal refere-se ao fato de que a ontologia deve ser legível por máquina; compartilhada reflete a noção de que uma ontologia captura um conhecimento consensual, isto é, não privativo de um indivíduo, mas aceito por um grupo.
BORST (1997)	Ontologias são definidas como uma especificação formal de uma conceituação compartilhada.

Autor	Definição
SWARTOUT (1997)	uma ontologia é um conjunto hierarquicamente estruturado de termos para descrever um domínio que pode ser usado como um esqueleto fundamental para uma base de conhecimentos
BERNARAS (1996)	Uma ontologia fornece significado para a descrever explicitamente uma conceituação atrás de um conhecimento representado em uma base de conhecimento

Fonte: Gómez-Pérez (1999)

De acordo com Noy e McGuinness (2004), uma ontologia é uma descrição formal e explícita de conceitos – classes – em um domínio do discurso, das propriedades de cada conceito, descrevendo diversas características e atributos de tal conceito – *slots* (também chamados de papéis ou propriedades), bem como das restrições nesses *slots* – também chamadas de restrições de papéis).

Uma ontologia, juntamente com um conjunto de instâncias individuais de classes, constitui uma base do conhecimento. Em relação à estrutura de uma ontologia, as autoras argumentam: a) não há uma forma correta de se modelar um domínio do conhecimento: há sempre diversas alternativas viáveis e a melhor solução sempre dependerá das características da aplicação ou serviço que se pretende desenvolver; b) o desenvolvimento de uma ontologia é necessariamente um processo iterativo; c) os conceitos em uma ontologia devem estar intimamente ligados a objetos físicos ou lógicos, e a suas inter-relações em um dado domínio do conhecimento. Tais conceitos provavelmente serão mais bem expressos por nomes ou verbos. Em resumo, uma ontologia refere-se à descrição exata de coisas e seus relacionamentos. Para a *web*, uma ontologia refere-se à descrição exata de uma informação disponível na rede e à descrição de seus relacionamentos com outras informações (W3Schools, 2004).

Do ponto de vista da representação do conhecimento, uma ontologia (não) deve ser concebida apenas como um vocabulário informal, ou mesmo como uma linguagem de termos estruturados – como um tesouro, por exemplo –, mas requer uma possibilidade de interpretação algorítmica dos seus significados e, por conseguinte, uma representação em uma linguagem formal, cujo processamento dos significados pode ser realizado por máquinas. Dito de outro modo: uma ontologia requer a explicitação lógico-formal de significados e palavras⁵², que devem ser expressos por meio de construtos matemáticos. A depender do domínio do conhecimento, as ontologias podem variar em termos estruturais, mas alguns construtos utilizados são independentes de domínio. Entre eles, Hyvonen (2002, p. 122), cita:

⁵² Em Terminologia, significados = conceitos; palavras = termos

a) classes – os conceitos genéricos de um vocabulário são definidos como classes. Por exemplo, a classe Tigre representa a categoria genérica de espécies de tigres;

b) relação de superclasse – as classes são organizadas em hierarquias conceituais. Exemplificando, a classe Tigre é uma subclasse da classe dos carnívoros;

c) propriedades de classe – classes podem ter propriedades, freqüentemente denominadas *slots* (pistas). As propriedades de uma subclasse podem ser herdadas de sua superclasse, o que leva a uma economia em termos de representação e que torna possível a realização de inferências simples. Exemplificando: uma vez que carnívoros comem carne e possuem dentes afiados, então os tigres também, visto que são carnívoros;

d) características de propriedade (facetas) – propriedades de classes podem, em si mesmas, possuir facetas que as restringem. Por exemplo: as facetas tipo de valor e cardinalidade podem ser associadas à propriedade *pais*, para denotar que é um tipo da classe Pessoa e que o número de valores deve ser de pelo menos 2;

e) indivíduos – um objeto individual sobre o qual trata uma ontologia é denominado de instância ou objeto. Cada objeto é uma instância de uma ou mais classes. Por exemplo: *Garfield* pode ser uma instância da classe *Gato* e também da classe *GatodeCartoon*. A relação de pertinência é freqüentemente denotada pela expressão “é um”;

f) axiomas e restrições – uma ontologia baseada em lógica formal pode ter axiomas e restrições adicionais ou regras de inferência. Axiomas e regras podem ser escritos por meio de termos, funções, predicados, operadores, quantificadores, entre outros.

Em suma, esclarece o autor: uma teoria ontológica define conceitos em um domínio do conhecimento, por meio de um vocabulário; uma aplicação faz uso dos objetos, por meio de axiomas e de regras lógicas; tais regras dizem como utilizar os conceitos referenciados, com vistas à solução de problemas em particular.

Segundo Gómez-Pérez (2000), ontologias fornecem um vocabulário comum para uma área e definem – com diferentes níveis de formalismo – o significado desses termos e as relações entre eles. Os conhecimentos em uma ontologia são formalizados com a utilização de cinco tipos de componentes: classes, relações, funções, axiomas e instâncias.

Classes (conceitos) – são utilizados em sentido amplo e podem ser abstratos ou concretos; elementares ou compostos; reais ou fictícios; uma classe pode ser um objeto ou uma tarefa, uma função, uma ação, uma estratégia, um processo de raciocínio;

Relações – representam um tipo de interação entre classes de um domínio; são formalmente definidas como qualquer subconjunto de um produto de n conjuntos;

Funções – são casos especiais de relações em que o n -ésimo elemento da relação é único para os $n-1$ elementos precedentes;

Axiomas – são usados para modelar sentenças que sempre são verdadeiras;

Instâncias – são usadas para representar elementos.

Em relação à elaboração de ontologias, a autora relaciona alguns princípios básicos que devem ser seguidos: a) clareza e objetividade – uma ontologia deve fornecer o significado dos termos definidos, fornecendo definições objetivas e também documentação em linguagem natural; b) completeza – uma definição expressa por condições necessárias e suficientes é preferível a uma definição parcial; c) coerência – para permitir inferências que sejam consistentes com as definições; d) maximização da extensibilidade monotônica – novos termos gerais ou especializados devem ser incluídos na ontologia de modo que não seja necessário realizar a revisão das definições já existentes; e) comprometimento ontológico mínimo – fazer a menor quantidade possível de declarações sobre o mundo que está sendo modelado, o que significa que a ontologia deve especificar, em um contexto tão pequeno quanto possível, os significados dos seus termos, dando liberdade à ontologia para especializar-se e instanciá-los; f) princípio da distinção ontológica – as classes em uma ontologia devem ser desmembradas: o critério utilizado para isolar o núcleo de propriedades considerado como invariante para uma instância de uma classe é chamado de critério de identidade; g) diversificação de hierarquias (mecanismos de herança múltipla) – se conhecimento suficiente é representado na ontologia, e tantos critérios de classificação quanto possível são utilizados, é fácil entrar com novos conceitos (desde que eles possam ser facilmente especificados a partir de conceitos e de critérios de classificação pré-existent) e também herdar propriedades a partir de diferentes pontos de vista; h) modularidade – para minimizar o acoplamento entre módulos; i) minimizar a distância semântica entre conceitos similares – conceitos similares são agrupados e representados como subclasses de uma classe e devem ser definidos utilizando-se as mesmas primitivas, enquanto conceitos menos similares são representados mais apartados na hierarquia; j) padronização – os nomes devem ser padronizados, quando possível;

Do ponto de vista do formalismo, Uschold & Gruninger (1996) distinguem quatro tipos de ontologias, de acordo com o modo como são construídas:

a) Altamente Informais – escritas em linguagem natural.;

b) Semi-informais – expressas em linguagem natural, mas de um modo restrito e estruturado, isto é, utilizando padrões;

c) Semi-formais – são definidas artificial e formalmente em uma linguagem definida;

d) Rigorosamente formais – são definidas em uma linguagem com semântica formal;

Podem ainda ser identificados os seguintes tipos de ontologias, quanto aos seus objetivos:

a) De representação do conhecimento – compreendem a representação de primitivas utilizadas para conhecimento formalizado em paradigmas de representação de conhecimento;

b) Gerais ou comuns – incluem vocabulário relativo a coisas, eventos, tempo, espaço, causalidade, comportamento, funções, etc.

c) Meta-ontologias – também chamadas de ontologias genéricas ou ontologias núcleo – são reutilizáveis em diversos domínios;

d) ontologias de domínio – são reutilizáveis em um dado domínio. Fornecem vocabulários sobre conceitos com um domínio e suas relações com as atividades que têm lugar nesse domínio e sobre teorias e princípios elementares que governam esse domínio.

O desenvolvimento de uma ontologia é diferente do planejamento de classes e relações na programação orientada a objetos. A programação orientada a objetos centra-se principalmente em torno de métodos em classes – um programador projeta decisões baseadas nas propriedades operacionais de uma classe. Em uma ontologia, o projetista toma tais decisões com base nas propriedades estruturais de uma classe. Como resultado, uma estrutura de classes e as relações entre classes em uma ontologia são diferentes da estrutura para um domínio similar em um programa orientado a objetos (NOY; MCGUINESS, 2004). Daconta (2003) partilha de tal orientação, argumentando que uma representação de classes e de relacionamentos entre classes para possibilitar inferências, requer formalismos rigorosos que estão muito além das convenções utilizadas nas atuais linguagens de programação orientadas a objetos, como Java, C e C#, por exemplo. No entanto, tais formalismos de hierarquias de classes, restrições de propriedades e relações entre classes podem facilmente ser representados por meio de ontologias.

Noy e McGuiness (2004) destacam algumas motivações para o desenvolvimento de ontologias: a) compartilhamento de um entendimento comum da estrutura da informação

entre humanos e agentes inteligentes; b) permitir o reuso do domínio do conhecimento; c) tornar explícitos os pressupostos de um domínio; d) separar o conhecimento sobre o domínio do conhecimento operacional; e) analisar o conhecimento sobre o domínio.

A análise de algumas afirmações de autores que tratam dos princípios teóricos da *web* semântica permite avaliar que há uma intenção de se estabelecer um certo nível de controle terminológico, em busca da interoperabilidade. Segundo Noy e McGuinness (2004), o compartilhamento de um entendimento comum acerca da estrutura da informação realizado entre humanos e computadores é um dos mais comuns objetivos do desenvolvimento de ontologias. Para ilustrar tal afirmação, as autoras argumentam que numa aplicação de comércio eletrônico, por exemplo, se os *sites* envolvidos em um domínio do conhecimento compartilharem e publicarem os mesmos termos contidos em uma ontologia que todos utilizem, então os agentes inteligentes serão capazes de extrair e agregar informações desses diferentes *sites* e, assim, utilizar tal informação para responder a questões de pesquisa de usuários, ou até mesmo como dados de entrada em outras aplicações.

A busca pela possibilidade de se reutilizar o conhecimento em dado domínio do conhecimento é uma das forças motrizes das pesquisas na área de ontologias. Se um grupo de pesquisadores desenvolve uma ontologia relativa a determinado domínio do conhecimento, outras pessoas poderão reutilizá-la mais tarde ou até mesmo integrá-la a outras ontologias para formar uma maior, que descreva mais profundamente certo domínio do conhecimento. Tornar explícitos os pressupostos de um domínio torna, também, possível modificar facilmente tais pressupostos, se o conhecimento sobre o domínio muda. Além disso, tal explicitação é muito útil para novos usuários, que devem compreender o que os termos de um domínio significam (Daconta, 2003).

2.4.3 Infra-estrutura da Web Semântica

Para Tim Berners-Lee (1998), os princípios da universalidade do acesso sem restrição de plataformas de *hardware* ou de *software*, infra-estrutura de rede, idioma, cultura, localização geográfica ou física são os valores núcleo do projeto da *web*. Além desses princípios, o autor destaca como “matéria-prima” da *web* a simplicidade e a modularidade; como “respiração da *web*”, a descentralização e a tolerância.

Essa arquitetura multiplataforma somente é possível na *web* devido ao uso do identificador universal de recursos – *Universal Resource Identifier*. URI⁵³ – como unidade fundamental de identificação de objetos na *web*, esse conceito é similar ao que todas as linguagens de programação ou de marcação utilizam e denominam como atributo identificador de um objeto. Dessa maneira, o uso de URIs permite que uma linguagem faça referência a objetos definidos em outra linguagem.

A expressão URI pode ser definida de acordo com os seus constituintes núcleos, da seguinte maneira (Berners-Lee, 1998a):

a) Uniforme – a noção de uniformidade em um URI associa-se a três princípios básicos: em primeiro lugar, deve ser possível a utilização de diferentes tipos de identificadores de recursos no mesmo contexto, ainda que os mecanismos usados para o acesso a tais recursos sejam diferentes; em segundo lugar, independentemente da forma como os identificadores são usados, deve ser possível a interpretação semântica de convenções sintáticas, mesmo que sejam usados diferentes tipos de identificadores de recursos; e em terceiro lugar, deve ser possível a reutilização dos identificadores em diferentes contextos o que, por conseguinte, permitirá a criação de novas aplicações ou protocolos.

b) Recurso – um recurso pode ser qualquer coisa que tenha identidade, como um arquivo eletrônico, uma imagem, um som, entre outras entidades. Nesse contexto, mesmo entidades que não sejam “recuperáveis” pela *web*, como por exemplo, seres humanos, empresas ou o acervo de uma biblioteca, são definidos como recursos. Um recurso, portanto, é o mapeamento conceitual para uma entidade ou conjunto de entidades, não necessariamente entidades que correspondem àquele mapeamento em alguma instância particular no tempo. Desta maneira, um recurso pode permanecer constante, mesmo quando seu conteúdo – a entidade à qual ele corresponde – muda com o tempo.

c) Identificador – um identificador é um objeto que pode atuar como uma referência a algo que possua identidade. No caso de um URI, o objeto é uma seqüência de caracteres com uma sintaxe restrita. Portanto, tendo identificado um recurso, um sistema

⁵³ Um *Uniform Resource Identifier* (URI) é um conjunto compacto de caracteres utilizado para a identificação de um recurso abstrato ou físico: um URI fornece ao recurso um significado simples e extensível. (BERNERS-LEE, 1998)

pode utilizá-lo para realizar operações tais como: acesso, atualização, substituição ou localização de atributos.

De acordo com Swartz (2002), um importante ponto a ser destacado é que um URI não é um conjunto de endereços dizendo ao computador como obter um arquivo específico na *web* (embora ela também faça isso). O autor lembra que um recurso é um nome para uma “coisa” que pode ou não ser acessível por meio da *Internet*. Portanto, é possível atribuir-se um URI para qualquer coisa e qualquer coisa que possua um URI pode ser considerada “na *web*”: um livro, um voo de avião, e qualquer outra coisa em que se possa pensar. Segundo Swartz o URI é o fundamento da *web*: enquanto qualquer outra coisa na *web* pode ser substituída, o URI não, pois é ele que mantém o resto da *web* junto. Um URI pode ou não fornecer um caminho para o computador obter mais informações acerca de um recurso; isto é, embora a possibilidade de um URI fornecer informações sobre outro URI seja importante para a *web* semântica. Swartz (2002) argumenta que um URI nada faz além de fornecer um identificador para um recurso.

Genericamente, um URI pode ser classificado como um localizador, um nome ou ambos. O termo *Uniform Resource Locator* – URL refere-se a um subconjunto do URI, que identifica recursos por meio da representação do seu mecanismo de acesso principal: o seu endereço, ao invés de identificá-lo pelo nome ou por outro atributo. Lembra Swartz (2002) que um URL é um tipo de URI que fornece um caminho para se obter informação sobre um recurso, ou talvez para se recuperar o próprio recurso. O termo *Uniform Resource Name* – URN refere-se ao subconjunto do URI que é requerido para permanecer globalmente único e persistente, mesmo quando o recurso deixar de existir ou estiver indisponível (Berners-Lee, 1998a).

Em sua idéia original sobre o que virá a ser a *web*, Berners-Lee (1998a) apresenta conceitos de *nós* e *links* que, no seu ponto de vista, permitem a melhor descrição de sistemas complexos. O autor compara esses conceitos aos conceitos de *círculos* e *setas*, que deixam qualquer pessoa livre para descrever as inter-relações dos objetos de uma maneira bastante flexível, impossível a outros sistemas como tabelas e diagramas hierárquicos. Os *círculos* seriam então como os *nós* e as *setas* como os *links*. Cada nó representa, portanto entidades como uma pequena nota, um sumário de um artigo; pode ser armazenado em forma gráfica, textual, ou em ambas; representa ou descreve uma pessoa ou objeto em particular. O autor sugere, desta maneira, exemplos de entidades representadas por nós: pessoas, módulos de *software*, grupos de pessoas, projetos, conceitos, documentos, tipos de *hardware*, objetos gráficos específicos, entre outras. Assim, dadas duas entidades A e B, os *links* (*setas*) entre uma e outra podem representar, por

exemplo que A: depende de B ; é parte de B; é feito de B; refere-se a B; usa B; é um exemplo de B.

Com essas concepções, ou autor desenvolveu um sistema já conhecido por todos, no qual qualquer tipo de informação pode ser armazenada e localizada por qualquer pessoa, mesmo que esta não saiba exatamente o que tenciona procurar. O sistema baseia-se, portanto, na utilização de tipos genéricos de links entre os itens (dependência, por exemplo), e tipos de nós (entidades: pessoas, coisas, documentos), sem a imposição de limites. O objetivo é disponibilizar um lugar para cada informação que alguém quiser tornar pública, de modo que tal informação seja encontrada posteriormente. Associadas a essas características, o autor anteviu também que o sistema seria suportado por diversas plataformas, além de ser extensível a novos formatos de dados.

A ilustração 44 apresenta o modelo proposto por Berners-Lee em 1989 para representar uma “web de objetos”, incluindo pessoas, organizações, tecnologias, documentos, tópicos, entre outras entidades. Assim, os links tipificados, como por exemplo ‘*wrote*’, ‘*unifies*’, ‘*includes*’ – escreveu, unifica, inclui – são utilizados para a representação do conhecimento sobre seus relacionamentos.

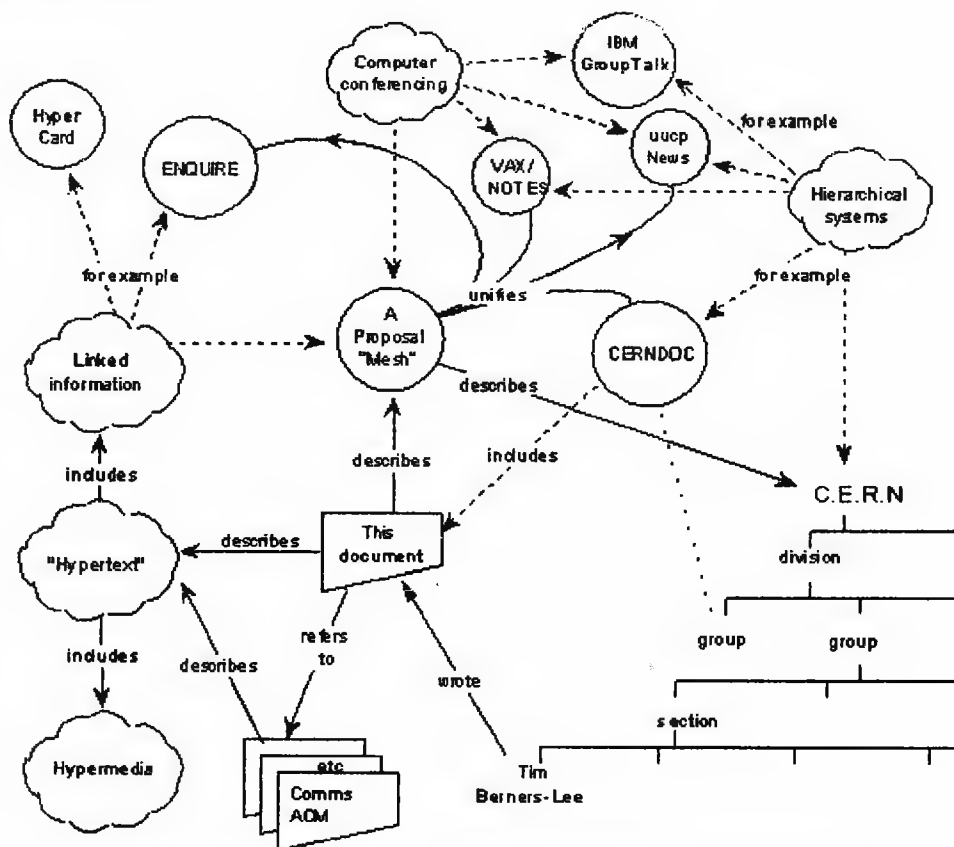


Ilustração 44 - Modelo proposto para a *World Wide web*, segundo Berners-Lee
Fonte: (Brickley, 2001)

A leitura do diagrama apresentado permite concluir que a idéia da *web* é bem mais ampla do que apenas fornecer meios para a localização de páginas, mas que, desde a sua concepção, o autor pensava em fornecer informações semânticas sobre os objetos. Por exemplo, é possível observar-se que: Tim Berners-Lee escreveu um “documento” → que descreve uma “proposta” → que “unifica procedimentos”; mas que também descreve “hipertexto” → que inclui “hipermídia”; mas que também inclui “informação ligada”, e assim por diante.

2.4.4 As gerações da web

Desde a sua concepção, a *web* passou por estágios evolutivos que podem ser classificados em três gerações. O esquema dado na ilustração 45 mostra essa evolução:

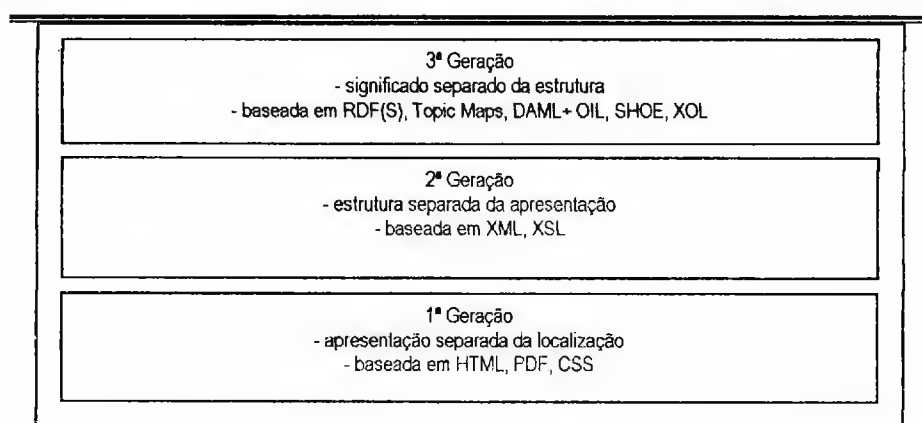


Ilustração 45: Gerações tecnológicas da *World Wide web* – adaptado de Hyvönen (2002).

A ilustração mostra as tecnologias utilizadas na *web* em cada geração. A primeira geração, baseada na linguagem HTML, tornou possível a exibição dos documentos independentemente de sua localização física; a segunda geração tornou possível o uso de diferentes formas de apresentação para a mesma estrutura de um documento, com base na linguagem XML; a terceira geração, a da *web* semântica, consiste na separação do significado de sua estrutura, isto é, estende a noção de “documento” com a possibilidade de “semântica legível por máquina”. Se a codificação semântica é muito mais centrada no autor humano, o computador pode auto ajudar-se, ajudando o homem a criar tais descrições. Nesta visão, a representação explícita do conhecimento formará a base para o desenvolvimento de serviços e aplicações de *web* mais inteligentes (HYVÖNEN, 2002).

As informações existentes na *web* atualmente são apresentadas (usando tecnologias como HTML, PDF e XML, por exemplo) de tal maneira que possam ser lidas e compreendidas

por pessoas. Entretanto, sem a interpretação de um leitor humano, essas tecnologias nada dizem a respeito do significado ou do conteúdo de um documento armazenado na *web*. O código, apresentado na ilustração 46, por exemplo, nada informa ao computador sobre o conteúdo do parágrafo em questão:

```
<p><b><font face="Arial">
LEI Nº 3211, DE 23 DE OUTUBRO DE 2003
</b></font></p>
```

Ilustração 46: o problema do significado com o uso da codificação HTML

A codificação apresentada na figura anterior apenas informa ao computador o que fazer em relação à forma do parágrafo, isto é: o tipo de letra que deve ser utilizado (*arial*); e o tipo de efeito aplicado a essa letra: **negrito**. Quanto ao conteúdo do parágrafo, é necessária a interpretação humana para identificar que o texto refere-se a uma lei que foi publicada em uma certa data. Essa codificação, portanto, não permite ao computador realizar inferências acerca do significado do texto em questão. De modo análogo, com a utilização da tecnologia XML, é possível ao computador reconhecer a estrutura de um certo registro em um documento, mas não dizer algo a respeito do significado dessa estrutura (BERNERS-LEE, 2001). Assim, a compreensão do seu significado é possível apenas para o leitor humano. A Ilustração 47 mostra essa situação:

```
<xml>
  <legislacao>
    <lei>
      <numero>3211</numero>
      <data>23/10/2003</data>
      <ementa>Dispõe sobre a estrutura orgânica da
        Secretaria....</ementa>
      .....
    </lei>
  </legislacao>
</xml>
```

Ilustração 47: O problema do significado com o uso da codificação XML

O computador é capaz de compreender, por exemplo, que o valor *3211* refere-se a um campo de dados, denominado *numero*, e que o valor *23/10/2003* refere-se a outro campo de dados, denominado *data*. Com a tecnologia XML, a realização de inferências acerca desses campos é possível apenas para humanos. Com as informações fornecidas no código da ilustração 47, o computador não é capaz de saber, por exemplo, que a data refere-se ao dia em que o documento foi aprovado em uma reunião de deputados; que a ementa corresponde a um resumo

do texto integral da lei; que o campo lei relaciona-se de algum modo com os termos *norma, código, legislação, direitos*.

Na prática, a maioria dos *sites* disponíveis na *Internet* utiliza, ainda, apenas as tecnologias da primeira geração. Do ponto de vista gerencial, as implicações desse uso são significativas, uma vez que tornam menos efetivas as operações de recuperação. Se, por um lado, a publicação se torna mais rápida apenas com o uso de textos integrais, por outro, a inexistência de metadados dificulta a representação e a organização da informação, o que torna mais baixa a precisão, no momento da recuperação. De tal modo a interpretação do significado e da relevância das informações fica a cargo dos humanos.

No entanto, a *Internet* não é utilizada apenas por pessoas. É cada vez maior o uso de robôs de busca, agentes de comércio eletrônico (*shopbots*), rastreadores (*web crawlers*) e outros artefatos, para os quais não é fácil interpretar ou “entender” a informação não estruturada disponível na *web* (HYVÖNEN, 2002). É com o pensamento nessas tecnologias que os pesquisadores estão desenvolvendo os fundamentos da *web* semântica.

2.4.4.1 A 1ª geração: HTML E XHTML

A *Hypertext Markup Language* surgiu como uma derivação da linguagem *Standard Generalized Markup Language* – SGML, linguagem originária dos anos 70 e que sempre foi conhecida no meio informático pela sua potencialidade como linguagem de marcação de textos. A primeira versão da linguagem HTML foi elaborada por Tim Berners-Lee, com o objetivo de suprir a *web*, uma vez que o seu projeto necessitava, para publicar informação para a distribuição global, de uma linguagem universalmente compreensível, como um tipo de esperanto⁵⁴, que todos os computadores pudessem potencialmente entender.

XHTML é um acrônimo para *eXtensible Hypertext Markup Language*. Essa linguagem é uma evolução da linguagem HTML e consiste em uma iniciativa para a imposição de regras semânticas às páginas escritas em HTML, com a utilização do padrão de descrição da linguagem XML (VALENTINE e MINNICK, 2001). Alguns dos benefícios esperados pela utilização dessa linguagem incluem: redução dos custos de autoria; melhor casamento entre aplicações de bancos de dados e *workflow*; solução modular para o problema de os navegadores possuírem cada vez maior número funcionalidades diferentes; solução para integrar, de modo transparente, documentos HTML com outras aplicações XML (W3C, 2003).

⁵⁴ Esperanto: Língua auxiliar de comunicação internacional, elaborada pelo médico e poliglota judeu-polonês Ludwig Zamenhof (1859-1917) e por ele divulgada em 1887, com o objetivo de servir como língua universal (FERREIRA, 1986).

2.4.4.2 A 2ª geração: XML, DTD e XML Schema

A linguagem eXtensible Markup Language – (XML) é um padrão de representação que permite a introdução de uma significativa quantidade de metadados para descrever, inclusive com grande nível de detalhamento, o conteúdo temático de um documento. É, portanto, um mecanismo importantíssimo para os sistemas de recuperação de informações, que permite, inclusive, a escolha mais diversificada quanto à forma (fichas, formulários contínuos, tabelas, entre outros).

De acordo com Valentine e Minnick (2001), a verdadeira finalidade de uma linguagem de marcação é descrever as partes de um documento sem se preocupar com a apresentação ou com a exibição do documento. As autoras argumentam que a marcação deve fornecer estrutura aos dados, permitindo que esses dados sejam usados de diversas maneiras, desde a exibição em um telefone celular até o armazenamento e recuperação em um banco de dados.

As autoras definem esses conjuntos de elementos XML, projetados para descreverem um determinado tipo de informação, como vocabulários XML; assim, uma grande quantidade de vocabulários XML já foi desenvolvida, como por exemplo: CML (*Chemical Markup Language*), para descrever informações relativas à química; OFX (*Open Financial Exchange*), para descrever informações financeiras.

XML é um conjunto de regras e convenções de sintaxe que podem ser utilizadas para a criação de conjuntos personalizados de elementos de marcação, com vistas à descrição do conteúdo de um documento eletrônico (VALENTINE e MINNICK, 2001). Essa linguagem foi desenvolvida para prover a necessidade de uma metodologia para a descrição da informação publicada na *Internet*.

A idéia de linguagens padrão de marcação não é muito recente. Iniciou-se por volta da década de 80 quando, depois de vários anos de trabalho a Organização Internacional para Padronização – (ISO), aprovou a *Standard Generalized Markup Language* (SGML) como um sistema – uma metalinguagem – para a criação de novas linguagens de marcação. O problema da SGML era sua complexidade para ser entendida pela grande maioria dos navegadores. Dessa maneira, um grupo de trabalho instituído pelo W3C⁵⁵ criou uma linguagem mais “enxuta”, sem todas as “firulas” da SGML, de modo que qualquer pessoa pudesse, a partir da XML, criar sua própria linguagem de marcação. Do mesmo modo que a HTML criou uma forma para cada

⁵⁵ Informações sobre os grupos de trabalho envolvidos com nas especificações da linguagem XML podem ser encontradas em: <http://www.w3.org/XML/>.

computador ler documentos na *Internet*, a XML torna possível, independentemente da torre de babel de sistemas de computadores incompatíveis entre si, criar uma espécie de esperanto, que todos podem ler e escrever (Bosak, 1999).

O autor argumenta, ainda, que diferentemente da maioria dos formatos de dados para computadores, a XML também pode fazer sentido para leitores humanos, uma vez que consiste em nada mais do que texto comum. Uma das forças da unificação da XML é a confiança no padrão UNICODE de codificação de textos, que suporta a utilização de caracteres de quase todos os principais idiomas do mundo. Dessa maneira, a XML possibilita o intercâmbio de informações não apenas entre diferentes sistemas computacionais, mas sobretudo, além das fronteiras dos países e de suas culturas. É natural, portanto, que desde a finalização de sua especificação pelo W3C, o padrão tenha se espalhado “como fogo selvagem” pela ciência e pela indústria, estendendo-se desde a medicina até o chão das fábricas.

Para que um documento XML possa interoperar na *web*, mantendo a característica da universalidade de plataformas e sistemas, duas condições devem ser atendidas: ele deve ser “válido” e “bem-formatado”. Diz-se que um documento XML é válido, se ele obedece as restrições sintáticas especificadas na estrutura da linguagem XML, conforme as especificações do W3C. Diz-se que um documento XML é bem-formatado, se a ele estão associadas regras de validação e restrições, de acordo com o objeto que se pretende representar. Conjuntos de regras de validação foram recomendados inicialmente pelo W3C como *Document Type Definitions* (DTD's) e essas DTD's proviam um conjunto de regras sintáticas às quais o documento deveria seguir para que fosse considerado válido. A desvantagem das DTD's é que permitem apenas a verificação sintática dos documentos, não sendo dotadas de nenhum tipo de controle semântico. Outro problema das DTD's é que a sua definição é feita em XML, mas numa notação própria (LIMA, 2003).

A autora argumenta que, como substituição às DTD's, foi introduzido o conceito de XML Schema, cuja principal contribuição foi a separação entre a descrição da estrutura do documento e a descrição dos mais de vinte tipos de dados que podem ser utilizados para a sua marcação. Essa linguagem é suficiente para a troca de dados entre colaboradores que utilizam o mesmo vocabulário, mas sua semântica é limitada no que se refere ao processamento automatizado de novos vocabulários.

2.4.4.2.1 XML

Diversos autores argumentam que a linguagem XML é a pedra fundamental da *web* semântica e tem sido amplamente aconselhada a sua adoção como instrumento para o

desenvolvimento de novas áreas de processamento de informação. Tal recomendação deve-se à flexibilidade da linguagem, por um lado, e às possibilidades de implementação de maior rigor sintático, por outro, dados os mecanismos e padrões fornecidos pelo XML-Schema (Daconta, 2003; Berners-Lee 2001, W3Schools, 2004, entre outros).

Daconta (2003), por exemplo, discute que a XML é uma plataforma importante no desenvolvimento da *web* semântica, entre outros motivos, porque: a) cria documentos e dados independentes de aplicativos; b) possui uma sintaxe padrão para meta dados; c) possui uma estrutura padrão, tanto para documentos quanto para dados; d) não é uma tecnologia nova⁵⁶; e) é independente de aplicativos, pois seus arquivos são criados em formato puramente textual, legível por humanos; f) fornece uma sintaxe padrão simplificada, de modo a permitir a representação de dados ou meta dados; g) fornece princípios fundamentais que podem ser utilizados para a codificação de informação semântica – desses princípios utilizam-se a RDF, o RDF-Schema entre outras tecnologias da *web* semântica. O autor define XML não como uma linguagem e sim, mais propriamente, “como um conjunto de regras sintáticas para a criação de linguagens de marcação semanticamente mais ricas, em um domínio particular”. Em outras palavras, isso quer dizer que XML é um padrão que pode ser aplicado para a criação de outras linguagens, às quais o autor denomina de “aplicações de XML” – são exemplos: MathML, InkML, CML, VoxML, MusicXML⁵⁷, entre outras (HAROLD, 1999).

A preocupação básica de uma linguagem de marcação é desenvolver métodos para a adição de informação semântica sobre o conteúdo de um documento, de modo a tornar tal conteúdo mais explícito. Esse procedimento tem sido efetuado manualmente, ao longo dos anos

⁵⁶ Como já discutido neste trabalho, XML é um subconjunto da linguagem *Standardized Generalized Markup Language* (SGML), que foi desenvolvida em 1969, por Charles Goldfarb, Ed Mosher e Ray Lorie. Assim, os conceitos e os princípios fundamentais para a XML foram planejados por mais de trinta anos, sendo continuamente aperfeiçoados, testados e amplamente implementados (Daconta, 2003). O advento da *web*, e dos seus problemas na representação da informação, naturalmente impulsionaram o surgimento da XML.

⁵⁷ InkML é um formato de dados baseado em XML, destinado à representação digital de marcações em que o dispositivo de entrada é uma caneta eletrônica. Informações mais detalhadas estão disponíveis em <<http://w3.org/2002/mmi/ink>>. MathML é uma especificação de baixo nível para realizar descrições matemáticas com uma base para comunicações máquina-a-máquina e que fornece a estrutura básica para a inclusão de expressões matemáticas em páginas da *web*. Para mais detalhes, recomenda-se uma visita a <<http://w3.org/Math/>>. VoxML é uma aplicação da empresa Motorola para mensagens faladas, cuja intenção é facilitar a interpretação pelas máquinas tanto de correio eletrônico de voz, como gravações de sistemas de auto-resposta. Detalhes adicionais podem ser obtidos no endereço <www.w3.org/Voice/1999/VoxML.pdf>. MusicXML é uma iniciativa em desenvolvimento pelo site Recordare.com e cuja intenção é promover a interoperabilidade entre diversos aplicativos utilizados para a gravação ou para a execução de músicas. A MusicXML inclui todas as notações musicais como notas, ritmos, claves, letra da música, entre outras. Mais detalhes podem ser encontrados no site <<http://www.recordare.com/>>. A *Chemical Markup Language* (CML) foi desenvolvida por Peter Murray-Rust e é, provavelmente, uma das primeiras aplicações da linguagem XML. O desenvolvimento da CML foi originalmente iniciado em SGML, mas gradualmente foi migrando para a XML; o objetivo da linguagem é a padronização de metadados sobre informações químicas na *web*. Para mais detalhes, recomenda-se visitar o site <<http://www.xml-cml.org/>>.

(Daconta, 2003, p. 32). Compartilhando dessa orientação, o tutorial do *site* W3Schools explicita que XML é uma ferramenta utilizada para transmitir informação em ambientes que envolvem diversas plataformas em termos de *hardware* e de *software*. XML será tão importante no futuro da *web* como a HTML tem sido até hoje; além disso, será a ferramenta mais importante para todo tratamento, manipulação e transmissão de dados, devido às seguintes características: a) foi desenvolvida para descrever em formato de texto puro, de modo que esses dados possam ser trocados entre sistemas de diferentes plataformas; b) como as marcações não são predefinidas, o projetista pode desenvolver seu próprio conjunto de marcações; c) as metodologias DTD ou XML-Schema permitem a escritura de regras de sintaxe para um conjunto de marcações criado pelo usuário; d) um documento XML é auto-descritivo.

O código apresentado na tabela 24 exemplifica a simplicidade, a flexibilidade e a natureza auto-descritiva da linguagem XML.

Tabela 24: estrutura de um documento escrito em XML

Bloco de Código	Função	Uso
<code><?xml version="1.0" encoding="ISO-8859-1" ?></code>	Linha de cabeçalho, define a versão da linguagem, bem como o conjunto de caracteres padrão que será utilizado no documento; neste exemplo, o código refere-se ao conjunto de caracteres latinos.	obrigatório
<code><lei></code>	Elemento raiz do documento	obrigatório
<code><numero>426</numero></code> <code><data></code> <code><dia>06</dia></code> <code><mes>04</mes></code> <code><ano>1993</ano></code> <code></data></code> <code><autor>Governo do Distrito Federal</autor></code> <code><ementa>Cria a Subsecretaria de Defesa do Consumidor PROCON na estrutura da Secretaria de Governo do Distrito Federal e dá outras providências.</ementa></code>	Elementos filhos, utilizados para a descrição dos meta dados da informação representada	definido pelo usuário
<code></lei></code>	Indicação de fim de arquivo	Obrigatório

Fonte: adaptado de W3Schools (2004)

A depender da necessidade e da conveniência, os elementos definidos pelo usuário podem possuir atributos. Assim, os dados podem ser representados por meio de elementos filhos ou por meio de atributos, como explicita a tabela 25:

Tabela 25: Representação de dados por meio de elementos filhos ou de atributos

Uso de atributos	Uso de elementos filhos
<pre><senador sexo="masculino"> <nome>Rui</nome> <sobrenome>Barbosa</sobrenome> </senador></pre>	<pre><senador > <sexo>masculino</sexo> <nome>Rui</nome> <sobrenome>Barbosa</sobrenome> </senador ></pre>

Fonte: adaptado de W3Schools (2004)

De acordo com o W3Schools (2004), não há regras definidas para a escolha quanto à representação por meio de atributos ou por meio de elementos filhos. No entanto, apesar de os atributos serem mais facilmente manipuláveis em HTML, em documentos XML eles devem ser evitados, para que a representação se torne mais explícita e, ainda, pelos seguintes motivos: a) atributos não podem conter valores múltiplos; b) atributos não são facilmente expansíveis, para mudanças futuras; c) atributos não podem descrever estruturas de dados; d) atributos são mais difíceis de serem manipulados automaticamente por aplicativos; e) a validação da sintaxe de atributos por DTD's ou XML-Schemas – que definem os elementos permitidos em um documento XML – é mais difícil, comparada com a validação de elementos filhos. Uma exceção a esta última recomendação deve ser considerada, quando houver necessidade de se atribuir um identificador unívoco para um registro de dados, como visto na ilustração 48.

Uso de elementos filhos
<pre><deputado identificador="d001"> <sexo>feminino</sexo> <nome>Cecília</nome> <sobrenome>Meireles</sobrenome> </deputado> <deputado identificador="d002"> <sexo>masculino</sexo> <nome>Jorge</nome> <sobrenome>Amado</sobrenome> </deputado></pre>

Ilustração 48: Utilização de atributo como identificador unívoco para um registro

Uma vez que os elementos em XML são flexíveis, é possível que haja conflitos quando dois documentos diferentes utilizarem os mesmos nomes para descreverem diferentes tipos de elementos. Se os elementos forem utilizados no mesmo documento, é certo que haverá conflitos de interpretação no momento da extração de informações por um aplicativo. Nos exemplos dados na tabela 26 têm-se, na coluna da esquerda, os elementos `<autor>` e `<data>`, representando informações pertinentes a uma coleção de documentos; na coluna da direita, os mesmos elementos podem representar informações relativas a uma das leis que integram a coleção.

Tabela 26: uso de nomes iguais em elementos que representam informações diferentes

Informações sobre o documento	Informações sobre um registro no documento
<code><legislacao></code>	...
<code><titulo></code> Legislação temática sobre meio-ambiente</titulo>	<code><lei></code>
<code><autor></code> Câmara Legislativa do DF</autor>	<code><autor></code> Augusto Ruschi</autor>
<code><data></code> 18/ago/2004</data>	<code><ementa></code> Define áreas de preservação ambiental no DF</ementa>
<code></legislacao></code>	<code><data></code> 12/01/1996</data>
	<code></lei></code>
	...

A resolução para os conflitos pode ser dada por meio da atribuição de prefixos que identifiquem univocamente cada elemento, como pode ser visto na tabela 27, em que foram utilizados os prefixos **d** (para documento) e **r** (para registro).

Tabela 27: adição de prefixos para desambiguar os elementos do código XML

Informações sobre o documento	Informações sobre um registro no documento
<code><d:legislacao></code>	<code><r:lei></code>
<code><d:titulo></code> Coletânea de Legislação Temática sobre Meio Ambiente</d:titulo>	<code><r:autor></code> Augusto Ruschi</r:name>
<code><d:autor></code> Câmara Legislativa do DF</d:autor>	<code><r:ementa></code> Define áreas de preservação ambiental no DF</r:ementa>
<code><d:data></code> 18/ago/2004</d:data>	<code><r:data></code> 12/01/1996</r:data>
<code></d:legislacao></code>	<code></r:lei></code>

Fonte: adaptado de W3Schools (2004)

Além dos prefixos, podem ser acrescentadas referências sobre o URI onde a determinação de tal prefixo é estabelecida. Então, ao invés de se utilizar apenas o prefixo, é possível se acrescentar o atributo **xmlns**, um *namespace*, que é também um identificador unívoco. Tal atributo integra a biblioteca de atributos reservados definida para a linguagem XML pelo consórcio W3C.

Informações sobre o documento	Informações sobre um registro no documento
<code><d:legislacao</code> <code>xmlns(d)="http://www.w3.org/TR/html4/"</code> <code><d:titulo></code> Coletânea de Legislação Temática sobre Meio Ambiente</d:titulo>	<code><r:lei</code>
<code><d:autor></code> Câmara Legislativa do DF</d:autor>	<code>xmlns(r)="http://www.cl.df.gov.br/legislacao"</code>
<code><d:data></code> 18/ago/2004</d:data>	<code><r:autor></code> Augusto Ruschi</r:autor>
<code></d:legislacao></code>	<code><r:ementa></code> Define áreas de preservação ambiental no DF</r:ementa>
	<code><r:data></code> 12/01/1996</r:data>
	<code></r:lei></code>

Ilustração 49: utilização de namespaces para qualificar os elementos e atributos em um documento

Em suma, os *namespaces* são implementados pela adição de um prefixo a cada elemento em um documento. Por sua vez, cada prefixo é mapeado para um URI, por meio do atributo `xmlns` (`xmlns:prefixo`). Assim, os *namespaces* têm dois propósitos: a) promover a distinção entre elementos e atributos que possuem o mesmo nome, mas que, por serem oriundos de

vocabulários diferentes, possuem significados diferentes: b) agrupar todos os elementos e atributos de uma aplicação XML, de tal sorte que qualquer *software* possa reconhecê-los (HAROLD; MEANS, 2002).

2.4.4.2.2 XML-Schema

XMLSchema é uma linguagem de definição de tipos complexos que permite a realização de restrições de vocabulário em documentos XML. O papel de um XMLSchema é similar ao dos esquemas de especificação de estruturas de dados – como nomes de campos e seus respectivos tipos – em um banco de dados relacional. A linguagem, inicialmente proposta pela Microsoft, ganhou *status* de recomendação do W3C em maio de 2001 (DACONTA, 2003; W3SCHOOLS, 2004)

Com discutido anteriormente, para garantir a integridade sintática e semântica de um documento XML, utiliza-se uma DTD ou um XMLSchema. O propósito de um XMLSchema ou de uma DTD é a especificação da estrutura de um documento em termos de elementos, atributos e tipos de dados que tal documento pode conter. A diferença entre uma *Document Type Definition* (DTD) e um XML Schema é que, embora uma (DTD) forneça insumos para a validação básica de um documento em termos de estruturação de elementos, de ocorrência de restrições, de atributos permitidos, de tipos de atributos e de valores padrão, não fornece um controle fino sobre o formato e os tipos de dados dos elementos e seus atributos. Como motivações principais para a utilização de XMLSchemas, ao invés de DTDs, estão os seguintes fatores: a) DTDs são escritas com sintaxe própria, diferente da XML; b) DTDs são muito limitadas com relação à capacidade de suporte a tipos de dados – suportam apenas 10 ~~de~~ tipos de dados, enquanto um XMLSchema pode suportar 44 ou mais tipos, já que é possível a criação de tipos personalizados; c) XMLSchema pode restringir ou estender um tipo de dados, por meio da derivação das definições de um novo tipo com base em um tipo antigo; por meio da definição de um conjunto em que elementos filhos podem ocorrer em qualquer ordem; por meio da definição de elementos substituíveis, isto é, que possuem os mesmos tipos de dados.

As tabelas 28 e 29 apresentam, respectivamente, as estruturas sintáticas para a declaração de elementos e atributos em DTDs.

Tabela 28: Declarações de elementos em uma DTD

Declarações de elementos em uma DTD	Descrição
Fórmula geral: <!ELEMENT nome-do-elemento lista-de-elementos ou restrições>	
<!ELEMENT nome-do-elemento EMPTY>	Declara um elemento vazio (que não possui subelementos)
<!ELEMENT nome-do-elemento (#PCDATA)>	Declara um elemento que somente pode possuir texto
<!ELEMENT nome-do-elemento ANY>	Declara um elemento que pode conter texto ou até mesmo subelementos
<!ELEMENT nome-do-elemento (sub1, ...,subn)>	Declara um elemento com todos os seus subelementos, bem como a ordem de entrada destes; nessa declaração, é definido que cada subelemento ocorrerá obrigatoriamente e apenas uma vez
<!ELEMENT nome-do-elemento (sub1,sub2+)>	Declara que o elemento filho, seguido do sinal +, ocorrerá pelo menos uma vez
<!ELEMENT nome-do-elemento (sub1*)>	Declara que o elemento seguido do sinal * poderá ocorrer uma ou mais vezes, ou não ocorrer
<!ELEMENT nome-do-elemento (sub1?) >	Declara que o elemento seguido do sinal ? poderá ocorrer apenas uma vez, ou não ocorrer
<!ELEMENT nome-do-elemento (sub1 sub2) >	Declara que uma lista subelementos opcionais, sendo que apenas um deles será utilizado
<!ELEMENT nome-do-elemento (sub1 sub2)* >	Declara que o elemento poderá possuir um dos dois elementos, um ou vários de um deles, vários de ambos, ou nenhum deles
<!ELEMENT nome-do-elemento (sub1 sub2)+ >	Declara que o elemento poderá possuir um dos dois elementos, um ou vários de um deles, vários de ambos; no entanto, a condição vazio não é permitida
<!ELEMENT nome-do-elemento (sub1 sub2)? >	Declara que o elemento pode ser vazio ou conter apenas uma ocorrência de um dos dois elementos

Fonte: Adaptado de Furgeri (2001).

Tabela 29: Declaração de atributos em uma DTD

Declaração de atributos de elementos em uma DTD	Descrição
Fórmula geral: <!ATTLIST nome-do-elemento nome-do-atributo tipo-do-atributo valor>	
CDATA	Qualquer tipo de texto, exceto caracteres reservados: "<", ">" e "&"
(valor1 valom)	Lista de valores
ID	Identificador do atributo
IDREF	Identificador de elemento relacionado
IDREFS	Lista de identificadores de elementos relacionados
NMTOKEN	Nome válido em XML
NMTOKENS	Lista de nomes válidos em XML
ENTITY	Entidade
ENTITYS	Lista de entidades
NOTATION	Notação
#DEFAULT	Valor padrão dado a um atributo
#REQUIRED	Indica a obrigatoriedade da inclusão do atributo no elemento
#IMPLIED	Indica a não obrigatoriedade da inclusão do atributo no elemento
#FIXED	Indica que o valor do atributo é fixo

Fonte: Adaptado de Furgeri (2001).

Um arquivo escrito em XML Schema utiliza sintaxe XML – e, conseqüentemente, é extensível – para declarar um conjunto de tipos simples ou complexos. Um tipo é um modelo que pode possuir um valor – tipo simples – ou múltiplos valores – tipo complexo. Então, cada tipo possui duas características-chaves: um nome e um conjunto de valores permitidos. Por exemplo, para a escrita do elemento *autor*, de modo que seja considerado válido em um documento XML, é necessário que para tal elemento se especifique um nome e um tipo de dados: `<xsd:element name="autor" type="xsd:string">` em um arquivo de esquema. Essa sintaxe garante a correção de uma declaração do tipo: `<autor>Rui Barbosa</autor>`, uma vez que especifica um nome válido para um elemento (*autor*) e tipos de valores aceitáveis (*string* – cadeia de caracteres).

Como arquivos XML que são, os XMLSchemas também aceitam sintaxes com uso de *namespaces*. Assim, usualmente, escreve-se `<xsd:identificador>`, onde **xsd** é um prefixo utilizado no lugar de um URI que indica onde está localizado o vocabulário aceito no XMLSchema, e *identificador* é um atributo identificador de meta dado. A declaração do *namespace* pode ser, por exemplo: `<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">`.

Nesse exemplo, é declarado um *namespace* para todos os elementos permitidos em um documento válido. Assim, o prefixo *xsd* corresponde ao URI `<"http://www.w3.org/2001/XMLSchema">`, que contém o vocabulário aceito. É importante que se destaque que o prefixo é meramente uma abreviação para o *namespace* que, na prática, é um URI. O prefixo *xsd* é um acrônimo para *XML Schema Definition* (W3SCHOOLS, 2004; DACONTA, 2003).

Em resumo, um esquema é uma descrição formal que compreende um documento válido, isto é, um XML Schema é um documento contendo uma especificação formal do que é um documento XML válido. A linguagem XML Schema, do W3C, é um esquema escrito em particular, cuja sintaxe é uma recomendação daquele Consórcio (HAROLD; MEANS, 2002; W3SCHOOLS, 2004).

Assim, diferentemente das DTD's, os esquemas podem forçar regras específicas sobre os conteúdos dos elementos e de seus atributos, de tal maneira que além de tipos simples previstos pela linguagem – como *string*, inteiro, decimal e datahora –, a linguagem de esquema permite a criação de novos tipos, derivados dos antigos, o reuso desses tipos por outros esquemas, bem como a possibilidade de se referenciar mais de um esquema no mesmo documento (HAROLD; MEANS, 2002; W3SCHOOLS, 2004). Segundo o tutorial do site W3Schools, um XML Schema define os elementos que podem aparecer em um documento,

define os atributos que podem aparecer em um documento; define que elementos são elementos filhos; define a ordem de precedência dos elementos filhos; define o número de elementos filhos; define se um elemento é vazio ou se pode conter texto; define os tipos de dados para elementos e atributos; define valores padrão para os elementos e atributos.

O código apresentado na tabela 30 exemplifica a utilização de um XMLSchema para a estruturação sintática de um documento XML.

Tabela 30: uso de um XMLSchema para a estruturação sintática de um documento XML

Arquivo XML	XML-Schema
<pre><<?xml version="1.0" encoding="ISO-8859-1"?> <lei> <numero>426</numero> <data>06/04/1993</data> <autor>Governo do Distrito Federal</autor> <ementa>Cria a Subsecretaria de Defesa do Consumidor PROCON na estrutura da Secretaria de Governo do Distrito Federal e dá outras providências. </ementa> <</lei></pre>	<pre><?xml version="1.0"?> <xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns="http://www.cl.df.gov.br"> <xsd:element name="lei"> <xsd:complexType> <xsd:sequence> <xsd:element name="numero" type=" xsd:string"/> <xsd:element name="data" type=" xsd:string"/> <xsd:element name="autor" type=" xsd:string"/> <xsd:element name="ementa" type=" xsd:string"/> </xsd:sequence> </xsd:complexType> </xsd:element> </xsd:schema></pre>

No esquema à direita, o elemento *lei* é do tipo *complexo*, porque contém outros elementos (*numero*, *data*, *autor*, *ementa*); esse documento faz uma referência ao URI do consórcio W3C (<http://www.w3.org/2001/XMLSchema>) em que se encontra a biblioteca de elementos válidos para garantir a integridade do próprio esquema. O esquema, por sua vez, destina-se a garantir a integridade de qualquer documento XML que utilize a estrutura definida (tabela 31).

Tabela 31: exemplo de estrutura de um XMLSchema

XML- Schema	Função	Uso
<?xml version="1.0" encoding="ISO-8859-1"?>	Linha de cabeçalho, define a versão da linguagem	Obrigatório
<xsd:schema	Declara o elemento raiz	Obrigatório
xmlns:xsd="http://www.w3.org/2001/XMLSchema"	Indica que os elementos a serem utilizados no documento são os definidos pelo W3C	Recomendado
xmlns="http://www.cl.df.gov.br">	Indica que o namespace http://www.cl.df.gov.br é o padrão	Recomendado
<xsd:element name="lei">		
<xsd:complexType>		
<xsd:sequence>		
<xsd:element name="numero" type="xsd:string"/>		
<xsd:element name="data" type="xsd:string"/>	Elementos definidos pelo usuário, conforme a necessidade de descrição das informações	Definido pelo usuário
<xsd:element name="autor" type="xsd:string"/>		
<xsd:element name="ementa" type="xsd:string"/>		
</xsd:sequence>		
</xsd:complexType>		
</xsd:element>		
</xsd:schema>	Indicação de final do arquivo de esquema	Obrigatório

Fonte: adaptado de W3Schools (2004)

A tabela 32 apresenta um exemplo de como uma chamada a esse esquema poderia ser feita em um arquivo XML.

Tabela 32: referência a um XMLSchema em um arquivo XML

Referência a um XMLSchema
<pre><?xml version="1.0" encoding="ISO-8859-1"?> <lei xmlns="http://www.cl.df.gov.br" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.cl.df.gov.br lei.xsd"> <numero>426</numero> <data>06/04/1993</data> <autor>Governo do Distrito Federal</autor> <ementa>Cria a Subsecretaria de Defesa do Consumidor PROCON na estrutura da Secretaria de Governo do Distrito Federal e dá outras providências. </ementa> </lei></pre>

A declaração `xmlns="http://www.cl.df.gov.br"` indica a qualquer aplicativo que for realizar uma validação no esquema que todos os elementos utilizados neste documento XML estão declarados nesse *namespace*. A declaração `xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"` indica que será utilizada uma instância de um esquema. A declaração

documentos na *web*. Na tabela 34, apresenta-se uma adaptação ao pensamento de Lima (2003) sobre as diversas linguagens envolvidas na construção da *web* semântica em suas inter-relações.

Qualquer que seja a linguagem utilizada na *web* semântica, seu objetivo é fornecer uma descrição para o conhecimento representado por meio de ontologias. Como já mencionado, uma ontologia é “um documento ou arquivo que formalmente define relações entre termos” (BERNERS-LEE, 2001) e, assim, “define os termos usados para descrever e representar uma área do conhecimento” (LIMA, 2003). Lima destaca, também, que a *web* semântica necessita de ontologias com relativo grau de complexidade, uma vez que destinam-se a especificar classes em diversos domínios de interesse, relacionamentos entre essas classes e propriedades que essas classes podem possuir⁵⁸.

Tabela 34: Linguagens da *web* semântica e suas funções

Linguagem	Função na <i>web</i> semântica
XML	fornecer sintaxe básica para a estruturação dos documentos; não há imposição de restrições semânticas
XML Schema	fornecer controle sintático à linguagem XML, por meio de restrições à estrutura, bem como da definição de tipos de dados para documentos XML
RDF	fornecer um modelo de dados para os objetos da <i>web</i> e seus relacionamentos, propiciando uma semântica simples, com possibilidades de representação do próprio modelo em XML
RDF Schema	fornecer um vocabulário capaz de descrever os objetos da <i>web</i> e suas relações hierárquicas de gênero-espécie
DAML+OIL e OWL	fornecer vocabulário adicional capaz de descrever outros tipos de relacionamentos entre os objetos da <i>web</i> , como relacionamentos de disjunção, cardinalidade, igualdade, enumeração; fornecem tipos de dados mais ricos que as linguagens anteriores.

Fonte: adaptado de (LIMA, 2003)

Outros autores apresentam também as suas definições para ontologias: “especificações compartilhadas de terminologia e conceituações necessárias em diferentes áreas de aplicação” (HYVÖNEN, 2002). “Ontologia é a descrição explícita e precisa de conceitos e relações que existem em um domínio particular tal como uma dada organização, um domínio de estudo, uma área de aplicação, etc”⁵⁹ (MOURA, 2002).

De acordo com Lassila (1997), pode-se dizer que a história dos meta dados no W3C começou com o desenvolvimento da Plataforma *Internet* de Seleção de Conteúdos (PICS),

⁵⁸ Nesta pesquisa, pode-se relacionar os termos classes, relacionamentos e propriedades, respectivamente a conceitos, relações e características.

⁵⁹ Por exemplo, a ontologia definida para um programa de computador deve incluir a especificação dos conceitos: programa, identificador, declaração, expressão, controles, operadores, etc. [um operador] deve estar relacionado por um tipo de relacionamento à classe de operadores relacionais. Identificadores devem ser definidos como string, etc” (MOURA, 2002).

um mecanismo para a classificação de conteúdos de páginas que permite a estruturação de meta dados. Inicialmente os PICS foram utilizados para o desenvolvimento de sistemas de filtro de conteúdo, para evitar que crianças ficassem expostas a conteúdos obscenos na *Internet*, informando restrições ao navegador. Tais sistemas são úteis quando é possível identificar todos os valores possíveis para os dados antecipadamente. Lassila afirma que a *Resource Description Framework* como uma estrutura geral de meta dados e, conseqüentemente, como um mecanismo de representação genérica do conhecimento para a *web*, foi inspirado nos PICS.

O significado é expresso em RDF como um conjunto de triplas, cada tripla correspondendo ao sujeito, ao verbo e ao objeto de uma sentença simples. Essas triplas podem ser escritas em XML e um documento RDF pode fazer declarações a respeito de coisas particulares (pessoa), que possuem suas propriedades (é irmã de), com certos valores (outra pessoa). RDF, afirma LIMA (2003), diz respeito a expressões processáveis por máquinas, pois fornece um mecanismo para a descrição de recursos, sem fazer nenhuma inferência sobre o domínio da aplicação.

Na concepção de Lassila (1997), em seu núcleo, o dado RDF consiste em nós (qualquer recurso *web*, como páginas, servidores e outras coisas que a que se possa atribuir uma URI, até outras instâncias de meta dados) e atributos anexados (propriedades dos nós e seus valores). A essência do RDF é o modelo de nós, atributos e valores, que devem ser representadas por grafos, de modo a serem comunicados de um agente automático para outro; isso é feito com o uso da linguagem XML.

Uma grande vantagem do RDF é que declarações feitas nesse modelo podem dizer praticamente qualquer coisa sobre um objeto, não importando quem diz; isto significa que não há um *site* oficial que diz absolutamente tudo sobre determinado assunto, mas qualquer *site* pode conter informações sobre tal assunto. Lassila (1997) conclui que isso leva a um importante princípio: em RDF, "qualquer coisa pode dizer qualquer coisa sobre qualquer coisa".

A despeito dessa flexibilidade, de acordo com o W3C, propriedades RDF podem ser pensadas como atributos de um recurso e, nesse sentido, correspondem aos tradicionais pares valor-atributo. No entanto, RDF não fornece mecanismos para descrição dessas propriedades e em para a descrição desses relacionamentos entre propriedades e outros recursos. Enfim, RDF oferece suporte básico para ontologias, mas é necessário maior rigor semântico.

Para descrever o vocabulário do RDF foi criado o padrão RDF Schema, que define propriedades que podem ser utilizadas para descreverem classes, propriedades e outros recursos (W3C, 2003). Então, a função do RDF'S é permitir a criação de vocabulário:

controlados, compartilháveis e extensíveis. Um esquema é um tipo de ontologia onde são definidos os termos que serão usados nos documentos RDF e o significado específico de cada termo. São esses esquemas que expressam a semântica dos documentos RDF (LIMA, 2003).

O uso de ontologias fornece uma grande possibilidade de se descrever objetos e suas relações com outros objetos. Entretanto, devido a restrições semânticas da arquitetura RDF, algumas iniciativas foram desenvolvidas com o objetivo de se solucionar tais problemas. A linguagem *Darpa Agent Markup Language* (DAML)⁶⁰, projeto iniciado em um encontro de especialistas em agosto do ano 2000, em Boston, é uma dessas iniciativas. A versão mais recente da linguagem é a DAML+OIL⁶¹, uma combinação das linguagens DAML e OIL, e fornece um conjunto de ferramentas com o qual é possível a criação de ontologias e a marcação de informação, de modo a viabilizar a leitura e compreensão automática por máquinas (DAML, 2003).

Face às restrições da arquitetura RDF, o W3C propôs a especificação de mais um padrão de linguagem, de modo a permitir maior rigor semântico à representação do conteúdo das páginas *web*. Assim, a *Web Ontology Language* (OWL), uma evolução da combinação DAML+OIL, foi projetada para uso por aplicações que necessitam processar conteúdo de informação, ao invés de apenas apresentar informações para humanos. De acordo com o W3C, a OWL facilita a maior interoperabilidade de conteúdo da *web* do que o suportado por XML, RDF e RDF Schema, fornecendo um vocabulário adicional e também uma estrutura de semântica formal. A OWL possui três sublinguagens, cada uma sendo uma extensão de sua predecessora mais simples:

a) OWL Lite – fornece suporte para aqueles usuários que necessitam primeiramente de uma hierarquia de classificação e de restrições simples;

b) OWL DL - destina-se àqueles usuários que querem o máximo de expressividade enquanto retendo completeza computacional (será garantido que todas as conclusões serão computadas) e confiabilidade nas decisões. A OWL DL inclui todos os construtos da OWL, mas eles podem ser utilizados somente sob certas restrições (por exemplo, embora uma classe possa ser uma subclasse de diversas classes, uma classe não pode ser uma

⁶⁰ Embora o Programa DAML não seja uma iniciativa do W3C, possui diversas interfaces com o consórcio, entre elas o fato de o próprio Tim Berners-Lee ter participado do seu desenvolvimento (OUELLET; OGBUJI, 2002).

⁶¹ O originária da cooperação de diversas instituições de pesquisa européias, a *Ontology Interchange Language* (OIL), é uma proposta de representação e uma camada de inferência para ontologias, que combina a modelagem de primitivas com semântica formal e serviços inteligentes fornecidos pela lógica descritiva. A OIL é compatível com RDF Schema (RDFS), e inclui uma semântica precisa, para a descrição do significado dos termos, bem como de informações implícitas (ONTOKNOWLEDGE, 2004).

instância de uma outra classe). O nome DL corresponde a *description logics*, um campo de pesquisa que tem estudado as especificações formais da OWL.

c) OWL Full - é útil para usuários que querem o máximo de expressividade e liberdade sintática em relação ao RDF, sem garantias computacionais. Por exemplo, na OWL Full, uma classe pode ser tratada simultaneamente como uma coleção de indivíduos e como o próprio indivíduo. OWL Full permite a uma ontologia aumentar o significado de um vocabulário pré-definido, seja ele RDF ou OWL.

Em resumo, a XML fornece uma sintaxe básica para a estruturação de documentos, mas não possui elementos que lhe permitam impor restrições semânticas ao significado de tais elementos. XML-Schema impõe restrições sintáticas à estrutura de documentos escritos em XML. RDE é um modelo de dados para objetos ou “recursos” e tal modelo pode ser representado utilizando-se a sintaxe da XML. Fisicamente, um modelo RDF – ou ontologia, como definem alguns autores – é um arquivo que formalmente define relações entre termos. RDF-Schema é um vocabulário para a descrição de propriedades e classes de um modelo RDF, com a utilização de elementos semânticos para a generalização de tais propriedades e classes. Complementando todas essas tecnologias, a OWL fornece mais vocabulário para a descrição de propriedades e classes, por meio de: a) relações entre classes – por exemplo: disjunção; b) cardinalidade – por exemplo: univocidade; c) igualdade; d) maior riqueza de propriedades para a especificação de tipos de dados; e) propriedades relativas a características – por exemplo: simetria; f) enumeração de classes (W3C, 2003). Esses são, também, os objetivos da linguagem DAML+OIL (OUELLET; OGBUJI, 2002).

2.4.4.3.1 RDF

O RDF é um esforço para a criação de convenções que controlem como a semântica, a sintaxe e a estrutura de metadados são formuladas em um domínio, de modo que, quando tais estruturas forem misturadas com um segundo ou terceiro domínio, as declarações originais mantenham sua clareza e legibilidade. Como já discutido na Revisão de Literatura, a linguagem XML é, por natureza, puramente estrutural e sintática e descreve dados apenas no nível de objetos. Por esse motivo, o padrão RDF foi projetado para permitir que *software* inteligente, como um *spider* ou um agente inteligente, consiga descobrir, catalogar e entender recursos em um *site*, uma vez que seus objetivos são: ter um modelo de dados simplificado; possuir semântica formal e provável inferência; utilizar um vocabulário extensível; utilizar sintaxe baseada em XML; dar suporte aos tipos de dados do XML Schema; permitir que qualquer pessoa possa fazer declarações acerca de qualquer recurso (W3C, 2001; STUCKENSCHMIDT; VAN

HARMELEN, 2003; POWERS, 2003; W3C, 2004). Além dessas características, o RDF pode utilizar valores representados em conformidade com os tipos de dados do XMLSchema, possibilitando, assim, o intercâmbio de dados com outras aplicações XML.

Um documento RDF pode ser utilizado para criar descrições de meta dados relativas não apenas a informações internas ao documento, mas também sobre metadados externos ao seu conteúdo ou até mesmo ocultos – como por exemplo, autor, data de criação, tipo de arquivo –, sendo particularmente útil na representação de conteúdo de arquivos de áudio e imagens (DACONTA, 2003, p. 85). Dessa maneira é possível a gravacão de dados em formato “compreensível-por-máquina”, favorecendo operações como intercâmbio, busca, catalogação, navegação e classificação, de maneira mais eficiente e sofisticada (POWERS, 2003).

São conceitos básicos do RDF: modelo de dados gráfico; vocabulário baseado em URI; tipos de dados; literais; sintaxe de serialização XML, expressão de fatos simples; implicações. O modelo gráfico baseia-se nos conceitos de sujeito, predicado e objeto que, tomados juntos, formam uma tripla. Um conjunto de tais triplas é denominado Grafo RDF. Cada tripla RDF é uma declaração e pode ser pensada, também, em termos de objetos, propriedades e valores de propriedades. (HAROLD, 1999; W3C, 2004). Por exemplo, na declaração “A Lei nº1/1917 é de autoria de Rui Barbosa”, o recurso é a “Lei nº1/1917”, a propriedade é a relação de “autoria” e o valor de tal propriedade é “Rui Barbosa”.

Como mencionado, em termos semânticos, a estrutura de um documento RDF é baseada no conceito de Triplas RDF, isto é, em três partes que compõem a informação e que são necessárias para a sua compreensão como uma unidade do conhecimento. Essa estrutura é o que torna possível a interpretação semântica de uma sentença, tanto por humanos quanto por máquinas. Dito de outro modo, a estrutura de triplas permite que a compreensão e a significação dadas por humanos para uma sentença sejam consistentemente interpretadas por máquinas (POWERS, 2003). Como partes integrantes do discurso humano, em língua inglesa, esses elementos são denominados como *sujeito, predicado e objeto*. O *sujeito*, no contexto gramatical, corresponde a um nome ou frase nominal que representa aquele que pratica uma ação – no campo da lógica, esse é o objeto sobre o qual se elabora alguma declaração. Em RDF, necessita-se de um URI para a identificação unívoca do conceito sobre o qual se elabora tal declaração; esse URI é também representativo de um *recurso*. O *predicado*, gramaticalmente, é a parte da sentença que modifica o sujeito e inclui uma frase verbal; em outras palavras, o *predicado* diz algo a respeito do *sujeito* – no campo da lógica, um predicado é uma função de um indivíduo (um tipo particular de sujeito) para valores verdade, com base em certo número de *argumentos*. Em RDF, um

predicado é uma *relação* entre o *sujeito* e o *objeto*. Gramaticalmente, um objeto é um nome que sofre a ação expressa por um verbo – para a lógica um objeto é influenciado pelo predicado. Em RDF, um objeto pode ser também um recurso, referenciado por um predicado ou por um valor literal (Daconta, 2003).

A ilustração 50 apresenta a estrutura de uma Tripla RDF.

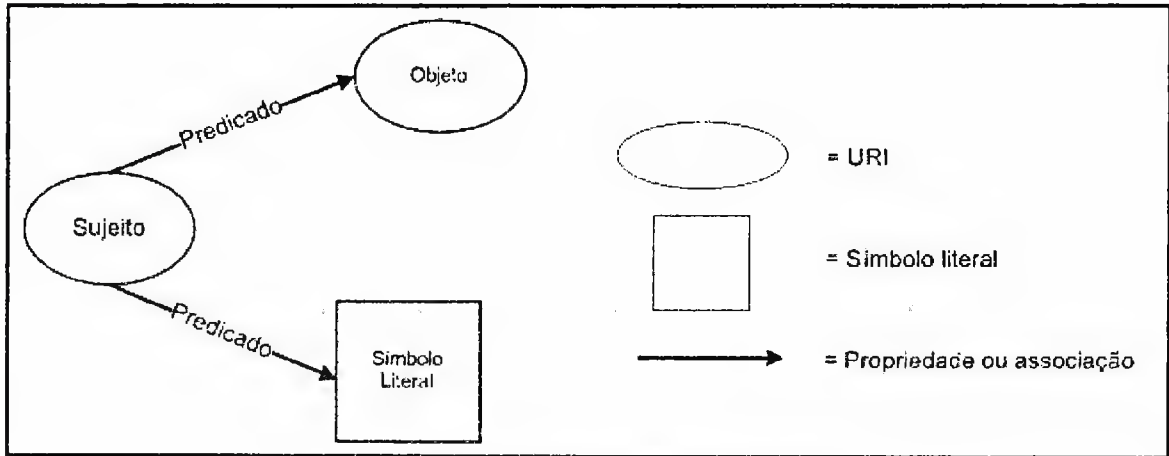


Ilustração 50: Tripla RDF (Daconta, 2003, p. 88)

Além do modelo lingüístico apresentado, é possível a construção de declarações RDF pensando-se em recursos, propriedades e valores, de tal modo que a cada recurso pode ser atribuída certa propriedade, com certo valor. O objetivo da arquitetura RDF é que, utilizando-se qualquer das abordagens, seja possível a realização de declarações como, por exemplo, “*Rui Barbosa é autor da Lei n.º 179*”. Assim, no modelo lingüístico RDF, *Rui Barbosa* é o sujeito, *é autor* o predicado e *Lei n.º 179* o objeto. Certamente, esse modo de interpretação para tornar-se diretamente aplicável à terminologia gramatical da Língua Portuguesa deveria sofrer alterações. Apenas para se ilustrar essa necessidade, no exemplo dado, a análise da declaração deveria ser feita da seguinte maneira:

“*Rui Barbosa*” → *sujeito* – elemento da oração sobre o qual se emite uma informação;

“*é autor da Lei n.º 179*” → *predicado* – elemento da oração que informa algo a respeito do sujeito;

O predicado, por sua vez, pode ser assim analisado:

“*é*” → *verbo de ligação* – verbo que não possui um conteúdo significativo, isto é, não traz nenhuma informação a respeito do sujeito, mas que serve como elemento de ligação entre o sujeito e o seu atributo (predicativo do sujeito);

“*autor da Lei nº 179*” → *predicativo do sujeito* – elemento que se refere ao sujeito mediante um verbo;

Em face do problema levantado nesse exemplo⁶², e devido à complexidade gramatical da Língua Portuguesa, o modelo *recurso-propriedade-valor* parece ser mais aplicável à documentação escrita em português. Desse modo, para o exemplo em questão, ter-se-ia:

“Rui Barbosa” → recurso;

“é autor” → propriedade;

“Lei nº 179” → valor

É importante notar que toda declaração é feita em um documento, e que todo documento possui um endereço, isto é, uma localização identificada por meio de um URI. Para a sentença dada, então, é possível se utilizar a seguinte notação:

<#Rui Barbosa> <#autor> <#Lei nº 179>

Estendendo-se sua significação, pode-se pensar que a referida lei possui uma página na Internet e, em consequência, um endereço; então, é válida a declaração:

<#Lei nº 179> <#tempagina> <<http://www.cl.df.gov.br/legislacao/lei179.htm>>

O símbolo #, em cada declaração anterior, significa que o URI dos conceitos referenciados é o próprio documento RDF que os contém. Essa é uma sintaxe abreviada para um URI absoluto, como por exemplo: <<http://www.cl.df.gov.br/legislacao/ontolegis>>. O mais usual – e recomendável – é a substituição do símbolo # por um prefixo que represente o *namespace* em que se localiza o documento RDF em questão. A formulação resultante é, então:

prefixo leis: <<http://www.cl.df.gov.br/legislacao/ontolegis>>

Desse modo, as declarações sobre a autoria de uma lei podem ser as seguintes:

<leis:Rui Barbosa><leis:autor><leis:lei nº 179>

Por certo, para tornar-se válida, essa declaração necessitaria ser inserida em um documento semântica e sintaticamente estruturado, isto é, em um documento válido. Como a arquitetura RDF possui suas bases na estrutura da linguagem XML, pode ser utilizada para a descrição de conteúdo com a implementação de todos os recursos desta linguagem, o que vale inclusive para os *namespaces*. É possível também a combinação com outros padrões de metadados,

⁶² O exemplo em questão foi fundamentado na obra Curso prático de gramática, de Terra (1991).

como, por exemplo, o Dublin Core, de modo que o arquivo se torne compreensível para aplicativos que fazem uso desses padrões, como os mecanismos de busca da *web*. A tabela 35 utiliza um exemplo proposto por Harold e Means (2002) para elucidar a aplicação de *namespaces* em um documento RDF.

Tabela 35: Estrutura de um arquivo RDF

Bloco de código	Função
<code><?xml version="1.0" encoding="ISO-8859-1"?></code>	Início do arquivo
<code><catalog></code>	elemento raiz
<code><RDF></code> <code> <Description</code> <code> about="http://www.cafeconleche.org/examples/impressionists.xml"></code>	descrição do tipo de conteúdo a que se refere o arquivo e do tipo de vocabulário utilizado, segundo a RDF
<code> <title> Obras impressionistas </title></code> <code> <creator> Elliotte Rusty Harold </creator></code> <code> <description></code> Contém uma lista de famosas obras impressionistas, organizadas por pintor e data <code> </description></code> <code> <date>22/08/2000</date></code> <code> </Description></code> <code></RDF></code>	Uso de elementos do Dublin Core, para descrever o documento
<code><painting></code> <code> <title>Memory of the Garden at Etten</title></code> <code> <artist>Vincent Van Gogh</artist></code> <code> <date>November, 1888</date></code> <code> <description></code> Two women look to the left. A third works in her garden. <code> </description></code> <code></painting></code>	Descrição de um registro referente a uma obra de arte
<code></catalog></code>	fim do elemento raiz e do arquivo

Fonte: adaptado de (HAROLD; MEANS, 2002).

Com a intenção de tornar o conteúdo conhecido para qualquer aplicativo que seja capaz de compreender padrões RDF e Dublin Core, os autores sugeriram utilizar no código o elemento *Description*, que pretence ao vocabulário da arquitetura RDF e os elementos *title*, *creator*, *description*, e *date*, que pertencem ao vocabulário do Dublin Core. Para que um aplicativo – um mecanismo de busca, por exemplo – seja capaz de reconhecer a informação, é necessário que os nomes dos elementos sejam preservados. No entanto, ocorre uma sobreposição dos elementos *title* e *description* que descrevem o documento com os elementos *title* e *description* que descrevem a obra de arte. Uma maneira de desambigüizar a descrição é utilizando *namespaces* para particionar os elementos em conjuntos diferentes de vocabulários. Elementos que possuem o mesmo nome, mas que apresentam URIs diferentes, são elementos distintos. Uma vez que URIs geralmente possuem caracteres especiais – como /, % e ~, para citar alguns não aceitos na sintaxe XML para nomear elementos –, recomenda-se o uso de prefixos curtos para a denominação dos *namespaces*.

Assim, cada prefixo é associado a um URI e, da mesma forma, nomes de elementos associados ao mesmo prefixo estarão no mesmo *namespace*. Para o código proposto anteriormente, na tabela 35, uma reescritura viável é exemplificada na tabela 36.

Tabela 36: inclusão de namespaces relativos a elementos Dublin Core em um arquivo RDF

Bloco de código	Função
<code><?xml version="1.0" encoding="ISO-8859-1"?></code>	início do arquivo
<code><catalog></code>	elemento raiz
<code><rdf:RDF xmlns:rdf="http://www.w3.org/TR/REC-rdf-syntax#"></code> <code><rdf:Description xmlns:dc="http://purl.org/dc/"</code> <code>about="http://www.cafeconleche.org/examples/impressionists.xml"></code>	Descrição do tipo de conteúdo a que se refere o arquivo e do tipo de vocabulário utilizado, segundo a RDF; uso de um prefixo para representar o namespace do Dublin Core
<code><dc:title> Obras impressionistas </title></code> <code><dc:creator> Elliotte Rusty Harold </creator></code> <code><dc:description></code> Contém uma lista de famosas obras impressionistas, organizadas por pintor e data <code></dc:description></code> <code><dc:date>22/08/2000</dc:date></code> <code></rdf:Description></code> <code></rdf:RDF></code>	Descrição do documento com elementos Dublin Core; uso do prefixo dc, para desambigüizar esses elementos em relação aos elementos que descrevem as obras
<code><painting></code> <code><title>Memory of the Garden at Etten</title></code> <code><artist>Vincent Van Gogh</artist></code> <code><date>November, 1888</date></code> <code><description></code> Two women look to the left. A third works in her garden. <code></description></code> <code></painting></code>	descrição um registro referente a uma obra de arte
<code></catalog></code>	fim do elemento raiz e do arquivo

Fonte: adaptado de (HAROLD; MEANS, 2002).

Para cada *namespace*, elementos ou atributos prefixados devem ser registrados com a seguinte estrutura: **prefixo:elemento (ou atributo)**. Para se conectar um prefixo a uma URI, utiliza-se o atributo **xmlns**, seguido do prefixo, da seguinte maneira: **xmlns:prefixo**. Analisando-se o código apresentado, tem-se: a) **xmlns:rdf** é usado como atributo para conectar o elemento RDF, representado por **rdf:RDF** ao URI do *namespace* **http://www.w3.org/TR/REC-rdf-syntax#**, que contém o vocabulário padrão definido pelo consórcio W3C; b) **rdf:Description** representa o elemento **Description** do vocabulário RDF definido pelo W3C; c) **xmlns:dc** é usado como atributo para conectar os elementos Dublin Core – *title*, *creator*, *description* e *date* – com o vocabulário padrão definido no URI **http://purl.org/dc/** e, embora tenha sido declarado no elemento **rdf:Description** – por questão de conveniência –, poderia ter sido declarado no elemento raiz.

A utilização desse tipo de construção, argumentam Harold & Means (2002), além de facilitar a implementação de esquemas como DTD's e XML-Schema, possibilita maior flexibilidade, caso seja necessária a redefinição de um prefixo ou sua atribuição a um URI diferente, posto que a declaração é feita apenas em um ponto do arquivo. Com essa estrutura, o arquivo poderá ser interpretado por qualquer aplicativo que se utilize dos padrões Dublin Core ou RDF para a descrição de conteúdo.

Um exemplo de aplicação da sintaxe RDF a uma coleção de documentos legais é apresentado na tabela 37.

Tabela 37: exemplo de aplicação da arquitetura RDF para um acervo legislativo

Bloco de Código	Função
<code><?xml version="1.0" encoding="ISO-8859-1"?></code>	Linha de cabeçalho, define a versão da linguagem, bem como o conjunto de caracteres padrão que será utilizado no documento; neste exemplo, o código refere-se ao conjunto de caracteres latinos
<code><rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"></code>	Indica que o conteúdo do documento é RDF. O prefixo rdf:RDF é utilizado para indicar que os elementos e atributos RDF aceitos nesse documento provêm da biblioteca definida pelo W3C (xmlns:rdf namespace)
<code>xmlns:lei="http://www.cl.df.gov.br/legislacao"></code>	Indica que os elementos e atributos definidos para a descrição de documentos legais provêm do URI referenciado
<code><rdf:Description rdf:about="http://www.cl.df.gov.br/legislacao/lei426"></code>	O elemento rdf:Description utiliza o atributo rdf:about para descrever cada lei
<code><lei:numero>426</lei:numero> <lei:data>06/04/1993</lei:data> <lei:autor>Governo do Distrito Federal</lei:autor> <lei:ementa>Cria a Subsecretaria de Defesa do Consumidor PROCON na estrutura da Secretaria de Governo do Distrito Federal e dá outras providências. </lei:ementa> </rdf:Description></code>	Os elementos numero, data, autor e ementa são referenciados pelo prefixo [lei:], indicando um repositório, como uma ontologia, em que tais elementos são válidos.
<code></rdf:RDF></code>	Indicação de final de arquivo

Retomando os conceitos de recurso, propriedades e valores, do exemplo dado na tabela 38, é possível expressar informações por meio de declarações como: “o **recurso** <<http://www.cl.df.gov.br/legislacao/lei426>> possui uma **propriedade autor**, cujo **valor** é *Governo do Distrito Federal*”; ou “o **recurso** <<http://www.cl.df.gov.br/legislacao/lei426>> possui uma **propriedade data**, cujo **valor** é *06/04/1993*”.

2.4.4.3.2 RDFS

Propriedades RDF podem ser pensadas como atributos de recursos e, nesse sentido, correspondem aos tradicionais pares atributo-valor utilizados em XML. Tais propriedades também representam relacionamentos entre recursos. Como é uma linguagem de descrição de vocabulário, é papel do RDFSchema descrever tais propriedades e suas relações com propriedades de outros recursos, o que é feito por meio da definição de classes e propriedades que podem ser utilizadas para descrever classes, propriedades e outros recursos.

A linguagem de descrição de classes e propriedades em RDF assemelha-se a linguagens de programação orientadas a objetos, uma vez que permite a definição de recursos como instâncias de classes ou como subclasses de classes. No entanto, enquanto em outras linguagens uma classe é definida em função das propriedades que uma instância pode possuir (por exemplo, a *classe livro* possui um *atributo autor* que é um *tipo de pessoa*). Em RDF, a descrição de uma propriedade é feita em termos de classes do recurso ao qual tal propriedade é aplicável (a *propriedade autor* possui um *domínio documento* e uma *extensão pessoa*). Em outras palavras, como abordagem RDF é centrada na propriedade, e não na classe, é possível estender-se a descrição de recursos, sem a necessidade de se redefinir a descrição original da classe.

Em suma, na especificação formal da arquitetura RDF, os recursos podem ser divididos em grupos chamados *classes*. Os membros de uma classe são denominados *instâncias*. As classes em si são também recursos e podem ser descritas por meio de *propriedades*. Ao conjunto de instâncias de uma classe denomina-se *extensão da classe*. Uma classe pode ser membro de sua própria extensão, isto é, pode ser uma instância de si própria (W3C, 2004b). Todas as classes em RDFSchema são agrupadas em uma classe principal denominada *rdfs:Class*. De tal maneira, se uma classe X é uma *subclasse* de Y, todas as instâncias de X serão também instâncias de Y e poderão ser declaradas com o uso da propriedade *rdfs:subClassOf*. O termo superclasse é utilizado para designar uma classe superior em relação a suas subclasses. No exemplo dado, Y é superclasse de X.

Coleções de recursos podem ser representadas em RDF pelo uso de *contêineres*, que podem ser de valores alternativos (*rdf:Alt*), de valores não ordenados (*rdf:Bag*) e de valores

seqüenciais (*rdf:Seq*). Exemplificando, o tipo de suporte de um documento pode ser representado por um contêiner de valores alternativos, para a definição dos tipos: em *papel*, em *cdrom* ou em *linha* (tabela 38). Palavras-chaves relativas a um documento podem ser representadas em uma seqüência alfabética ou definida pelo indexador, com o uso do contêiner *rdf:Seq* (tabela 39). Do mesmo modo, tais palavras-chaves podem ser representadas por um contêiner *rdf:Bag* (tabela 40), que não determina uma ordem de descrição.

Tabela 38: contêiner para descrever recursos com valores alternativos

```
...
<rdf:Description
  rdf:about="http://www.cl.df.gov.br/legislacao/lei001">
  <lei:suporte>
    <rdf:Alt>
      <rdf:li>Papel</rdf:li>
      <rdf:li>CD-ROM</rdf:li>
      <rdf:li>Em linha</rdf:li>
    </rdf:Alt>
  </lei:suporte>
...

```

Tabela 39: contêiner para descrever recursos com valores não ordenados

```
...
<rdf:Description
  rdf:about="http://www.cl.df.gov.br/legislacao/lei001">
  <lei:palavrachave>
    <rdf:Seq>
      <rdf:li>Clinica</rdf:li>
      <rdf:li>Direito a saúde</rdf:li>
      <rdf:li>Hospital</rdf:li>
      <rdf:li>Saúde</rdf:li>
    </rdf:Seq>
  </lei:palavrachave>
...

```

Tabela 40: contêiner para descrever recursos com valores seqüenciais

```
...
<rdf:Description
  rdf:about="http://www.cl.df.gov.br/legislacao/lei001">
  <cd:palavrachave>
    <rdf:Bag>
      <rdf:li>Saúde</rdf:li>
      <rdf:li>Clínica</rdf:li>
      <rdf:li>Hospital</rdf:li>
      <rdf:li>Direito a saúde</rdf:li>
    </rdf:Bag>
  </cd:palavrachave>
...

```

Além de *classes* e de *contêineres*, a arquitetura RDF propicia diversas *propriedades* e *atributos* que podem ser utilizadas para a descrição de recursos. As tabelas 41, 42 e 43 apresentam, respectivamente, as classes, propriedades e atributos admitidos pela estruturas RDF e RDFS.

Tabela 41: classes RDF

Nome da classe	Descrição da classe
rdf:Alt	Descreve contêineres de valores alternativos.
rdf:Bag	Descreve contêineres de valores não ordenados
rdf:List	Descreve listas de valores
rdf:Property	Descreve propriedades RDF. É uma subclasse de rdfs:Class.
rdf:Seq	Descreve contêineres de valores sequenciais
rdf:Statement	Utilizada para realizar declarações RDF
rdf:XMLLiteral	É usada para descrever valores literais em XML.
rdfs:Class	É a classe das classes em RDF. Uma classe rdfs:Class pode ser uma instância de rdfs:Class (uma instância de sua própria classe)
rdfs:Container	Descreve contêineres RDF. É superclasse de rdf:Alt, rdf:Bag e rdf:List
rdfs:Datatype	É utilizada para a definição de tipos de dados em RDF. Cada instância de rdfs:Datatype é uma subclasse de rdfs:Literal
rdfs:Literal	É utilizada para a descrição de valores literais, que podem ser alfabéticos ou numéricos. É uma subclasse de rdfs:Resource
rdfs:Resource	Todas as coisas descritas em RDF são recursos e são instâncias da classe rdfs:resource. Todas as outras classes são subclasses desta classe. A classe rdfs:resource é uma instância da classe rdfs:Class.

Fonte: adaptado de W3C (2004a)

Tabela 42: Propriedades RDF

Propriedade	Descrição	Domínio	Extensão
rdf:type	Indica que o recurso (sujeito) é uma instância de uma classe	rdfs:Resource	rdfs:Class
rdfs:subClassOf	Indica que o recurso (sujeito) é uma subclasse de uma classe	rdfs:Class	rdfs:Class
rdfs:subPropertyOf	Indica que o recurso (sujeito) é uma subpropriedade de uma propriedade	rdf:Property	rdf:Property
rdfs:domain	Indica o domínio da propriedade sujeito (ou de um recurso)	rdf:Property	rdfs:Class
rdfs:range	Indica a extensão da propriedade domínio (ou de um recurso)	rdf:Property	rdfs:Class
rdfs:label	Fornecer um rótulo legível por humanos para um recurso (sujeito)	rdfs:Resource	rdfs:Literal
rdfs:comment	Fornecer uma descrição para o recurso (sujeito)	rdfs:Resource	rdfs:Literal
rdfs:member	Indica que um recurso é membro de outro	rdfs:Resource	rdfs:Resource
rdf:first	Indica que o recurso deve ser o primeiro em uma lista RDF.	rdf:List	rdfs:Resource
rdf:rest	Indica os recursos compreendidos em uma lista RDF, apresentados após o primeiro recurso.	rdf:List	rdf:List
rdfs:seeAlso	Propicia uma relação de remissiva (ver também) para o recurso em questão	rdfs:Resource	rdfs:Resource
rdfs:isDefinedBy	Fornecer uma definição para o recurso em questão	rdfs:Resource	rdfs:Resource
rdf:value	Descreve valores estruturados	rdfs:Resource	rdfs:Resource
rdf:subject	Indica o sujeito em uma declaração RDF	rdf:Statement	rdfs:Resource
rdf:predicate	Indica o predicado em uma declaração RDF	rdf:Statement	rdfs:Resource
rdf:object	Indica o objeto em uma declaração RDF	rdf:Statement	rdfs:Resource

Fonte: adaptado de W3C (2004a)

Tabela 43: Atributos RDF

Elemento	Descrição
Rdf:about	Especifica que recurso está sendo descrito
Rdf:Description	Contêiner para a descrição de um recurso
Rdf:resource	Define um recurso para especificar uma propriedade
Rdf:datatype	Define o tipo de dados de um elemento
Rdf:ID	Define um ID (atributo identificador) para um elemento
Rdf:li	Define uma lista
Rdf:_n	Define um nó
Rdf:nodeID	Define o ID de um elemento nó
Rdf:parseType	Define como um elemento deve ser analisado sintaticamente (parsed)
Rdf:RDF	Declara o elemento raiz RDF em um documento

De modo similar ao XMLSchema o RDFSschema foi concebido para fornecer rigor sintático e semântico a um documento e funciona como uma extensão para o RDF, fornecendo a arquitetura para a descrição de classes e propriedades específicas para certa aplicação. A tabela 44.

Tabela 44: exemplo de aplicação de RDFSschema

Código	Descrição
<?xml version="1.0"?>	indicação de início de arquivo
<pre><rdf:RDF xmlns:rdf= "http://www.w3.org/1999/02/22-rdf- syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:base= "http://www.cl.df.gov.br/legislacao"></pre>	declaração de que o arquivo é do tipo RDF; o conteúdo de <i>xmlns:rdf</i> indica o uso do vocabulário permitido pelo RDF; o conteúdo de <i>xmlns:rdfs</i> indica o uso do vocabulário permitido pelo RDFSschema; <i>xml:base</i> , indica o URI dos elementos permitidos em uma estrutura definida pelo autor do arquivo
<pre><rdf:Description rdf:ID="norma"> <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf- schema#Class"/> </rdf:Description></pre>	descreve o elemento norma, como uma classe RDF
<pre><rdf:Description rdf:ID="leicomplementar"> <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf- schema#Class"/> <rdfs:subClassOf rdf:resource="#norma"/> </rdf:Description></pre>	descreve o subelemento leicomplementar como uma subclasse RDF da classe norma
<pre><rdf:Description rdf:ID="leiordinaria"> <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf- schema#Class"/> <rdfs:subClassOf rdf:resource="#norma"/> </rdf:Description></pre>	descreve o subelemento leiordinaria como uma subclasse RDF da classe norma
</rdf:RDF>	indica o final de arquivo

Fonte: adaptado de W3Schools (2004)

Pelo código da tabela 44, um mecanismo de inferência pode deduzir que uma *lei complementar* é um tipo de *norma*, ou que uma *norma* pode ser uma *lei complementar* ou uma *lei ordinária*. O mesmo exemplo pode ser abreviado pelo uso da classe *rdfs:Class*, ao invés de

rdf:Description e pela eliminação da informação *rdf:type*. O código resultante pode ser visto na tabela 45.

Tabela 45: exemplo de aplicação de RDFSschema (simplificado)

Código	Descrição
<code><?xml version="1.0"?></code>	indicação de início de arquivo
<code><rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:base="http://www.cl.df.gov.br/legislacao"></code>	declaração de que o arquivo é do tipo RDF; o conteúdo de <i>xmlns:rdf</i> indica o uso do vocabulário permitido pelo RDF; o conteúdo de <i>xmlns:rdfs</i> indica o uso do vocabulário permitido pelo RDFSschema; <i>xml:base</i> , indica o URI dos elementos permitidos em uma estrutura definida pelo autor do arquivo
<code><rdfs:Class rdf:ID="norma" /></code>	descreve o elemento norma, como uma classe RDF
<code><rdfs:Class rdf:ID="leicomplementar"> <rdfs:subClassOf rdf:resource="#norma"/> </rdfs:Class></code>	descreve o subelemento leicomplementar como uma subclasse RDF da classe norma
<code><rdfs:Class rdf:ID="leiordinaria"> <rdfs:subClassOf rdf:resource="#norma"/> </rdfs:Class></code>	descreve o subelemento leiordinaria como uma subclasse RDF da classe norma
<code></rdf:RDF></code>	indica o final de arquivo

Fonte: adaptado de W3Schools (2004)

Essa representação torna o código mais facilmente legível para humanos, mas em nada difere em termos de significação para um agente inteligente; a diferença básica está na sintaxe, pela redução do número de linhas a serem interpretadas, e não no significado.

2.4.4.3.3 OWL

Como já abordado no capítulo de revisão de literatura, a *Web Ontology Language* (OWL) é uma linguagem de construção de ontologias e que possui suas bases na arquitetura RDF. A OWL destina-se ao processamento inteligente de informações na web e foi projetada para ser interpretada por computadores. Em comparação com a RDF, pode-se dizer que as linguagens são muito parecidas, com a diferença de que a OWL é bem mais robusta em termos de interpretabilidade por máquinas do que RDF, por causa do seu vocabulário mais extenso. W3Schools (2004).

A tabela 46 apresenta a estrutura básica de um documento escrito em OWL, que deve possuir: a) uma marcação de início de arquivo, com a indicação de que é um tipo de XML; b) um cabeçalho de indicação do esquema ou dos esquemas utilizados para a validação sintático-semântica – em geral são usados esquemas RDFS e XMLS, além do próprio RDF; c) um corpo, iniciado pelo elemento raiz *<ontology>*, que indica a localização da ontologia principal, definida

pelo usuário – que do ponto de vista terminológico, pode se pensada como um arquivo em que estão definidas as relações conceituais num sistema de conceitos; d) indicação de final de arquivo, dada pelo fechamento do elemento `</ontology>`.

Tabela 46: Estrutura básica de um documento OWL (sintaxe XML)

Código	Descrição
<code><?xml version="1.0"?></code>	indicação de início de arquivo
<code><!DOCTYPE Ontology [<!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >]></code>	indicação de que o arquivo é uma ontologia que utiliza um vocabulário XMLSchema definido pelo W3C
<code><owlx:Ontology owlx:name="http://www.cl.df.gov.br/leis" xmlns:owlx="http://www.w3.org/2003/05/owl- xml"></code>	elemento raiz, marca o início das declarações sobre as classes; "owlx:name=..." referencia uma ontologia definida pelo autor, de onde serão validados os nomes dos elementos; xmlns:owlx faz referência ao conjunto de elementos definidos pelo W3C para a linguagem OWL.
cabecinhos de elementos, bem como números de classes, e indivíduos	declarações a respeito das classes e propriedades
<code></owlx:Ontology></code>	indicação de final de arquivo.

Fonte: Adaptado de (W3C, 2003c)

Um documento OWL pode ser escrito com o uso de uma sintaxe puramente baseada em XML, como apresentado na tabela 46, ou em uma sintaxe que utiliza RDF e RDFSschema. A indicação do tipo de sintaxe que será usada em um documento é especificada pelo namespace referenciado, bem como pelo prefixo usado – a convenção é o uso do prefixo *owlx* para a sintaxe XML e do prefixo *owl*, para a sintaxe XML/RDF. A tabela 47 indica essas diferentes formas.

Tabela 47: Namespaces e prefixos OWL

Prefixo	Namespace	Descrição
<i>owlx</i>	"http://www.w3.org/2003/05/owl-xml"	indicação da sintaxe XML
<i>owl</i>	"http://www.w3.org/2002/07/owl"	indicação da sintaxe XML/RDF

Como já abordado no capítulo de Revisão de Literatura, na seção que tratou sobre a Web Semântica, conforme o escopo de classes e propriedades a ser utilizado, a OWL pode ser classificada em OWL Lite, OWL DL e OWL Full, que possuem níveis de complexidade crescente, nessa mesma ordem. A indicação da versão da linguagem que se pretende utilizar é

ddada pela referência, por meio de um namespace, ao esquema que a ela corresponde eespecificamente.

Do mesmo modo que em RDF, a linguagem OWL utiliza estruturas de classes pppara a organização dos recursos. Assim, cada classe pode fornecer um mecanismo de abstração pppara o agrupamento de recursos e está associada a um conjunto de objetos individuais, a que se ddenomina *extensão* da classe. Cada objeto individual em uma classe é denominado instância de tal cclasse e possui um significado intensional. Devido a essa concepção, todos os recursos em um dddocumento OWL são descritos como classes, subclasses ou propriedades de classes e subclasses. Isso garante, também, à linguagem, a possibilidade de representar significações com base em eelementos que atuam como funções em algoritmos e que permitem a realização de inferências pppor agentes inteligentes.

A ilustração 51 apresenta a indicação do conjunto de elementos utilizados na liliinguagem OWL.

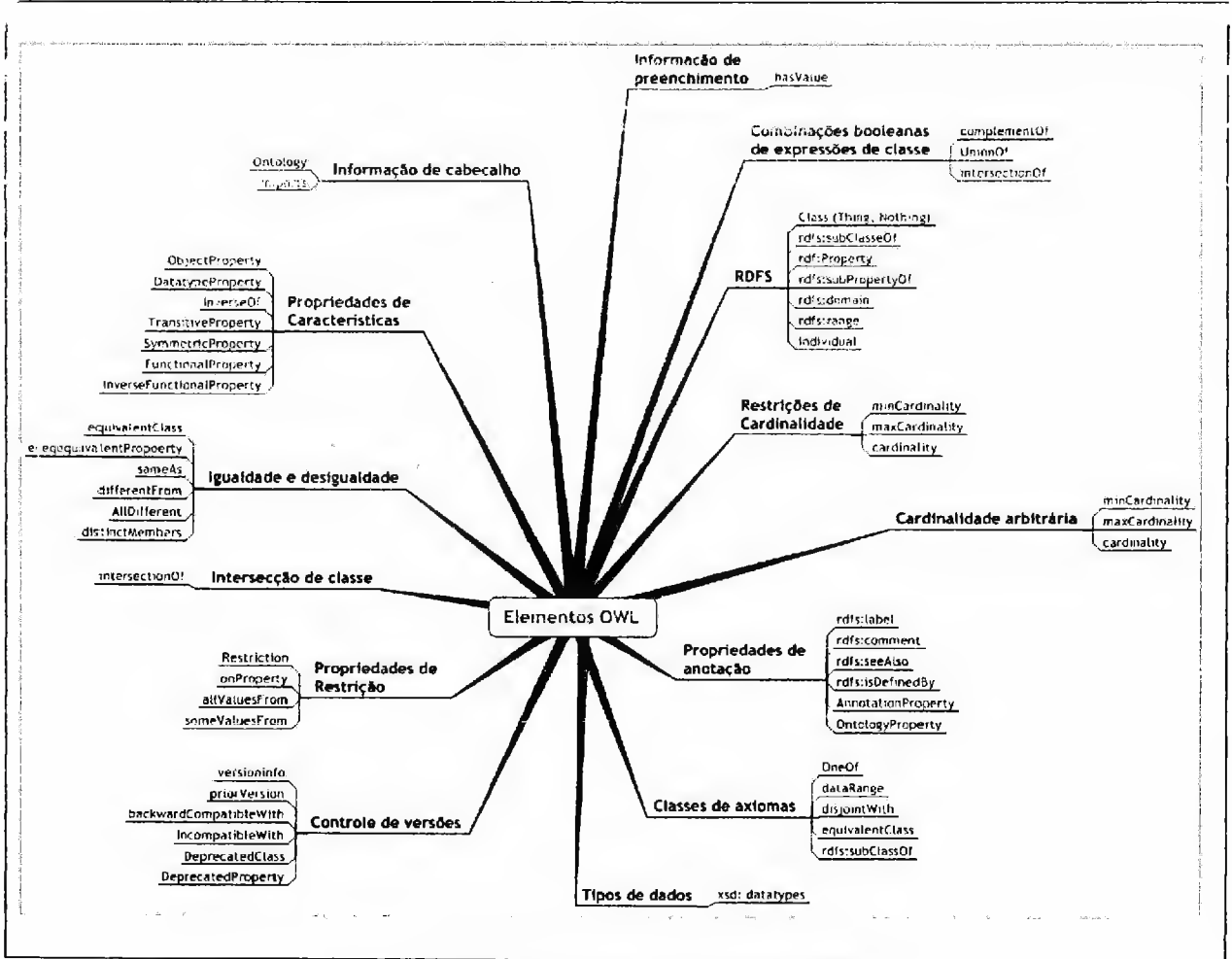


Ilustração 51: Elementos da linguagem OWL

Como se pode observar pela ilustração 51, a OWL possui grande quantidade de elementos, cuja característica principal é serem destinados ao estabelecimento de relações entre classes e subclasses, por meio de propriedades e restrições de propriedades. Entre esses elementos há um conjunto importado das arquiteturas RDF e RDFS, a saber: *Class*, *subClassOf*, *Property*, *subPropertyOf*, *domain*, *range* e *Individual*.

O elemento *Class* define um grupo de objetos individuais que possuem características em comum. Classes podem ser organizadas hierarquicamente por meio do uso do elemento *subClassOf*. Com o elemento *rdfs:subClassOf* pode-se criar hierarquias para a escrita de uma ou mais declarações de que uma classe é subclasse de outra. Isso pode permitir construções como: uma *lei* é uma subclasse de *norma*. De tal modo, um agente inteligente pode deduzir que se um objeto é uma *lei*, então é uma *norma*. O elemento *rdf:Property* pode ser utilizado para estabelecer relacionamentos entre indivíduos ou entre indivíduos e valores de dados. Ex.: Na declaração uma *lei* possui *autor*, os indivíduos *lei* e *autor* podem ser ligados por uma *propriedade temaautor*. O elemento *rdfs:subPropertyOf* é utilizado para indicar que um indivíduo pertence a uma classe de um domínio, pelo fato de possuir uma propriedade que se inscreve em tal classe. O elemento *rdfs:range* limita os valores de uma propriedade em relação a certo domínio. O último elemento do conjunto, *Individual*, refere-se a um objeto individual como uma instância de uma classe e suas propriedades são utilizadas para estabelecer suas relações com outros indivíduos da mesma classe

Os elementos de comparação são: *equivalentClass*, *equivalentProperty*, *sameAs*, *differentFrom* e *AllDifferent*. Duas classes são ditas equivalentes se possuem as mesmas instâncias. O elemento *equivalentClass* pode ser utilizado para criar classes sinônimas. Por exemplo, a classe *Deputado* pode ser definida como *equivalentClass* em relação a *Parlamentar*, dessa forma qualquer agente inteligente poderá concluir que um *Deputado* é também um *Parlamentar* e vice-versa. Propriedades equivalentes são utilizadas para relacionar um indivíduo com outro na mesma classe. Dito de outro modo, o elemento *equivalentProperty* pode ser utilizado para declarar propriedades sinônimas. O elemento *sameAs* é usado quando se deseja especificar que dois indivíduos devem ser tratados como um só. Por exemplo, o indivíduo <Rui Barbosa> é o mesmo que <SenadorRui>. O elemento *differentFrom* é usado para definir que um indivíduo é diferente de outros indivíduos. *AllDifferent* indica que os indivíduos de uma lista são mutuamente diferentes, quando tais indivíduos são comparados entre si.

Entre os elementos que definem propriedades de características, estabelecendo comparações entre elas ou restrições, estão: *inverseOf*, *TransitiveProperty*, *SymmetricProperty*,

FunctionalProperty, *InverseFunctionalProperty*, *allValuesFrom* e *someValuesFrom*. O elemento *inverseOf* indica que uma propriedade é o inverso de outra. Por exemplo, se a propriedade *autorde* é inversa à propriedade *foiescritapor*, diante da declaração <Jorge Amado> *autorde* <Lei nº 1> um agente pode deduzir que <Lei nº 1> *foiescritapor* <Jorge Amado>. *TransitiveProperty* é utilizada para indicar a associatividade entre três ou mais propriedades. Por exemplo, se um par (x,y) é uma instância de P, e o par (y,z) é uma instância de P, então o par (x,z) é, também, uma instância de P. *SymmetricProperty* é utilizada para indicar a comutação entre duas propriedades. Por exemplo, se o (x,y) é uma instância de P, então o par (y,x) é também uma instância de P. *FunctionalProperty* indica um valor único para uma propriedade. Uma propriedade declarada como *FunctionalProperty* não pode ter mais de um valor para cada indivíduo em uma classe. Entretanto, pode ter valores nulos para um indivíduo. *InverseFunctionalProperty* indica um valor unívoco para uma propriedade, isto é, se uma propriedade como *possuiCPF* for declarada do tipo *InverseFunctionalProperty*, não poderá haver duas instâncias individuais que possuam CPFs idênticos. O elemento *allValuesFrom* indica restrição de valores de propriedades em relação a uma classe, isto é, essa propriedade, em uma classe particular, possui uma extensão restrita. Dito de outro modo, todos os valores admitidos para essa propriedade vêm de uma classe definida. Diferentemente de *allValuesFrom*, uma propriedade definida como *someValuesFrom* não restringe todos os valores admissíveis a uma mesma classe. Isto significa que alguns dos valores admitidos podem advir de uma classe determinada.

Os elementos de restrição de cardinalidade da linguagem são: *minCardinality*, *maxCardinality* e *cardinality*. O elemento *minCardinality* indica os valores mínimos (0 ou 1) que podem ser assumidos por uma propriedade. Em outras palavras, indica se o valor de uma propriedade é no mínimo 0 ou no mínimo 1. O elemento *maxCardinality* indica os valores máximos (0 ou 1) que podem ser assumidos por uma propriedade. Em outras palavras, indica se o valor de uma propriedade é no máximo 0 ou no máximo 1. Finalmente, *cardinality* indica um valor exato que pode ser assumido por uma propriedade. Nas versões OWL DL e OWL Full, esses elementos são expandidos, e permitem o uso de números inteiros positivos para a indicação de relações de cardinalidade. Enquanto em OWL Lite as cardinalidades podem ser indicadas por *ao menos*, *no máximo* e *exatamente* 0 ou 1, nas versões DL e Full, outros valores podem ser indicados.

Além desses elementos, outros elementos que podem ser destacados no escopo da linguagem OWL são aqueles que definem relações de intersecção (*intersectionOf*), enumeração (*oneOf*), valoração (*hasValue*), disjunção (*disjointWith*) e união (*unionOf*). A formulação de declarações que denotam a intersecção entre determinadas classes é feita com o elemento

intersectionOf e indica que certas instâncias de uma classe estão em outra e vice-versa. Combinações entre elementos como *unionOf*, *complementOf* e *intersectionOf* podem ser utilizadas, nas linguagens OWL DF e OWL Full, para a construção de frases booleanas. O elemento *oneOf* pode ser utilizado para descrever uma classe por meio da enumeração dos seus indivíduos e indica que os membros de uma classe são exatamente o conjunto de indivíduos enumerados, nem mais, nem menos. O elemento *hasValue* pode ser utilizado para declarar um valor específico para uma propriedade. Por exemplo, *hasValue* pode ser usado para indicar que, para um indivíduo da classe *Automovel*, o valor esperado para o valor da propriedade *propulsao* é “a motor”. Classes podem ser declaradas distintas umas das outras por meio do elemento *disjointWith*. Por exemplo, *Homem* e *Mulher* podem ser declarados como classes distintas entre si. Assim, um mecanismo de inferência pode identificar que se um indivíduo é uma instância de uma classe, não poderá ser instância da outra. Em outras palavras, se *Cecilia Meireles* é uma instância de *Mulher*, então não é uma instância de *Homem*.

A ilustração 52 apresenta uma esquematização dos elementos OWL segundo os aspectos sintáticos. O elemento, raiz é o elemento *ontology*, ao qual estão subordinados outros grupos de elementos que compõem a estrutura da linguagem. Os elementos de cabeçalho são aqueles utilizados para descrever o documento OWL em relação a características como versão, compatibilidade com outras tecnologias, bibliotecas que estão sendo importadas ou outros comentários que se julgue necessários ao arquivo.

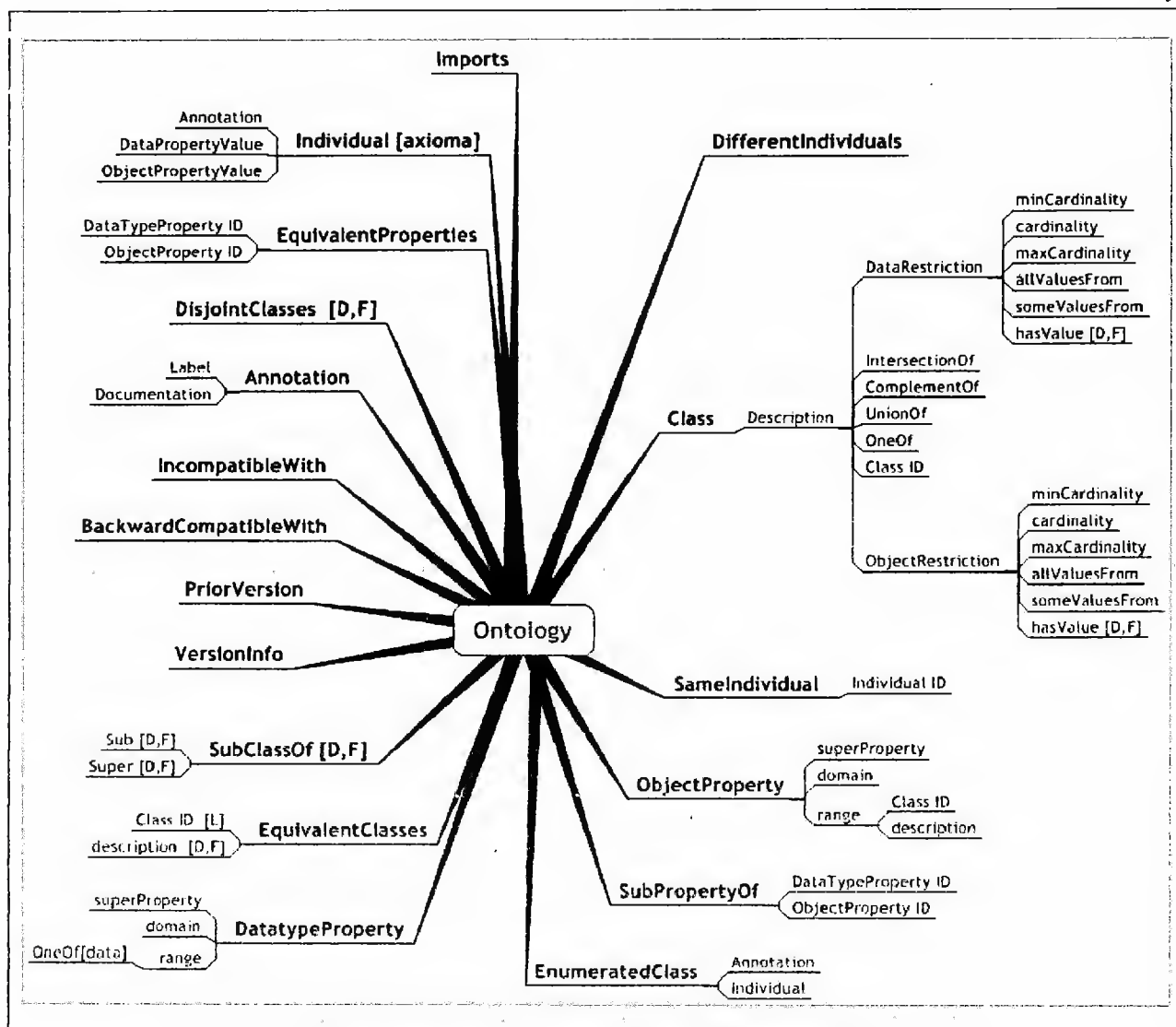


Ilustração 52: Conjunto de elementos da linguagem OWL

Fonte: (W3c, 2003c)

2.5 Informação Legislativa

A documentação jurídica pode ser dividida em três partes: legislação, jurisprudência e doutrina. A legislação pode ser definida como o conjunto de atos normativos emanados do Poder Executivo ou do Legislativo e compreende as seguintes espécies: leis, decretos, decretos-leis, atos, resoluções, portarias, projetos de leis ou de decretos legislativos ou de resoluções legislativas, ordens internas, circulares, exposições de motivos, entre outros. A jurisprudência é o conjunto de decisões judiciais referentes a casos semelhantes e que não tenham ocorrido de forma isolada, mas sim de maneira uniforme e constante. A jurisprudência compreende as seguintes espécies de documentos: acórdãos, pareceres, recursos, decisões, entre

outros. A doutrina é o conjunto de publicações teóricas a respeito de um determinado assunto e reúne idéias, opiniões, conceitos que os autores expõem e defendem no estudo e no ensino do direito. Tais conceitos servem de sustentação para teorias e interpretações da ciência jurídica e têm força de verdade científica. A doutrina constitui uma das fontes do Direito, suprimindo a lei e o costume, como norma interpretativa que a jurisprudência tende a seguir na aplicação de uma lei (ATIENZA, 1979; PEREIRA, 2004; HOUAISS, 2004). O objeto de estudo desta pesquisa compreende apenas a documentação legislativa, especificamente a emanada do poder legislativo.

2.5.1 O Poder Legislativo no Brasil

Conforme a doutrina de Montesquieu (apud CUNHA, 1987), o Estado integra-se por meio de três poderes: “o de fazer leis, o de executar as resoluções públicas e o de julgar os crimes ou desavenças entre pessoas específicas”. Desses poderes, o Poder Legislativo é o que detém a função primordial da produção das leis, atribuindo aos indivíduos determinados comportamentos, no intuito de adequarem-se às normas legislativas por ele produzidas (PEREIRA, 2000). Lei é uma palavra que origina-se do latim: *lex* ou *legis*, do verbo *legere*, que significa ler e escolher (o texto da lei era lido ao povo pelo magistrado romano, durante comícios, para a sua aprovação). De acordo com Amaral (2003), Montesquieu define as leis, no seu significado mais lato, como relações necessárias que derivam da natureza das coisas. Há, portanto, uma razão primitiva para a sua existência, e elas leis são as relações que se encontram entre os vários seres.

Assim, cabe ao Poder Legislativo a função legislativa, que corresponde à edição de normas gerais e abstratas (as leis), seja para regular os demais atos estatais, seja para regular a vida dos cidadãos. Pereira (2000), relata que o poder legislativo moderno tem suas origens na Inglaterra da idade média, quando representantes da nobreza, da igreja e do povo procuraram limitar a autoridade absoluta dos reis.

O Poder Legislativo pode ser classificado em Poder Legislativo Originário e Derivado. O Poder Legislativo Originário refere-se às leis básicas fundamentais da organização estatal e é exercido pela Assembléia Constituinte. Uma vez instituído o Estado, a função legislativa passa a ser exercida pelo Poder Legislativo Derivado, que produz as normas válidas para toda a sociedade e para o Estado, à luz das normas fundamentais.

De acordo com o país, o Poder Legislativo pode receber diferentes denominações. Exemplificando, o órgão legislativo é referido como Estado Geral, na França; Estamento, na Sardenha; Parlamento, na Sicília; e Corte, na Espanha (PEREIRA, 2000). Em relação à organização legislativa, são diversas também as fórmulas em que se organiza o poder.

legislativo nos diferentes países. As fórmulas mais comuns são o unicameralismo e o bicameralismo. A tabela 19 apresenta, com base no texto de Pereira (2000), os argumentos em favor de cada tipo de cameralismo. O sistema bicameral, geralmente utilizado por países com menor contingente populacional e menor extensão territorial, possui uma única Câmara Legislativa. O sistema bicameral, em que há duas câmaras, pode ser classificado, conforme a natureza dessas câmaras em:

a) aristocrático, no qual uma câmara representa o povo e a outra, de menor influência jurídica, mas com maior força política, representa a nobreza; esse tipo de bicameralismo foi mais freqüente no século XIX (PEREIRA, 2000);

b) federal, no qual uma câmara representa o povo (câmara baixa) e a outra representa os estados (câmara alta);

c) sistemático, no qual a segunda câmara tem as funções moderadoras de “refrear” os impulsos da câmara do povo; um exemplo desse tipo de bicameralismo é o poder legislativo francês, no qual a função moderadora é exercida pelo senado;

d) clássico, no qual a Segunda Câmara tem funções apenas de assessoria.

Segundo Pereira (2000), a técnica de organização legislativa adotada no Brasil remonta à constituição de 1824, que adotou o sistema bicameral, havendo, desde então, o processo de revisão legislativa entre as Casas do Congresso Nacional. Essa mesma orientação é mantida no texto da Constituição de 1988 que expressa, no Art. 44, que o Poder Legislativo é exercido pelo Congresso Nacional, composto por suas duas Casas: a Câmara dos Deputados e o Senado Federal. O autor relata, ainda, que o processo de revisão legislativa entre essas instituições justifica-se pelos seguintes argumentos: a) o trabalho revisado pelas Casas Legislativas dá maior segurança jurídica à sociedade, uma vez que corresponde a um trabalho mais bem elaborado, no que tange aos aspectos jurídicos, constitucionais e políticos; b) cabe, tanto ao Senado, quanto à Câmara dos Deputados, desempenhar um papel de revisão das matérias originárias da outra Casa, o que proporciona maior ponderação e perfeição ao trabalho legislativo.

Tabela 48: Argumentos em defesa do unicameralismo ou do bicameralismo

Argumentos em defesa do unicameralismo	Argumentos em defesa do bicameralismo
sendo a lei a expressão da vontade geral, deve receber expressão formal única	o sistema representa melhor e mais completamente a opinião pública
uma câmara legislativa atua com maior rapidez	a dualidade é garantida, ao contrário da possibilidade de despotismo de uma única assembleia
uma câmara legislativa é mais econômica	duas câmaras podem proporcionar maior ponderação e perfeição ao trabalho legislativo
uma câmara legislativa é mais progressista, democrática e popular	o sistema bicameral mitiga os conflitos entre os poderes executivo e legislativo
no sistema bicameral, se apenas um partido dominar as duas casas, o que é feito em uma repetir-se-á na outra;	o bicameralismo consolida a opinião parlamentar
no sistema bicameral, se cada câmara for dominada por um partido diferente, poderá haver conflitos insolúveis entre elas	a segunda câmara pondera para evitar reformas impetuosas e arriscadas
no sistema bicameral, a segunda câmara poderá ter um caráter muito conservador ou reacionário	a segunda câmara pode abrigar os interesses corporativos e econômicos

Fonte: Adaptado de Pereira (2000)

Ao Congresso Nacional cabe deliberar sobre matérias de competência da União, como: sistema tributário, arrecadação e distribuição de rendas; plano plurianual, diretrizes orçamentárias, orçamento anual, operações de crédito, dívida pública e emissões de curso forçado; fixação e modificação do efetivo das Forças Armadas; planos e programas nacionais, regionais e setoriais de desenvolvimento; limites do território nacional, espaço aéreo e marítimo e bens do domínio da União; incorporação, subdivisão ou desmembramento de áreas de Territórios ou Estados, ouvidas as respectivas Assembleias Legislativas; transferência temporária da sede do Governo Federal; concessão de anistia; organização administrativa, judiciária, do Ministério Público e da Defensoria Pública da União e dos Territórios e organização judiciária, do Ministério Público e da Defensoria Pública do Distrito Federal; criação, transformação e extinção de cargos, empregos e funções públicas; criação, estruturação e atribuições dos Ministérios e órgãos da administração pública; telecomunicações e radiodifusão; matéria financeira, cambial e monetária, instituições financeiras e suas operações; moeda, seus limites de emissão, e montante da dívida mobiliária federal (BRASIL, 1988).

À Câmara dos Deputados, relacionam-se matérias como: autorizar a instauração de processo contra o Presidente da República, o Vice-Presidente e seus Ministros; tomar as contas do Presidente da República quando este não o fez no prazo constitucional; elaborar o seu regimento interno e deliberar sobre sua organização, funcionamento, serviços e assuntos correlatos; e eleger membros ao Conselho da República.

Ao Senado Federal, foram dispostas matérias cujos conteúdos se referem às questões ligadas à federação; questões ligadas à função jurisdicional, no caso de processo e

juízo do Presidente da República, Vice-Presidente e Ministros de Estado e dos Ministros do Supremo Tribunal Federal, nos crimes de responsabilidade; questões ligadas às funções executivas, especialmente na aprovação de chefes de missão diplomática e outros temas (PEREIRA, 2000).

Aos Estados da Federação e ao Distrito Federal é concedido o direito de regerem-se por suas próprias constituições e leis, observados os princípios da Constituição Federal.

Para materialização das competências dessas instituições, a Constituição Federal prevê as seguintes espécies normativas: a) emenda à Constituição; b) leis complementares; c) leis ordinárias; d) leis delegadas; e) medidas provisórias; f) decretos legislativos e; g) resoluções. Essa classificação baseia-se nos critérios de iniciativa e de hierarquia.

2.5.2 Poder Legislativo no Distrito Federal

Para Szklarowsky (2003), a autonomia do Distrito Federal tem sido discutida desde os tempos do Brasil Império. O autor relata que a Constituição Imperial de 1824 e a Lei 16 de 1834, faziam menção à Capital do Império como uma instituição política independente, isto é, a autoridade legislativa da província em que a Corte se localizasse não teria poderes sobre a Corte e nem sobre o seu Município. O mesmo postulado manteve-se posteriormente, com o advento da República, quando o antigo Município Imperial passa a constituir o Distrito Federal, continuando a ser a capital da União, mas sem pertencer ou subjugar-se a nenhum estado. Apesar de ser equiparado aos Estados e Territórios – uma vez que, juntamente com ambos, fazia parte da união indissolúvel dos Estados Unidos do Brasil –, com a reforma constitucional de 1937, o Distrito Federal passou a ser administrado por um prefeito nomeado pelo Presidente da República, com a aprovação do Conselho Federal. Nove anos mais tarde, com a Constituição de 1946, o Distrito Federal ganhou autonomia relativa: recebeu o direito do voto direto para eleger os membros de uma Câmara de Vereadores; mas seu prefeito continuava sendo indicado pelo Presidente da República, com a aprovação do Senado Federal. Em 1956, com a emenda constitucional nº. 02, o Distrito Federal ganhou o direito de eleger, pelo voto direto, o seu prefeito, como já o fazia com os vereadores. Com a transferência da capital para o Planalto Central, nova emenda à constituição foi aprovada e o cenário político do Distrito Federal modificou-se novamente: acabaram as eleições diretas para prefeito, passando este a ser nomeado pelo Presidente da República, com a aprovação do Senado Federal; e, em contrapartida, previu-se a eleição de representantes para o Senado Federal, bem como para a Câmara dos Deputados; criou-se, também, a Câmara do Distrito Federal. A data das eleições, no entanto, não foi estipulada, motivo pelo qual o Distrito Federal ficou sem seus representantes antes do Golpe de

1964 e durante todo o Governo Militar. Em tempos de ditadura militar, a Constituição de 1967, retirou do Distrito Federal a autonomia, passando suas funções legislativas a serem exercidas pelo Senado Federal. Instituiu-se novamente o sistema de nomeação do Prefeito pelo Presidente da República, com anuência do Senado. Com a Constituição de 1969, o Prefeito do Distrito Federal recebeu o status de Governador. Essa Constituição criou, também, a Comissão do Distrito Federal, que atuava no Senado Federal nas matérias legislativas relativas ao DF.

Com a abertura política originária do movimento da Nova República, ocorrido na primeira metade da década de 80, diversas entidades civis, representações políticas, associações e organizações passaram a reivindicar a autonomia política do Distrito Federal e finalmente, a partir de 1986, Brasília teve assegurado o direito de eleger seus representantes na Câmara dos Deputados e no Senado Federal (ROCHA, 2001). O instrumento normativo que deu origem a essa autonomia foi a Emenda Constitucional nº. 25, de maio de 1985, que nos seus artigos 39 a 41 instituiu a representação política para o DF. A emenda foi posteriormente publicada na forma de lei, recebendo o nº. 7376/85 (GOMES, 1995).

A Constituição de 1988, avançando nessa direção mais democrática, conferiu ao Distrito Federal autonomia político-administrativa, na mesma medida que a possuem os estados e territórios. A partir dessa Carta, ao Distrito Federal foi dado o direito de elaborar sua Lei Orgânica, bem como de eleger, pelo voto direto, seus representantes na Câmara Legislativa e no Governo do Distrito Federal. Brasília passou, então, a ser a capital da República e a sede do Governo do Distrito Federal.

Em razão da natureza jurídica do Distrito Federal, que possui características de Estado e de Município, a Câmara Legislativa acaba conjugando atribuições de Assembléia Legislativa (órgão legislativo dos Estados) e de Câmara Municipal (órgão legislativo dos municípios).

A Lei Orgânica do Distrito Federal, promulgada em 1993, consolidou a divisão dos poderes executivo e legislativo, e concedeu à Câmara Legislativa do DF a responsabilidade de organizar e de exercer o poder legislativo nesta unidade federativa. Dessa forma, cabe à Câmara Legislativa o estudo e a proposição de leis sobre todas as matérias de competência do Distrito Federal, especialmente sobre educação; saúde; previdência; habitação; cultura; ensino; desporto; segurança pública; uso do solo; serviços públicos no Distrito Federal; proteção aos deficientes, às crianças, jovens e idosos; e, ainda, sobre sistema de emprego. São também competências da Câmara Legislativa a fiscalização contábil, financeira, orçamentária, operacional e patrimonial do Distrito Federal e o acompanhamento de planos e programas governamentais, em especial

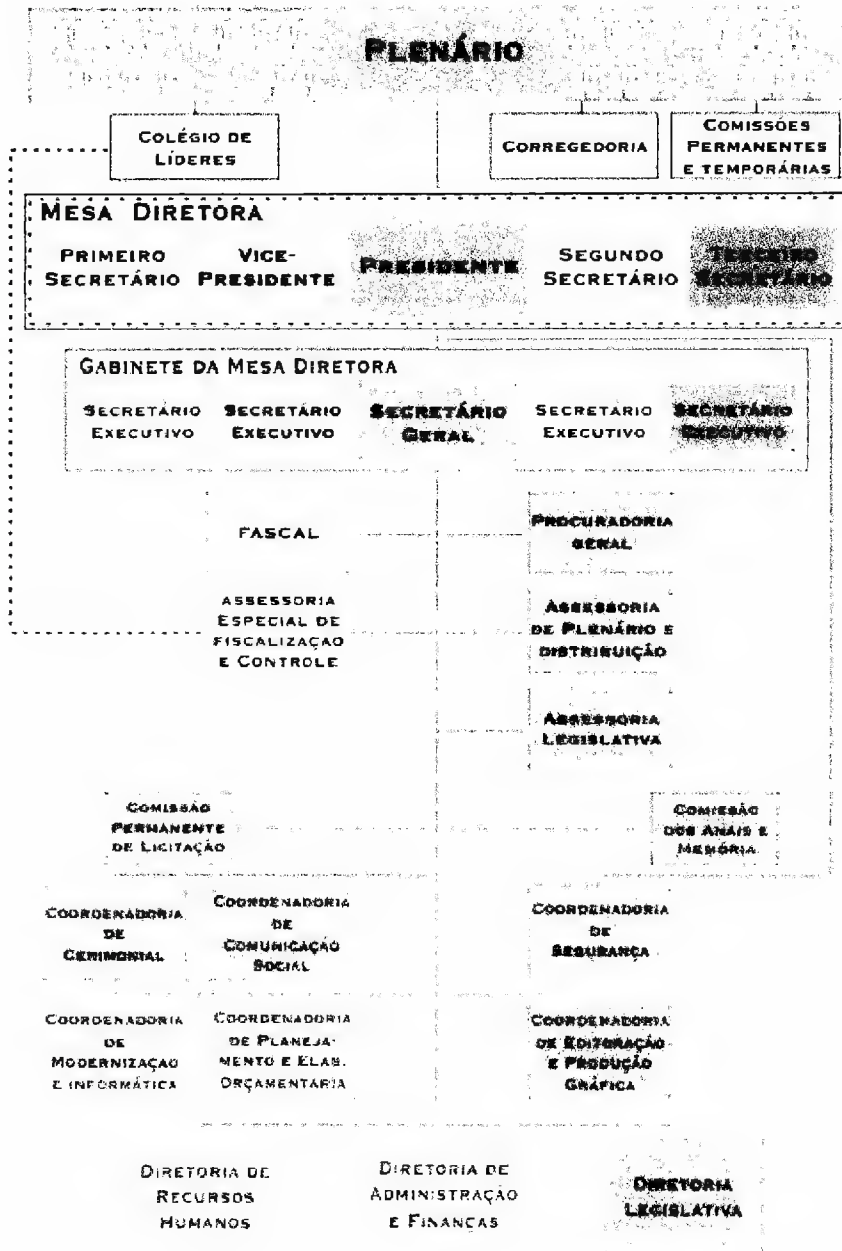
aqueles relacionados com defesa do consumidor; política industrial; política de incentivo à agropecuária; política hídrica e minerária; política de meio ambiente; política de desenvolvimento do turismo; política econômica; estudos; pesquisas e programas de desenvolvimento da ciência e tecnologia; energia; telecomunicações e informática; defesa civil; defesa dos direitos individuais e coletivos e, finalmente, direitos humanos.

2.5.2.1 Estrutura da Câmara Legislativa do DF

A Câmara Legislativa do Distrito Federal é composta por 24 deputados, eleitos em intervalos de quatro anos. Seu órgão diretor colegiado é a Mesa Diretora, composta do Presidente, Vice-Presidente, Primeiro, Segundo e Terceiro Secretários, bem como de três Suplentes de Secretário. O Colégio de Líderes é constituído pelos Líderes dos Partidos Políticos, dos Blocos Parlamentares e do Governo. À Corregedoria da Câmara Legislativa é dado zelar pelo decoro parlamentar, pela a ordem e pela disciplina no âmbito da Casa, e apurar infrações disciplinares ao Código de Ética e Decoro Parlamentar. As Comissões Permanentes possuem caráter técnico-legislativo ou especializado, e têm por finalidade apreciar os assuntos e proposições submetidos ao seu exame e sobre eles emitir parecer. A essas comissões cabe, também, no âmbito de seu respectivo campo temático e áreas de atuação, exercer o acompanhamento de planos e programas governamentais e o controle dos atos dos Poderes Executivo e Legislativo, bem como exercer a fiscalização orçamentária do Distrito Federal. As Comissões Temporárias, são aquelas criadas para apreciarem um assunto em particular; essas comissões extinguem-se ao término da legislatura, ou antes dele, quando tiverem alcançado seu objetivo ou expirar-se o prazo de sua existência. A ilustração 53 apresenta o organograma da Câmara Legislativa.

2.5.2.2 Processo Legislativo

O exercício da atividade legislativa materializa-se com a elaboração de espécies normativas - atos legislativos - que formalizam a ação parlamentar e consolidam a iniciativa política (DISTRITO FEDERAL, 1996). A construção dessas leis implica um processo sistemático com fases articuladas e interdependentes denominado Processo Legislativo. No âmbito do Distrito Federal, as espécies normativas que compõem o processo legislativo são: emendas à Lei Orgânica, Leis Complementares, Leis Ordinárias, Decretos Legislativos e Resoluções.



Situação em fevereiro de 2001

Ilustração 53: Estrutura Organizacional da CLDF.

Fonte: www.cl.df.gov.br

Em relação às leis (ilustração 54), o processo de legislativo compreende as seguintes etapas:

- apresentação do projeto;
- tramitação (submissão à análise das comissões temáticas);
- parecer e emendas;
- submissão para aprovação ou rejeição pelo plenário;

e) encaminhamento ao poder executivo (para sanção ou veto);

f) caso haja o veto total ou parcial, o projeto retorna à Câmara Legislativa (para manutenção ou derrubada do veto);

g) promulgação.

As demais normas, isto é, os decretos legislativos, as resoluções e as propostas de emendas à Lei Orgânica, tramitam somente no âmbito da própria Câmara Legislativa.

Os projetos de Lei Complementar, que tratam apenas de assuntos determinados pela Lei Orgânica do Distrito Federal, podem ser iniciados por parlamentares, pelas comissões, pelo Governador do Distrito Federal, pelo Tribunal de Contas do Distrito Federal ou por qualquer cidadão. Podem ser objeto de Leis Complementares: organização do TCDF; estatuto dos servidores públicos civis do DF; organização da procuradoria geral do DF; sistema tributário do DF; sistema de educação do DF; plano diretor de ordenamento territorial do DF (PDOT); atribuições do vice-governador do DF; organização da previdência dos servidores públicos do DF (DISTRITO FEDERAL, 1996).

Os Projetos de Leis Ordinárias destinam-se a regular matérias de competência do Poder Legislativo, com a sanção do Governador do DF. Sua proposição pode ser manifestada pelas mesmas personalidades que podem iniciar um projeto de Lei Complementar. Leis Ordinárias tratam sobre matérias como: educação, saúde, previdência, habitação, cultura, ensino, desporto, segurança pública, regiões administrativas, serviços públicos, imóveis do Governo do Distrito Federal, entre outras.

Os Projetos de Decretos Legislativos destinam-se a regular matérias de caráter privativo da Câmara Legislativa, sem a sanção do Governador do Distrito Federal. Um Decreto Legislativo é promulgado pela Mesa Diretora da Câmara Legislativa e pode versar sobre: sustação de atos do GDF; aprovação de contas do executivo; autorização para o governador e o vice viajarem por mais de quinze dias; fixar a remuneração do governador, vice e deputados; concessão de títulos honorários (DISTRITO FEDERAL, 1996).

O Projeto de Resolução destina-se a regular, com eficácia de lei, matérias de competência da Câmara Legislativa, sem a sanção do Governador do Distrito Federal. Apresentado pela Mesa Diretora, por um ou mais deputados ou por uma comissão, ao contrário dos decretos legislativos, seus efeitos são apenas internos à Câmara. Trata de matérias de caráter processual, legislativo ou administrativo: perda de mandato de deputado, criação ou conclusão de Comissão Parlamentar de Inquérito e outros assuntos administrativos.

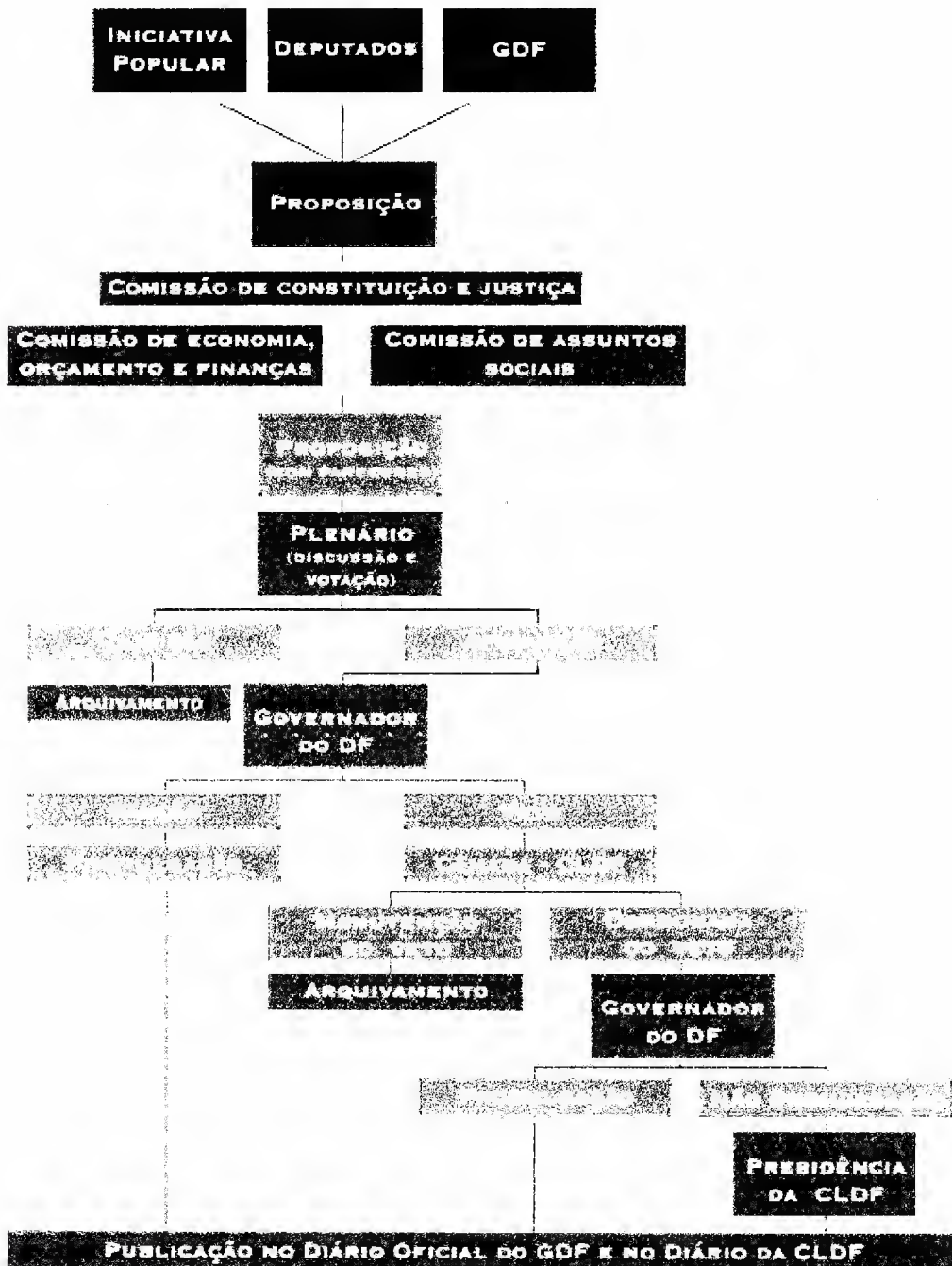


Ilustração 54: Fluxograma do Processo Legislativo na CLDF.

Fonte: www.cl.df.gov.br

Independente de sua espécie normativa, é importante que o projeto de ato legislativo seja redigido de forma sistematizada, de modo que se possa garantir a coerência e a harmonia interna de seus dispositivos, bem como a sua adequada inserção no ordenamento jurídico. A tabela 49 dada a seguir relaciona os elementos constituintes de um projeto:

Tabela 49: Estrutura dos projetos de atos legislativos

Item	Características ou Elementos constitutivos
I - cabeçalho ou preâmbulo	o cabeçalho compõe-se de epígrafe, autoria e ementa; epígrafe: tipo da norma + número + data (Ex. Projeto de Lei nº. de...) autoria: (nome do autor do projeto) ementa: resumo do projeto; quando tiver detalhamento de mais de um objetivo, usa-se "e dá outras providências"
II - fórmula de promulgação	a) identifica o órgão legiferante b) pode usar os verbos: decreta (decretos de lei, decretos legislativos), resolve (resoluções), promulga (emendas à lei orgânica)
III - texto ou corpo	é composto por artigos, que podem se desdobrar em parágrafos e incisos. O texto do artigo é denominado <i>caput</i> ; parágrafos e incisos podem se desdobrar em alíneas; alíneas podem se desdobrar em itens, identificados por números. o parágrafo é o complemento explicativo ou restritivo do <i>caput</i> do artigo; seu assunto depende do assunto tratado no artigo; inciso é o complemento discriminativo do <i>caput</i> do artigo, destina-se à enumeração; artigos pertinentes ao mesmo assunto podem ser agrupados em seções e estas em capítulos; recebe o nome de título o conjunto de capítulos; ao conjunto de títulos, denomina-se livro.
IV - cláusula de vigência	É o artigo que dispõe sobre a entrada em vigor do ato legislativo: "esta lei entra em vigor na data de sua publicação"
V - cláusula de revogação	é o artigo que dispõe sobre a revogação de dispositivos até então vigentes; a revogação pode ser geral ou específica; por exemplo: revogação geral: "ficam revogadas as disposições em contrário" revogação específica: ficam revogados os arts... da(lei, decreto, etc...)
VI - justificção	é o texto que acompanha o projeto e no qual o autor justifica a necessidade da proposição.
VII - fecho	compreende: a) local b) data c) nome do autor (autor do ato de publicação)

Fonte: adaptado de Distrito Federal (1996)

2.6 Sistematização da Fundamentação Teórica e Metodológica

Na concepção dos diversos autores, Língua Geral é definida como um conjunto de palavras ou expressões que pertencem à linguagem cotidiana e que são combinadas logicamente para a formação do discurso, e que não se referem a uma atividade humana especializada (RONDEAU, 1984; WRIGHT; BUDIN, 1997). Em relação ao conceito de Língua de Especialidade compreende apenas o conjunto de termos atribuídos a conceitos utilizados em dado campo ou domínio do conhecimento e pode receber as denominações: *língua de especialidade*, *linguagem especial*, *linguagem especializada*, *linguagem especialista* ou *linguagem para propósitos especiais*. Tais línguas se inserem em universos que alguns autores (RONDEAU, 1984; ARNTZ; PICHT, 1995)

definem como zonas de linguagem ou como níveis de especialização. Rondeau (1984) assim as define: zona mediana – mais próxima da língua geral; zona central – língua de especialidade; e zona ultra-especializada – apenas termos muito especializados da língua de especialidade. Já Arntz e Picht consideram cinco níveis (de A a E), cujas características variam, conforme: grau de abstração; representação lingüística; área ou âmbito e relações entre emissor e receptor. Em todas as abordagens, os autores consultados mencionam vocabulários restritos e em que há a valorização das funções cognitiva e referencial, cuja intenção é a comunicação não ambígua.

Os autores consultados apresentam diferentes visões a respeito da definição de conceito. Alguns autores e normas, embasados na argumentação de que na realidade conceitos não existem fora da mente humana, apresentam abordagens em que o conceito é visto como *unidade do pensamento*. Nessa ótica, o homem vive em um mundo de objetos materiais ou imateriais que existem mais ou menos independentemente dele como um ser pensante. Tais objetos são designados por nomes próprios ou por termos que são imediatamente atribuídos a conceitos individuais, isto é, a representações mentais (ver tabela 50). Um ponto de vista interessante é o de Sager, que em sua obra, devido à diversidade de definições e tendo em vista os propósitos aplicados da Terminologia, preferiu deixar a expressão por definir. Embora também conceba conceito como uma *unidade do pensamento*, esse autor o considera uma primitiva axiomática, a partir da qual se pode estruturar o conhecimento e expressar percepções sobre o mundo.

Outros autores citados no capítulo de Revisão de Literatura, como Dahlberg (1978), Krieger e Finatto (2004), Gomes e Campos (1990), argumentam em favor de outra abordagem em que o conceito é visto como *unidade do conhecimento*. Para Gomes e Campos (1990) há uma contradição na definição estabelecida pela norma ISO 1087, que define conceito como uma “*unidade de pensamento*”. Para as autoras, ratificando o pensamento de Dahlberg (1978), melhor expressão seria “*unidade de conhecimento*”, uma vez “*pensamento*” é uma operação mental individual e que, portanto, não há como uniformizá-la – unidade é o que é comum a todos e o sentido de “*unidade de pensamento*” seria inadequado. Por outro lado, já que o conhecimento pressupõe uma compreensão comum de um dado objeto observável, seria cabível a expressão “*unidade do conhecimento*”, compreendendo afirmações verificáveis acerca de dado item de referência (conceito material ou imaterial) representado por uma forma verbal (termo) (Dahlberg, 1978).

Tabela 50: Sistematização dos fundamentos teóricos relativos a conceito

Autor	Definição	Características/Observações	Fonte
Rondeau	representação abstrata composta por um conjunto de traços comuns essenciais a um grupo de entidades (objetos ou idéias) e obtida pela extração das características individuais de tais entidades	- a Terminologia analisa os conceitos partindo de um significado para encontrar a melhor denominação que o represente (orientação contrária à da lingüística).	(RONDEAU, 1984, p. 24)
Wright & Budin			(WRIGHT; BUDIN 1997, p. 330)
Felber	representação mental, por abstração, de um objeto individual ou de um conjunto de objetos individuais que possuem características em comum. Pode representar qualidades (expressas por nomes), ações (expressas por verbos ou nomes), localizações, situações ou relações (expressas por advérbios, preposições conjunções ou nomes)	<ul style="list-style-type: none"> - um conceito, é um elemento do pensamento e necessita de um símbolo lingüístico (termo) para ser expresso ou comunicado; - o conjunto das características de um conceito constitui a sua intensão – no caso da agregação de objetos individuais, tal extensão é denominada classe; - o conjunto de todos os conceitos subordinados a um conceito dado constitui a sua extensão; 	(FELBER, 1984)
Picht	unidade do pensamento que serve para o reconhecimento dos objetos, para a comunicação sobre os objetos e para a organização mental de tais objetos	<ul style="list-style-type: none"> - a totalidade das características que permitem uma síntese mental de objetos individuais e a delimitação mútua dos conceitos constitui a compreensão de um conceito (intensão); - a extensão de um conceito é a totalidade de objetos individuais que possuem todas as suas características; - classe é definida como a totalidade de objetos que correspondem a um conceito; - não é possível a modificação do número de características de um conceito em um dado momento sem, ao mesmo tempo, modificar-se o próprio conceito; 	(ARNTZ; PICTH, 1995).
Cabré			

Autor	Definição	Características/Observações	Fonte
Sager	construto mental – abstração, ou qualquer unidade do pensamento – usada na classificação de objetos individuais do mundo interior ou exterior.	<ul style="list-style-type: none"> - um conceito é uma primitiva axiomática; - objetos podem ser campos do conhecimento ou atividade humana, coisas, suas propriedades, qualidades, fenômenos, ou outros tipos de objetos, são representados por meio de conceitos; - o conceito é uma unidade do pensamento e é utilizado para estruturar o conhecimento e a percepção do mundo circundante; - o conceito necessita ser expresso ou explicitado, uma vez que corresponde a uma referência a atividades mentais, seja como um resultado ou como um processo; - a soma das características utilizadas para a descrição de um conceito é chamada de intensão; - conceitos são organizados em classes tais como entidades físicas, métodos científicos de análise e descrição, propriedades dos objetos, entidades abstratas ou concretas, produtos manufaturados, processos industriais; 	
ISO 1087	Qualquer unidade do pensamento, geralmente expressa por um termo, símbolo literal ou por qualquer outro símbolo	<ul style="list-style-type: none"> - classe é a totalidade de objetos aos quais um conceito se refere; - extensão é a totalidade de conceitos específicos incluídos em um conceito genérico; - a adição de características a um conceito diminui sua extensão, surgindo conceitos mais específicos; - novos conceitos têm sua origem na evolução de outros já conhecidos por meio de determinação (enriquecimento da intensão), de conjunção (fusão de intensões), de disjunção (fusão de extensões) ou de integração (combinação ou fusão de constituintes). 	
ISO 704.1	unidade do conhecimento que retrata ou corresponde a um conjunto de objetos e é representada na língua por meio de uma designação ou definição	<ul style="list-style-type: none"> - o conceito é não apenas uma unidade do pensamento, mas também uma unidade do conhecimento; - objetos concretos são ditos objetos materiais e objetos abstratos são ditos imateriais 	

Os autores consultados sustentam que a definição de um conceito está estreitamente associada às suas características. As características constituintes de um conceito definem a sua *intensão*. A totalidade dos objetos a que um conceito se refere possui certas características em comum e compreende a *extensão* de tal conceito. As características inerentes ao conceito são ditas *intrínsecas* e as características que determinam sua origem ou seu propósito são ditas *extrínsecas*. Um apanhado sobre a visão dos diversos autores acerca de características é apresentada na tabela 51.

Tabela 51: Sistematização dos fundamentos teóricos relativos a características

Autor	Definição	Características/Observações	Fonte
Felber	elemento de um conceito que serve para descrever ou identificar certa qualidade de um objeto individual; é também, por si só, um conceito	<ul style="list-style-type: none"> - características intrínsecas – inerentes ao objeto (forma tamanho, material, cor, entre outras propriedades); - características extrínsecas de propósito – designam o propósito do objeto (aplicação, funcionamento, performance, localização e posicionamento na montagem); - características extrínsecas de origem – designam a origem do objeto (método de manufatura, descobridor, descritor, inventor, produtor, país de origem, fornecedor); - característica simples – refere-se a uma única propriedade de um objeto individual; - característica complexa – refere-se a uma combinação de duas ou mais propriedades do objeto individual; - características são utilizadas como elementos de classificação para a colocação de um conceito em um sistema de conceitos; - características equivalentes podem ser substituíveis entre si sem expressiva modificação no sentido do termo em questão; 	
Sager	elemento conceitual ou elemento do conhecimento; é também um conceito e serve para a determinação das relações entre o conceito a que pertence e outros conceitos num sistema de conceitos	<ul style="list-style-type: none"> - características representam as propriedades dos objetos e podem ser determinadas com base na observação, na medição, na atribuição, entre outros processos; - características servem para diferenciar um conceito de outros em uma dada área de especialidade; - características determinam da intensão de um conceito; - características servem como fundamento para a formação de termos; - características servem à estruturação de sistemas de conceitos, por meio da determinação de equivalências; - a modificação de uma ou mais características de um conceito resulta na formação de um novo conceito 	

Autor	Definição	Características/Observações	Fonte
ISO 704	propriedade específica de um objeto ou comum a um conjunto de objetos; pode ser combinada com outras características para a criação de conceitos	<ul style="list-style-type: none"> - são características intrínsecas: forma, tamanho, material constituinte, cor, posição; - são características extrínsecas: origem, performance, localização, descobridor, inventor; - características são usadas na análise dos conceitos para a construção de sistemas de conceitos, para a formulação de definições e para a formação termos; - o conjunto de características de um conceito é a sua intensão; - o conjunto de objetos que possuem intensões similares compreende a extensão de um conceito; - características essenciais de um conceito são aquelas consideradas indispensáveis sua compreensão e cuja ausência o altera fundamentalmente; - características não-essenciais são aquelas cuja ausência não altera fundamentalmente o conceito que caracterizam. 	

A reunião das características de um conceito, em comparação com outros conceitos, fornece insumos para o entendimento de suas relações em um sistema de conceitos mais amplo. Da análise e comparação das intensões dos conceitos em questão, é possível se obter as semelhanças e diferenças entre eles. A compreensão dos conceitos e de suas relações, tanto no campo da Terminologia como no campo da Linguagem Documentária é indispensável para o estabelecimento das estratégias necessárias à realização do controle da terminologia utilizada em dado campo do conhecimento.

As visões dos autores acerca dos tipos de relações variam, em grande medida, de acordo com os critérios de classificação, tendo sido encontrados os seguintes tipos principais de classificações para essas relações: lógicas, ontológicas, hierárquicas, não-hierárquicas, associativas e relações de efeito. As tabelas 52 e 53 apresentam tais classificações em dois níveis de detalhamento e análise. No primeiro nível, procurou-se apresentar as classificações mais gerais. No segundo nível, foram definidos e comparados todos os tipos de relações que cada autor admite.

Como resultado dessa análise foi possível observar que as relações lógicas são concebidas como relações de semelhança, isto é, que ocorrem entre conceitos que possuem similaridades em relação às suas intensões; nesse tipo de relação ocorre sempre uma sobreposição de intensões, de modo a se estabelecer uma hierarquia entre os conceitos. A relação lógica mais freqüentemente abordada pelos autores foi a do tipo genérico-específica (CABRÉ, 1993; FELBER, 1984; IBICT, 1984; GOMES, 1990). Picht (1984) apresenta as relações e lógicas como subdivisão das relações hierárquicas.

As relações ontológicas são dadas pelas interações de um conceito no tempo e no espaço; são admitidas por Cabré (1993), Felber (1984), pela norma ISO 1087 (1990), por Gomes (1990) e pela norma ISO WD704.1 (1996). Dessa categoria, as relações mais importantes são as relações todo-parte. As relações ontológicas são classificadas, segundo Picht (1984) e segundo a norma ISO WD 704.1 (1996) como subtipos das relações hierárquicas.

Felber (1984) concebe as relações de efeito para expressão de conceitos que se assemelham devido a fenômenos como causalidade, instrumentação entre outros. No âmbito das relações tesaurais, essa abordagem é adotada também por Gomes (1990).

As relações hierárquicas são admitidas por Picht (1984), pela norma ISO WD 704.1 (1996), pelas normas ISO 2788 (1974) e Unisist (1975), por Aitchison e Gilchrist (1979), por Lancaster (1987) e por Cintra (1994). Nesse tipo de relação presume-se a existência de conceitos superordenados e conceitos subordinados. Os autores que adotam essa orientação

classificam nesse tipo as relações genérico-específicas e também as relações partitivas. Entre esses autores, as relações não-hierárquicas são admitidas apenas por Picht (1984), por Aitchison e Gilchrist (1979) e por Cintra (1994) e são definidas como relações em que não existe a subordinação de um conceito sobre outro, mas sim uma contigüidade no espaço ou no tempo.

As relações de equivalência são admitidas pelas normas ISO 2788 (1974), pelas diretrizes do Unisist (1975), por Aitchison e Gilchrist (1979) e pela norma do IBICT (1984) e são definidas como aquelas em que os conceitos se relacionam por sinonímia ou quase sinonímia.

As relações associativas são admitidas pelas normas ISO 2788 (1974), pelas diretrizes do Unisist (1975), pela norma do IBICT (1984), por Lancaster (1987) e pela norma ISO WD 704.1 (1996) e são definidas como relações que não admitem hierarquia ou equivalência, mas sim algum tipo de conexão temática entre os conceitos, permitindo o estabelecimento de relações de dependência entre eles, com base na sua proximidade no tempo ou no espaço.

Pelo aspecto filosófico, a Terminologia utiliza métodos da Lógica e da Ontologia (RONDEAU, 1984, p. 11). O autor argumenta que a Terminologia relaciona-se com a Ontologia, uma vez que o significado de um termo – conceito, para a Terminologia – corresponde a entidades concretas ou abstratas que são consideradas inicialmente pela sua existência para, em seguida, buscar a forma lingüística que as representa ou que deveria representar.

Comparativamente com os modelos de triplas da arquitetura RDF, e de classes da linguagem OWL, em Terminologia tem-se as relações entre conceitos. Assim, a declaração RDF que envolve sujeito, predicado e objeto, poderia ser expressa por meio de uma relação conceitual: conceito A, relaciona-se com o conceito B, onde a relação pode variar, de acordo com os diversos tipos de propriedades e classes que integram a gramática formal RDF. Em outras palavras, propriedades RDF podem ser pensadas como atributos de recursos e também podem representar relações entre recursos, já que é a descrição das propriedades de um recurso que o diferencia dos demais recursos em uma ontologia.

O mesmo raciocínio pode ser utilizado em relação à linguagem OWL, destacando-se a diferença de que a gramática formal OWL estabelece uma quantidade de tipos de relacionamentos bem mais ampla do que a RDF, incluindo mais funções, como o aperfeiçoamento das relações de cardinalidade, a inclusão de expressões para a combinação booleana de classes e a disponibilidade de classes para a descrição de propriedades de classes e para o estabelecimento de relações que equivalência entre classes.

Em termos da especificação formal da arquitetura RDF e da linguagem OWL, os recursos podem ser divididos em grupos chamados *classes*. Os membros de uma classe são denominados *instâncias*. As classes em si, são também recursos e podem ser descritas por meio de *propriedades*. Ao conjunto de instâncias de uma classe denomina-se *extensão da classe*. Essa abordagem é semelhante à abordagem da Terminologia em relação a *características* e *conceitos*. Em Terminologia, uma característica é também um conceito; Em OWL/RDF, uma *instância* é também uma *classe*.

Tabela 52: Sistematização dos fundamentos teóricos relativos a relações conceituais (nível de análise e detalhamento 1)

Autor/Norma	Lógicas	Ontológicas	De efeito	Hierárquicas	Não-hierárquicas	Polivalentes	Complexas	Associativas
CABRÉ	são estabelecidas entre conceitos que apresentam semelhanças pelo fato de possuírem uma ou mais características em comum	Interessam-se pela natureza dos seres e pelas interações entre eles; preocupam-se com a proximidade situacional dos elementos da realidade, isto é, pelo contato	----	----	----	----	----	----
FELBER	relações de similaridade; conceitos similares possuem certas características em comum, ou seja, sua certa parte de sua intensão é comum	Relações indiretas; contigüidade ou justaposição de conceitos no espaço ou no tempo; ou por conexões de causa e efeito	Envolvem causalidade, instrumentação descendência	----	----	----	----	----

Autor/Norma	Lógicas	Ontológicas	De efeito	Hierárquicas	Não-hierárquicas	Polivalentes	Complexas	Associativas
PICHT	As relações de abstração são também denominadas relações lógicas ou genéricas. Esse tipo de relação ocorre entre um conceito superordenado e o seu correspondente subordinado	----	----	Parte-se da combinação de conceitos originais para a formação de um novo; classificam-se em lógicas e ontológicas	São relações que envolvem contigüidade no espaço ou no tempo	----	----	----
SAGER	----	----	----	----	----	Um conceito é colocado em mais de uma hierarquia	A relação entre os conceitos se dá em função de suas classes conceituais e não de hierarquias	----
ISO WD 704	----	----	----	Os conceitos são organizados em níveis em que o conceito superordenado é subdividido pelo menos em um nível	----	----	----	Ocorre quando se pode estabelecer uma conexão temática entre dois conceitos ou por proximidade entre eles no tempo ou no espaço

Autor/Norma	Lógicas	Ontológicas	De efeito	Hierárquicas	Não-hierárquicas	Polivalentes	Complexas	Associativas
ISO 1087	Baseiam-se na semelhança entre os conceitos e originam relações genérico-específicas	Baseiam-se na contigüidade dos conceitos; no contato entre eles no tempo ou no espaço	----	----	----	----	----	----

Tabela 53: Sistematização dos fundamentos teóricos sobre relações conceituais (nível de análise e detalhamento 2)

Relações	Subclassificações	CABRÉ	FELBER	PICHT	SAGER	ISOWD704	ISO1087	Definições
Lógicas	Conjunção	<input checked="" type="checkbox"/>						Combinação entre dois conceitos para formarem um terceiro, específico de ambos
	Coordenação	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>					Dois conceitos específicos compreendem as mesmas características do conceito genérico
	Determinação	<input checked="" type="checkbox"/>						um conceito simples é determinado, tendo sua intensão modificada por outro e resultando num terceiro conceito complexo
	Diagonal		<input checked="" type="checkbox"/>					ocorre em uma relação genérico-específica, quando dois conceitos não estão relacionados nem por coordenação e nem por subordinação
	Disjunção	<input checked="" type="checkbox"/>						terceiro conceito é formado a partir dos traços diferenciadores de cada um dos conceitos geradores
	Genérico-específicas		Gen					os conceitos mais específicos herdam todas as características do conceito mais amplo e, além delas, cada conceito específico possui alguma que lhe é peculiar
	Hiponímia Lógica		Gen					um conceito mais amplo (hiperônimo) cede todas as suas características a um conceito mais específico (hipônimo)
	Intersecção			<input checked="" type="checkbox"/>				um conceito possui apenas parte da intensão de outro, isto é, nem todas as características são as mesmas
	Ligação	Determinação		<input checked="" type="checkbox"/>				um segundo conceito é integrado como característica do primeiro; deste modo a intensão do primeiro conceito é aumentada por pelo menos uma característica adicional; o conceito resultante é uma espécie do primeiro conceito
	Conjunção		<input checked="" type="checkbox"/>				ocorre quando se dá a integração das intensões de dois conceitos membros; o conceito resultante será então uma espécie de cada um dos conceitos originais	

Relações	Subclassificações	CABRÉ	FELBER	PICHT	SAGER	ISOWD704	ISO1087	Definições	
	Disjunção		<input checked="" type="checkbox"/>					ocorre quando se dá a integração das extensões de dois conceitos; o conceito formado passa a ser um superordenado em relação aos conceitos originais	
	Subordinação		Gen					um conceito (espécie) possui todas as características de outro (gênero) e pelo menos uma característica adicional	
	Coordenação	Par						descrevem o relacionamento entre um todo e suas partes e entre as distintas partes de um mesmo todo.	
	Encadeamento	Suces						referem-se à sucessão dos objetos no tempo, isto é, a eventos de causa-efeito.	
	Material/Produto		Mat					referem-se aos diferentes estágios em um processo de produção, partindo da matéria-prima ao produto final	
	No espaço						Esp	contigüidade dos conceitos, isto é, no contato entre eles, que pode se dar no espaço	
	No tempo						Suces	contigüidade dos conceitos, isto é, no contato entre eles, que pode se dar no tempo	
Ontológicas	Coordenação		Par					ocorre entre dois objetos que representam partes de um mesmo todo	
	Diagonal		<input checked="" type="checkbox"/>					objetos fazem partes do mesmo inteiro, mas não estão relacionados nem por subordinação nem por coordenação	
	Intersecção		<input checked="" type="checkbox"/>					ocorre entre dois objetos que possuem algumas partes em comum	
	Partitivas	Ligação		<input checked="" type="checkbox"/>					ocorre a partir da integração de dois ou mais objetos em uma nova entidade. (não é uma combinação de dois ou mais conceitos, mas dois ou mais objetos individuais que pertencem a esses conceitos)
		Subordinação		<input checked="" type="checkbox"/>					se um objeto individual consiste das mesmas partes de outro objeto individual, mais pelo menos uma parte, então aquele objeto (todo) compreende este objeto como membro (parte)
	Sucessão		Suces					dois conceitos que encontram-se ordenados por contigüidade, no tempo	
De efeito	Causalidade		Caus					envolvem fenômenos como <i>causalidade</i> , <i>instrumentação</i> , <i>descendência genealógica</i> , <i>descendência ontológica</i> ou <i>descendência entre estágios de substâncias</i>	
	Instrumentação		Inst						
	Descendentes	Estágios de Substâncias		ESub					
		Genealogia		Genet					
	Ontológica		<input checked="" type="checkbox"/>						
Hierárquicas	Genéricas						Gen	a intensão do conceito <i>subordinado</i> (conceito específico) inclui a intensão do conceito <i>superordenado</i> (conceito genérico) e pelo menos mais uma característica delimitadora	

Relações	Subclassificações	CABRÉ	FELBER	PICHT	SAGER	ISOWD704	ISO1087	Definições	
	Ontológicas (partitivas ou todo parte)			Par		Par		o conceito <i>superordenado</i> (conceito abrangente) representa uma unidade inteira, enquanto os conceitos <i>subordinados</i> (conceitos partitivos) representam partes do todo	
	Causais			Caus				há a ocorrência de causa e efeito	
	Cronológicas			Suces				há a coexistência ou sucessão no tempo	
	Dependência			<input checked="" type="checkbox"/>				entre pupilo e tutor	
	Funcionais			<input checked="" type="checkbox"/>				entre argumento e função	
Não-hierárquicas	Seqüenciais			Genet				se estabelece uma relação entre produtor e produto	
	Instrumentais			Inst				entre instrumento e a utilização de tal instrumento	
	Obrigação Jurídica			<input checked="" type="checkbox"/>				entre credor e devedor	
	Produção			Mat				se estabelece uma relação entre material e produto	
	Transmissão			<input checked="" type="checkbox"/>				se estabelece uma relação entre emissor e receptor	
	Pragmáticas			<input checked="" type="checkbox"/>				baseiam-se na proximidade temática entre os conceitos, mas não podem ser classificadas como relações hierárquicas e nem como relações seqüenciais. No campo dos tesouros, tais relações são denominadas <i>relações associativas</i>	
Genéricas					Gen			estabelece uma ordem hierárquica, que identifica os conceitos como pertencentes à mesma categoria e na qual há um conceito amplo (<i>genérico</i>), denominado <i>superordenado</i> e um conceito mais restrito (<i>específico</i>), denominado <i>subordinado</i>	
Partitivas					Par			indica a conexão entre conceitos consistindo de mais de uma parte e suas partes constituintes	
Polivalentes					<input checked="" type="checkbox"/>			um conceito é colocado em mais de uma hierarquia para um dado campo de assunto	
Complexas	Atividade/lugar				Esp			subconjuntos de relações colocando-se os conceitos em classes conceituais de referência como, por exemplo, objetos, métodos, propriedades, qualidades, estados e processos. Dessa classificação resulta que o relacionamento entre dois conceitos pode se dar apenas em função de suas classes conceituais; por exemplo, a relação material-produto só poderá ocorrer entre duas entidades materiais	
	Causa-Efeito				Caus				
	Fenômeno/Medida				<input checked="" type="checkbox"/>				
		Estado				ESub			
	Material	Produto				Mat			
		Propriedade				Carac			
	Objeto	Característica				Carac			
Elemento oposto					<input checked="" type="checkbox"/>				

Relações	Subclassificações	CABRÉ	FELBER	PICHT	SAGER	ISOWD704	ISO1087	Definições												
	Forma							Carac												
	Material							Carac												
	Operação							<input checked="" type="checkbox"/>												
	Qualidade							<input checked="" type="checkbox"/>												
	Recipiente							<input checked="" type="checkbox"/>												
	Instrumento							Inst												
	Método							<input checked="" type="checkbox"/>												
	Paciente							<input checked="" type="checkbox"/>												
	Produto							<input checked="" type="checkbox"/>												
	Ação e agente							<input checked="" type="checkbox"/>												
	Ação e lugar ou localização							Esp												
	Ação e objeto							<input checked="" type="checkbox"/>												
	Causa e efeito							Caus												
	Evento e tempo							Suces												
Associativas	Item concreto e forma							Carac												
	item concreto e material							Carac												
	Matéria ou substância e propriedade							Carac												
	Material/estado							Esub												
	Matéria-prima/produto							Mat												
Gen	Genérica	Caus	Causa e efeito	Par	Partitiva	Inst	Instrumento	Suces	Sucessão	Carac	Característica	Esp	Espaço	Esub	Estado	Substância	Mat	Material/produto	Genet	Genética

Diz-se que há uma relação *associativa* entre conceitos quando se puder estabelecer uma conexão temática entre tais conceitos ou quando se puder estabelecer uma dependência entre eles com base na sua proximidade no espaço ou no tempo

Tabela 54: Sistematização dos fundamentos teóricos relativos a relações conceituais em tesouros (nível de análise e detalhamento 1)

Autor/Norma	Lógicas	Ontológicas	De efeito	Hierárquicas	Não-hierárquicas	Polivalentes	Equivalência	Associativas
Unisist e ISO 2788	----	----	----	expressam relações entre conceitos superordenados e conceitos subordinados e dividem-se em relações genéricas e relações todo-parte	----	----	termos são considerados similares, isto é, possuem quase o mesmo sentido sendo associados ao mesmo conceito	São geralmente empregadas para cobrirem outras relações entre conceitos que não as hierárquicas ou as de equivalência
IBICT	pode ser hierárquico, lateral, ou de oposição. Um relacionamento hierárquico é o que mais distingue um tesouro sistemático de outras listas organizadas de termos, pois exprime os graus ou níveis de superordenação e de subordinação entre os conceitos. Tais relacionamentos podem ser classificados em: genérico, partitivo e polierárquico	----	----	----	----	----	O relacionamento de equivalência ocorre entre dois termos que designam o mesmo conceito	O relacionamento associativo ocorre entre termos que não são equivalentes e nem formam uma hierarquia, mas que estão ligados de alguma maneira

Autor/Norma	Lógicas	Ontológicas	De efeito	Hierárquicas	Não-hierárquicas	Polivalentes	Equivalência	Associativas
Aitchison e Gilchrist	----	----	----	que incluem as do tipo gênero/espécie e coisa/tipo; c) relação todo-parte, que os autores não consideram como relação hierárquica, a não ser quando se trata de localizações geográficas – França/Paris; d) relações polierárquicas,	não hierárquicas, que compreendem vários tipos de relações como: coisa-propriedade; coisa-processo; coisa-coisa como atributo; propriedade-processo; propriedade-propriedade como atributo; processo-agente; processo-propriedade;	----	relações de equivalência, que incluem os sinônimos e os quase sinônimos	----
Gomes	são relações de similaridade, isto é, são explicitadas por meio da verificação de características comuns entre os conceitos	relações que ocorrem de maneira indireta entre os conceitos e podem caracterizar-se pela contigüidade no tempo e no espaço ou pela conexão de causa-efeito.	relacionamentos de efeito podem ser classificados em três tipos: de causalidade, de instrumentalidade e de descendência	----	----	----	----	----

Autor/Norma	Lógicas	Ontológicas	De efeito	Hierárquicas	Não-hierárquicas	Polivalentes	Equivalência	Associativas
Lancaster	----	----	----	relações hierárquicas abrangem as relações genéricas, partitivas e exemplares	----	----	----	ocorrem entre termos inter-relacionados em dado contexto que não envolva sua hierarquização.
Cintra	----	----	----	entre conceitos subordinados em um ou diversos níveis, de acordo com uma certa ordem	não-hierárquicas definem-se como todas aquelas em os conceitos não podem ser hierarquicamente organizados	----	----	----

Tabela 55: Sistematização dos fundamentos teóricos relativos a relações conceituais em tesouros (nível de análise e detalhamento 2)

Relações	Subclassificações	SO2788/UNISIST	IBICT	Aitchison e Gilchrist	Gomes	Lancaster	Cintra	
Associativas	Antonímia	<input type="checkbox"/>						
	Coordenação	<input type="checkbox"/>						
	Genética	<input type="checkbox"/>						
	Uso simultâneo	<input type="checkbox"/>						
	Causa e efeito	<input type="checkbox"/>						
	Instrumental	<input type="checkbox"/>						
	Material	<input type="checkbox"/>						
	Similaridade	Física	<input type="checkbox"/>					
		Material	<input type="checkbox"/>					
De processos		<input type="checkbox"/>						

Relações	Subclassificações	ISO2788/UNISIST	IBICT	Aitchison e Gilchrist	Gomes	Lancaster	Cintra	
	Disciplina / objeto de estudo		<input type="checkbox"/>					
	Processo ou operação / agente ou instrumento		<input type="checkbox"/>					
	Ação / resultado da ação		<input type="checkbox"/>					
	Ação / paciente		<input type="checkbox"/>					
	Coisa / propriedades		<input type="checkbox"/>					
	Pessoas ou coisas / origem		<input type="checkbox"/>					
	Dependência causal		<input type="checkbox"/>					
	Coisa / contra-agente		<input type="checkbox"/>					
	Sincategoremática		<input type="checkbox"/>					
	Equivalência	Deslocamento genérico		<input type="checkbox"/>				
Quase-sinônimos				<input type="checkbox"/>				
Sinonímia		Origem lingüística diferente		<input type="checkbox"/>				
		raízes diferentes		<input type="checkbox"/>	<input type="checkbox"/>			
		Nomes populares / nomes científicos		<input type="checkbox"/>	<input type="checkbox"/>			
		Substantivos comuns / nomes comerciais		<input type="checkbox"/>	<input type="checkbox"/>			
		Variantes para conceitos emergentes		<input type="checkbox"/>	<input type="checkbox"/>			
		Grafias diversas		<input type="checkbox"/>				
		Termos correntes / termos antigos		<input type="checkbox"/>	<input type="checkbox"/>			
		Termos de culturas diferentes		<input type="checkbox"/>				
		Termo abreviado / termo extenso		<input type="checkbox"/>				
		Forma composta / forma fatorada		<input type="checkbox"/>				
		Uso local / internacional		<input type="checkbox"/>	<input type="checkbox"/>			

Relações	Subclassificações		ISO2788/UNISIST	IBICT	Aitchison e Gilchrist	Gomes	Lancaster	Cintra	
Lógicas	Hierárquicas	Genéricas		<input type="checkbox"/>					
		Partitivas	Sistemas e órgãos do corpo		<input type="checkbox"/>				
			Localidades geográficas		<input type="checkbox"/>				
			Disciplinas ou áreas do discurso		<input type="checkbox"/>				
			Estruturas sociais hierarquizadas		<input type="checkbox"/>				
		Polierárquicas		<input type="checkbox"/>					
	Laterais		<input type="checkbox"/>						
	De oposição	Contraditória		<input type="checkbox"/>		<input type="checkbox"/>			
		Contrária		<input type="checkbox"/>		<input type="checkbox"/>			
		Positivo / indiferente / negativo		<input type="checkbox"/>		<input type="checkbox"/>			
	Genérico-específicas				<input type="checkbox"/>				
	Analíticas				<input type="checkbox"/>				
	Ontológicas	Material/Produto					<input type="checkbox"/>		
Partitivas		Sistemas e órgãos do corpo				<input type="checkbox"/>			
		Localidades geográficas				<input type="checkbox"/>			
		Disciplinas ou áreas do discurso				<input type="checkbox"/>			
		Estruturas sociais hierarquizadas				<input type="checkbox"/>			
Sucessão				<input type="checkbox"/>					
De efeito	instrumental					<input type="checkbox"/>			

Relações	Subclassificações		ISO2788	UNISIST	Aitchison e Gilchrist	Gomes	Lancaster	Cintra
	Descendência	Estágios de Substâncias				<input type="checkbox"/>		
		Genealogia				<input type="checkbox"/>		
		Ontogenética				<input type="checkbox"/>		
	Causalidade					<input type="checkbox"/>		
Hierárquicas	Genéricas		<input type="checkbox"/>				<input type="checkbox"/>	<input type="checkbox"/>
	Partitivas		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
	Polierárquicas				<input type="checkbox"/>			
	Hierárquicas verdadeiras	Coisa / tipo			<input type="checkbox"/>			
		Gênero / espécie			<input type="checkbox"/>			
Exemplares						<input type="checkbox"/>		
Não hierárquicas	Coisa	Coisa / parte			<input type="checkbox"/>			
		Coisa / processo			<input type="checkbox"/>			
		Coisa / propriedade			<input type="checkbox"/>			
		Coisa / coisa como atributo			<input type="checkbox"/>			
		Coisa / coisa como aplicação			<input type="checkbox"/>			
	Propriedade	Propriedade / processo			<input type="checkbox"/>			
		Propriedade / propriedade como atributo			<input type="checkbox"/>			
	Processo	Processo / coisa			<input type="checkbox"/>			
		Processo / propriedade			<input type="checkbox"/>			
	Quase-sinônimos				<input type="checkbox"/>			
	Atributiva							<input type="checkbox"/>
	Campo de estudos / objeto							<input type="checkbox"/>

Relações	Subclassificações	ISO2788	UNISIST	Aitchison e Gilchrist	Gomes	Lancaster	Cintra
			IBICT				
	Processo ou operação / agente ou instrumento						<input type="checkbox"/>
	Influência						<input type="checkbox"/>
	Matéria-prima / produto						<input type="checkbox"/>
	Ação / resultado da ação						<input type="checkbox"/>
	Causa / consequência						<input type="checkbox"/>
	Coisa / aplicação						<input type="checkbox"/>
	Efeito / causa						<input type="checkbox"/>
	Atividade / agente						<input type="checkbox"/>
	Atividade / propriedade						<input type="checkbox"/>
	Atividades complementares						<input type="checkbox"/>
	Opostos						<input type="checkbox"/>
	Ação / paciente						<input type="checkbox"/>
	Coisa ou atividade / propriedades ou agentes						<input type="checkbox"/>
	Coisa / contra-agente						<input type="checkbox"/>
	Atividade / produto						<input type="checkbox"/>
	Pessoa ou coisa / origem						<input type="checkbox"/>
	Associação implícita						<input type="checkbox"/>
	Expressões de categorização / substantivos						<input type="checkbox"/>
	Interfaceta						<input type="checkbox"/>

Levantamento e Análise dos Sistemas de Informação Legislativa - SIL

- 3.1 Levantamento e Análise dos SIL na Internet
- 3.2 Caracterização dos Sistemas: NJUR, SAPL, NJUR-DF, LEGIN

Este capítulo apresenta os levantamentos e análises resultantes do estudo dos SIL realizado nesta pesquisa. Tal estudo, como relatado no capítulo de Introdução, na seção de Metodologia, teve como objetivo a identificação das características dos SIL quanto a aspectos técnicos de armazenamento, de processamento e de controle terminológico. Os levantamentos de dados foram realizados por meio da Internet e, também, por meio de entrevistas com técnicos que trabalham em sistemas cuja sede está localizada no Distrito Federal.

3.1 Levantamento e análise dos SIL na Internet

Dados do IBGE, em agosto de 2004, apontam que no Brasil existem aproximadamente 5.560⁶³ municípios. Tendo em vista que haveria dificuldades para a obtenção de informações nesse universo, decidiu-se pesquisar apenas as instituições legislativas dos níveis federal e estadual. Assim, esta pesquisa contemplou 29 casas legislativas, a saber: Senado Federal, Câmara dos Deputados, Câmara Legislativa do Distrito Federal e as 26 Assembleias Legislativas Estaduais. Optou-se por um estudo qualitativo e não quantitativo, tendo em vista que os objetivos desta pesquisa concentram-se no plano de fundamentação teórica, culminando com a apresentação de uma proposta metodológica de organização e representação de informação em sistemas legislativos. Para isso, considerou-se suficiente o estudo de sistemas nas instituições citadas, uma vez que o modelo foi desenvolvido de tal modo que possa ser aplicado também à organização da legislação municipal.

3.1.1 Sistema operacional e aplicativo de serviços web

As informações de identificação dos sistemas operacionais e dos aplicativos de serviços web utilizados nos SIL foram obtidas com o apoio do mecanismo de pesquisa do sítio Netcraft⁶⁴, que fornece uma descrição a respeito da configuração do servidor de *web* correspondente a dado domínio na *web*. Os dados apresentados nas ilustrações 55 e 56 foram coletados no dia 27/07/2004 e reconfirmados no dia 15/12/2004. Foram coletadas informações relativas aos sítios da Câmara dos Deputados, do Senado Federal e de todas as Assembleias Legislativas Estaduais, bem como da Câmara Legislativa do Distrito Federal. Os dados coletados referem-se apenas ao computador dedicado aos serviços da *web* e não são indicativos de que toda a rede de computadores de uma casa legislativa utilize a mesma plataforma. Pode haver casos, a

⁶³ Para mais detalhes, sugere-se a leitura do artigo: IBGE. *Brasil já tem mais de 180 milhões de habitantes*. Disponível em: <http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=207&cid_pagina=1>. Acesso em: 16 dez. 2004.

⁶⁴ Consulte a seção *What's this site running?* disponível em: <http://www.netcraft.com/>

exemplo da Câmara Legislativa do Distrito Federal, em que a rede utiliza outras plataformas, como Linux e Unix em outros serviços da rede.

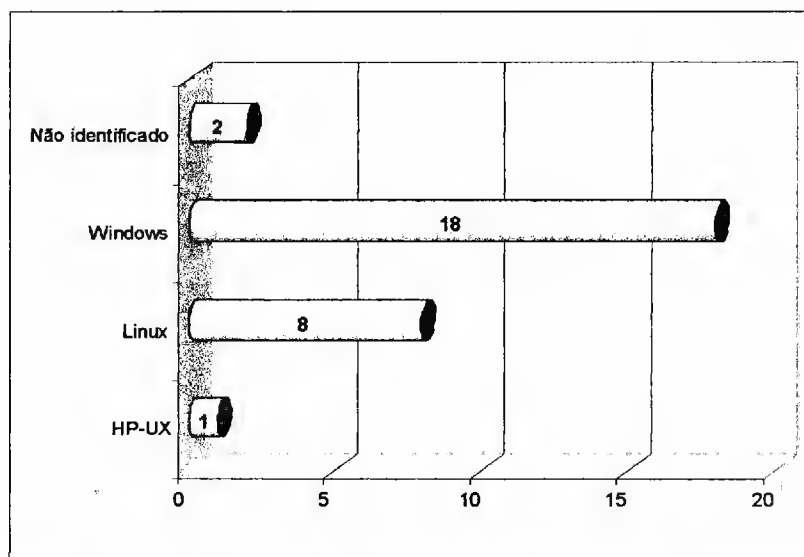


Ilustração 55: Levantamento do uso de sistemas operacionais pelas 29 instituições legislativas
Fonte: pesquisa realizada por meio do recurso "What's that site running", disponível no sítio do Netcraft.com

Os dados apresentados nas ilustrações 55 e 56, referentes ao universo de 29 instituições legislativas em todo o Brasil, indicam o predomínio de servidores que utilizam sistema operacional Microsoft Windows – nas versões NT, 2000 ou XP – conjugado com o aplicativo de serviços *web* Microsoft *Internet Information Server* – nas versões 4.0, 5.0 e 6.0. Em termos de predomínio, o segundo sistema operacional de maior ocorrência é o Linux, conjugado com o aplicativo de serviços *web* Apache. Foi identificada apenas uma casa legislativa que utiliza o sistema Unix (HP-UX) em seu servidor e, neste caso, o aplicativo de serviços *web* é o Netscape Enterprise 3.6.

Com relação aos aplicativos de serviços *web* (ilustração 56), comparados com levantamentos mais genéricos, esses números parecem seguir uma tendência predominante na *Internet*, que é o maior uso dos produtos Apache e Microsoft.

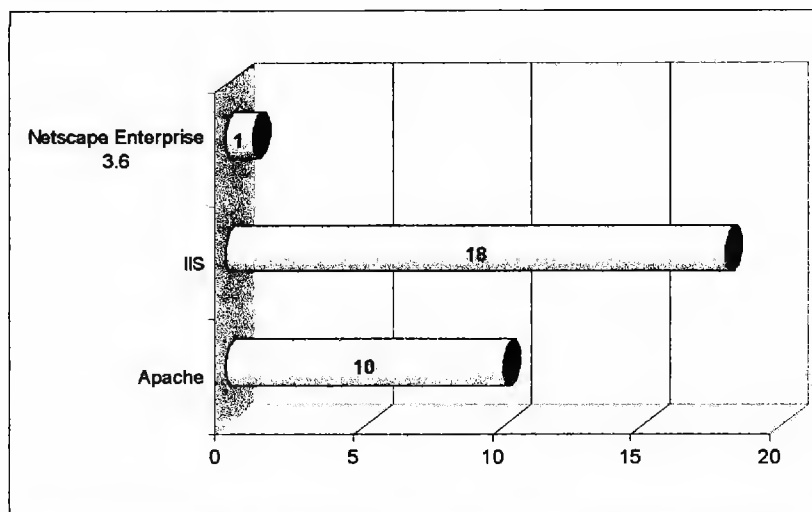


Ilustração 56: Levantamento do uso de aplicativos de serviços para web pelas 29 instituições legislativas
 Fonte: pesquisa realizada por meio do recurso "What's that site running", disponível no sítio do Netcraft.com

A diferença observada nos sítios analisados é o predomínio de aplicativos Microsoft sobre as diferentes versões do Apache – o relatório do Netcraft indica que a tendência prevalente na *Internet* é o uso do Apache. Para elucidação, verifique-se o relatório de pesquisa do *site* Netcraft (ilustração 57). Em termos numéricos, a tabela 56 aponta as tendências, segundo o *site* NetCraft, quanto ao uso de servidores web.

Tabela 56: evolução do uso de servidores web nov/dez 2004

Fabricante	Nov 2004	%	Dez 2004	%	Varição
Apache	38.028.642	67,77	38.614.673	67,84	0,07
Microsoft	11.923.566	21,25	12.062.761	21,19	-0,06
Sun	1.761.705	3,14	1.812.966	3,18	0,04
Zeus	739.006	1,32	687.508	1,21	-0,11

Fonte: http://news.netcraft.com/archives/web_server_survey.html

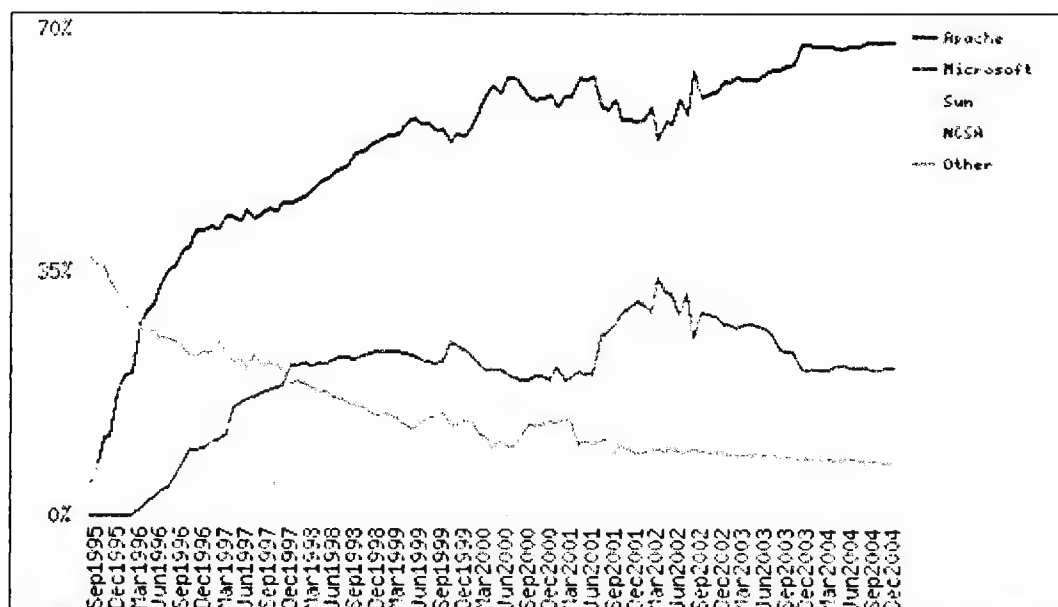


Ilustração 57: mercado dos servidores web mais populares em todos os domínios Ago/95 – Dez/04

Fonte: http://news.netcraft.com/archives/web_server_survey.html

Os percentuais apresentados nesta seção referem-se apenas aos 29 *sites* analisados. Cabe investigar se as proporções se repetem também para os *sites* de assembleias municipais, o que não foi objetivo desta pesquisa.

3.1.2 Tipo de indexação

Com relação a aspectos da indexação dos *sites* dos sistemas legislativos, foram utilizadas duas estratégias: a análise do uso de metadescritores no código HTML da página de entrada – com o apoio do sistema SimSpider⁶⁵ – e a interação com os sistemas de cada um dos *sites* das casas legislativas, por meio de visitação direta. Os dados foram coletados no dia 27/07/04 e reconfirmados em 15/12/04. A ilustração 58 apresenta a situação de uso de metadescritores nas páginas de entrada em cada sistema legislativo. A análise da ilustração permite observar que apenas uma pequena parte dos sistemas utiliza metadescritores como apoio para a indexação pelos mecanismos de busca. Embora, como visto no capítulo de Revisão de Literatura, na seção Uso de metadados na web, os autores e os especialistas em técnicas de promoção de *sites* para a localização por mecanismos de busca, recomendem o uso de pelo menos algumas marcações de metadados, pode-se observar, que parece não haver essa preocupação, na maioria dos sistemas pesquisados.

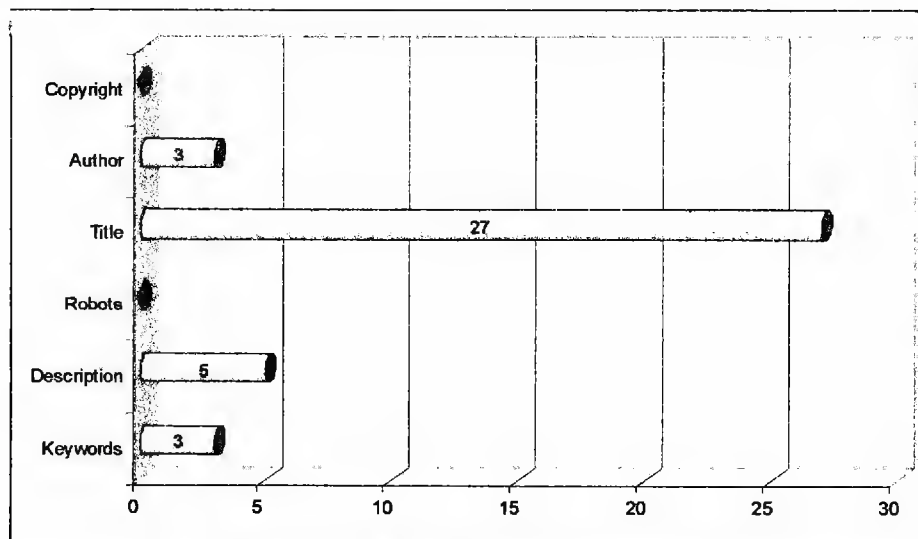


Ilustração 58: Uso de metadescritores para a indexação por mecanismos de busca da web
 Fonte: levantamento realizado com o sistema SimSpider

Outro fator que parece denotar a falta de preocupação com aspectos da indexação e da organização de metadados para a descrição de conteúdos nas páginas é o uso inadequado dos títulos, das descrições e das palavras chaves das páginas. Foram registrados casos em que o

⁶⁵ Sugere-se consultar o endereço < http://www.searchengineworld.com/cgi-bin/sim_spider.cgi >

título da página inicial do *site* nada descreve acerca do seu conteúdo, como por exemplo: “*principal*”; “=*Razão Social da Sua Entidade*”; “*SIG 6.2*”. Além dessas palavras, foram identificados, ainda, caracteres como *pontos* e *dois-pontos* precedendo ou sucedendo as palavras do título, o que em nada favorece a indexação. Na tag `<description>`, utilizada para fornecer descrições acerca do conteúdo da página a aplicativos de indexação, foram encontrados conteúdos que não possuem nenhuma relação com os objetivos do *site*, como: “*Página inicial da Intranet*”; “*FW MX DW MX HTML*”; “*Este site contém informações sobre a entidade SIG Sistemas Interativos*”; “*cadastro de usuários Links Web*”. Tais ocorrências demonstram que não existe preocupação expressa quanto a favorecer a localização e a descrição do conteúdo dos *sites* em questão por sistemas de busca da web ou até mesmo outros sistemas, como indexadores automáticos que pudessem estar disponíveis no próprio *site*. Em alguns casos de *sites* desenvolvidos com o apoio de gerenciador de conteúdo desenvolvido pelo Instituto de Tecnologia do Legislativo (ITECLEGIS), foi utilizada a tag `<meta http-equiv=refresh content="0; url=endereço...">`, cuja função é o redirecionamento do navegador do usuário para uma outra URL, em que o título e o autor são corretamente fornecidos. No entanto o uso do sistema SimSpider, que simula o comportamento de um mecanismo de busca, recupera o título genérico “=*Razão Social da Sua Entidade*” e nenhuma descrição de conteúdo ou de autoria. A ocorrência de apenas três casos de utilização da tag `<keywords>` demonstra que parece não haver preocupações quanto a se fornecer aos indexadores da web informações mais precisas sobre o conteúdo do *site*, o que certamente fará com que tais mecanismos indexem cada página com base apenas em seu conteúdo textual.

Em nenhum dos sistemas analisados constatou-se preocupações com relação ao uso da meta tag `<copyright>`, para registrar informações quanto aos direitos autorais em relação às páginas. A indicação de autoria `<author>` também não foi registrada no código HTML, o que significa que não pode ser identificada automaticamente por nenhum aplicativo. Desse modo, apenas humanos ou sistemas capazes de realizarem pesquisa textual na página, como os grandes mecanismos de busca, poderão extrair tal informação, se ela estiver presente no texto.

Nenhum sistema analisado adota o uso da tag `<robots>` ou do arquivo *robots.txt*, o que pode denotar a ausência de uma política de controle de acesso pelos mecanismos de busca. A especificação de valores para essa meta tag ou a escrita de um arquivo *robots.txt* permitem que se controle o nível da indexação realizada pelos mecanismos de busca da internet, que podem em certos casos, representar uma ameaça à segurança e à integridade das informações de um *site*. Uma pesquisa realizada no sistema Google, no dia 3/01/2005, retornou apenas 81 resultados para o termo `<allinurl:robots.txt site:.br>`, cuja função é a localização de arquivos com o nome *robots.txt* em endereços de *sites* em qualquer domínio brasileiro. Considerando-se que há no Brasil

mais de 700 mil domínios, como visto no capítulo de introdução desta pesquisa, o número recuperado na busca acima mencionada representa cerca de 0,01% de *sites* que adotam alguma política em relação à indexação pelos mecanismos de busca com o uso do arquivo *robots.txt*. Quando se repetiu a pesquisa acrescentando-se o termo *gov* para domínios de governo: `<allinurl:robots.txt site:gov.br>`, o número caiu para 1, que representa 0,12% dos 831 domínios de governo registrados⁶⁶. Embora diante desses resultados, não se pode afirmar categoricamente, sem uma pesquisa mais aprofundada, que os domínios brasileiros não utilizem outras formas de garantir sua segurança contra indexações indesejadas, como o próprio uso da meta tag `<robots>`. No entanto, o questionamento é pertinente. Em relação aos Sistemas de Informação Legislativa pesquisados, consultas a alguns domínios, utilizando-se o sistema Google, revelaram algumas preocupantes falhas de segurança. Foram recuperados cerca de 3.460 arquivos de bancos de dados nos formatos *Data Base Format* (DBF) e *Microsoft Access* (MDB) cujo conteúdo revela resultados de votações em certa casa legislativa – é possível que tais resultados sejam de domínio público, mas não foram encontradas indicações expressas. De qualquer modo, um arquivo de banco de dados sempre traz informações reveladoras para hackers que dominam as técnicas de SQL Injection⁶⁷, como nomes de campos, tamanhos e tipos de dados aceitos. Em consultas similares, foram encontrados também outros tipos de arquivos como documentos do Microsoft Word contendo informações não institucionais (links para páginas pessoais contendo músicas); um tutorial destinado à orientação de usuários internos para alteração de senhas na rede local e configuração de contas de correio eletrônico; além de um arquivo no formato PDF contendo as assinaturas de parlamentares em um projeto de lei. Entende-se que tais tipos de arquivos não deveriam estar tão facilmente disponíveis e que, mesmo sendo público o seu conteúdo, é recomendável sua apresentação em formatos mais seguros.

3.1.3 Disponibilidade de acervo e de módulo de pesquisa

O levantamento de informações para a identificação da disponibilidade de acervo e de sistema de busca foi realizado por meio da visitação direta. Como já mencionado na seção de Metodologia desta pesquisa, o conceito de *acervo* aqui adotado refere-se exclusivamente às normas legislativas já publicadas. Como também já definido, *módulo de pesquisa* refere-se ao conjunto de formulários interativos que permitem ao usuário o envio, pela Internet ou intranet, de termos de

⁶⁶ Para informações atualizadas sobre o número de domínios registrados no Brasil, sugere-se uma consulta à seção de estatísticas do *site* `<http://registro.br>`.

⁶⁷ SQL Injection é uma técnica que utiliza códigos da linguagem Structured Query Language (SQL) para a invasão a bancos de dados. Tal técnica explora falhas de arquitetura na organização de estruturas de consulta. Mais detalhes sobre o assunto podem ser localizados em:

`<http://www.sqlsecurity.com/DesktopDefault.aspx?tabindex=2&tabid=3>`

consulta ao Sistema de Informações Legislativas automatizado. A ilustração 59 apresenta a disponibilidade de acervos legislativos, indicando que a maioria das casas legislativas (25) possuem acervos disponíveis na Internet, enquanto somente 4 não possuem. Das 25 casas que possuem acervo disponível, 20 possuem módulo de pesquisa (ilustração 60).

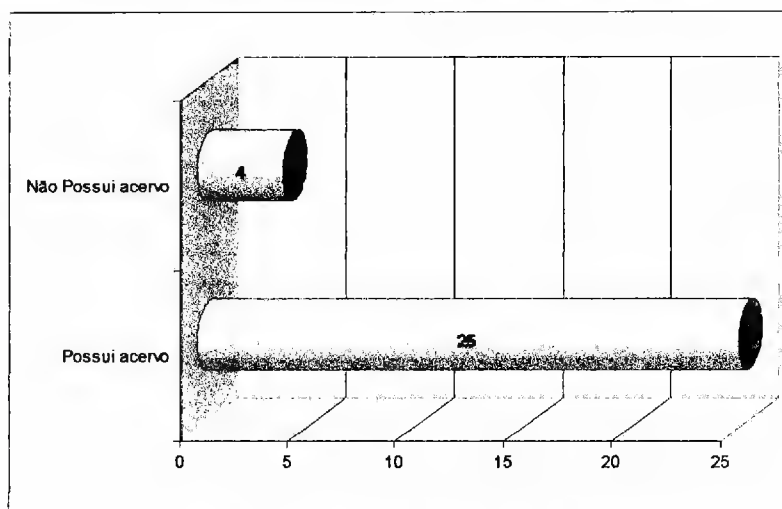


Ilustração 59: disponibilidade de acervo legislativo

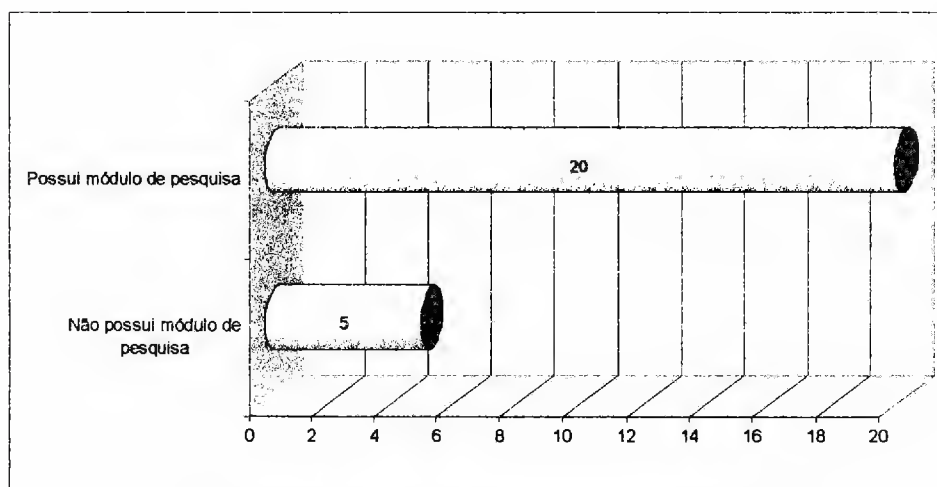


Ilustração 60: disponibilidade de módulo de pesquisa

A ausência de módulos de busca em alguns sistemas pode ser representativa da necessidade de propiciar maior facilidade à pesquisa de informações, inclusive porque há leis que são mais conhecidas pelo seu nome popular ou apelido. Apenas para ilustrar, em uma lista de 40 solicitações de pesquisa encaminhadas ao Setor de Pesquisa e Recuperação da Informação da Câmara Legislativa do Distrito Federal, no período de abril a maio de 2003, apenas 10 fazem menção expressa ao número da lei; nas demais solicitações, os usuários referem-se aos assuntos, datas, origem, ou a outras informações sobre a norma solicitada (Anexo B).

3.1.4 Linguagem de recuperação

Em relação à linguagem de recuperação, verificou-se que a maioria dos sistemas utiliza os operadores booleanos mais amplamente conhecidos: E, OU e NÃO, além de operadores de frase exata e de truncamento (ilustração 61). Foi registrado o não uso do operador de frase exata ou do operador E por alguns sistemas. Acredita-se isso pode acarretar dificuldades para os usuários no momento da busca, trazendo informações de baixa precisão. Foi registrado o uso de operadores como o XOU (OU exclusivo), que permite encontrar documentos que contenham OU o primeiro OU o segundo termo, de forma exclusiva, isto é, não podem ambos ocorrer em um mesmo documento; como o operador COM, que recupera documentos em que ambos os termos aparecem na mesma sentença ou campo.

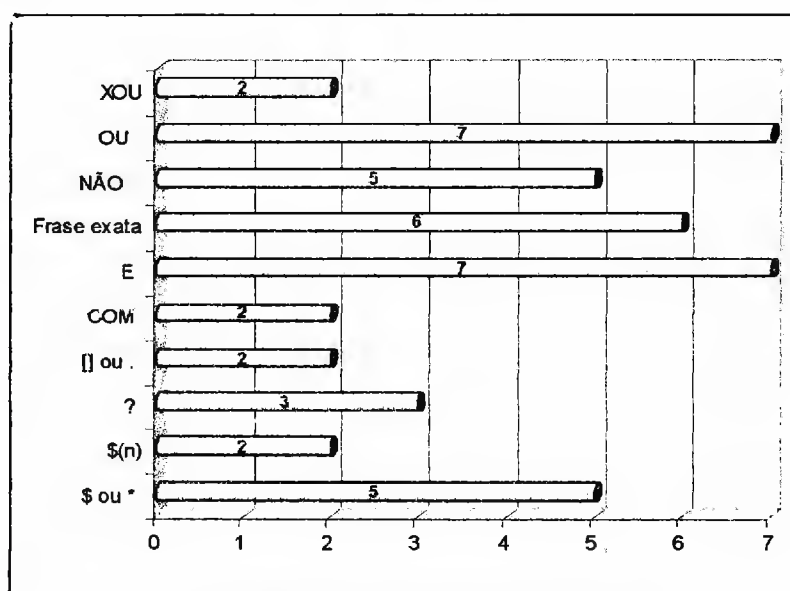


Ilustração 61: uso de operadores de busca

Além desses, foi registrado o uso de operadores de delimitação [], que se prestam à localização de informações em um campo dado; e de operadores de truncamento, como \$ ou *, que permitem o truncamento - vários caracteres; \$(n), que permite o truncamento de de uma quantidade definida de caracteres; e ?, que permite o truncamento de um caractere. A ilustração 62 apresenta apenas os operadores mais freqüentes.

Em relação aos campos de busca, os mais freqüentemente encontrados foram: tipo, número, ano, e assunto (ilustração 62). Como visto no capítulo de revisão de literatura, os sistemas de busca, como Google, Kartoo, Teoma, Scirus, entre outros, têm evoluído cada vez mais para o aprimoramento das pesquisas por termos livres, em alguns casos, retornando termos correlatos. Acompanhando essa tendência, seria desejável que todos os sistemas de informação

legislativa fossem capazes de propiciar buscas de termos livres nos textos integrais dos documentos legislativos, o que somente foi observado em dois desses sistemas.

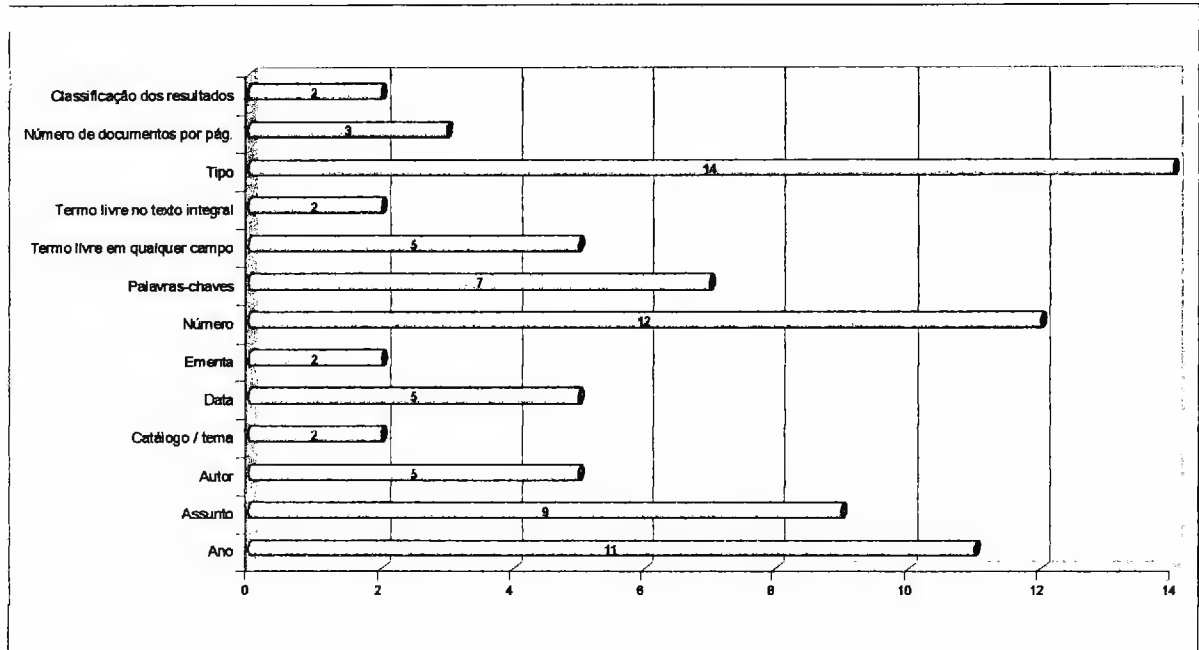


Ilustração 62: Tipos de campos de busca mais frequentes

3.1.5 Tipos de normas

A ilustração 63 apresenta os tipos de normas mais frequentemente encontrados nos sistemas, merecendo destaque entre eles: Leis Ordinárias, Leis Complementares, Emendas Constitucionais, Regimentos Internos, Decretos Legislativos, Constituições Estaduais e Constituição Federal. É interessante observar que dois terços das casas legislativas disponibilizam a Constituição Federal, sendo que em alguns casos, o arquivo encontra-se disponível para download, denotando duplicidade de esforços, já que a iniciativa de publicação da Constituição é do Congresso Nacional, merecendo destaque o trabalho desenvolvido no âmbito do Senado Federal, que mantém uma versão em formato de hipertexto com possibilidades de consultas atualizadas e retrospectivas, favorecendo a localização das diversas alterações que já sofridas por aquela norma⁶⁸ até a última, realizada em junho de 2004.

⁶⁸ Para uma consulta mais detalhada, recomenda-se visitar o endereço:
<<http://www.senado.gov.br/sf/legislacao/const/>>

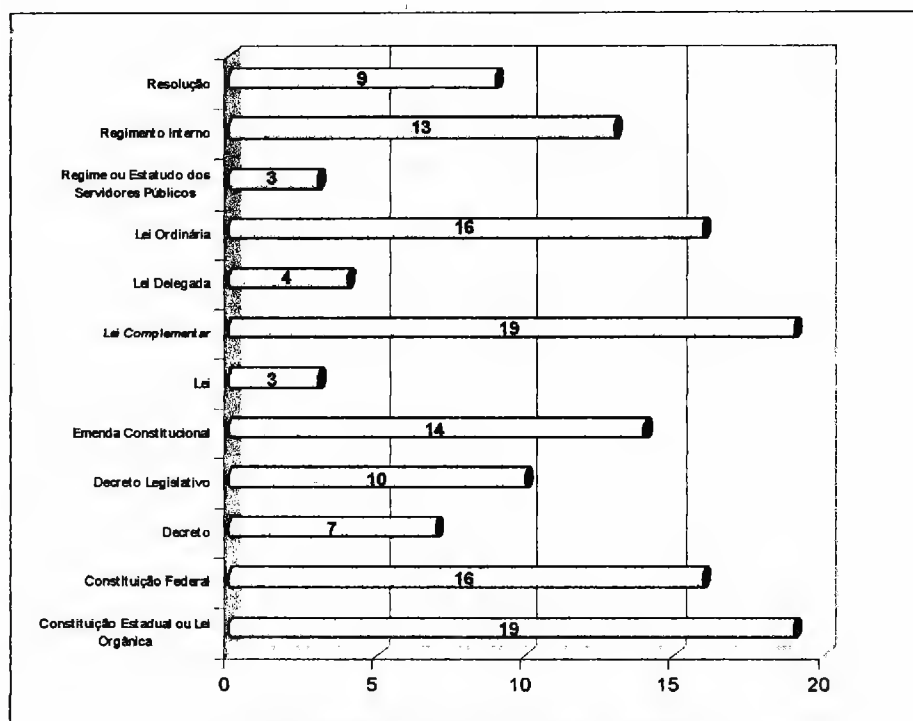


Ilustração 63: Tipos de normas mais frequentes

Tabela 57: Fontes do Direito, segundo a CDDir

340.3	Fontes do Direito
340.32	Leis. Direito positivo
340.321	Hierarquia das leis
340.3211	Lei constitucional
340.3212	Emenda à Constituição
340.3213	Lei complementar
340.3214	Lei ordinária
340.3215	Lei delegada
340.3216	Medida Provisória
340.3217	Decreto-lei
340.3218	Lei auto-aplicável e lei regulamentável
340.3219	Regulamento. Decreto Regulamentar. Decreto autônomo
340.322	Obrigatoriedade das leis. Eficácia
340.323	Eficácia da lei no tempo. Direito intertemporal. Conflitos de leis no tempo
340.3231	Vigência
340.3232	Revogação
340.3233	Desuso
340.3234	Retroatividade e irretroatividade das leis
340.324	Eficácia da lei no espaço
340.326	Interpretação da lei. Hermenêutica
340.328	Elaboração, redação, alteração e consolidação das leis. Técnica legislativa
340.34	Outras fontes de Direito além das leis. Direito usual. Direito consuetudinário

Fonte: (CARVALHO, 2002)

Embora a ilustração 63 apresenta apenas os mais frequentes, no levantamento realizado nesta pesquisa foram encontrados 51 tipos de normas, demonstrando a diversidade existente no Brasil, nas esferas de governo pesquisadas. Essa diversidade, no entanto, não é recente. Em 1948, Carvalho (2002) elaborou um detalhamento da classe 340 da Classificação

Decimal de Dewey (CDD) para adequá-la às necessidades brasileiras. De tal detalhamento surgiu a Classificação Decimal do Direito (CDDir), revista e reeditada em 1953, em 1977 e em 2002. A CDDir prevê, na classe 340.3, de Fontes do Direito, as diversas subclasses apresentadas na tabela 57. A intenção do trabalho de Carvalho foi a de apresentar uma sistematização de assuntos para a classe 340 como um todo, e dessa classificação acabaram resultando, inclusive, as 22 subclasses da classe 340.3.

Recentemente, Andreta (2004) apresentou uma proposta de reformulação dessa classificação, para estruturar a subclasse 340.3 conforme a tabela 58, ampliando ainda mais a lista das subclasses.

Tabela 58: Proposta de reestruturação da subclasse Fontes do Direito da CDDir

340.	DIREITO	340.322213	Item
340.3	Fontes do Direito	340.322214	Alínea
340.32	Leis. Direito positivo. Legislação	340.322215	Parágrafo
340.321	Tipologia e hierarquia dos atos normativos	340.3222151	Parágrafo único
340.3211	Constituição	340.3223	Agrupamentos de artigos
340.32111	Emenda constitucional	340.32231	Seção
340.3212	Leis	340.32232	Capítulo
340.32121	Lei complementar	340.32233	Título
340.32122	Lei delegada	340.32234	Livro
340.32123	Lei ordinária	340.3224	Cláusulas especiais
340.3213	Medidas provisórias	340.32241	Cláusula financeira
340.32131	Decreto-lei	340.32242	Cláusula de vigência
340.3214	Decretos	340.32243	Cláusula revogatória
340.32141	Decreto	340.322431	Cláusula revogatória expressa
340.32142	Decreto legislativo	340.322432	Cláusula revogatória tácita
340.32143	Decreto reservado	340.32244	Disposições transitórias
340.32144	Decreto sem número	340.323	Protocolo final ou Fecho
340.3215	Resoluções	340.3231	Data tópica ou lugar
340.3216	Portarias	340.3232	Data cronológica
340.3217	Atos normativos específicos (constitutivos)	340.3233	Subscrição
340.32171	Códigos	340.3234	Precação ou testemunhos
340.32172	Consolidações	340.324	Eficácia da lei
340.32173	Estatutos	340.3241	Eficácia da lei no tempo
340.32174	Regulamentos	340.32411	Retroatividade da lei
340.32175	Regimentos	340.32412	Irretroatividade da lei
340.3218	Atos normativos internos (de trâmite)	340.32413	Desuso da lei
340.32181	Portarias	340.3242	Eficácia da lei no espaço
340.32182	Ordens de serviço	340.325	Vigência da lei
340.32183	Circulares	340.3251	Vigência imediata
340.32184	Despachos	340.3252	Vacatio legis
340.32185	Deliberações	340.326	Elaboração das leis
340.3219	Projetos de atos normativos	340.3261	Proposição
340.32191	Anteprojetos	340.3262	Discussão
340.32192	Projetos	340.3263	Emendas
340.322	Estrutura dos atos normativos	340.3264	Votação

340.32201	Redação parlamentar	340.3265	Promulgação
340.3221	Protocolo inicial. Preâmbulo	340.3266	Sanção
340.32211	Epigrafe	340.3267	Veto
340.32212	Ementa	340.327	Interpretação da lei. Hermenêutica
340.32213	Fórmula de promulgação	340.3271	Interpretação gramatical ou filológica
340.322131	Autoria	340.3272	Interpretação sistemática
340.322132	Fundamento legal de autoridade	340.3273	Interpretação extensiva
340.322133	Ordem de execução	340.3274	Interpretação estrita
340.3222	Corpo ou texto	340.3275	Interpretação histórico-evolutiva
340.32221	Artigo	340.3276	Interpretação finalística ou teleológica
340.322211	Caput	340.328	Aplicação da lei
340.322212	Inciso		

Fonte: (ANDRETTA, 2004)

No entanto, ambas as propostas misturam as espécies normativas com suas temáticas e até mesmo com suas características como, por exemplo: Leis, Constituição, hierarquia das leis, obrigatoriedades, vigência e técnica legislativa; ou com suas partes constitutivas, como se vê na proposta de Andretta (2004), que prevê subclasses como *artigo*, *caput*, *inciso*, cuja denominação, em si mesma, não parece adequada à representação de um assunto, mas, mais apropriadamente de um elemento constitutivo de uma espécie normativa. Apesar dessas considerações, as propostas de Andretta e de Carvalho, entre outras, como será discutido no capítulo seguinte, adicionadas dos resultados obtidos neste levantamento, foram utilizadas como insumos para a proposição de um RDFSschema destinado à organização e representação da informação legislativa.

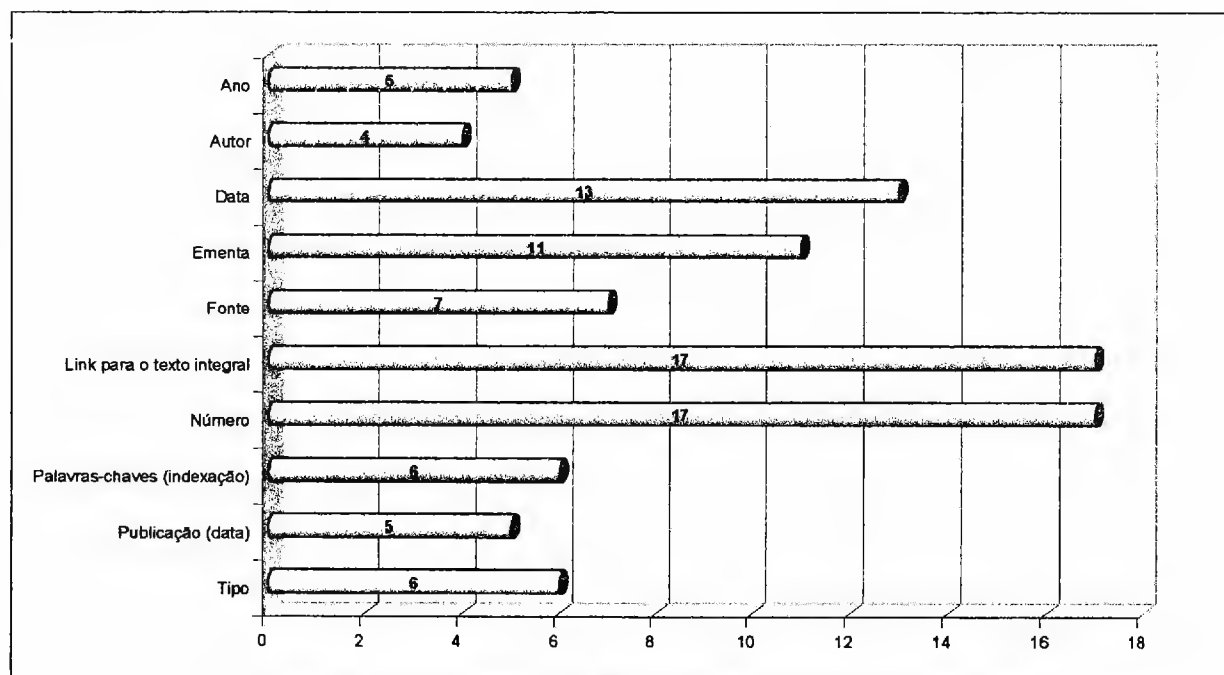


Ilustração 64: Formato dos registros documentários – campos mais frequentes

Como se pode observar na ilustração 64, os metadados mais freqüentemente utilizados para a descrição de documentos nos sistemas legislativos estudados são número, data, ementa, fonte, palavras-chaves e tipo da norma. Considerando que esses dados fornecem a mínima informação sobre um texto legislativo, seria esperado que estivessem presentes em todos os formatos adotados. Como mencionado anteriormente, muitos atos normativos são amplamente divulgados pelos seus nomes populares (apelidos), a exemplo da Lei Pelé, Lei Zico, Lei Sarney, Lei da Propriedade Intelectual, Lei do Silêncio, entre outras. Nesta pesquisa, no entanto, não foi encontrado, em nenhum dos sistemas estudados, um campo destinado ao registro do nome popular do texto legislativo. Seria esperado, também, que os campos número e palavras-chaves estivessem presentes em todos os sistemas, uma vez que representam importantes descritores dos documentos em sistemas informatizados.

Com relação aos formatos de apresentação, foram registrados os tipos HTML, PDF, ZIP e formatos de imagens. Os três últimos formatos citados apresentaram a desvantagem de não tornarem possível a busca textual pelo próprio sistema pesquisado. No caso de arquivos de imagens, mesmo havendo um esforço para a sua representação temática por meio de metadados, é sempre requerido um significativo nível de especialização do profissional da informação para o seu tratamento, de modo a favorecer a localização efetiva de informações.

3.1.6 Módulo de ajuda

Em relação à ocorrência do módulo de ajuda, a metade dos sistemas que disponibilizam módulos de pesquisa não apresenta também ajuda, o que seria esperado e desejado para facilitar a interação dos usuários com o sistema. A ilustração 65 apresenta os resultados do levantamento em que, dos 20 sistemas analisados, dois estavam indisponíveis no momento da pesquisa.

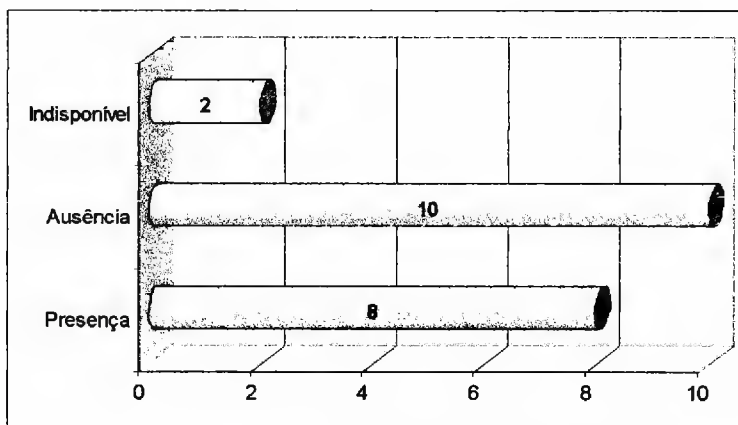


Ilustração 65: disponibilidade de módulo de ajuda ao usuário

Em geral, as informações relativas ao módulo de ajuda referem-se à linguagem de recuperação e aos operadores aceitos como padrão pelo sistema. Um dos sistemas apresenta também uma lista de palavras não-indexadas, mas não apresenta ao usuário como utilizá-la. Dos sistemas que dispõem de módulo de ajuda, os mais bem estruturados são aqueles que implementaram a versão do Sistema de Apoio ao Processo Legislativo, implementado pelo Programa Interlegis.

3.1.7 Linguagem de Programação

Entre as linguagens de programação utilizadas pelos sistemas estudados (ilustração 66), a de maior predomínio foi a Active Server Pages (ASP), desenvolvida por volta de 1998 e bastante difundida pela Microsoft. No entanto, com o advento da arquitetura .Net, no ano de 2002, novas implementações à linguagem ASP foram descontinuadas por aquela empresa, tendo a linguagem sido substituída pela ASP.NET, cuja estrutura é bastante diferenciada, uma vez que essa linguagem, ao contrário da ASP, é totalmente voltada para implementações baseadas na técnica de orientação a objetos⁶⁹. O amplo uso da linguagem ASP pelos sistemas legislativos pode ser um indicativo de que, em pouco tempo, serão necessárias mudanças em suas estruturas para que possam suportar novas tecnologias difundidas na web. Por outro lado, como a linguagem ASP foi bastante difundida e é considerada de fácil aprendizagem, esse é também um indicativo de que os sistemas poderão dispor de técnicos especializados ainda por mais algum tempo. Entre os sistemas consultados, somente dois haviam, até o momento da pesquisa, implementado a tecnologia ZOPE, que está sendo adotada e amplamente incentivada pelo Programa Interlegis, do Senado Federal. A plataforma ZOPE é um servidor de aplicações de Internet, totalmente baseado em código aberto (software livre), destinada à implementação de sistemas de gerenciamento de conteúdo de *sites*, portais, intranets e aplicações personalizadas⁷⁰.

⁶⁹ Detalhes sobre a plataforma .Net e sobre a linguagem ASP.NET estão disponíveis em <http://msdn.microsoft.com>.

⁷⁰ Mais detalhes sobre a plataforma ZOPE podem ser obtidos no *site*. <http://www.zope.org/>

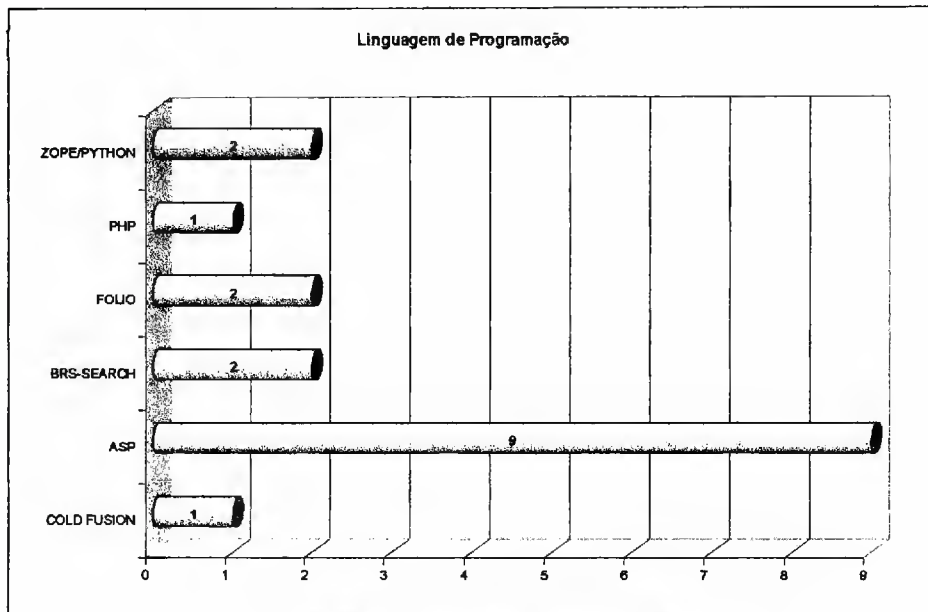


Ilustração 66: Linguagem de programação

3.2 Caracterização dos sistemas: NJUR, SAPL, NJUR-DF, LEGIN

Para o detalhamento da caracterização dos sistemas que possuem sede no Distrito Federal, como relatado no capítulo da metodologia, foram realizadas entrevistas com técnicos que atuam nos sistemas SAPL, LEGIN, NJUR e NJUR-DF. A partir dessas entrevistas, obteve-se acesso a manuais de uso e a outros documentos que permitiram mais profundidade na avaliação das características. Os resultados de tais entrevistas são apresentados a seguir.

3.2.1 Sistema de Normas Jurídicas do Senado Federal (NJUR)

O sistema de Normas Jurídicas do Senado Federal contempla a Legislação Federal Brasileira promulgada a partir de 1946. A base é produto do trabalho de catalogação e indexação realizado pela Subsecretaria de Informações do Senado Federal desde 1972. Estão cadastradas informações: fonte, origem, indexação, vigência e hyperlinks para normas correlatas. A partir do ano 2000 essa base passou a ser disponibilizada na Internet, integrando um sistema mais amplo, denominado Legislação Brasileira. Tal sistema permite a realização de pesquisas textuais em uma base de dados de referências que conta com mais de 120.000 referências de normas jurídicas. A base referencial

A cobertura da base de dados é o conjunto de todas as normas de hierarquia superior, incluindo: Constituição Federal; Emendas Constitucionais de Revisão; Emendas Constitucionais; Leis Complementares; Leis Ordinárias; Leis Delegadas; Medidas Provisórias; Decretos-Lei; Decretos Legislativos; Resoluções do Senado Federal; Resoluções do Congresso Nacional; Atos Declaratórios; Decretos expedidos pelo Poder Executivo Federal; Regimento

Interno da ANC; Atos da Mesa da ANC; Resoluções da Revisão Constitucional; Atos da Mesa Diretora do Congresso Nacional; Atos Convocatórios e Aditamentos ao Ato Convocatório. (LIMA; SOUZA, 2001). Além dessas normas, integram também a base algumas normas abrangentes de hierarquia inferior.

Como instrumento de controle de vocabulário para a indexação dessas normas foi criado o Thesaurus do Senado Federal (THES), também por volta de 1972. Além dessas normas, o Thes é utilizado para a indexação de documentos relativos a jurisprudência e a pronunciamentos de parlamentares. O tesouro é utilizado tanto em operações de indexação, quanto de recuperação de documentos processados nos seguintes bancos de dados que integram o Sistema de Informação do Congresso Nacional (SICON): Normas Jurídicas Federais (NJUR); Discursos de Senadores e Deputados (DISC); e Matérias em Tramitação no Congresso Nacional (MATE). O Vocabulário Controlado Básico (VCB) é uma linguagem documentária desenvolvida para a indexação e recuperação de documentos em outros bancos de dados, a saber: Monografias (BIBR); Publicações seriadas e artigos de periódicos (PERI); Artigos de jornais (JORN); e Obras raras (OBRAS). Os termos, em ambos os instrumentos, são apresentados em ordem alfabética e podem ser: termos autorizados, ou seja, termos escolhidos para representar os conceitos; ou termos não autorizados, aqueles que não representam os conceitos pertinentes ao domínio coberto por essas linguagens documentárias, isto é, os sinônimos ou quase sinônimos e as formas alternativas, cujo uso na indexação não está autorizado.

A tabela 59 apresenta uma síntese da metodologia de representação das relações conceituais no VCB e no THES.

Tabela 59: - Representação das relações conceituais no VCBS e no THES

Tipo de relação conceitual/nota	simbologia
o termo em questão deve ser utilizada em lugar de outro que não é aceito como termo padronizado no sistema de conceitos	Usado para (UP)
Relação associativa ou de equivalência, que pode também indicar uma remissiva a outro termo (ver também)	Termo Relacionado (TR)
Relação hierárquica genérico específica (somente no VCBS)	Termo Genérico (TG) ou Termo Específico (TE)
Nota explicativa indicando o contexto ou os limites de utilização do termo a que se refere	Nota Explicativa (NE)

Entre as formas de pesquisa dos termos é utilizado um índice *Key Word Out of Context* (Palavra-chave fora do contexto - KWOC), com a finalidade de permitir ao usuário uma pesquisa alfabética às palavras que compõem descritores ou não descritores, ou seja, de termos nos quais as palavras utilizadas na busca estejam presentes. Segundo Guimarães (1990), o THES

apresenta apenas as relações de equivalência e de associação (USE, UP e TR), sem qualquer hierarquização de termos; os termos são inseridos no sistema a partir de documentos indexados, mas sem qualquer tipo de estruturação. Essa informação, embora date de cerca de 15 anos, é confirmada por Gonçalves (2004), que afirma, ainda, não existirem normas formais ou diretrizes para a manutenção do instrumento. Desse modo, a manutenção do Thes é feita com base na experiência dos técnicos e em outras normas ou diretrizes mais conhecidas, como a norma ISO 2788 ou as normas do IBICT.

Como mencionado anteriormente, a partir do ano 2000, a base de dados NJUR passou a integrar o sistema Legislação Brasileira, disponível para consulta pública na Internet. Essa base de dados integra também a nova versão do SICON. As ilustrações 67, 68 e 69 apresentam as telas de acesso e consulta ao sistema.

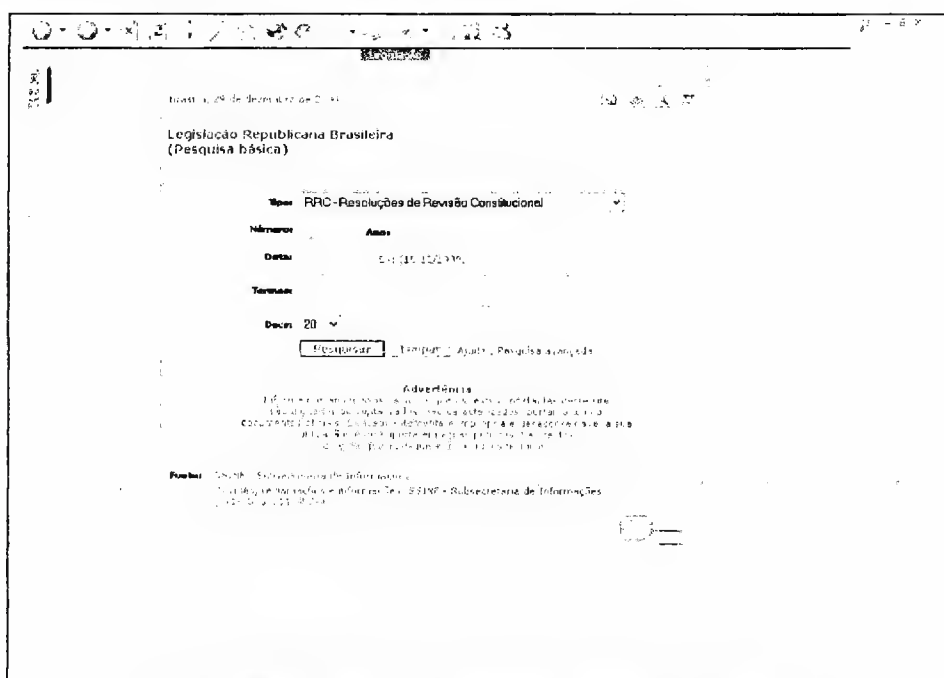


Ilustração 67: Tela inicial do sistema Legislação Federal

Fonte: <http://www.senado.gov.br/sf/legislacao/legisla/>

A tela inicial do sistema dá acesso direto ao formulário de busca que contém os campos *tipo*, *número*, *data*, *termos* e *docs* (quantidade de documentos por página). Uma crítica importante ao sistema é que não existe nas suas páginas, nenhuma documentação histórica a seu respeito, nem a respeito das bases que o integram e nem da unidade orgânica do Senado Federal que o mantém. A ilustração 68 apresenta o formulário de pesquisa avançada, que inclui os campos *ementa*, *origem*, *observação*, *vide*, *indexação* e *catálogo*, além dos anteriormente mencionados. Uma pesquisa ao acervo retorna a tela apresentada na ilustração 69, em que se pode escolher, por meio de um formato de saída intermediário, a norma ou as normas que se deseja consultar.

Acionando-se um ou vários dos links retornados, é possível ter acesso ao formato detalhado do registro. O formato intermediário apresenta o *número* e o *tipo da norma*, além de sua *ementa* e *data de publicação*. O formato detalhado acrescenta *fonte*, *indexação*, *origem*, *catálogo* e *link* para o texto integral (ilustração 69).

Legislação

Brasília, 29 de dezembro de 2004

Legislação Republicana Brasileira
(Pesquisa avançada)

Tipo: Todos

Número: _____ Ano: _____

Data: Ex: 15/12/1993

Ementa: _____

Origem: _____

Observações: _____

Vida: _____

Indexação: _____

Catálogo: _____

Termos: _____

Busc: 20

Presquisar | Ajuda | Pesquisa básica

Fonte: SSINF - Subsecretaria de Informações
 2. www.senado.gov.br/netacgi/nph-brs.exe?sect2=NJURNEW&id2=1640doc&id3=841-845-841-cm&idTab=21

Ilustração 68: Formulário de pesquisa avançada
 Fonte: <http://www.senado.gov.br/sf/legislacao/legisla/avancada.asp>

Senado Federal
Subsecretaria de Informações

Pesquisa Legislação Brasileira

Número de Documentos : 1 de 20 de 175

Visualizar Desmarcar

<input type="checkbox"/>	LEI-8.069-4	12/11/2008	ALTERA A LEI 8.069, DE 18 DE JULHO DE 1990, QUE DISPÕE SOBRE O ESTATUTO DA CRIANÇA E DO ADOLESCENTE E DA OUTRAS PROVIDÊNCIAS.
<input type="checkbox"/>	LEI-8.069-2	15/04/2002	ESTENDE A MÃE ADOTIVA O DIREITO A LICENÇA-MATERNIDADE E AO SALÁRIO-MATERNIDADE, ALTERANDO A CONSOLIDAÇÃO DAS LEIS DO TRABALHO, APROVADA PELO DECRETO-LEI 5.452, DE 1 DE MAIO DE 1942, E A LEI 8.213, DE 24 DE JULHO DE 1991.
<input type="checkbox"/>	LEI-8.069-4	08/11/2000	DA PRIORIDADE DE ATENDIMENTO AS PESSOAS QUE ESPECÍFICA, E DA OUTRAS PROVIDÊNCIAS.
<input type="checkbox"/>	LEI-8.069-2	22/06/2000	ACRESCENTA ARTIGO A LEI 8.069, DE 18 DE JULHO DE 1990, QUE DISPÕE SOBRE O ESTATUTO DA CRIANÇA E DO ADOLESCENTE.
<input type="checkbox"/>	LEI-8.069-70	17/05/2000	INSTITUI O DIA 18 DE MAIO COMO O DIA NACIONAL DE COMBATE AO ABUSO E À EXPLORAÇÃO SEXUAL DE CRIANÇA E ADOLESCENTES.
<input type="checkbox"/>	LEI-8.069-2	06/07/1994	DISPÕE, EM CARÁTER EMERGENCIAL, SOBRE A PRESTAÇÃO DE SERVIÇOS POR ENTIDADES DE ASSISTÊNCIA SOCIAL, ENTIDADES BENEFICENTES DE ASSISTÊNCIA SOCIAL E ENTIDADES DE FINS FILANTRÓPICOS E ESTABELECE PRAZOS E PROCEDIMENTOS PARA O RECADASTRAMENTO DE ENTIDADES JUNTO AO CONSELHO NACIONAL DE ASSISTÊNCIA SOCIAL E DA OUTRAS PROVIDÊNCIAS.
<input type="checkbox"/>	LEI-8.069-2	07/12/1993	DISPÕE SOBRE A ORGANIZAÇÃO DA ASSISTÊNCIA SOCIAL E DA OUTRAS PROVIDÊNCIAS.
<input type="checkbox"/>	LEI-8.069-42	31/03/1993	DISPÕE SOBRE A INSTITUIÇÃO DO PROGRAMA NACIONAL DE ATENÇÃO INTEGRAL À CRIANÇA E AO ADOLESCENTE - PROTECA E DA OUTRAS PROVIDÊNCIAS.
<input type="checkbox"/>	LEI-8.069-2	06/11/1992	CRIA A SECRETARIA NACIONAL DE PROJETOS EDUCACIONAIS ESPECIAIS E DÁ OUTRAS PROVIDÊNCIAS.

Concluído

Windows Media Player | Windows Media Player | Senado Federal | SICON - Sistema... | Normas Jurídicas... | 20:50

Ilustração 69: Formato intermediário para consulta a um registro
 Fonte: <http://www.senado.gov.br/netacgi/nph-brs.exe?sect2=NJURNEW>

Em relação à linguagem de recuperação, o sistema utiliza os operadores listados na tabela 60.

Tabela 60: Linguagem de recuperação no sistema Legislação Brasileira (que contém a base NJUR)

Descrição do operador
\$ ou * (truncamento - vários caracteres)
\$(n) (truncamento de uma quantidade definida de caracteres)
? (truncamento de um caractere)
[] ou . (define um campo de dados específico para a busca)
[-] ou .. (exclui um campo de dados específico da busca)
{ } (usa máscaras ou padrões de cadeias de letras)
ADJ (adjacência com ordem direta)
ADJ(n) (ADJ especificando-se um número de termos)
COM (localiza termos na mesma sentença ou no mesmo campo)
E
MESMO (localiza termos no mesmo parágrafo)
NÃO
NÃO ADJ (não adjacente na ordem direta)
NÃO ADJ(n) (NÃO ADJ especificando um número de termos)
NÃO COM (não na mesma sentença)
NÃO MESMO (não no mesmo parágrafo)
NÃO PROX (não adjacente em qualquer ordem)
NÃO PROX(n) (NÃO PROX especificando um número de termos)
OU
PROX (adjacência em qualquer ordem)
PROX(n) (PROX especificando-se um número de termos)
XOU (ou exclusivo;

Fonte: Tela de ajuda do sistema

As ilustrações 70 e 71 apresentam as telas de pesquisa do Sistema SICON, que, confirmando informações de Gonçalves (2004), passou a apresentar de maneira mais interativa a reunião das seguintes bases: Bibliotecas da Rede RVBI, Discursos de Senadores, a base NJUR, atualmente (2004), denominada Legislação Federal e a base de Matérias com tramitação no Senado.

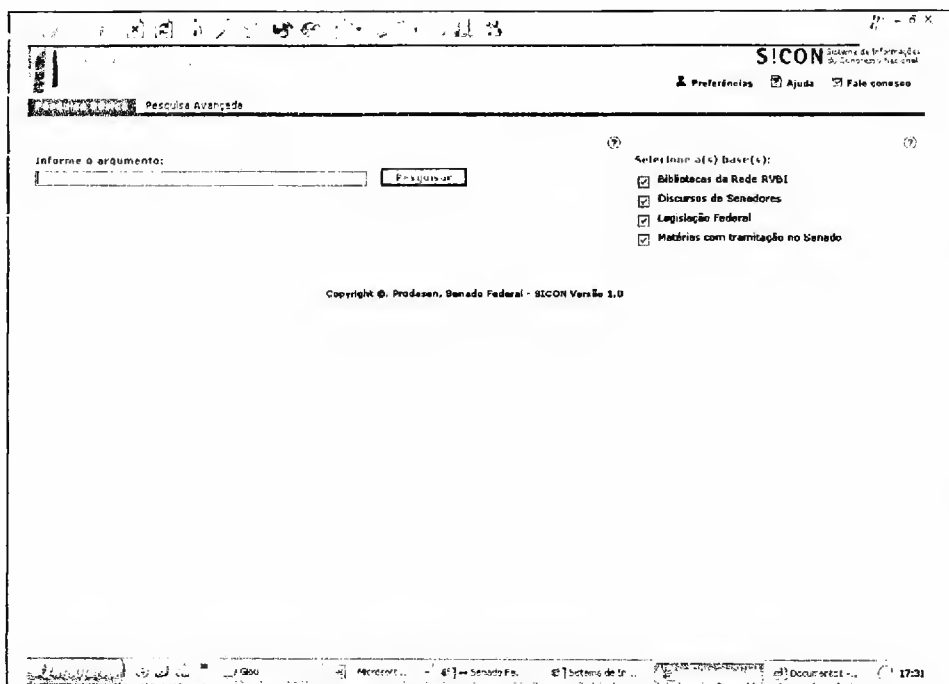


Ilustração 70: Tela de pesquisa básica do sistema SICON

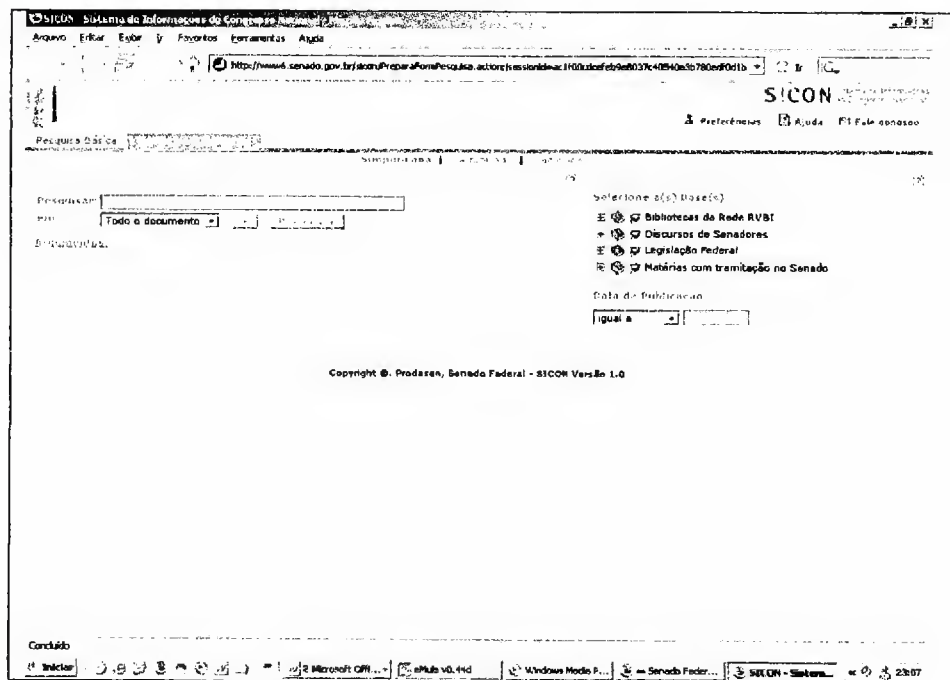


Ilustração 71: Tela de pesquisa avançada do SICON

3.2.2 Sistema de Apoio ao Processo Legislativo (SAPL)

O Sistema de Apoio ao Processo Legislativo (SAPL), desenvolvido no ano de 2001, pelo Programa Interlegis, está atualmente (2004) em sua sétima versão, que recebeu o número 2.0.2. O objetivo inicial do projeto era propiciar apoio às Casas Legislativas na organização das informações relativas ao processo legislativo, com vistas a sua divulgação na Internet. De acordo com Ciciliati Jr. (2004), no entanto, à medida que o sistema se popularizou entre as casas legislativas conveniadas ao Programa Interlegis, surgiram algumas necessidades de ajustes, e o sistema tem se tornado cada vez mais eficiente. Entre as principais características do sistema estão: a) a possibilidade de o parlamentar elaborar e enviar eletronicamente suas proposições legislativas – em versões anteriores à 2.0.2 apenas os funcionários da Casa Legislativa possuíam privilégios de inclusão de proposições; b) incorporação automática do texto integral das proposições; c) indexação de texto integral das normas jurídicas, com pesquisa por tipo de norma, número, ano e assunto ou pesquisa textual livre; d) incorporação de editor de proposições em XML, que auxilia na elaboração de documentos com estrutura correta; e) versão atualizada para a plataforma ZOPE 2.7; f) possibilidade de se cadastrarem comissões permanentes, temporárias e especiais com dados básicos e complementares; g) criação de pauta de sessões e com atualização diária; h) cadastro de parlamentares com dados de identificação e de qualificação; i) possibilidade de cadastro de normas legislativas das esferas federal, estadual e municipal, bem como seus textos integrais.

Conforme relata Ciciliati Jr. (2004), o desenvolvimento do SAPL iniciou-se a partir de um levantamento no sistema NJUR, mantido pelo PRODASEN, comparando-se as suas características com as necessidades das câmaras municipais e das assembleias legislativas estaduais. Como a iniciativa é totalmente embasada na filosofia de *software* livre⁷¹, o Programa Interlegis motivou a criação de uma comunidade para a sustentação do projeto. Tal comunidade atua em cooperação, por meio de uma lista de discussão e de um *site* denominado Portal de Desenvolvimento Colaborativo Interlegis (COLAB)⁷², com o objetivo de aprimorar o SAPL, bem como de compartilhar outros aplicativos desenvolvidos pela comunidade, a exemplo do Sistema de Gravação de Áudio (SGA), destinado à Taquigrafia de seções plenárias e do Sistema de Recursos Humanos (SRH). Outro aplicativo disponível no portal é o *Structured Document Editor* (SDE), que é um editor de documentos XML para a plataforma Zope. O programa foi

⁷¹ Informações sobre *software* livre podem ser encontradas no sítio da *Free Software Foundation*, disponível em: <<http://www.gnu.org/>>, e no *Free Software Directory* <<http://directory.fsf.org/>>. Detalhamentos sobre os tipos de licenças estão disponíveis em <<http://www.gnu.org/philosophy/license-list.html>>.

⁷² Para acesso ao Colab, visite: <<http://genesis.interlegis.gov.br/>>

desenvolvido pelos técnicos do Interlegis e, a partir da definição do modelo da estrutura do documento, é capaz de guiar o usuário na criação de documentos totalmente compatíveis com a linguagem XML. O editor é personalizável e compatível com os navegadores Mozilla 1.5 e Internet Explorer 6.0 e está disponível em linha para testes⁷³.

A intenção dos técnicos com essa iniciativa é habilitar os usuários finais ao desenvolvimento de documentos XML, mesmo se não detiverem nenhum conhecimento sobre a linguagem. O SDE foi incorporado à nova versão do sistema SAPL, para organizar a entrada das informações na etapa de registro denominada "Elaboração de Proposições".

Além desses e de outros aplicativos, os membros da comunidade estão trabalhando no desenvolvimento de um modelo XML para a representação dos documentos legislativos brasileiros, de modo que possam ser validados com XML-Schemas e convertidos em HTML com XSLT. A ilustração 72 e a tabela 61 apresentam alguns elementos e subelementos que já foram definidos.

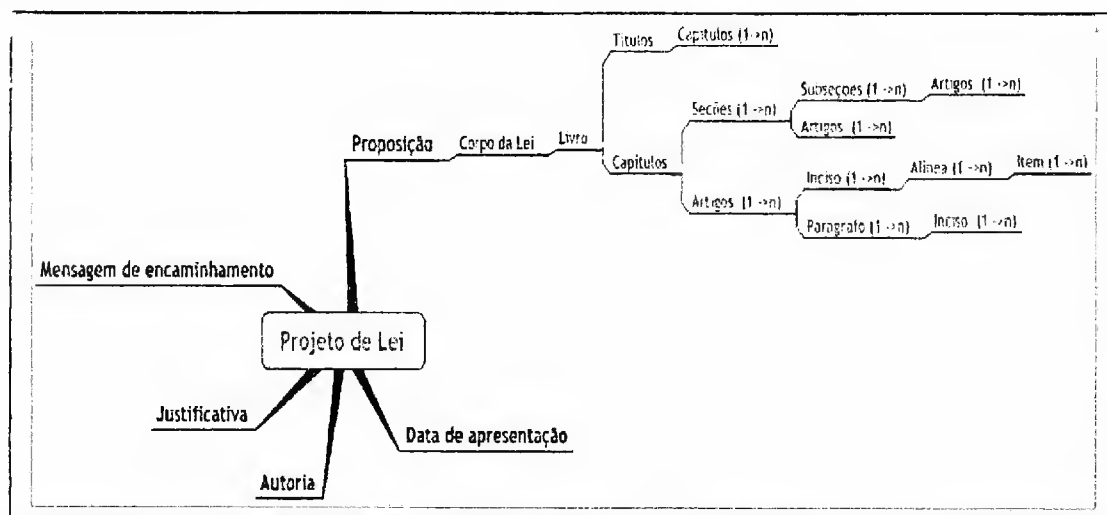


Ilustração 72: estrutura de Projeto de Lei elaborada pela Comunidade Interlegis para serialização com XML
Fonte: adaptado de Colab (2004)

Tabela 61: descrição dos elementos

Elemento	Descrição
Justificativa	Exposição de motivos redigida pelo autor a respeito do mérito da proposição
Data de Apresentação	Data em que o projeto foi apresentado, normalmente no formato: 'Sala das Sessões, dia de mês de ano'
Autores	Autores que subscrevem a proposição
Mensagem de Encaminhamento	Quando o autor do Projeto de Lei é o Poder Executivo
Proposição Legislativa	Texto da lei
Epígrafe	Título designativo da matéria
Ementa	Resumo que apresenta o objetivo da Lei
Preâmbulo	Determina a forma de legitimação

Fonte: adaptado de Colab (2004)

⁷³ Para acesso ao ambiente de testes do SDE, visite: <<http://genesis.interlegis.gov.br/colab/Editor>>

As ilustrações 73 a 76 apresentam algumas telas de interação com o sistema SAPL. Mais detalhes sobre as características do sistema são apresentados nas tabelas 64 a 68 (seção 3.2.5 – comparação entre os sistemas), em comparação com os demais sistemas estudados.

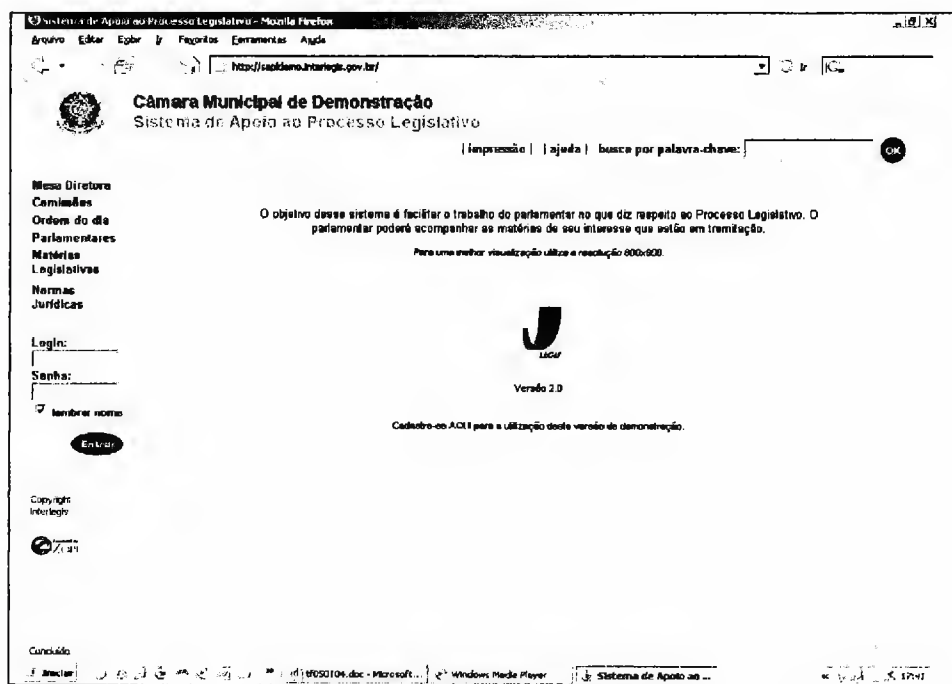


Ilustração 73: Tela inicial do sistema SAPL

Fonte: <http://sapdemo.interlegis.gov.br/>

A tela inicial do sistema apresenta opções de navegação à esquerda, entre as quais a opção *Normas Jurídicas*, que dá acesso à tela de pesquisa. Essa é a tela que está disponível para os usuários externos, por meio da *Internet* (ilustração 73). A pesquisa pode ser realizada por *tipos de normas*, *número*, *ano*, *assunto* ou por *texto* (ilustração 74). Não há no sistema de ajuda uma explicação para o campo *texto*, em *Pesquisa Textual* e nem para a opção *busca por palavra chave*, no canto superior da tela. Para usuários cadastrados, estão disponíveis os campos de *login* e *senha*, que dão acesso aos módulos de cadastro. As ilustrações 75 e 76 apresentam respectivamente os formatos de saída intermediário e detalhado para a localização e exame dos registros recuperados após uma consulta na *Internet*.

Câmara Municipal de Demonstração
Sistema de Apoio ao Processo Legislativo

Impressão | Ajuda | Busca por palavra-chave:

Mesa Diretora
Comissões
Ordem do dia
Parlamentares
Métricas Legislativas
Normas Jurídicas

Logim
Senha:
 lembrar nome
Entrar

Tipo da Norma: Anex:

Número:

Assento:

Texto:

Copyright Interlegis

Concluído

Ilustração 74: Formulários de consulta a normas jurídicas
Fonte: <http://sapldemo.interlegis.gov.br>

Câmara Municipal de Demonstração
Sistema de Apoio ao Processo Legislativo

Impressão | Ajuda | Busca por palavra-chave:

Mesa Diretora
Comissões
Ordem do dia
Parlamentares
Métricas Legislativas
Normas Jurídicas

Logim
Senha:
 lembrar nome
Entrar

Copyright Interlegis

Concluído

IND 752 2002 - Indicação
NO SENTIDO DE IMPLANTAR UM LOCAL - CIDADE DA CRIANÇA AO LADO DO CAMPO DE FUTEBOL DO BOM PASTOR.
Autor: João Rocha
Localização Atual: Executivo Municipal
Situação: Aprovado por unanimidade
Última Ação: Encaminhado pelo nº no 1 45802

PLO 174 2002 - Projeto de Lei Ordinária
DECLARA DE UTILIDADE PÚBLICA O PROGRAMA BENEFICENTE CRIANÇA CIDADÃO DO FUTURO.
Autor: Prefeito Municipal - Prof. Félix Sábido Júnior
Localização Atual: Executivo Municipal
Situação: Aprovado por unanimidade
Última Ação: Aprovado em 2ª discussão por unanimidade. Autógrafo nº

PLO 193 2002 - Projeto de Lei Ordinária
AUTORIZA O PODER EXECUTIVO MUNICIPAL A DOAR UM IMÓVEL DE SUA PROPRIEDADE AO PROGRAMA CRIANÇA CIDADÃO DO FUTURO E DAS OUTRAS PROVIDÊNCIAS
Autor: Prefeito Municipal - Prof. Félix Sábido Júnior
Localização Atual: Executivo Municipal
Situação: Devolvido ao autor
Última Ação: A pedido do Vereador Marcos Crippa.

RFO 646 2001 - Requerimento
SOLICITANDO AO PREFEITO MUNICIPAL DOAÇÃO DE TERRENO AO PROGRAMA CRIANÇA - CIDADÃO DO FUTURO.
Autor: José Henrique Montagner
Localização Atual: Plenário
Situação: Aprovado por unanimidade
Última Ação: Encaminhado pelo of. nº 58001

REQ 927 2001 - Requerimento
SOLICITANDO AO PREFEITO MUNICIPAL DOAÇÃO DE ÁREA AO PROGRAMA CRIANÇA CIDADÃO DO FUTURO

Ilustração 75: Formato intermediário de saída para a escolha do registro pelo usuário
Fonte: <http://sapldemo.interlegis.gov.br>

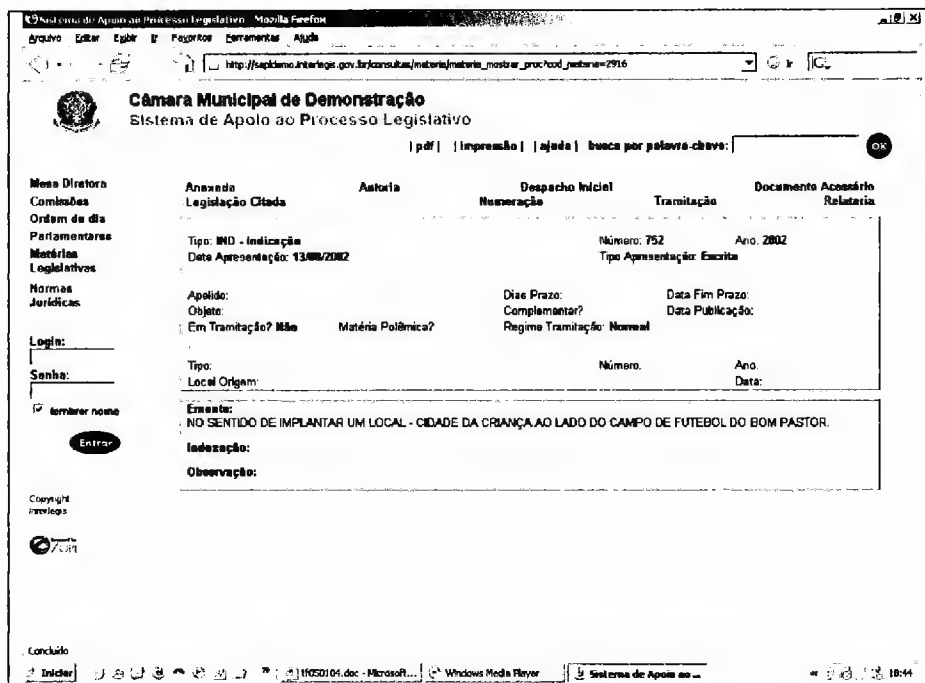


Ilustração 76: Formato detalhado de saída para o exame do registro pelo usuário

Fonte: <http://sapldemo.interlegis.gov.br>

O sistema permite, ainda, visualização e a impressão do texto integral da norma. No entanto, após a realização de uma consulta a estrutura do sistema é falha, pois não apresenta uma barra de navegação, para a redefinição ou refinamento dos argumentos da consulta.

3.2.3 Sistema de Normas Jurídicas do Distrito Federal (NJUR-DF)

A iniciativa de disponibilizar a informação legislativa originária da Câmara Legislativa em meio eletrônico e hipertextual partiu da Coordenadoria de Modernização e Informática daquela Casa quando, no ano de 1997, publicou em um disquete a Lei Orgânica do Distrito Federal e suas emendas. O documento, elaborado no formato de arquivos de ajuda do sistema operacional MS Windows 95, foi ampla e gratuitamente distribuído à população. Posteriormente, com o a criação do *site* da Câmara na *Internet*, iniciaram-se os trabalhos para a publicação das normas legislativas emanadas daquele órgão.

Paralelamente a esses esforços, o Tribunal de Contas do Distrito Federal institucionalizava, por meio da portaria número 290/98, o Sistema Integrado de Legislação do TCDF (SILEGIS), cujo objetivo era “reunir as normas jurídicas de interesse do Tribunal em um banco de dados único, destinado a consultas por meio informatizado” (TCDF, 1998). Sob a coordenação da Divisão de Documentação daquele órgão, o sistema foi implementado e iniciou suas operações reunindo, entre outras normas, a Lei Orgânica do Distrito Federal, as Leis Complementares e as Leis Ordinárias. No ano de 1999, interesses comuns motivaram a Coordenadoria de Modernização e Informática da Câmara Legislativa e a Divisão de

Documentação do Tribunal de Contas a proporem o estabelecimento de um acordo de cooperação entre as duas casas, no sentido de que o sistema pudesse ser aprimorado. A partir de então, toda a legislação oriunda da Câmara Legislativa passou a ser publicada no próprio *site* da Casa, contando com a colaboração de servidores do Tribunal de Contas (ilustração 77). Tal medida foi interessante para o Tribunal, por um lado, pois propiciou um alívio na carga de trabalho dos seus recursos humanos envolvidos no projeto; e para a Câmara Legislativa por outro, uma vez que facilitou o processo de publicação, tornando possível disponibilizar as leis quase imediatamente à sua aprovação em plenário.

Ao longo do tempo o sistema foi incrementado com novas rotinas, mas se mantém em funcionamento com o uso das mesmas tecnologias adotadas desde 1999. Atualmente (2004) a base do sistema é composta pelo seguinte acervo: Lei Orgânica do Distrito Federal e Emendas; Leis Complementares; Leis Ordinárias (essas para acesso público) e Resoluções (para acesso apenas pela rede interna da Casa).

A despeito da sua grande utilidade e necessidade para a comunidade do Distrito Federal, o sistema padece restrições técnicas, que comprometem o seu desempenho, a sua qualidade e, conseqüentemente, o cumprimento da sua missão: lentidão na composição automática dos índices; perda sistemática de índices; inexistência de controle terminológico em relação a número, gênero, abreviaturas, siglas e outros tipos de relações entre os termos; baixa precisão na recuperação textual.

Câmara Legislativa do Distrito Federal
Bem-vinda ao Portal CLDF

Fale com a Câmara

Legislação

Distrito Federal

NOTA IMPORTANTE
Fale conosco se os textos integrantes das bases de dados receberam conferência, e que não houve a ocorrência de falhas. No caso de ser identificada alguma impropriedade, solicitamos sua comunicação por meio do e-mail webmaster@cl.df.gov.br.

Leis Ordinárias: textos completos e redações atualizadas

Leis Complementares: textos completos e redações atualizadas

Leis Distritais: textos completos

Resoluções do CLDF: textos completos disponíveis na base de documentos da CLDF

Permite por Assunto, permite a consulta por palavras-chaves

Busca por número, emite a consulta por números de leis complementares ou Resoluções

Ilustração 77: Sistema de Informações Legislativas da CLDF

Fonte: www.cl.df.gov.br

Anteriormente ao acordo de cooperação mantido com o TCDF para a publicação dos textos integrais, outra unidade orgânica da CLDF, o Setor de Pesquisa e Recuperação da Informação, mantinha, desde o início das atividades daquela Casa, a base de dados NJUR, devido a um convênio com o PRODASEN. Até o ano de 2003 a base NJUR funcionava como uma rede cooperativa que recebia a inclusão de referências de normas jurídicas de diversos órgãos de Governo, entre eles a Câmara Legislativa. No entanto, houve uma iniciativa do PRODASEN para descontinuar a rede e a CLDF viu-se obrigada a migrar suas informações para outro sistema. Como o sistema SILEGIS não estava preparado para a inclusão de informações referenciais, pois é totalmente baseado no armazenamento de textos integrais em formato HTML e indexados automaticamente, a Câmara viu-se diante da necessidade do desenvolvimento de novo aplicativo que foi informalmente denominado NJUR-DF (ROCHA, 2004). A data prevista para a implantação do novo sistema é março de 2005.

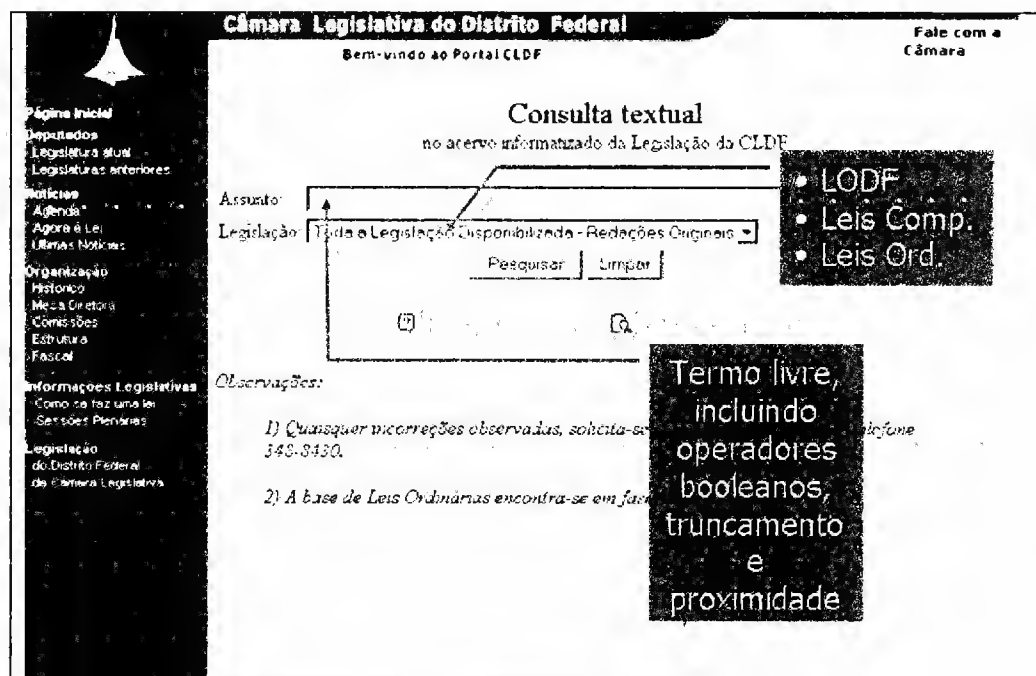


Ilustração 78: Tela de consulta textual do Sistema de Informações

Fonte: www.cl.df.gov.br

Não obstante a iniciativa de se desenvolver o NJUR-DF, o sistema SILEGIS mantém-se em atividade, propiciando aos usuários duas modalidades de pesquisa: uma textual e outra por número da norma legislativa (ilustrações 78 e 79). Na consulta textual é possível a escolha entre as redações originais e as redações atualizadas da Lei Orgânica do Distrito Federal, das Leis Complementares e das Leis Ordinárias. São admitidos os operadores E, OU, NÃO, NEAR e truncamento à direita.

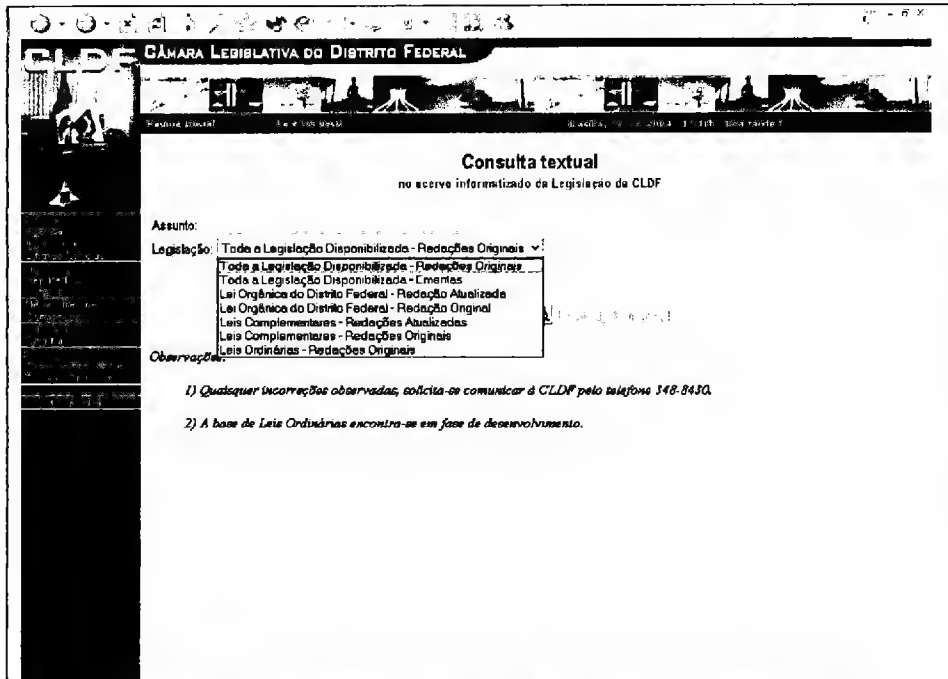


Ilustração 79: Tela de consulta textual no sistema de informações da CLDF
Fonte: www.cl.df.gov.br

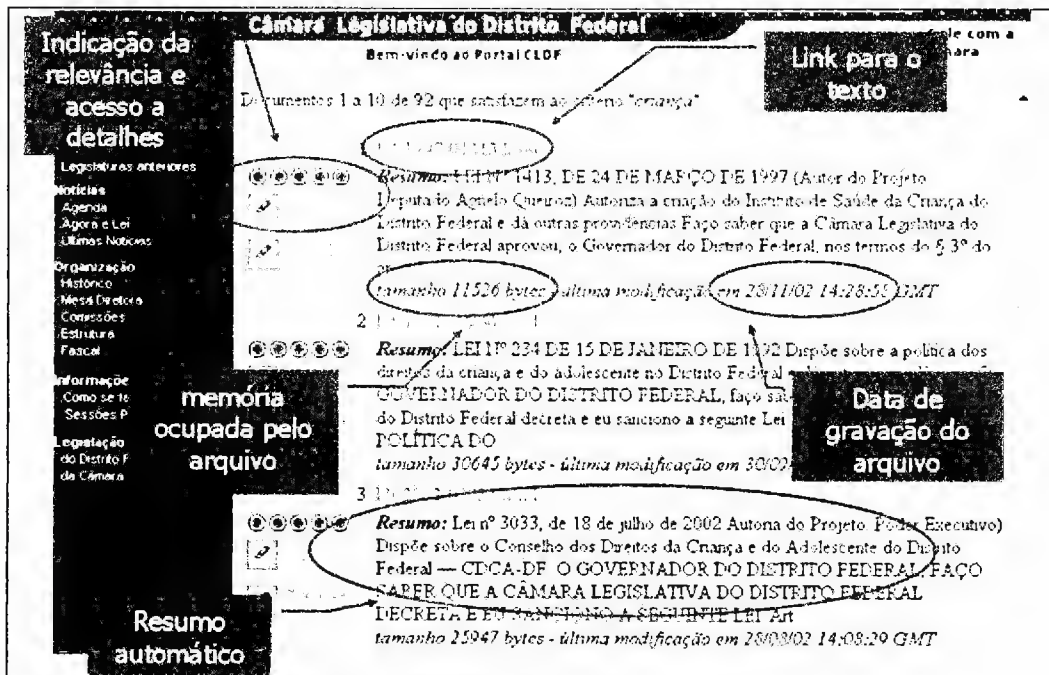


Ilustração 80: Formato de saída de um registro
Fonte: www.cl.df.gov.br

A ilustração 80 apresenta os resultados de uma pesquisa por assunto no sistema. O registro possui as seguintes informações: a) resumo automático, obtido a partir da extração de certa quantidade de parágrafos do texto integral; b) data e hora da gravação do texto; c) tamanho em bytes; d) link para o texto integral; e) indicação da relevância do texto com base na frequência

de ocorrência do termo. O indexador utilizado é o Microsoft Index Server, serviço incorporado ao sistema operacional desde a versão NT 4.0. A linguagem de consulta é Active Server Pages (ASP). A ilustração 81 apresenta o detalhamento da busca. Uma vez que a indexação é feita automaticamente, não controle terminológico com relação a aspectos como sinonímia, variações de número e gênero, polissemia, entre outros fenômenos lingüísticos.

Câmara Legislativa do Distrito Federal
Bem-vindo ao Portal CLDF

Fale com a Câmara

LEI Nº 1413, DE 24 DE MARÇO DE 1997
(Autor do Projeto: Deputado Agnelo Queiroz)

Autoriza a criação do Instituto de Saúde da Criança do Distrito Federal e dá outras providências

Para saber que a Câmara Legislativa do Distrito Federal aprovou, o Governador do Distrito Federal, a seguinte Lei:

Lei

Art. 1º Fica o Poder Executivo autorizado a criar, na estrutura da Secretaria de Saúde do Distrito Federal, o Instituto de Saúde da Criança

Art. 2º O Instituto de Saúde da Criança tem os seguintes objetivos:

I - construir, organizar e manter em funcionamento, no Distrito Federal, um hospital infantil que atenda todas as crianças de até 14 (quatorze) anos de idade, com diagnóstico de doenças crônicas e graves, e de doenças infecciosas e parasitárias;

Art. 2º O Instituto de Saúde da Criança tem os seguintes objetivos:

I - construir, organizar e manter em funcionamento, no Distrito Federal, um hospital infantil que atenda todas as crianças de até 14 (quatorze) anos de idade, com diagnóstico de doenças crônicas e graves, e de doenças infecciosas e parasitárias;

Se o termo procurado fosse "crianças", esta lei não apareceria nos resultados

O termo procurado aparece em destaque

Ilustração 81: detalhamento da ocorrência de um termo no texto pesquisado

Fonte: www.cl.df.gov.br

De acordo com ROCHA (2004) e LIMA (2004) o advento do novo sistema deverá resolver o problema do controle terminológico – ao menos quanto à padronização das entradas – uma vez que, ao contrário do SILEGIS, utilizará indexação intelectual. Os técnicos prevêm para a linguagem de recuperação, o uso dos operadores lógicos *E*, *OU*, *NÃO*, *XOU*, bem como dos operadores contextuais (tabela 62), além de operadores de comparação de cadeias de caracteres (tabela 63).

Tabela 62: operadores contextuais previstos para o NJUR-DF

Operadores Contextuais	
Adj	termo adjacente (termos juntos e na ordem dada)
Near	termos próximos, não necessariamente na mesma ordem.
With	pesquisa palavras na mesma sentença ou frase (engloba c near)
And	pesquisa os termos em qualquer contexto

Fonte: (ROCHA, 2004)

Tabela 63: operadores de comparação

Operadores de comparação de cadeias de caracteres	
=(EQ)	igual a
NE	diferente de
>(GT)	maior que
<(LT)	menor que
NG	menor ou igual
NL	maior ou igual
WL	entre limites; limites separados por vírgula
OL	Fora de limites; limites separados por vírgula

Fonte: www.ci.df.gov.br

3.2.4 Sistema Legislação Informatizada (LEGIN)

O sistema Legislação Informatizada, da Câmara dos Deputados, foi desenvolvido pelo Centro de Documentação e Informação, por meio da Seção de Legislação Brasileira, em conjunto com o Centro de Informática, com o objetivo de atender rápida e eficazmente as necessidades de informação legislativa dos deputados, servidores e demais usuários. O sistema inicialmente denominado Sistema de Legislação Brasileira (SLB), foi iniciado no ano 2000, e contempla as seguintes normas: a) as de hierarquia superior como leis, decretos, decretos-lei, medidas provisórias, decretos legislativos; b) resoluções da Câmara dos Deputados; c) Resoluções do Congresso Nacional; d) legislação inferior relevante aos objetivos institucionais (marginália) (CÂMARA DOS DEPUTADOS, 2000).

O registro de normas no sistema é feito a partir da leitura dos documentos de Imprensa Oficial, a saber: Diário Oficial da União, Diário da Justiça, Diário do Congresso Nacional, Diário da Câmara dos Deputados e Diário do Senado Federal, além do Boletim da Câmara dos Deputados. Após a leitura é feita a indexação e os atos selecionados são incluídos no sistema. O objetivo da instituição é recuperar toda a legislação brasileira retroativamente ao início do registro das atividades legislativas brasileiras. No entanto, até o momento presente (2004) foi possível a conversão para formato eletrônico das normas de hierarquia superior de 1946 a 2004. Para garantir o sucesso desse trabalho, a Câmara dos Deputados realizou a contratação de uma empresa para a digitação e conferência dos atos normativos publicados nos veículos de Imprensa

Oficial retroativamente a 1946. Com relação à atualização diária, os textos são obtidos por meio dos *sites* institucionais, como, por exemplo o *site* do Departamento de Imprensa Nacional, em formato PDF. Uma vez localizada a norma, o usuário copia seu conteúdo para a memória e o insere no formulário de entrada do LEGIN. Em operações dessa natureza a perda de formatação é freqüente, portanto, foi desenvolvido um módulo para a correção e reformatação dos textos, de modo que se tornem apresentáveis na forma que determina a técnica legislativa (ALVES, 2004).

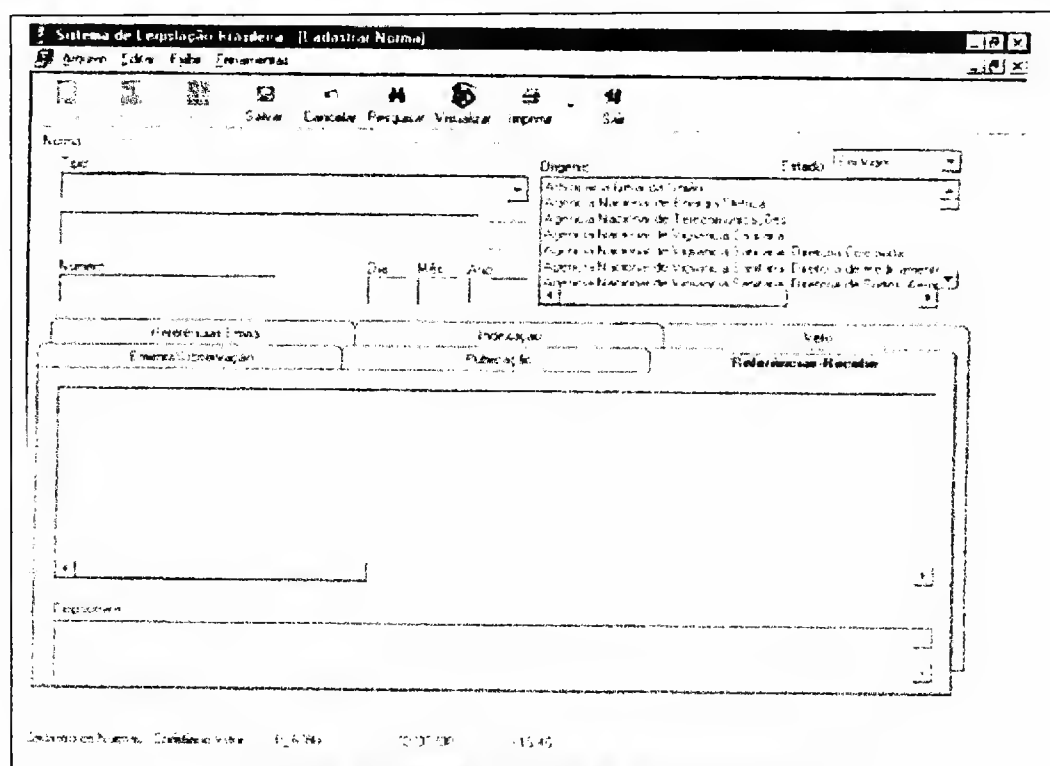


Ilustração 82: Sistema Legin - tela de cadastro de normas
 Fonte: (CÂMARA DOS DEPUTADOS, 2000)

A ilustração 82 apresenta a tela de cadastro de normas do sistema. Em relação ao módulo de pesquisa, o sistema dispõe de uma tela de pesquisa estruturada (ilustração 83), que permite a busca por *tipo de norma, número, ano, origem, ementa, indexação, estado e data de publicação*. A linguagem de recuperação admite os operadores *E* e *OU* e é feita por campos. Na pesquisa avançada é possível o cruzamento de valores de campos por meio do uso de operadores de comparação (*contém, igual, não contém*).

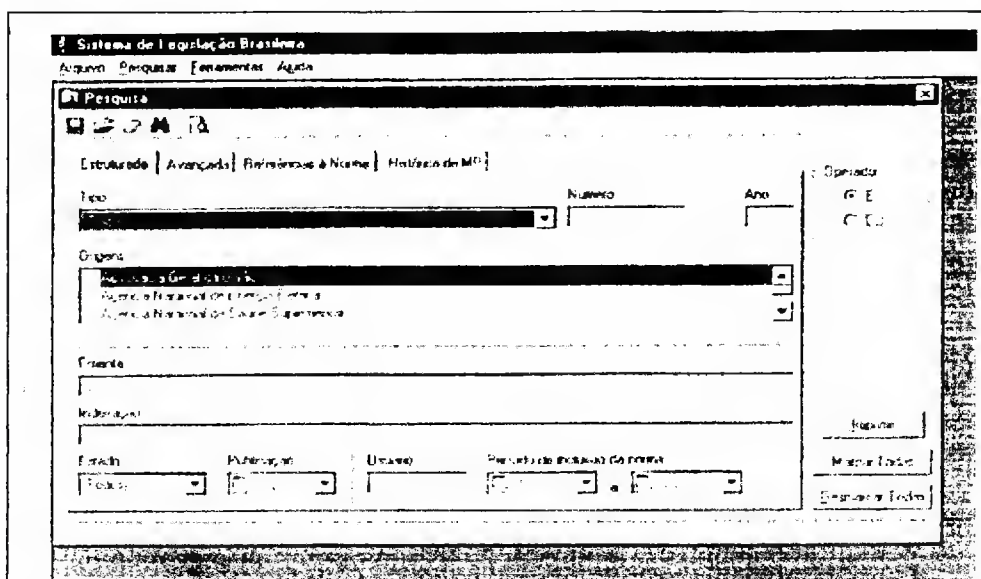


Ilustração 83: Sistema Legin – tela de pesquisa estruturada
 Fonte: (CÂMARA DOS DEPUTADOS, 2000)

Devido a uma série de revisões de conteúdo da base de dados e também das características tecnológicas do sistema, a aplicação só está disponível para o público interno da Câmara dos Deputados, por meio da Intranet.

3.2.5 Comparação entre os sistemas

As tabelas 64 a 68 apresentam dados que permitem comparação entre os sistemas estudados, apontando suas principais características em relação a: tipo de acervo; formato do registro documentário; tipo de indexação; estágio de desenvolvimento; público alvo; objetivos; sistema operacional; linguagens de programação; sistema de armazenamento do acervo; linguagem de recuperação; caracterização do módulo de pesquisa; uso tecnologias da web semântica ou similares.

Tabela 64: Características de identificação do acervo -

Característica	Sistema				
	SAPL	NJUR/Constituição Federal	NJUR-DF	LEGIN	
Acervo (tipos de normas que o sistema abrange)	Varia de acordo com as necessidades da casa legislativa	Ver tabela 59	Ver tabela 59	Ver tabela 59	
Formato do registro documentário	Ver tabela 60	Ver tabela 60	Ver tabela 60	Ver tabela 60	
Indexação	Controle de vocabulário	Não utiliza, pois os usuários finais ainda não possuem habilidades necessárias à indexação. É prevista a implementação de suporte a tesouros no sistema	Utiliza Thesaurus do Senado Federal, mas não há diretrizes escritas, estabelecidas formalmente. Dispõe também de um índice KWOC. O controle de sinonímia, ambigüidade, polissemia e a lematização são realizados com base na experiência dos técnicos. Termos não-autorizados são indicados por meio de remissivas	Pretende-se utilizar uma versão impressa do Thesaurus do Senado Federal e novos termos estão sendo incluídos manualmente. O controle de vocabulário será baseado nas diretrizes elaboradas para o sistema da Rede RVBI ⁷⁴ .	Utiliza o Thesaurus do Senado Federal. Foi constituída uma equipe para o desenvolvimento do tesouro da Câmara
	Método	Se houver indexação pelo usuário final, é feita manualmente	Manual	Manual (há intenção do desenvolvimento de um aplicativo para o gerenciamento do tesouro)	Manual. Está em fase de estudos a implementação de um módulo para gerenciamento de tesouros

Fonte: sistematização do autor com base nas entrevistas concedidas pelos técnicos dos SIL

⁷⁴ Para mais informações sobre a Rede Virtual de Bibliotecas do Senado Federal, recomenda-se uma consulta ao *site* <http://www.senado.gov.br/sf/biblioteca/rvbi_historico.asp>

Tabela 65: Características de identificação do Sistema

Característica	Sistema			
	SAPL	NJUR/Constituição Federal	NJUR-DF	LEGIN
Nome	Sistema de Apoio ao Processo Legislativo	Banco de dados Normas Jurídicas, que integra o Sistema de Informações do Congresso Nacional. Para a Constituição, foi desenvolvido o sistema Constituição Federal	Sistema de Normas Jurídicas do Distrito Federal	Sistema de Legislação Informatizada
Situação	Em produção desde jan/02 (como é baseado em software livre, está em constante desenvolvimento)	Em nov/2004 entrou em funcionamento uma nova versão do SICON, que integra as bases de dados de discursos, matérias em tramitação, do NJUR (1972) e da biblioteca do SF. Além do SICON, está disponível o sistema Constituição Federal, desde out/2003. O tesouro e o vocabulário controlado foram convertidos para meio eletrônico em 1998	Em desenvolvimento	Em produção desde jan/2001. Novos módulos estão sendo implementados
Público alvo	Casas legislativas estaduais e municipais	Parlamentares, magistrados, advogados, estudantes e sociedade	Parlamentares, assessores parlamentares, demais órgãos do Distrito Federal e sociedade	Deputados, Senadores, Órgãos internos e sociedade. Atualmente o sistema está disponível apenas na intranet
Objetivos		Divulgação e democratização da informação legislativa, compreendendo uma visão histórica da evolução do texto constitucional e o acervo de normas de hierarquia superior	Manter um acervo das normas jurídicas no âmbito do DF, incluindo texto integral, indexação técnica, suporte à pesquisa e registro da correlação entre as normas. Suprir a carência deixada pelo fato de o PRODASEN haver descontinuado a rede NJUR	Democratizar o acesso à informação legislativa, incluindo toda a legislação de hierarquia superior, desde 1946.
Usuários	Entre 50 e 100 casas legislativas utilizam. Como a filosofia é de software livre, há casas que utilizam mas não mantêm contato com o Programa Interlegis	Milhares de usuários, já que o sistema está disponível na Internet	Milhares de usuários, já que o sistema estará disponível na Internet	Milhares de usuários, já que o sistema estará disponível na Internet

Fonte: sistematização do autor, com base nas entrevistas com os técnicos dos SIL

Tabela 66: características tecnológicas do sistema

Característica	Sistema				
	SAPL	NJUR	NJUR-DF	LEGIN	
Sistema operacional e rede	Sistema Operacional	Embora seja portátil em Windows, por questões ideológicas, seu uso é incentivado apenas no linux. Como a linguagem Python ⁷⁵ é um padrão Linux, o sistema pode rodar em qualquer máquina	Diversos, incluindo main frame e plataformas de microinformática	Windows, podendo ser portátil em outros ambientes	Windows
	Abrangência	Totalmente baseado em web, pode servir à Intranet e a intranets	O cadastro das normas utiliza plataforma cliente servidor em rede local LAN; a consulta é feita por meio da intranet e da Internet	O cadastro de normas utilizará ambiente Cliente-Servidor e a pesquisa será feita por meio da intranet e da Internet	Arquitetura Cliente-Servidor para cadastro e manutenção; Intranet e Internet (ainda inoperante) para pesquisa
	Servidor web	ZOPE	Para a Constituição Federal, usa-se IIS 5.0 e para as normas do NJUR, Apache. Quanto ao novo SICON, há previsão de implementação no Oracle-Application-Server	Internet Information Services	Internet Information Services
Linguagens de programação	Python e XML implementados em ambiente ZOPE	Java e XML	Java, JSP e Struts	Plataforma .NET, incluindo implementações em VB.NET, ASP.NET e JavaScript	
Sistema de armazenamento do acervo	Texto integral em formato XML. Usa XMLSchema para controlar a sintaxe e XSLT para apresentar o conteúdo em HTML	Atualmente o acervo de normas está armazenado no banco de dados Full Access, desenvolvido por técnicos da Casa para rodar em um computador de grande porte. Há previsão da migração para Oracle em 2005. A constituição armazena texto integral em formato XML e utiliza XSLT para a conversão em HTML.	Banco de dados relacional Informix ⁷⁶	Banco de dados relacional Oracle. Uso da linguagem PL/SQL para a execução de procedimentos predefinidos (stored procedures). Disponibilidade de texto integral.	

⁷⁵ Segundo Peter Norvig (apud PYTHON, 2004), a Linguagem tem tido participação importante no Google, desde seu início e essa conjuntura permanece, à medida que o sistema se desenvolve. Mais detalhes sobre a linguagem estão disponíveis no site: <<http://www.python.org/>>

⁷⁶ Para mais informações sobre o banco de dados Informix, recomenda-se uma visita ao site <<http://www-306.ibm.com/software/data/informix/pubs/>>

Característica	Sistema			
	SAPL	NJUR	NJUR-DF	LEGIN
Linguagem de recuperação	Utiliza a tecnologia TextIndexNG ⁷⁷ que permite recuperação por radicais, por proximidade, dá suporte a listas de termos não-indexados (stop words), indexação textual, truncamento à esquerda, pesquisa de subcadeias (substrings), além dos operadores booleanos tradicionais	Utiliza os operadores WITH, ADJ, AND, OR, NOT, SELECT()	É prevista a utilização de E, OU, NÃO, XOR, ADJ, NEAR, WITH, entre outros operadores. Prevê a utilização do Excalibur, um módulo de recuperação textual do Informix	São aceitos os booleanos E, OU, NÃO, os operadores de truncamento à esquerda e à direita, além do operador NEAR()
Módulo de pesquisa	O formulário de busca parte dos próprios campos do registro documentário. Não há busca por frases booleanas.	No sistema Constituição Federal não há módulo de busca, apenas navegação hipertextual. Quanto ao sistema de busca de normas jurídicas, a pesquisa é feita por campos e pode ser básica ou avançada	Será feita por campos (indexação, catálogo, lista de nomes próprios, observação, ementa, entre outros) Será possível a pesquisa livre em em texto integral.	
Uso de tecnologias da web semântica ou similares	Como os objetivos do sistema ainda estão relacionados com a inclusão digital das diversas casas legislativas brasileiras, esta ainda não é uma preocupação dos técnicos envolvidos	Estão sendo iniciados estudos para o uso de Topic Maps e do padrão Dublin Core	Esta ainda não é uma preocupação dos técnicos envolvidos	Esta ainda não é uma preocupação dos técnicos envolvidos

Fonte: sistematização do autor com base nas entrevistas concedidas pelos técnicos dos SIL

⁷⁷ Para mais informações, recomenda-se consultar o *site*: <<http://www.zopyx.com/OpenSource/TextIndexNG>>

Tabela 67: Tipos de normas legislativas disponíveis nos sistemas estudados

Tipos de normas legislativas disponíveis	NJUR		NJURDF		LEGIN		SAPL	Freq
	A	B	A	B	A	B	A	
Aditamento a Atos Convocatórios	<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>		2
Ato Convocatório	<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>		2
Ato da Mesa	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		3
Ato da Mesa da ANC						<input checked="" type="checkbox"/>		1
Ato da Mesa Diretora do Congresso Nacional	<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>		2
Ato das Disposições Constitucionais Transitórias	<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>		2
Ato Declaratório	<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>		2
Ato do Comando Supremo da Revolução de 1964						<input checked="" type="checkbox"/>		1
Ato do Presidente		<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		3
Ato Institucional						<input checked="" type="checkbox"/>		1
Constituição Estadual ou Lei Orgânica				<input checked="" type="checkbox"/>				1
Constituição Federal						<input checked="" type="checkbox"/>		1
Constituições Federais anteriores a 1988						<input checked="" type="checkbox"/>		1
Decisão Normativa da Presidência		<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>		2
Decreto						<input checked="" type="checkbox"/>		1
Decreto do Conselho de Ministros						<input checked="" type="checkbox"/>		1
Decreto Legislativo				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		2
Decreto Legislativo do Congresso						<input checked="" type="checkbox"/>		1
Decreto Reservado						<input checked="" type="checkbox"/>		1
Decreto-Lei						<input checked="" type="checkbox"/>		1
Deliberação		<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>		2
Emenda à Lei Orgânica		<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>				2
Emenda Constitucional					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2
Emenda Constitucional de Revisão						<input checked="" type="checkbox"/>		1
Lei						<input checked="" type="checkbox"/>		1
Lei Complementar				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	3
Lei Constitucional						<input checked="" type="checkbox"/>		1
Lei de Responsabilidade Fiscal						<input checked="" type="checkbox"/>		1
Lei Delegada						<input checked="" type="checkbox"/>		1
Lei Ordinária				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		2
Medida Provisória						<input checked="" type="checkbox"/>		1
Moção								0
Ordem de Serviço						<input checked="" type="checkbox"/>		1
Portaria		<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>		2
Regimento Interno						<input checked="" type="checkbox"/>		1
Regimento Interno da ANC		<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>		2
Resolução						<input checked="" type="checkbox"/>		1
Resolução da Câmara dos Deputados		<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>		2
Resolução de Revisão Constitucional						<input checked="" type="checkbox"/>		1
Resolução do Congresso Nacional						<input checked="" type="checkbox"/>		1
Resolução do Senado Federal						<input checked="" type="checkbox"/>		1

Fonte: sistematização do autor com base nas entrevistas concedidas pelos técnicos dos SIL

Tabela 68: Formato do registro documentário

Campos de Pesquisa	NJUR		NJURDF		LEGIN		SAPL	Freq
	A	B	A	B	A	B	A	
Ano				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	4
Assunto				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	4
Autor				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	3
Catálogo / tema	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>						3
Data				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		4
Ementa	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		4
Indexação	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	6
Número				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	4
Número da ADIN								
Número da Proposição					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2
Observação		<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		4
Órgão				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		3
Origem	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>			3
Palavras-chaves				<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		2
Poder						<input checked="" type="checkbox"/>		1
Publicação (data)					<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2
Republicação (ano)				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		3
Texto Integral				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		3
Tipo				<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	4
Vide	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		3
Identificação	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>						2
Fonte	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>			3
Situação	<input checked="" type="checkbox"/>						<input checked="" type="checkbox"/>	2
Lista		<input checked="" type="checkbox"/>						1
Vetos					<input checked="" type="checkbox"/>			1
Preâmbulo					<input checked="" type="checkbox"/>			1
Sessão					<input checked="" type="checkbox"/>			1
Anexos					<input checked="" type="checkbox"/>			1
Tipo de autor							<input checked="" type="checkbox"/>	1
Partido do autor							<input checked="" type="checkbox"/>	1
Relator							<input checked="" type="checkbox"/>	1

Fonte: sistematização do autor com base nas entrevistas concedidas pelos técnicos dos SIL

A análise das tabelas permite a formulação das seguintes observações:

a) em relação a características do acervo legislativo – existe grande diversidade de normas legislativas sendo cobertas pelos sistemas com relação à tipologia, mas no âmbito do sistemas LEGIN e NJUR, essa diversidade não se aplica com relação aos documentos em si. A análise dos objetivos e dos acervos desses sistemas permite observar-se uma duplicidade de esforços da Câmara dos Deputados e do Senado Federal, na cobertura dos textos legislativos. Ou

aspecto intrigante, é o fato de que as instituições recorrem aos arquivos da legislação em formato PDF para a leitura por humanos e posterior alimentação de suas bases de dados. Esse fato ocorre nos sistemas LEGIN, NJUR e NJUR-DF. Considerando-se que a origem das leis está nas proposições legislativas, formuladas pelos parlamentares no próprio âmbito de cada casa, parece um contra-senso que os textos sejam obtidos de fontes externas e em formatos de difícil tratamento. Nesse aspecto, é louvável a iniciativa dos técnicos do sistema SAPL, cuja estratégia é a coleta dos textos ainda em estágio de proposição e em formato XML – um padrão de interoperabilidade universal, compatível com qualquer máquina e qualquer sistema operacional. Essa estratégia certamente garante bons níveis na racionalização de procedimentos, bem como na interoperabilidade com outras instituições.

b) em relação a características tecnológicas – segundo os entrevistados, os sistemas LEGIN, NJUR e NJUR-DF utilizam ou irão utilizar tecnologias proprietárias para o armazenamento dos acervos. Por um lado essa estratégia é favorável, uma vez que há garantia de suporte tecnológico pelos fornecedores, durante a vigência dos contratos. Por outro lado, aspectos negativos como o alto custo dos sistemas e restrições quanto a aspectos de interoperabilidade devem ser considerados. Embora haja diferenças nos acervos com relação à denominação das normas contempladas, a elaboração de tais normas segue técnicas legislativas similares, do que resultam estruturas que podem ser utilizadas como padrão em busca de formatos de representação baseados na linguagem XML. A formulação de padronizações para a estrutura de documentos legislativos pode originar grandes ganhos do ponto de vista da organização e representação da informação legislativa como, por exemplo o surgimento de uma linguagem de marcação extensível. Cabe destacar-se aqui que os técnicos do Programa Interlegis e do PRODASEN, há alguns anos, têm iniciado as discussões nesse âmbito (LIMA; SOUZA, 2001; Interlegis, 2002; COLAB, 2004).

Proposta Metodológica de Organização e Representação de Informação Legislativa

- 4.1 Formato de Marcação de Terminologias
(FORMATER)
- 4.2 Formato de Marcação Legislativa
(FORMALEGIS)

A proposta apresentada nesta pesquisa é fundamentada no losango semiótico, uma adaptação do triângulo semiótico proposto por Ogden & Richards (1972, p. 32) e estendido para ajustar-se aos fundamentos da *Web Semântica*. Os autores basearam sua abordagem de representação do conhecimento no *simbolismo* que, como definem, é o estudo do papel desempenhado pela linguagem e pelos mais diversos símbolos, nas transações humanas. O simbolismo estuda também a influência desses fatores sobre o pensamento.

Nessa perspectiva, as palavras, por si mesmas, nada significam, mas assumem certo significado quando um pensamento as utiliza numa dada acepção, para referir-se a um objeto – “coisa”. Assim, os autores propõem o diagrama em forma de triângulo, colocando, em cada vértice, um dos fatores envolvidos – *símbolo*, *pensamento* e *objeto* – na representação do significado, sendo as relações existentes entre esses fatores representadas pelos lados do triângulo. Tem-se, assim, que entre um *pensamento* e um *símbolo* são mantidas relações causais, o que equivale a dizer que quando alguém exprime (falando ou escrevendo) ou interpreta (ouvindo ou lendo) uma idéia, estará motivando outras pessoas ou sendo motivado por elas, respectivamente, a uma associação de símbolos com suas próprias atitudes de pensamento. Entre um *pensamento* e um *objeto* existe também uma relação em que o pensamento é uma representação mental do objeto e o objeto é conceituado como um pensamento. Entre o *símbolo* e o *objeto* há também uma relação, uma vez que alguém sempre utilizará um símbolo para representar um objeto. A ilustração 84 representa a proposta de Ogden e Richards (1972, p. 32).

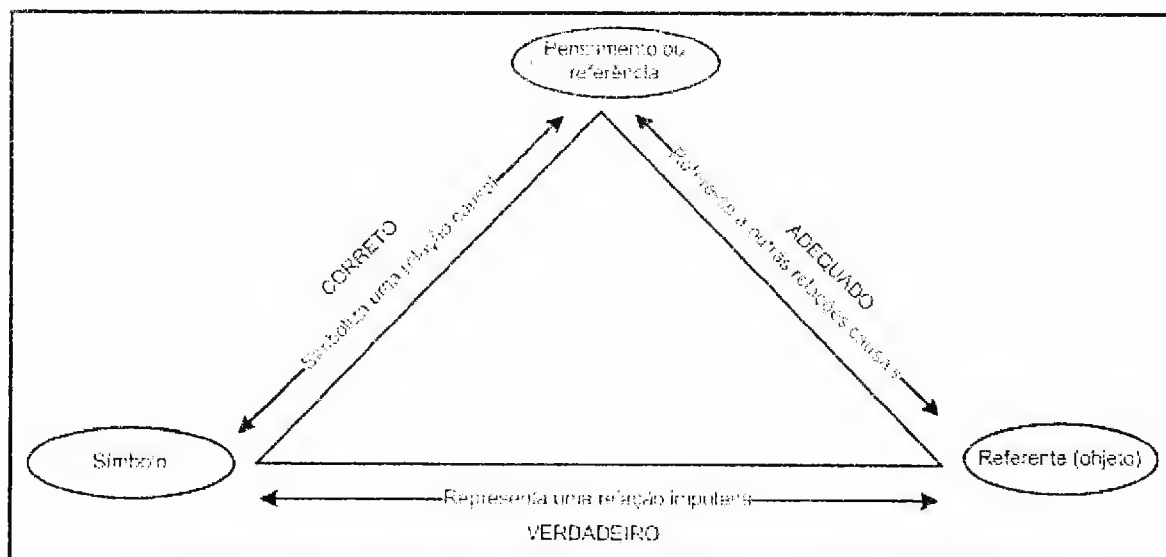


Ilustração 84: Representação dos conceitos segundo Ogden e Richards (1972, p. 32)

Um símbolo verdadeiro registra corretamente uma referência adequada. Um símbolo incorreto é aquele que, num dado universo do discurso, provoca num intérprete uma referência diferente da simbolizada pelo emissor da mensagem.

Como já relatado no capítulo de revisão de literatura, a representação de Ogden e Richards foi estudada por diversos autores, que propuseram atualizações. Esse modelo serviu como base também para a estruturação da Teoria Geral da Terminologia, introduzida por Wüster, que estabelece duas relações básicas: *objeto individual* – *conceito* e *símbolo* – *significado* (FELBER, 1984).

Nesta pesquisa, o losango semiótico, apresentado na ilustração 85, introduz a *localização* como mais um fator envolvido na representação do significado.

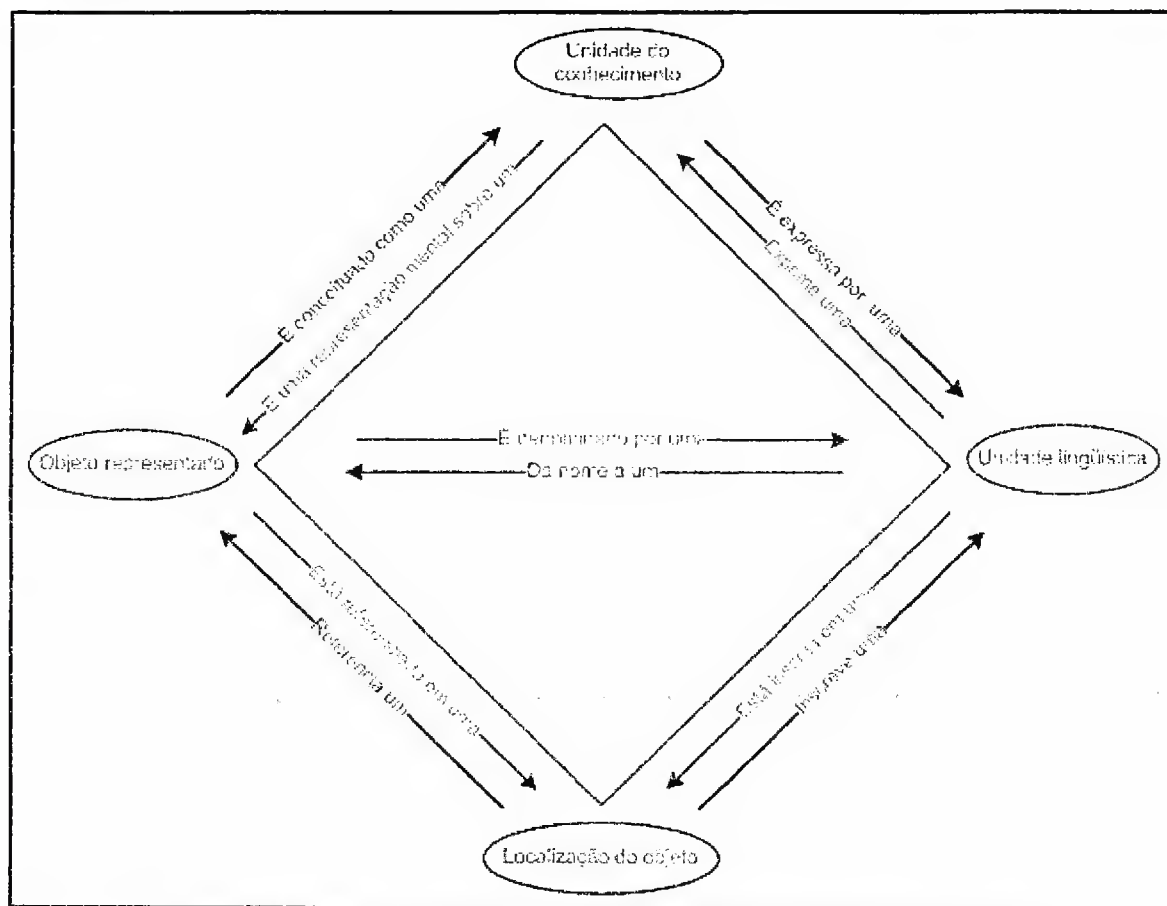


Ilustração 85: Losango semiótico para representação de conceitos na web semântica

Para se adequar o triângulo semiótico de Ogden e Richards aos princípios da Terminologia, o fator *Pensamento* ou *Referência*, foi substituído aqui, pelo fator *Unidade do Conhecimento*, que se refere a *conceito*; o fator *símbolo* foi substituído pelo fator *Unidade Linguística*, que se refere a *termo*; o fator *referente* foi substituído pelo fator *objeto*, que se refere ao *objeto* que está sendo conceituado. O fator *localização* foi introduzido na representação, convertendo o triângulo em um losango, com o objetivo de se estabelecer uma relação entre o objeto conceituado, o símbolo que o representa e a localização em que se inscreve, numa realidade concreta ou abstrata. Têm-se, assim, as seguintes relações:

- a) a *unidade do conhecimento* é expressa por uma *unidade lingüística* e uma *unidade lingüística* exprime uma *unidade do conhecimento*;
- b) uma *unidade do conhecimento* é uma representação mental sobre um *objeto* e um *objeto* é conceituado como uma *unidade do conhecimento*;
- c) uma *unidade lingüística* nomeia um *objeto* e um *objeto* é denominado por uma *unidade lingüística*;
- d) um *objeto* está referenciado em uma *localização* e uma *localização* referencia um *objeto*;
- e) uma *unidade lingüística* está inscrita em uma *localização* e uma *localização* inscreve uma *unidade lingüística*.

A introdução do fator *localização* foi motivada pelo fato de que, na *web*, um dos objetos de estudo desta pesquisa, os objetos podem estar referenciados ou até mesmo armazenados em computadores. Assim, a *localização* fornece a situação do objeto ou do termo no espaço. O modelo pressupõe também a existência de objetos que não estejam fisicamente armazenados em computadores, mas que, existindo no mundo real abstrato ou concreto, tenham sido referenciados por meio de um termo em alguma página da *web*, como, por exemplo, no plano concreto: pessoas, carros, roupas, residências, livros, entre outros; ou no plano abstrato: sentimentos, idéias, mitos, entre outros. De tal maneira, uma ontologia da *web* pode ser utilizada para a realização de controle terminológico. Para exemplificar, nas expressões “*banco da praça*” e “*banco de investimentos*” o conceito “*banco*” pode ser inambigüamente identificado pelo uso de um atributo em uma ontologia localizada em um URI: <http://www.cl.df.gov.br/ontologias/financas/#banco>, que pode se referir sempre a uma única definição de *banco*.

Graficamente Berners-Lee (1989) propõe a representação da *web* por meio de diagramas de círculos e setas, nos quais os círculos (ou nós) representam objetos ou entidades e setas (ou links) representam as relações entre esses objetos (ilustração 86). Comparando-se esse modelo com os sistemas de conceitos, é possível se admitir que os nós representem os conceitos e que as setas representem as relações entre eles. Os nomes dados aos nós podem ser admitidos como termos – unidades lingüísticas. Caso uma unidade lingüística relativa a um objeto esteja presente na *web*, ela poderá ser representada por um identificador uniforme de recurso (URI).

A expressão URI, como já discutido no capítulo de Revisão de Literatura e também no capítulo de Fundamentação Teórica e Metodológica, define-se como unidade fundamental de identificação de objetos na *web*. Um recurso é qualquer coisa física ou abstrata

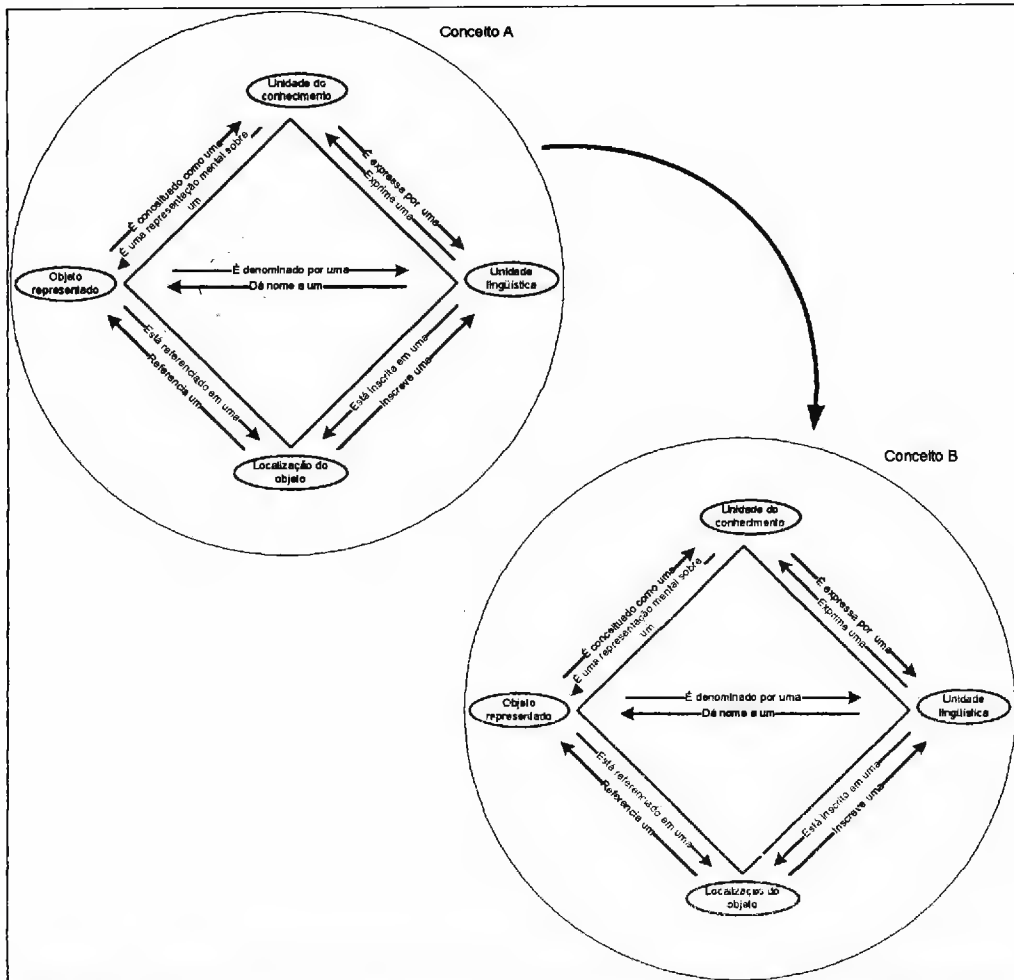


Ilustração 87: representação gráfica da web introduzindo-se conceitos, termos, objetos e localizações

Na concepção vista na ilustração 87, os objetos passam a ser analisados não mais como objetos, puramente, mas como conceitos que podem ser localizados em algum ponto do espaço cibernético. A seta representa uma *relação conceitual*. Um conjunto de conceitos inscritos no espaço representa um *sistema de conceitos*. Retomando-se o enunciado de *definição* adotado por Arntz e Picht (1995, p. 87), pode-se, ainda, admitir que o modelo dado (ilustração 86) seja aplicável à formulação de definições a respeito de um *conceito*. Tal pressuposto se justifica, uma vez que a *definição* de um *conceito* toma como base as relações existentes entre o conceito em questão e outros conceitos, para delimitá-lo de modo inequívoco no sistema de conceitos. A ilustração 88 apresenta a delimitação do campo semântico de alguns conceitos e um sistema aplicando o losango semântico.

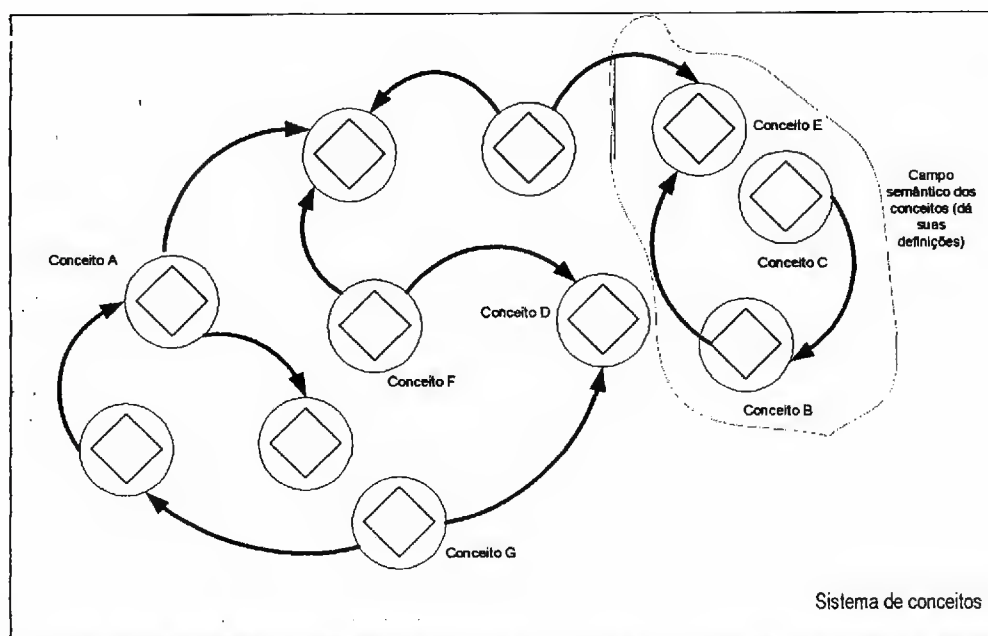


Ilustração 88: Delimitação do campo semântico de conceitos em um sistema para suas definições

Do ponto de vista da terminografia, o modelo é aplicável, desde que se introduza no registro terminológico a informação sobre a localização do conceito ou, em outras palavras, seu URI. A indicação das relações do conceito em um sistema é dada pela sua definição, que também é imprescindível ao registro. Como foi exhaustivamente estudado, os conceitos relacionam-se entre si de diversas maneiras e a proposta aqui colocada pressupõe a representação dessas relações no ambiente web, para a organização de informações. Para tanto, é possível a utilização de um formato de dados terminológico, cujos registros forneçam informações detalhadas acerca do sistema de conceitos e dos termos utilizados em tal sistema. Esta pesquisa apresenta, na seção seguinte, uma proposta de formato terminológico a ser utilizado para a organização e representação da informação legislativa, de modo a contemplar o modelo introduzido pelo losango semiótico.

4.1 Formato de Marcação de Terminologias (FORMATER)

Esta seção apresenta uma proposta de formato de marcação de terminologias, a que se denominou FORMATER e cuja base tecnológica é a linguagem XML. A base teórica do FORMATER advém do estudo dos diversos formatos de registro de dados terminológicos abordados no capítulo de Revisão de Literatura, na seção de Terminografia (FELBER e PICHT, 1984; FAULSTICH, 1995; KRIEGER e FINATTO, 2004; ABNT, 1997a; SÁNCHEZ, 1999; TTT.ORG, 2004). Além desses formatos, adotou-se como referência a norma NISO Z.3985 (2001), que define o conjunto de elementos do padrão Dublin Core, para estruturar a descrição

dos elementos. A estrutura de descrição dos elementos prevista naquela norma inclui: *nome do elemento, rótulo, definição e comentário*. Na proposta FORMATER, essa estrutura é expandida, incluindo-se: *ocorrência, tipo, atributos, subelementos*.

A despeito das iniciativas internacionais para a elaboração de formatos aplicáveis a terminologias, como o MARTIF e a norma ISO 12.620, a proposição do FORMATER é justificável tendo em vista os seguintes aspectos:

- a) é uma proposta cujas aplicações são direcionadas às peculiaridades do idioma Português;
- b) é uma proposta de sistematização de outros formatos e de adequação às tecnologias da web semântica.
- c) é motivado pela necessidade de se cobrirem algumas lacunas da norma ABNT 13.789 (1997a) – que dispõe sobre princípios e métodos para a elaboração de normas de terminologia –, no que tange a aspectos de informatização de registros terminológicos;

A aplicação dos procedimentos teóricos da Terminologia envolve as seguintes atividades: – identificar os conceitos e as relações conceituais pertinentes a um certo campo do conhecimento; – após a identificação dos conceitos e de suas relações, desenvolver sistemas de conceitos; – elaborar definições para os conceitos; – atribuir termos a conceitos (ISO:WD 740.1,1996). Em complemento a essa abordagem, é importante que a terminologia e o trabalho terminográfico tenham suas atividades registradas de maneira organizada e consistente, para se adequarem a necessidades tais como: a) descrição sistemática do conjunto de termos fundamentais à formação do discurso de um domínio do conhecimento socialmente distinto; b) transmissão e difusão dos conhecimentos de áreas temáticas específicas, por meio de sua terminologia; c) normas que se apliquem à formação dos termos e ao arcabouço teórico na formação de uma obra terminológica (FAULSTICH,1990)

Conforme apresentado no capítulo de Revisão de Literatura, há diferentes formatos para a elaboração de registros terminológicos. Como tais formatos variam de acordo com as orientações dos autores, uma proposta de interoperabilidade entre eles seria uma tarefa não muito trivial, a menos que esteja sustentada em padrões amplamente aceitos e difundidos. Nesse sentido, esta proposta se apresenta como uma alternativa para as questões destacadas. O formato pretende contribuir para a definição de um conjunto de elementos e atributos passíveis de ocorrência em um registro de dados terminológico com vistas à sua escrita em XML. Além

disso, fornecer insumos para o desenvolvimento de instrumentos que permitam a um terminólogo ou bibliotecário leigo em informática elaborar registros destinados a armazenamento em meio eletrônico.

O que se espera é que o uso do formato propicie a organização dos registros de modo consistente e facilmente recuperável. Tal organização é dada pela definição de regras e restrições claras com relação a elementos, subelementos, atributos e seus tipos de dados. Do ponto de vista da Web Semântica, metadados são o ponto de partida para o processamento e representação semântica, o que é garantido pela sintaxe universal da linguagem XML (DACONTA, 2003). Acredita-se, assim, que a descrição da terminologia de um domínio específico conforme metodologias e fundamentos da Terminologia possa utilizar tecnologias da *web* semântica como OWL ou RDF, com o objetivo de propiciar a obtenção do controle conceitual e terminológico na organização e recuperação de informações.

4.1.1 Conjunto de elementos do FORMATER

Nesta seção são apresentados os elementos que compõem o conjunto FORMATER. Com o objetivo de propiciar rigor sintático aos elementos, nos moldes da linguagem XML, optou-se por registrar seus rótulos apenas com letras minúsculas⁷⁸ e sem caracteres especiais. Em relação à origem, o conjunto de elementos, apresentado na ilustração 89, é composto por dois subconjuntos:

a) elementos de definição própria – constitui-se de todos os subelementos do elemento registro terminológico;

b) elementos importados – compreendem os elementos *Title*, *Creator*, *Subject*, *Description*, *Publisher*, *Date*, *Format*, *Language* e *Rights*, definidos no padrão de metadados Dublin Core, cuja intenção de uso é facilitar a localização de coleções por mecanismos de inferência que utilizam padrões internacionais.

Os elementos importados foram traduzidos com os nomes de *Título*, *Autor*, *Palavra-chave*, *Descrição*, *Instituição*, *Data*, *Formato*, *Idioma* e *Copyright*, respectivamente. No entanto, para o seu uso, sugere-se que a sintaxe do padrão Dublin Core seja respeitada (em HTML: *<meta*

⁷⁸ Embora a XML aceite o uso de maiúsculas, essa linguagem faz distinção entre elementos nomeados com a mesma palavra, mas com diferenciação de caixa alta e baixa; assim, o elemento *<Título>* é diferente do elemento *<título>*. Sugere-se, assim, para os elementos de definição própria, o uso apenas de letras minúsculas com o objetivo de facilitar a edição de documentos e para evitar problemas motivados por erros de digitação. Os elementos importados do Dublin Core devem ser registrados conforme a sintaxe estipulada pelo padrão; nesse caso, com iniciais maiúsculas.

`name="DC.Elemento" content="descrição-do-elemento">` e em XML: `<dc:Elemento>descrição-do-elemento</dc:Elemento>`). Um exemplo do primeiro caso é:

```
<meta name="DC.Title" content="Coleção de termos sobre legislação">
```

E, para o segundo caso:

```
<dc:Title> Coleção de Termos Legislativos </dc:Title>.
```

Como exemplificado, recomenda-se a utilização do prefixo *dc*, para referência ao namespace `xmlns:dc="http://purl.org/dc/"`, que define o vocabulário daquele formato.

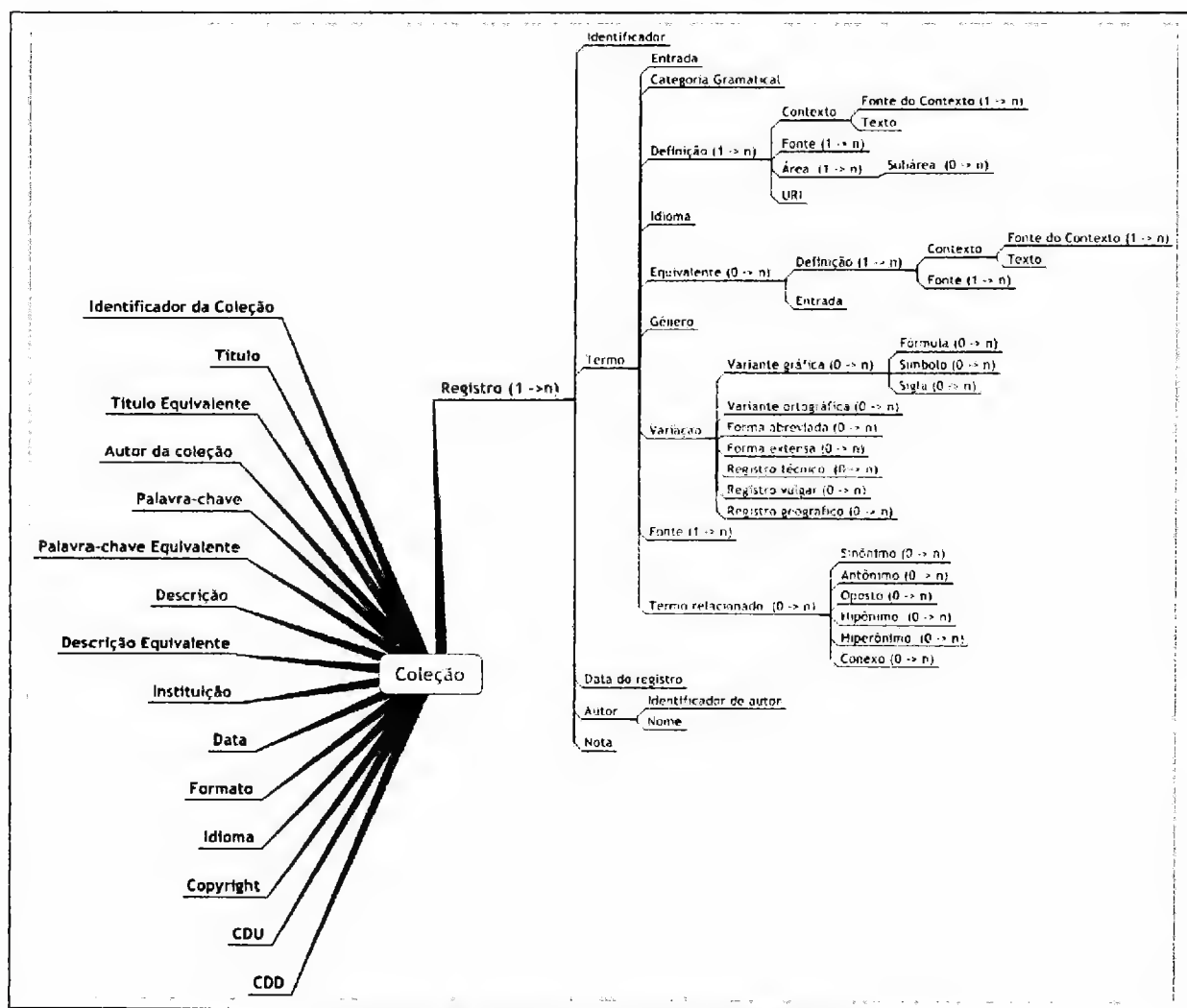


Ilustração 89: conjunto de elementos do FORMATER

Os elementos referem-se à descrição de dois grupos de informações principais: a *Coleção* de termos e cada *Registro Terminológico* individual. Os elementos que identificam a coleção são aqueles importados do padrão Dublin Core (*Título*, *Autor*, *Palavra-chave*, *Descrição*, *Instituição*, *Data*, *Formato*, *Idioma* e *Copyright*), além dos seguintes: *Código de Classificação Decimal* (CDU), código

de *Classificação Decimal de Dewey (CDD)*, *Título Equivalente*, *Descrição Equivalente*, *Palavra-chave Equivalente*. Os elementos que identificam o *Registro Terminológico* são os de definição própria do FORMATER, a saber: *Identificador do Registro*, *Termo*, *Data*, *Autor* e *Nota*.

4.1.1.1 Elementos identificadores da Coleção

As tabelas 69 a 83 apresentam o detalhamento das características dos elementos que descrevem a coleção de dados terminológicos, a saber: *Título*, *Título Equivalente*, *Autor*, *Palavra-chave*, *Palavra-chave Equivalente*, *Descrição*, *Descrição Equivalente*, *Instituição*, *Data*, *Formato*, *Idioma* e *Copyright*. Em relação à ocorrência, os elementos são ser classificados em obrigatórios, recomendados e opcionais.

Tabela 69: elemento Coleção

Nome: Coleção		Código: C1
Rótulo	coleccion	
Definição	conjunto de registros de dados terminológicos em um sistema de conceitos	
Ocorrência	única e obrigatória	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	é o elemento obrigatoriamente utilizado como raiz de um documento escrito em XML	
Atributos	Identificador	
Subelementos	Título, Título Equivalente, Autor, Palavra-chave, Palavra-chave Equivalente, Descrição, Descrição Equivalente, Instituição, Data, Formato, Idioma, Copyright, CDD, CDU e Registro	

* Um elemento do tipo complexo, para efeitos deste formato é aquele que possui subelementos

Tabela 70: elemento Título

Nome: Título		Código: C2
Rótulo	dc:Title	
Definição	um nome ou expressão atribuída à coleção ao recurso coleção	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 71: elemento Título Equivalente

Nome: Título Equivalente		Código: C3
Rótulo	tituloeq	
Definição	um nome ou expressão atribuída à coleção em outro idioma	
Ocorrência	opcional e ilimitada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é utilizado em complemento ao elemento Título, para armazenar o título em outro idioma	
Atributos	—	
Subelementos	—	

Tabela 72: elemento Autor da coleção

Nome: Autor da coleção		Código: C4
Rótulo	dc:Creator	
Definição	uma entidade ou pessoa responsável pela manutenção da coleção	
Ocorrência	obrigatória, no mínimo uma vez	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 73: elemento Palavra-chave

Nome: Palavra-chave		Código: C5
Rótulo	dc:Subject	
Definição	indicação de palavra-chave que descreve o conteúdo da coleção	
Ocorrência	obrigatória, no mínimo uma vez	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core.	
Atributos	—	
Subelementos	—	

Tabela 74: elemento Palavra-chave Equivalente

Nome: Palavra-chave Equivalente		Código: C6
Rótulo	pchaveq	
Definição	indicação de palavra-chave que descreve o conteúdo da coleção em outro idioma	
Ocorrência	opcional e ilimitada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core.	
Atributos	—	
Subelementos	—	

Tabela 75: elemento Descrição

Nome: Descrição		Código: C7
Rótulo	dc:Description	
Definição	resumo descritivo do conteúdo da coleção	
Ocorrência	obrigatória, no mínimo uma vez	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core. Pode ser utilizado mais de uma vez, caso se deseje descrever a coleção em outro idioma	
Atributos	—	
Subelementos	—	

Tabela 76: elemento Descrição Equivalente

Nome: Descrição Equivalente		Código: C8
Rótulo	descricaoeq	
Definição	resumo descritivo do conteúdo da coleção	
Ocorrência	opcional e ilimitada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento complementa o uso do elemento Descrição. Pode ser utilizado mais de uma vez, caso se deseje descrever a coleção em outro idioma	
Atributos	—	
Subelementos	—	

Tabela 77: elemento Instituição

Nome: Instituição		Código: C9
Rótulo	dc:Publisher	
Definição	instituição responsável pela publicação da coleção na web	
Ocorrência	obrigatória e ilimitada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 78: elemento Data de atualização

Nome: Data de atualização		Código: C10
Rótulo	dc:Date	
Definição	data associada com a última publicação da coleção	
Ocorrência	única e obrigatória	
Tipo	formato de data (recomenda-se o armazenamento no padrão internacional yyyy-dd-mm, com vistas à recuperação automática por mecanismos de busca ou outros agentes; no entanto, recomenda-se a conversão para dd-mm-yyyy no momento da exibição ao usuário de língua portuguesa)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 79: elemento Formato

Nome: Formato		Código: C11
Rótulo	dc:Format	
Definição	indicação do formato físico de armazenamento da coleção em linha	
Ocorrência	única e opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 80: elemento Idioma

Nome: Idioma		Código: C12
Rótulo	dc:Language	
Definição	indicação do idioma da coleção de dados terminológicos	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string) recomenda-se seguir o padrão internacional adotado para os símbolos de idiomas dado pela norma ISO 639 e pela RFC 3066 ⁷⁹ , preferencialmente com dois caracteres	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 81: elemento Copyright

Nome: Copyright		Código: C13
Rótulo	dc:Rights	
Definição	indicação dos direitos de propriedade intelectual sobre a coleção.	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 82: elemento CDU

Nome: CDU		Código: C14
Rótulo	cdu	
Definição	indicação do código na Classificação Decimal Universal	
Ocorrência	única e recomendada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento destina-se a facilitar a localização de uma coleção em meio eletrônico por agentes inteligentes. Visa também a organização prévia da coleção para suportes que não estejam em rede, como papel, CDROM e DVD; seu registro deve obedecer às regras previstas na norma de classificação	
Atributos	—	
Subelementos	—	

⁷⁹ Para mais detalhes sobre as recomendações RFC, sugere-se uma visita a: <<http://www.faqs.org/rfcs/>>

Tabela 83: elemento CDD

Nome: CDD		Código: C15
Rótulo	cdd	
Definição	indicação do código na Classificação Decimal de Dewey	
Ocorrência	única e recomendada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento destina-se a facilitar a localização de uma coleção em meio eletrônico por agentes inteligentes. Visa também a organização prévia da coleção para suportes que não estejam em rede, como papel, CDROM e DVD; seu registro deve obedecer às regras previstas na norma de classificação	
Atributos	—	
Subelementos	—	

4.1.1.2 Elementos identificadores do Registro Terminológico

As tabelas 84 a 89 apresentam o detalhamento das características de cada um dos elementos que identificam o *Registro Terminológico*, a saber: *Identificador do Registro*, *Termo*, *Data do Registro*, *Autor* e *Nota*.

Tabela 84: elemento Registro

Nome: Registro		Código: R1
Rótulo	registro	
Definição	conjunto de todas as informações a respeito de um conceito, incluindo termo, definição, variações, termos equivalentes, termos relacionados e suas definições	
Ocorrência	obrigatória	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Identificador do Registro, Termo, Data, Autor e Nota	

* Um elemento do tipo complexo, para efeitos deste formato é aquele que possui subelementos

Tabela 85: elemento Identificador do Registro

Nome: Identificador do Registro		Código: R2
Rótulo	idregistro	
Definição	elemento identificador unívoco de cada registro terminológico	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string) ou inteiro (recomendado quando for utilizado recurso de autonumeração)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 86: elemento Data do Registro

Nome: Data do Registro		Código: R3
Rótulo	dtregistro	
Definição	indica a data em que o registro foi inserido no banco de dados terminológico	
Ocorrência	obrigatória e única	
Tipo	formato de data (recomenda-se o armazenamento no padrão internacional yyyy-dd-mm, com vistas à recuperação automática por mecanismos de busca ou outros agentes; no entanto, recomenda-se a conversão para dd-mm-yyyy no momento da exibição ao usuário de língua portuguesa)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 87: elemento Autor do Registro

Nome: Autor do Registro		Código: R4
Rótulo	auregistro	
Definição	indica a pessoa que realizou a inclusão do registro	
Ocorrência	obrigatória	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Identificador do Autor, Nome	

* Um elemento do tipo complexo, para efeitos deste formato é aquele que possui subelementos

Tabela 88: elemento Nota

Nome: Nota		Código: R5
Rótulo	nota	
Definição	indica uma observação ou comentário que o autor julgue necessária, acerca do registro	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 89: elemento Termo

Nome: Termo		Código: R6
Rótulo	termo	
Definição	contém todas as informações relativas ao termo, incluem definição, categoria gramatical, idioma, equivalentes, gênero, entre outras	
Ocorrência	obrigatória	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Entrada, Categoria Gramatical, Definição, Idioma, Equivalente, Gênero, Variação, Fonte, Termo Relacionado	

* Um elemento do tipo complexo, para efeitos deste formato é aquele que possui subelementos

4.1.1.2.1 Elementos identificadores do Termo

O elemento *Termo* subdivide-se nos subelementos *Entrada*, *Categoria Gramatical*, *Definição*, *Idioma*, *Equivalente*, *Gênero*, *Variação*, *Fonte* e *Termo Relacionado*. Desses subelementos, *Definição*, *Equivalente*, *Variação* e *Termo Relacionado* também apresentam subdivisões. *Definição* subdivide-se em *Contexto*, *Fonte*, *Área* e *URI*. *Contexto*, por sua vez, subdivide-se em *Texto* e *Fonte do Contexto*. *Equivalente* subdivide-se em *Entrada* e *Definição*. *Variação* subdivide-se em *Variante Gráfica*, *Variante Ortográfica*, *Forma Abreviada*, *Forma Extensa*, *Registro Técnico*, *Registro Vulgar*, *Registro Geográfico*. *Variante Gráfica* subdivide-se, por sua vez, em *Fórmula*, *Símbolo* e *Sigla*. *Termo Relacionado* subdivide-se em *Sinônimo*, *Antônimo*, *Oposto*, *Hipônimo*, *Hiperônimo* e *Conexo*. Os subelementos do Termo totalizam 43. Por esse motivo, sua descrição foi acrescentada no apêndice D deste trabalho.

Embora a preocupação básica do FORMATER seja a organização e representação de informações para serem veiculadas por meio da web. As regras de definição dos elementos permitem também o seu uso para orientar o desenvolvimento de bancos de dados relacionais. Dessa maneira, o formato pode ser utilizado para orientar o desenvolvimento do modelo de Entidade x Relacionamento. No entanto, no caso específico desse uso, sugere-se nomear os campos importados do padrão Dublin Core iniciando-os por *dc*, para garantir a legibilidade por mecanismos e inferência e agentes inteligentes da web. Ter-se-ia, assim, para um banco de dados relacional, o elemento *dc:dcdata*, pertencente à tabela *coleção*, e o elemento *dc:data*, pertencente à tabela *registro*. Tal procedimento poderá facilitar a conversão futura dos dados para o formato XML se houver essa intenção.

A tabela 90 apresenta um exemplo de codificação XML que utiliza o FORMATER em uma coleção de dados terminológicos relativos à legislação.

Como este código foi escrito em linguagem XML, é possível a criação de um arquivo XMLSchema para o controle da sintaxe dos rótulos dos campos, de modo que possa ser utilizado para validar o código no momento em que o autor estiver editando o documento coleção. É possível também a escrita do mesmo código utilizando-se a arquitetura RDF ou a linguagem OWL.

Em termos estruturais, essa metodologia garante a expansão do modelo sempre que houver necessidade de se conferir maior rigor no controle terminológico. Por exemplo, no código apresentado na tabela 90, foi acrescentado mais um elemento `<iplural>` que é um subelemento de `<variacao>`. O rótulo *iplural* é atribuído ao elemento *Forma Plural*, cuja função é indicar a grafia do termo de entrada no plural (variação de número). Mesmo que não esteja

presente no conjunto inicial de elementos, é possível a sua inclusão. Em casos como esse, obviamente, é necessária a atualização do arquivo XMLSchema anexado, para que o elemento possa integrar a biblioteca. No código apresentado na tabela 90, uma referência ao XMLSchema não atualizado faria com que o novo elemento *<vplural>* fosse rejeitado, isto é, ele seria considerado não válido.

Tabela 90: Representação de dados terminológicos utilizando FORMATER

descrição	código
indicação de início de arquivo	<?xml version="1.0" encoding="ISO-8859-1" ?>
indicação das informações sobre a coleção, incluindo os campos título, autor, palavra-chave, descrição, instituição, data, idioma, copyright e cdu (observe-se que os campos importados foram referidos com seus rótulos originais do Dublin Core, para facilitar a localização por mecanismos de busca)	<pre> <colecão id="1"> <dc:Title>Terminologia Legislativa</dc:Title> <tituloeq>Legislative Terminology</tituloeq> <dc:Creator>Setor de Pesquisa e Recuperação da Informação</dc:Creator> <dc:Subject>Terminologia, Legislação, Poder Legislativo</dc:Subject> <dc:Description>Coleção de dados terminológicos utilizados de acordo com o acervo de normas legislativas da CLDF</dc:Description> <dc:Publisher>Câmara Legislativa do Distrito Federal</dc:Publisher> <dc>Date>2004-21-11</dc>Date> <dc:Language>pt</dc:Language> <dc:Rights>Direitos reservados à Câmara Legislativa do Distrito Federal</dc:Rights> <cdu>001.4</cdu> </pre>
indicação das informações sobre o registro	<pre> <registro> <idregistro>1</idregistro> <dtregistro>2004-21-11</dtregistro> <auregistro> <idautor>15</idautor> <nomeautor>Olavo Bilac </nomeautor> </auregistro> <termo> <entrada>adolescente</entrada> <ctgramatical>substantivo</ctgramatical> <definicao> <dtexto> individuo na idade da adolescência</dtexto> <darea>direito de infância e adolescência</darea> <uri>#legislacao/termos/dia</uri> </definicao> <idioma>pt</idioma> <genero>indiferente</genero> <variacao> <vplural>adolescentes </vplural> </variacao> <conexo> criança </conexo> <oposto> adulto </oposto> <trelacionado> </termo> </registro> </pre>
indicação de outro registro	<registro>...</registro>
indicação de final de arquivo	</colecão>

4.2 Formato de Marcação Legislativa (FORMALEGIS)

O levantamento das características dos acervos e dos SIL estudados nesta pesquisa propiciou a identificação de diferentes formatos de registro documentário, bem como de diferentes tipos de normas legislativas. Nos diversos sistemas estudados observou-se que a maioria dos registros documentários utiliza elementos do texto como número, data, ementa, entre outros, para realizar a descrição de conteúdo do texto legal. No entanto, na maioria dos diversos sistemas estudados, essa descrição é feita pela entrada manual em um sistema informatizado. Assim, uma pessoa especializada faz a leitura do texto, extrai esses valores e digita-os em um registro informatizado. Entre os técnicos entrevistados, dois já iniciaram propostas para o desenvolvimento de metodologias mais inteligentes, que permitam a descoberta da estrutura do por agentes de busca. Essas iniciativas são a Constituição Federal, disponível no *site* do Senado Federal desde 1998 em formato XML (LIMA; SOUZA, 2001), e o sistema SAPL, ao qual foi adicionado um editor de XML para a coleta das proposições legislativas já com o uso dessa linguagem (CICILIATI JR, 2004). Ambos os sistemas partiram da definição estrutural da norma com vistas à proposição de um modelo de esquema ou DTD para a validação do vocabulário XML aceito.

A estrutura básica utilizada em ambos os casos coincidente com as orientações dadas pela Lei Complementar nº 95/1998, que dispõe sobre a elaboração de textos normativos:

Art. 3º A lei será estruturada em três partes básicas:

I - parte preliminar, compreendendo a epígrafe, a ementa, o preâmbulo, o enunciado do objeto e a indicação do âmbito de aplicação das disposições normativas;

II - parte normativa, compreendendo o texto das normas de conteúdo substantivo relacionadas com a matéria regulada;

III - parte final, compreendendo as disposições pertinentes às medidas necessárias à implementação das normas de conteúdo substantivo, às disposições transitórias, se for o caso, a cláusula de vigência e a cláusula de revogação, quando couber.

Desse modo, a lei define que os textos das normas sejam estruturados em: parte preliminar, parte normativa e parte final. Essas estruturas são organizadas em outros elementos menores, a saber: a) ementa – subdivide-se em identificação numérica, espécie normativa e ano de promulgação; b) preâmbulo – indica o órgão competente para promulgar o ato, bem como o fundamento legal; c) parte normativa – organizada em artigos, que desdobram-se em parágrafos ou incisos; os incisos subdividem-se em alíneas, e essas em itens. É comum em atos normativos, o agrupamento de artigos para constituírem elementos denominados subseções. O agrupamento

de subseções constitui a seção; o de seções, o capítulo; o de capítulos, o título; o de títulos, o livro; e o de livros, a parte. Partes podem ser classificadas como parte geral ou parte especial, conforme o caso (LEI 95/1998; DECRETO 4176/2002).

Com base na legislação citada, que trata sobre as diretrizes para a elaboração de atos normativos, e no estudo dos formatos levantados nos diversos SIL durante esta pesquisa, apresenta-se, nesta seção, a proposta de Formato de Marcação de Normas Legislativas (FORMALEGIS). A exemplo do Formato de Marcação de Terminologias, o FORMALEGIS foi concebido para ser representado em XML sendo, contudo, aplicável a outras tecnologias, como os bancos de dados relacionais. O objetivo da proposta é favorecer a organização e a representação de informações em acervos legislativos informatizados.

4.2.1 Conjunto de elementos do FORMALEGIS

Nesta seção são apresentados os elementos que compõem o conjunto FORMALEGIS. Da mesma maneira como no FOMATER, optou-se por registrar rótulos dos elementos com letras minúsculas e sem caracteres especiais. Em relação à origem, o conjunto de elementos, apresentado nas ilustrações 90 a 97, é composto por dois subconjuntos:

a) elementos importados – compreendem os elementos *Title*, *Creator*, *Subject*, *Description*, *Publisher*, *Date*, *Format*, *Language* e *Rights*, definidos no padrão de metadados Dublin Core, cuja intenção de uso é facilitar a localização de normas por mecanismos de busca que utilizam padrões internacionais.

b) elementos de definição própria – constitui-se de todos os demais elementos do formato;

Os elementos importados foram traduzidos com os nomes de *Título*, *Responsável*, *Palavra-chave*, *Descrição*, *Instituição*, *Data do registro*, *Formato*, *Idioma* e *Copyright*, respectivamente. No entanto, como explicitado na seção anterior, quanto ao seu uso, sugere-se a adoção da sintaxe do padrão Dublin Core.

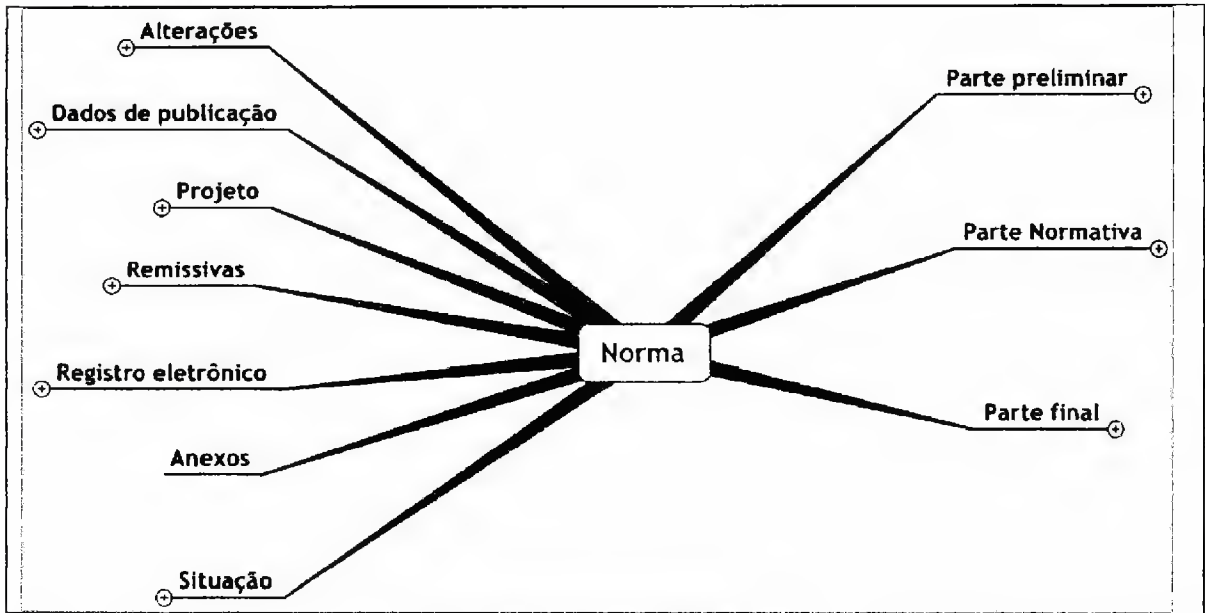


Ilustração 90: Elementos da Norma

Os elementos da norma referem-se a: a) projeto inicial (Projeto); b) parte textual da norma, antes das alterações (Parte preliminar, Parte normativa, Parte final); c) registro eletrônico; d) alterações da norma; e) dados de publicação; f) dados de remissivas; g) anexos; h) situação (ilustração 90).

Cada um desses elementos, por sua vez subdivide-se em subelementos, cujas características variam de acordo com o tipo de informação que será armazenada. O elemento registro eletrônico contém informações relativas à descrição da norma armazenada em meio eletrônico e inclui todos os elementos importados do padrão Dublin Core. A ilustração 91 apresenta a estrutura do subelemento registro eletrônico. Os elementos Dublin Core *Title*, *Description*, *Creator*, *Subject*, *Format*, *Publisher*, *Language* e *Rights* foram traduzidos como *Título*, *Descrição*, *Responsável*, *Palavra-chave*, *Formato*, *Instituição*, *Idioma* e *Copyright*.

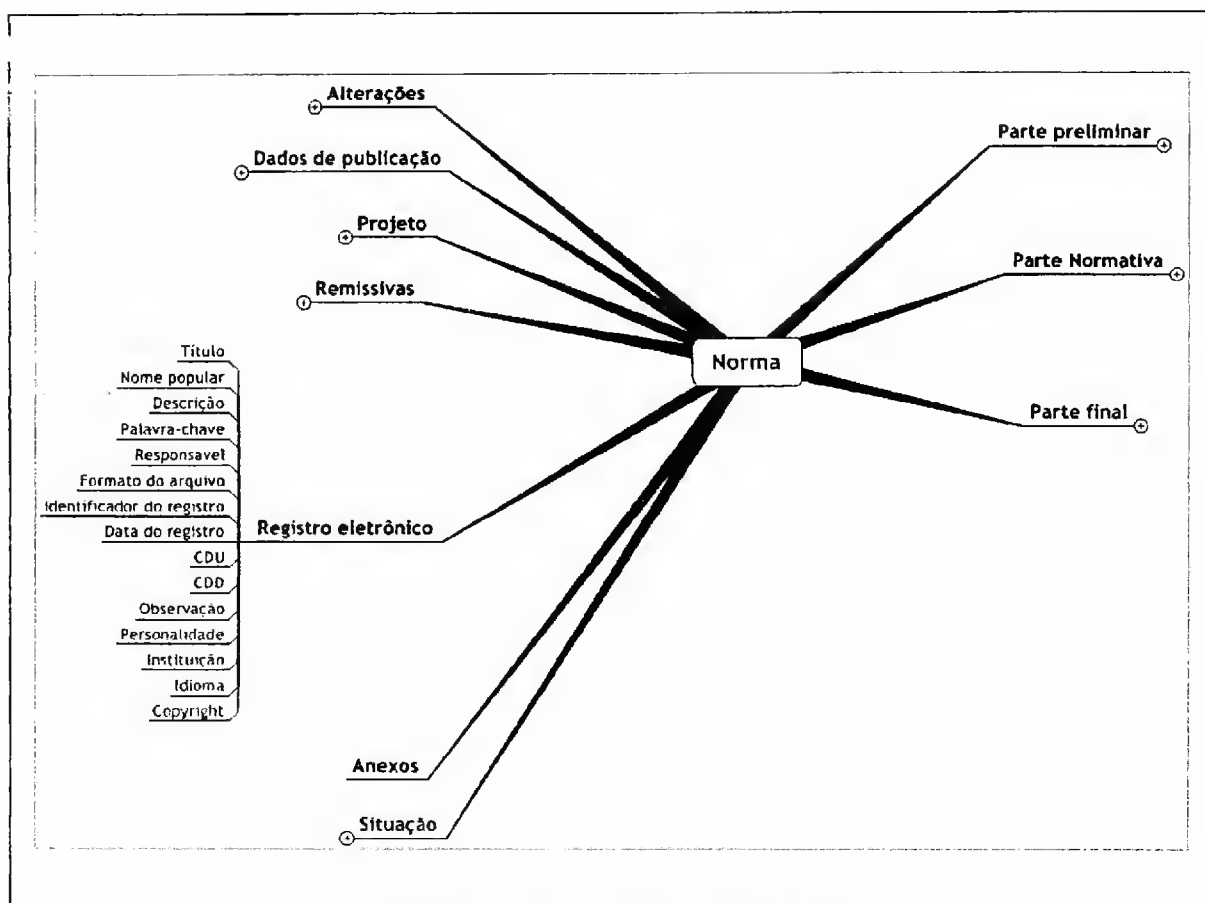


Ilustração 91: Subelementos do Registro eletrônico

As tabelas 91 a 105 apresentam o detalhamento das características de cada um dos elementos que identificam o *Registro eletrônico*, a saber: *Título*, *Nome popular*, *descrição*, *Palavra-chave*, *Responsável*, *Formato do Arquivo*, *Identificador do registro*, *Data do registro*, *CDU*, *CDD*, *Observação*, *Personalidade*, *Instituição*, *Idioma* e *Copyright*.

Tabela 91: elemento Registro eletrônico

Nome: Registro eletrônico		Código: R1
Rótulo	registro	
Definição	conjunto de todas as informações referenciais sobre a norma, incluindo termo, definição, variações, termos equivalentes, termos relacionados e suas definições	
Ocorrência	obrigatória	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Título, Nome Popular, Descrição, Palavra-chave, Responsável, Formato do Arquivo, Identificador do registro, Data do registro, CDU, CDD, Observação, Personalidade, Instituição, Idioma e Copyright	

* Um elemento do tipo complexo, para efeitos deste formato é aquele que possui subelementos

Tabela 92: elemento Título

Nome: Título		Código: R2
Rótulo	dc:Title	
Definição	um nome ou expressão atribuída à norma	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 93: elemento Nome Popular

Nome: Nome Popular		Código: R3
Rótulo	npopular	
Definição	um nome ou expressão popular atribuída à norma	
Ocorrência	recomendada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 94: elemento Descrição

Nome: Descrição		Código: C7
Rótulo	dc:Description	
Definição	resumo descritivo do conteúdo da norma	
Ocorrência	obrigatória, no mínimo uma vez	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core.	
Atributos	—	
Subelementos	—	

Tabela 95: elemento Palavra-chave

Nome: Palavra-chave		Código: C5
Rótulo	dc:Subject	
Definição	indicação de palavra-chave que descreve o conteúdo da norma	
Ocorrência	obrigatória, no mínimo uma vez	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core.	
Atributos	—	
Subelementos	—	

Tabela 96: elemento Responsável

Nome: Responsável		Código: C4
Rótulo	dc:Creator	
Definição	uma entidade ou pessoa responsável pela manutenção da norma	
Ocorrência	obrigatória, no mínimo uma vez	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 97: elemento Instituição

Nome: Instituição		Código: C9
Rótulo	dc:Publisher	
Definição	instituição responsável pela publicação da norma na web	
Ocorrência	obrigatória, no mínimo uma vez	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 98: elemento Data do registro

Nome: Data do registro		Código: C10
Rótulo	dc:Date	
Definição	data associada com a última publicação da norma	
Ocorrência	única e obrigatória	
Tipo	formato de data (recomenda-se o armazenamento no padrão internacional yyyy-dd-mm, com vistas à recuperação automática por mecanismos de busca ou outros agentes; no entanto, recomenda-se a conversão para dd-mm-yyyy no momento da exibição ao usuário de língua portuguesa)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 99: elemento Formato

Nome: Formato		Código: C11
Rótulo	dc:Format	
Definição	indicação do formato físico de armazenamento da norma em linha	
Ocorrência	única e opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 100: elemento Copyright

Nome: Copyright		Código: C13
Rótulo	dc:Rights	
Definição	indicação dos direitos de propriedade intelectual sobre o registro da página norma.	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento é importado do conjunto de elementos permitidos no padrão Dublin Core	
Atributos	—	
Subelementos	—	

Tabela 101: elemento CDU

Nome: CDU		Código: C14
Rótulo	cdu	
Definição	indicação do código na Classificação Decimal Universal	
Ocorrência	única e recomendada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento destina-se a facilitar a localização de uma coleção em meio eletrônico por agentes inteligentes. Visa também a organização prévia da coleção para suportes que não estejam em rede, como papel, CDROM e DVD; seu registro deve obedecer às regras previstas na norma de classificação	
Atributos	—	
Subelementos	—	

Tabela 102: elemento CDD

Nome: CDD		Código: C15
Rótulo	cdd	
Definição	indicação do código na Classificação Decimal de Dewey	
Ocorrência	única e recomendada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	este elemento destina-se a facilitar a localização de uma coleção em meio eletrônico por agentes inteligentes. Visa também a organização prévia da coleção para suportes que não estejam em rede, como papel, CDROM e DVD; seu registro deve obedecer às regras previstas na norma de classificação	
Atributos	—	
Subelementos	—	

Tabela 103: elemento Identificador do Registro

Nome: Identificador do Registro		Código: R2
Rótulo	idregistro	
Definição	elemento identificador único de cada registro legislativo	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string) ou inteiro (recomendado quando for utilizado recurso de autonumeração)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 104: elemento Observação

Nome: Observação		Código: R5
Rótulo	observ	
Definição	indica uma observação ou comentário que o autor julgue necessária, acerca do registro	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 105: elemento Personalidade

Nome: Personalidade		Código: R6
Rótulo	personalidade	
Definição	contém informações relativas a uma pessoa ou entidade sobre a qual trata a norma e cujo registro pode auxiliar na recuperação da norma	
Ocorrência	obrigatória	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Entrada, Categoria Gramatical, Definição, Idioma, Equivalente, Gênero, Variação, Fonte, Termo Relacionado	

* Um elemento do tipo complexo, para efeitos deste formato é aquele que possui subelementos

Os elementos que identificam a parte textual da norma são aqueles que compõem a parte preliminar, a parte normativa e a parte final. Conforme a ilustração 92, a *Parte Preliminar* é composta por: *Epígrafe* (especie normativa, data de promulgação, letra, número); *Fórmula*; *Preâmbulo* (Título, Autoridade, Fundamento legal, ordem de execução); *Enunciado do Objeto*; *Âmbito de aplicação*.

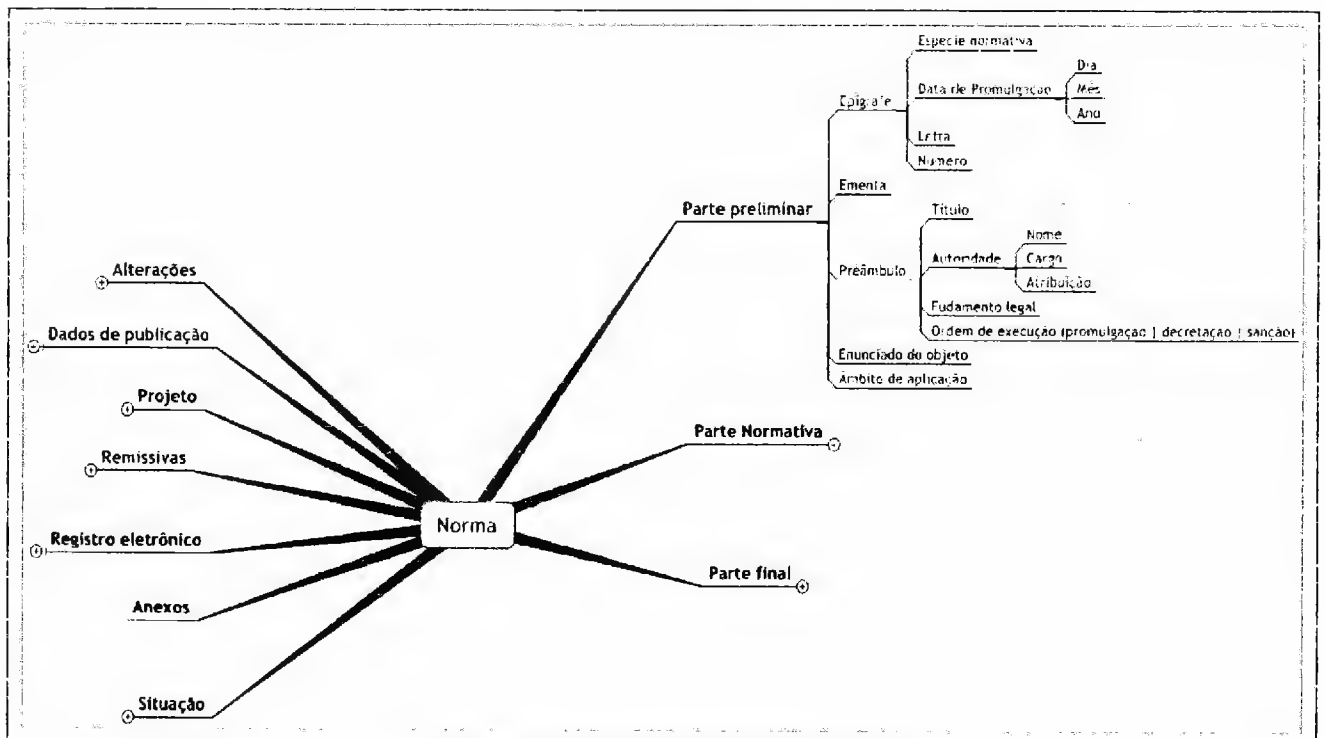


Ilustração 92: Subelementos da Parte preliminar

Conforme a ilustração 93, os elementos que identificam a parte normativa são: *artigo* (número, letra, caput, parágrafos, incisos); *parágrafo* (incisos, número, texto); *inciso* (alíneas, número, texto); *alínea* (itens, letra, texto) e *item* (número, texto); *subseção* (nome, número, artigos) *seção* (nome, número, subseções); *capítulo* (nome, número, seções); *título* (nome, número, capítulos); *livro* (nome, número, títulos); *parte* (nome, tipo, número, livros).

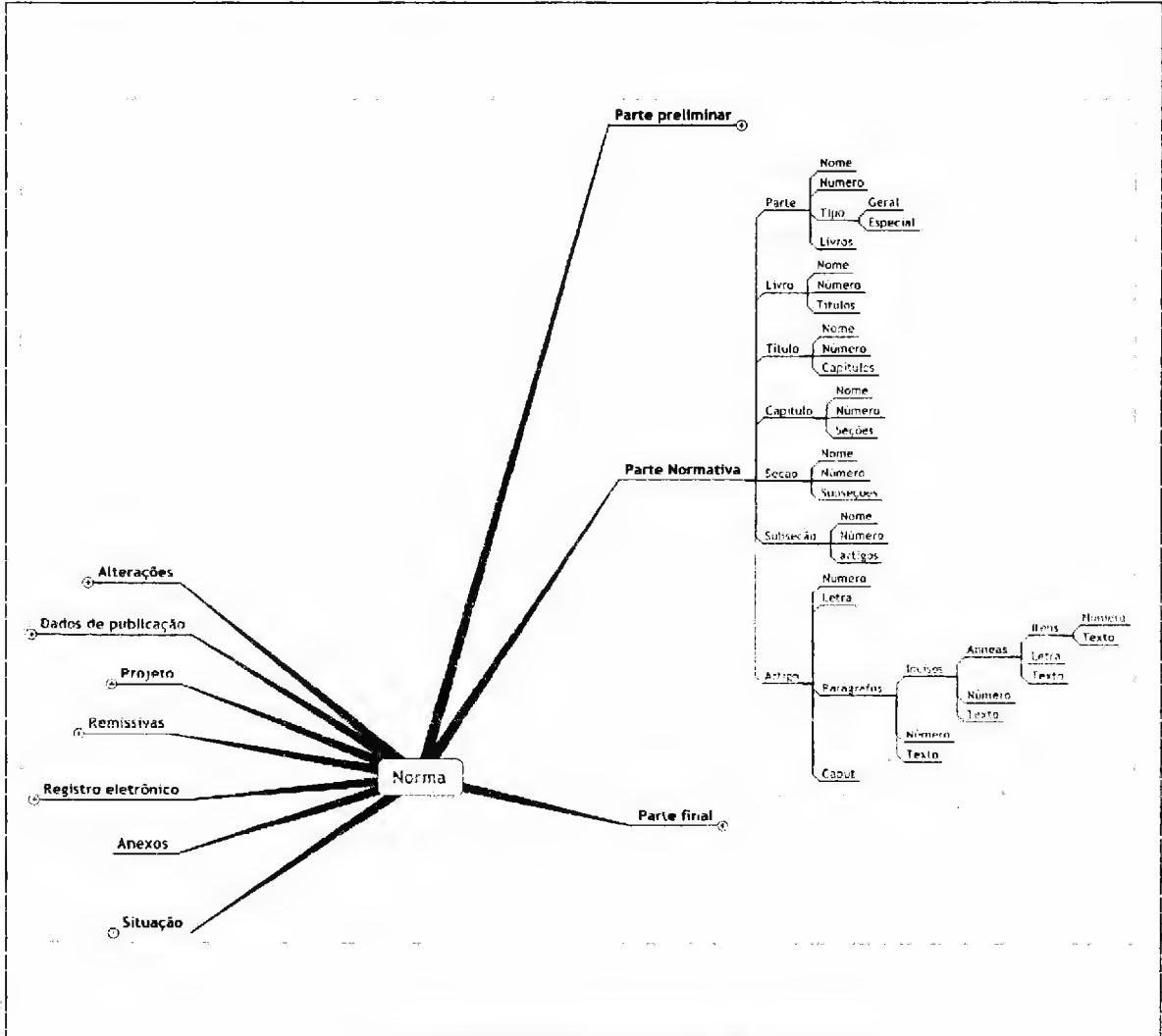


Ilustração 93: Subelementos da Parte normativa

Conforme a ilustração 94, os subelementos da parte final incluem: *medidas para implantação*, *disposições transitórias*, *cláusula de vigência*, *cláusula de revogação* (número, data, artigos); *fecho* (local, data, assinatura, ano da república, ano da independência).

A estruturação dos demais elementos, a saber: alterações, dados de publicação, projeto, remissivas, anexos e situação, pode ser observada nas ilustrações 95 a 97 e pode seguir o mesmo tipo de organização dada para os elementos abordados anteriormente.

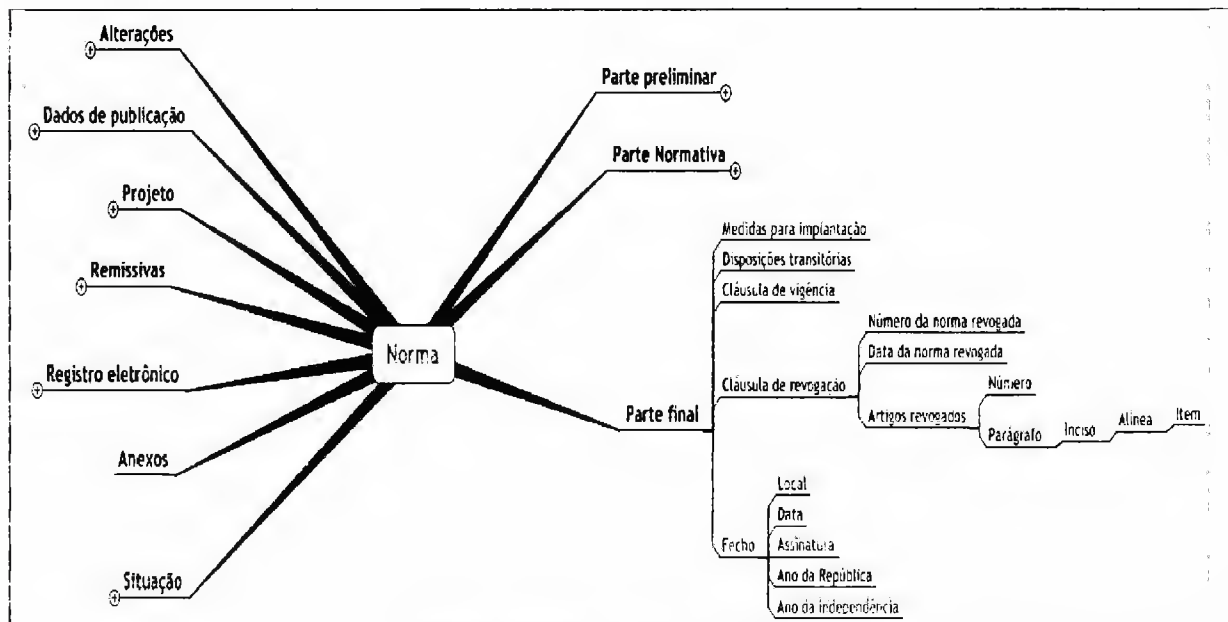


Ilustração 94: Subelementos da Parte Final

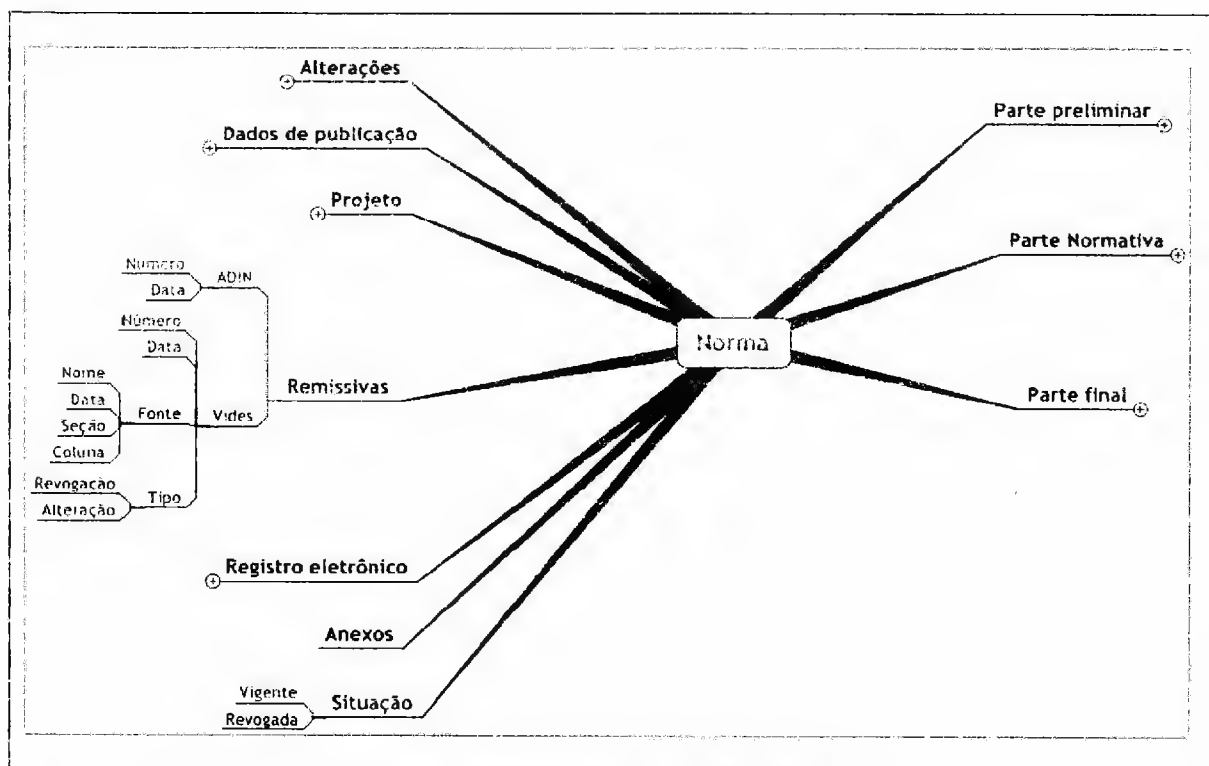


Ilustração 95: Subelementos de Remissivas

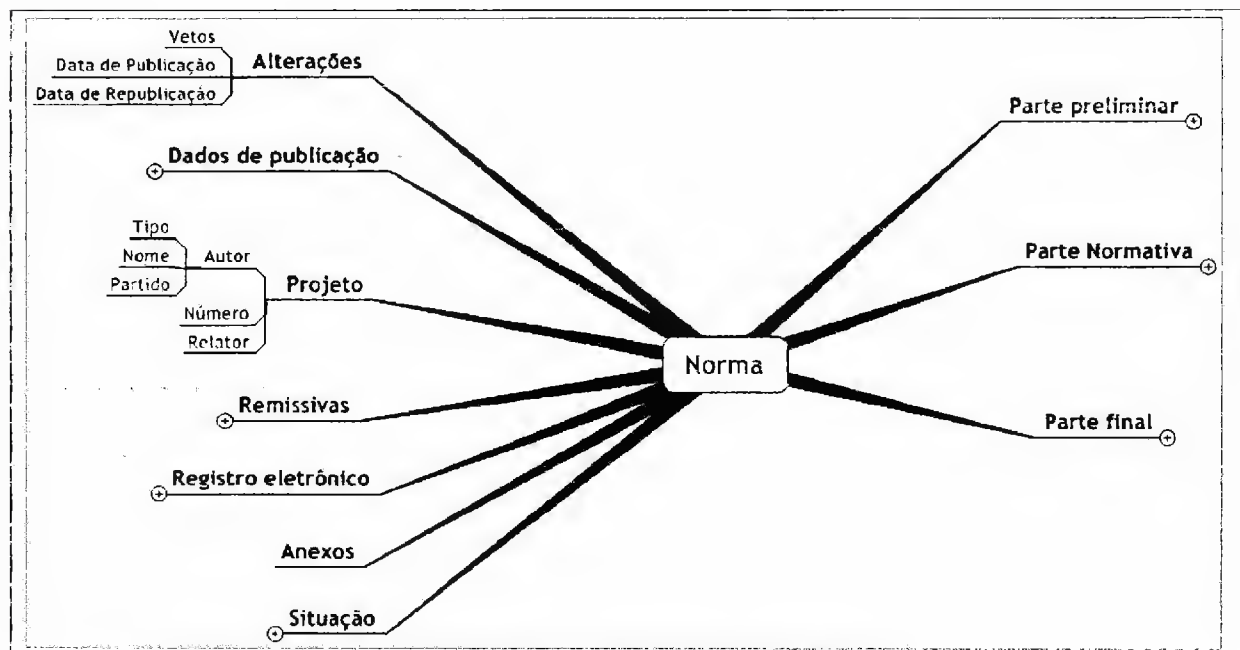


ilustração 96: Subelementos do Projeto

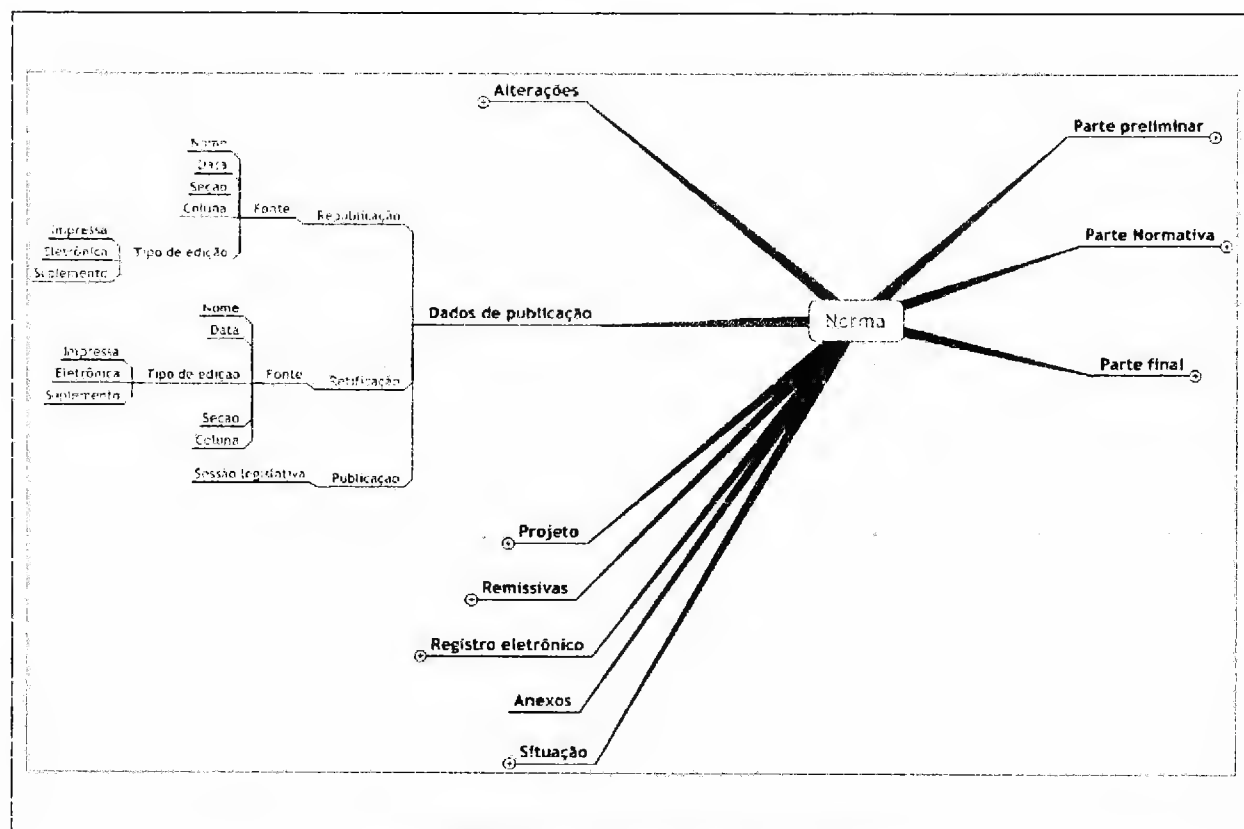


ilustração 97: Subelementos de Dados de publicação

Analogamente ao FORMATER, uma coleção de registro baseada no FORMALEGIS pode ser escrita em qualquer linguagem derivada da XML, incluindo

XMLSchema, RDF, RDFS, OWL. A tabela 106 apresenta uma representação de registros de uma norma legislativa com base na estrutura FORMALEGIS, em formato XML.

Tabela 106: Estrutura de uma coleção de normas utilizando FORMALEGIS

descrição	código
indicação de início do arquivo	<?xml version="1.0" encoding="ISO-8859-1"?>
indicação das informações sobre a norma, incluindo os campos título, autor, palavra-chave, descrição, instituição, data, idioma, copyright e cdu (observe-se que os campos importados foram referidos com seus rótulos originais do Dublin Core, para facilitar a localização por mecanismos de busca)	<pre> <norma id="1"> <registroeletronico> <dc:Title>LEI N° 1065 DE 06 DE MAIO DE 1996</dc:Title> <npopular>Lei do Silêncio</npopular> <dc:Creator>Setor de Pesquisa e Recuperação da Informação</dc:Creator> <dc:Subject>Meio ambiente, Poluição Sonora, Sinalização, Silêncio </dc:Subject> <dc:Description>Dispõe sobre normas de preservação ambiental quanto a poluição sonora</dc:Description> <dc:Publisher>Câmara Legislativa do Distrito Federal</dc:Publisher> <dc>Date>2004-21-11</dc>Date> <dc:Language>pt</dc:Language> <dc:Rights>Direitos reservados à Câmara Legislativa do Distrito Federal</dc:Rights> <cdu>001.4</cdu> </registroeletronico> <ptpreliminar> <epigrafe> <tpnorma>lei</tpnorma> <data>1996-05-06</data> <numero>1065</numero> </epigrafe> <ementa>Dispõe sobre normas de preservação ambiental quanto a poluição sonora</ementa> <preambulo> <autoridade>O Governador do Distrito Federal</autoridade> <ordexecucao>A Câmara Legislativa do Distrito Federal decreta e eu sanciono</ordexecucao> <objeto>preservação ambiental quanto à poluição sonora</objeto> </preambulo> </ptpreliminar> <ptnormativa> <articulado> <artigo> <numero>1</numero> <caput>Esta Lei estabelece as normas de preservação ambiental quanto à poluição sonora, fixando níveis máximos de emissão de sons e ruídos, de acordo com o local e a duração da fonte</caput> <paragrafos> <parnumero>1</parnumero> <partexto>Considera-se poluição sonora qualquer som indesejável, principalmente quando interfere em atividades humanas ou ecossistemas a serem preservados </partexto> </paragrafos> </artigo> </articulado> </ptnormativa> <ptfinal> <clvigencia> Esta Lei entra em vigor na data de sua publicação </clvigencia> <clrevogacao>Revogam-se as disposições em contrário </clrevogacao> <data>1996-05-06</data> <assinatura>Joaquim Domingos Roriz</assinatura> </norma> </pre>
indicação das informações sobre o registro	

Com a utilização dos formatos FORMATER e FORMALEGIS na organização e representação da informação legislativa, acredita-se ser possível melhorar a recuperação sob dois aspectos principais. Por um lado, o uso de um banco de dados terminológicos, em que são estabelecidas as relações entre os conceitos e são cuidadosamente escolhidos os termos adequados a sua representação em certo domínio do conhecimento ou área, torna bem mais rigoroso o controle terminológico, uma vez que o formato prevê o registro não apenas do termo de entrada, como também das variações, definições, e termos relacionados, inclusive com a especificação do tipo de relação entre eles. Por outro lado, a representação legislativa sob o formato FORMALEGIS poder trazer o benefício de incluir, em um mesmo documento, tanto a informação relativa às referências sobre a norma, quanto o próprio texto integral. Desse modo, a representação acredita-se na possibilidade de maior economia de tempo e recursos de processamento eletrônico, o que inclui não apenas as operações de entrada, mas também as de saída. O uso de elementos importados de um padrão internacional, a saber o Dublin Core, pode representar também um grande ganho na localização de registros disponíveis na Internet. Por fim, o fato de os formatos serem previstos para escrita em XML torna-os compatíveis com qualquer estrutura de armazenamento de dados, uma vez que a linguagem adota um padrão de interoperabilidade aceito em qualquer plataforma.

A ilustração 99 sugere um exemplo de sistema em que se tenha os acervos informatizados e que utilize ambos os formatos. O sistema representado prevê o uso de um servidor de banco de dados e de um servidor de páginas web para a Internet ou para uma intranet.

As atividades relativas à organização e representação da informação legislativa, são previstas em três categorias:

1) rotinas de atualização do acervo terminológico a ser utilizado para a organização da terminologia pertinente ao campo do conhecimento, campo profissional, ou campo técnico em questão; no caso da ilustração, o domínio é o legislativo.

2) realização de consultas de no acervo terminológico, com vistas à indexação de documentos, com a utilização de termos de entrada controlados;

3) atualização do acervo de documentos, incluindo-se versões eletrônicas de textos integrais, indexadas intelectualmente.

As atividades de busca e recuperação pelo usuário final podem ser realizadas compreendendo as seguintes etapas:

4) preenchimento de um formulário disponível em linha para o envio do termo de busca ao sistema. Nessa etapa, o mecanismo de pesquisa pode realizar a busca do termo informado pelo usuário tanto no acervo de textos integrais dos documentos, quanto no acervo terminológico. Destaque-se que se o acervo estiver armazenado no formato textual FORMALEGIS, o sistema poderá pesquisar também as informações de referência bibliográfica do registro da norma. Quanto ao acervo terminológico, o sistema pode retornar o termo solicitado, ou termos correlatos, de acordo com a consistência das informações armazenadas. A ilustração 98 sugere as possibilidades desse tipo de tratamento. Por exemplo, supondo uma pesquisa pelo termo *criança* no banco de dados terminológico, o sistema poderia retornar termos relacionados, como *adolescente*, *trabalho infantil*, *direitos humanos*, *infância*, entre outros informados previamente pelo especialista em informação, de acordo com o domínio do conhecimento em questão.

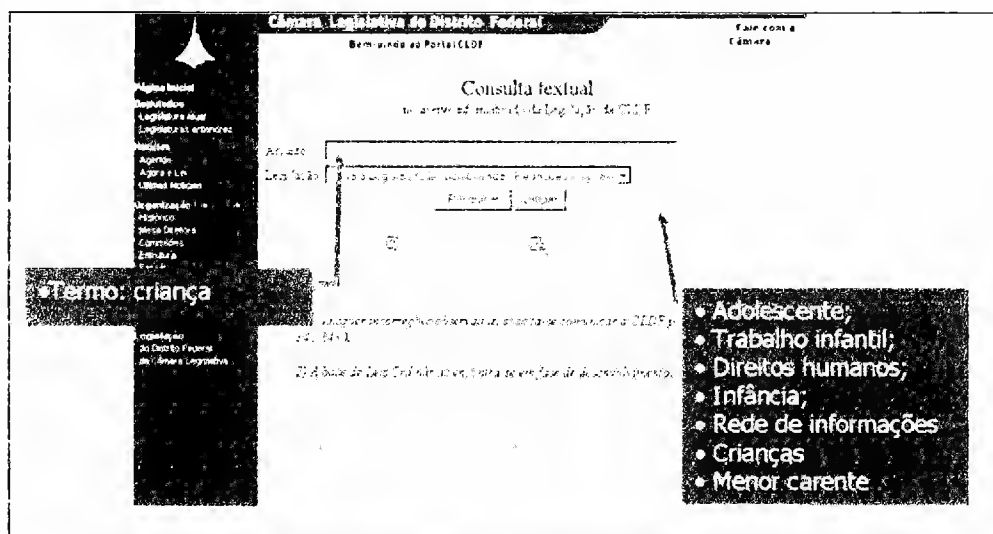


Ilustração 98: Exemplo de busca com apoio de banco de dados terminológico

5) nesta etapa (ilustração 99), o sistema poderia retornar os resultados para o usuário, que teria disponibilidade de escolha de links para o texto integral ou de refazer a busca utilizando termos relacionados advindos do acervo terminológico;

6) nesta etapa, a escolha de um link pelo usuário enviaria uma requisição ao servidor web, para acesso ao texto integral;

7) nesta etapa, o sistema poderia apresentar o texto integral para que o usuário tenha acesso à informação desejada ou que a imprima, converta para outros formatos, envie-a por correio eletrônico, por exemplo.

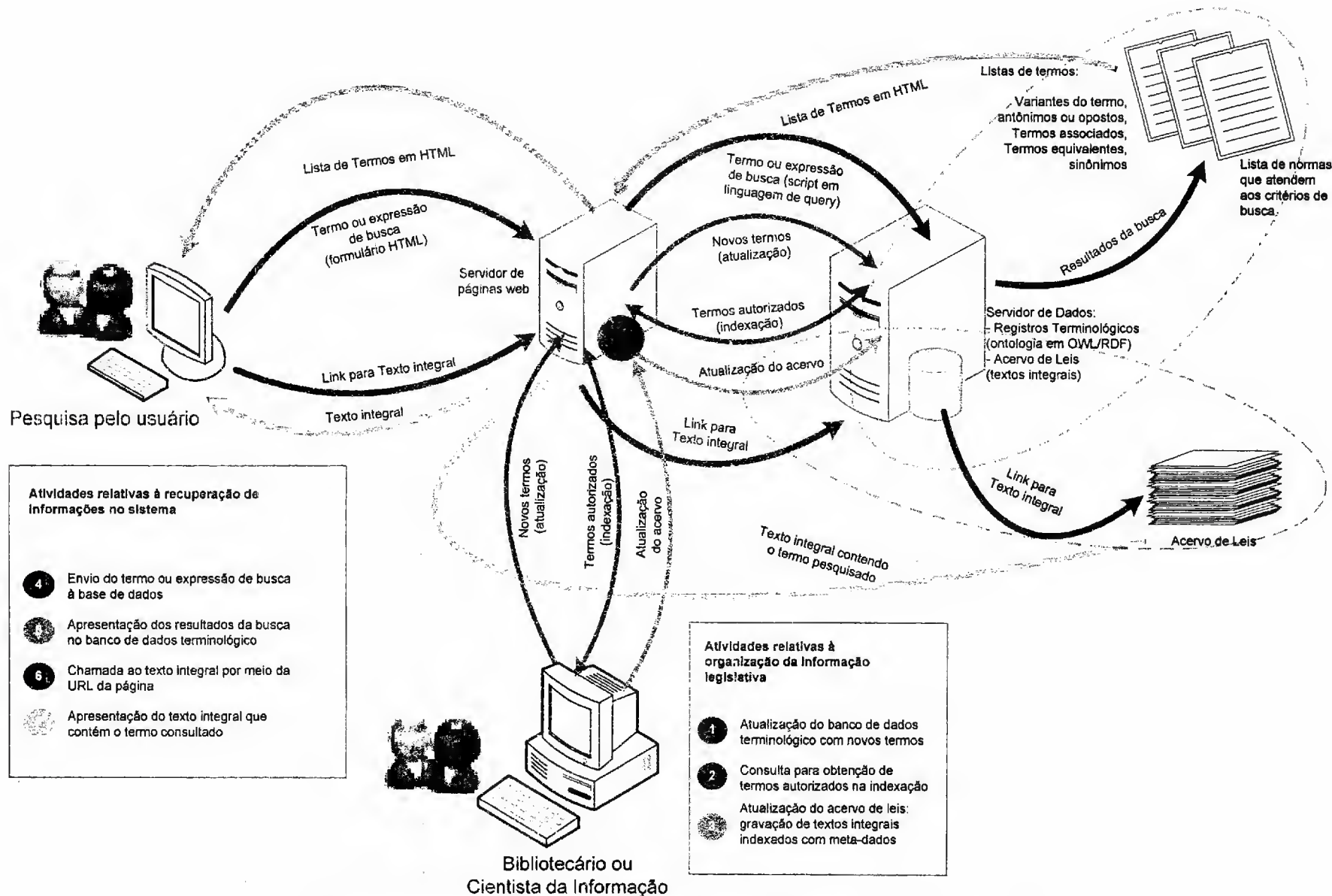


Ilustração 99: Sistema de organização e representação de informações baseado em banco terminológico

Conclusões e sugestões de pesquisa

Esta pesquisa investigou as interfaces teóricas e aplicadas das áreas de Terminologia Linguagens Documentárias e *Web Semântica*, com vistas à utilização em Sistemas de Informação Legislativa. O objetivo foi identificar fundamentos teóricos e metodológicos, bem como princípios e diretrizes referentes a essas áreas do conhecimento, cuja sistematização pudesse fornecer insumos para a elaboração de uma proposta metodológica de organização e representação da informação legislativa. Para o alcance desse objetivo foram realizados um estudo na literatura e um levantamento de características de sistemas legislativos. Quanto ao estudo teórico, realizou-se o levantamento em fontes selecionadas, de acordo com sua pertinência e relevância para o tema tratado na pesquisa e, sem seguida, a sistematização da fundamentação teórica e metodológica. Quanto ao levantamento junto aos Sistemas de Informação Legislativa, foi realizado em duas etapas: a primeira, de âmbito mais amplo, utilizou os *sites* institucionais de 29 casas legislativas brasileiras, para investigar suas características técnicas em relação a armazenamento, processamento informatizado e controle terminológico dos acervos; a segunda etapa, de caráter mais específico, realizada por meio de entrevistas, buscou o estudo de dois sistemas que organizam acervos de Legislação Federal (NJUR e LEGIN), de um sistema baseado em código aberto e destinado à inclusão digital de acervos legislativos em todo o país (SAPL) e, por fim, de um sistema que possui grande relevância para a sociedade local, o NJUR-DF. As características levantadas na primeira etapa, pela Internet, serviram como insumo para a formulação do instrumento usado como roteiro nas entrevistas da segunda etapa. Entre essas características, destacam-se o formato dos registros documentários e a tipologia das normas, utilizados como insumos para a elaboração da proposta metodológica apresentada nesta pesquisa.

Com fundamento no estudo teórico e nos levantamentos realizados, formulou-se a proposta metodológica para a organização e representação da informação legislativa, contemplando dois instrumentos: o Formato de Marcação de Terminologias (FORMATER) e o Formato de Marcação de Legislativa (FORMALEGIS). A proposta elaborada utiliza o conceito de losango semiótico, uma adaptação do modelo de triângulo semiótico de Ogden e Richards (1972), introduzida nesta pesquisa em face dos fundamentos teóricos e metodológicos da Terminologia e da Web Semântica.

Do estudo teórico, foi possível identificar fundamentos teóricos e metodológicos que indicam a interdisciplinaridade entre os três campos do conhecimento estudados: Terminologia, Linguagem Documentária e Web Semântica. A Terminologia possui interdisciplinaridade com o campo de Linguagem Documentária, uma vez que ambas as disciplinas utilizam métodos destinados ao estudo e ao desenvolvimento de instrumentos de controle de vocabulário, com vistas ao aperfeiçoamento da comunicação em um âmbito

específico do conhecimento, seja ele científico, técnico ou profissional. No campo da Linguagem Documentária, o instrumento mais relevante para esta pesquisa é o tesouro. Como já discutido nos capítulos de Revisão de Literatura e de Fundamentação Teórica, a base de construção dos tesouros encontra-se no estudo e no estabelecimento das relações entre os termos, cuja nomenclatura da área denomina como descritores. A literatura estudada indica ainda que, para essa disciplina, a organização das informações fundamenta-se no relacionamento entre os termos e que essas relações podem ser hierárquicas (ISO:2788, 1974; UNISIST, 1975; AITCHISON; GILCHRIST, 1979; LANCASTER, 1987; CINTRA, 1994), de equivalência (ISO:2788, 1974; UNISIST, 1975; AITCHISON; GILCHRIST, 1979; IBICT, 1984) ou associativas (ISO:2788, 1974; UNISIST, 1975; IBICT, 1984; LANCASTER, 1987; ISO:WD704.1,1996). Em geral as relações hierárquicas são classificadas como genérico-específicas e partitivas, e as associativas em uma profusão de tipos.

Embora utilize nomenclaturas diferenciadas, a Terminologia, que busca organizar o mundo real ou abstrato em sistemas de conceitos, parte da análise de tais conceitos para especificar relações entre eles e, assim, sistematizar o conhecimento de determinada área. Em Terminologia, a análise dos conceitos é realizada em um domínio específico do conhecimento ou área profissional, e compreende o uso de palavras que pertencem a um vocabulário restrito a esse domínio. Esse conjunto de palavras, destinado à expressão de significados também específicos, *adquire uma língua de especialidade*. Como indicado pela literatura estudada, as relações conceituais em uma Língua de Especialidade também recebem classificações diversas (CABRÉ, 1993; FELBER, 1984; PICHT, 1984; ARNTZ;PICHT, 1995; SAGER, 1990). A análise dos enunciados dados a essas relações permite deduzir que embora os autores utilizem nomenclaturas diferenciadas, várias de suas definições se assemelham. O estudo de relações conceituais em Terminologia parte da análise de características próprias de um conceito (intrínsecas), ou de características externas ao conceito em questão e que permitem estabelecer a delimitação do seu significado em relação a outros conceitos, num domínio específico (extrínsecas). Assim, é possível definir os relacionamentos de um conceito em um sistema com base na sua *intensão* (conjunto das características intrínsecas) e na sua *extensão* (conjunto das características extrínsecas). Pode-se concluir, então, que tanto para a disciplina Linguagem Documentária quanto para a disciplina Terminologia, a representação e a organização do conhecimento dão-se pelo uso de um vocabulário restrito (vocabulário controlado/língua de especialidade), que os significados são representados por um símbolo lingüístico (descriptor/termo) e que tais significados são identificados e definidos com base nas características de determinados objetos

individuais. Em Linguagem Documentária, um objeto individual é referenciado por meio de um significado, ao passo que, na Terminologia, o é por um conceito.

A Terminologia também encontra fundamentos na lógica e é uma “*uma ciência cujo objeto é de ordem lingüística, mas que é essencialmente pluridisciplinar e participa ao mesmo tempo da lingüística, da lógica, da ontologia, da teoria da classificação e da informática*” (Rondeau, 1984). Em relação à Web Semântica, a Terminologia e a Linguagem Documentária também possuem interfaces, posto que as ontologias, como definidas por aquele campo do conhecimento, procuram estabelecer relações entre objetos e classes de objetos, tomando para análise suas propriedades constitutivas (intensão). Pode-se dizer que uma relação de interdisciplinaridade entre essas duas áreas é bastante presumível, já que ambas fundamentam suas teorias na Lógica.

As ontologias da Web Semântica procuram descrever os objetos de uma mesma classe em função do conjunto de suas propriedades comuns, a que denominam extensão. Assim, na Web Semântica a expressão formal dos relacionamentos entre classes ou subclasses é feita por meio de uma declaração a respeito de como as propriedades de uma classe se relacionam com as propriedades de outra classe em um domínio específico do conhecimento ou campo profissional (uma ontologia). Em outras palavras, a expressão dos relacionamentos é feita por meio de um axioma formalmente definido em uma gramática de uma linguagem de programação. Tal axioma é capaz de fornecer a um agente inteligente a instrução necessária para estabelecer comparações entre duas ou mais as classes. Partindo dessas comparações, um mecanismo pode realizar inferências sobre tais classes. No campo da Terminologia, como já mencionado, a análise dos conceitos é feita por meio das suas características que, em si, são também conceitos. Assim, o estudo revelou que os três domínios, Terminologia, Web Semântica e Linguagem Documentária, em suas intenções de garantir a comunicação, por meio da organização e representação da informação (cada uma com seus instrumentos), possuem princípios, métodos, procedimentos e diretrizes passíveis de serem combinados. Visando à comunicação entre pessoas de uma mesma comunidade ou âmbito do conhecimento, a Terminologia, por meio dos seus princípios, procura estabelecer um conjunto de termos (*terminologia*) em um domínio específico do conhecimento (*Lingua de Especialidade*) de modo que os objetos materiais ou imateriais sejam comunicados de maneira inambígua, por meio do seu significado (*conceito*) expresso em uma declaração (*definição*). A Linguagem Documentária, visando à comunicação entre autores de informação e usuários dessa informação, estabelece um conjunto de palavras (*descriptores*) que podem ser utilizados em um sistema comunicativo (*vocabulário controlado*) com vistas ao armazenamento e à recuperação de informações em um sistema, na maioria das vezes informatizado. A Web Semântica, visando à comunicação não apenas entre pessoas, mas também entre máquinas, estabelece um conjunto de

tecnologias e de linguagens para a formulação de declarações a respeito de objetos materiais ou imateriais, localizados ou não em uma página web. Essas declarações reunidas, formam uma ontologia que pode ser utilizada para a extração do conhecimento e que refere-se a determinado domínio do conhecimento científico, profissional, técnico, ou até mesmo de uma comunidade.

Essas considerações, até este ponto, referem-se às disciplinas estudadas no plano teórico. No plano aplicado, relativamente às características dos diversos Sistemas de Informação Legislativa estudados, destacam-se alguns pontos importantes. Em primeiro lugar, em uma esfera mais ampla, que compreende os sistemas estudados por meio da Internet, observou-se que há várias diferenças quanto à organização da informação legislativa. O fato de haver instituições que ainda não possuem acervo disponível para o acesso público pela Internet pode ser um indicativo da necessidade de se realizarem eventos ou atividades que propiciem a transferência de tecnologias entre as diversas casas legislativas brasileiras.

Outro ponto a ser destacado é que, pelo estudo realizado, parece não haver preocupações por parte dos administradores dos sistemas com relação à indexação automática das páginas pelos mecanismos de busca da web. Isso não apenas em termos de promoção ou marketing das informações dos sistemas, mas também em relação à segurança das informações armazenadas em servidores que estão conectados à Internet. O estudo mostrou que há diferenças em relação à adoção de tecnologias, sendo a mais significativa o fato de que alguns sistemas, como o SAPL e o sistema da Assembleia Legislativa da Paraíba, adotam posturas cooperativas e orientadas para a promoção do *software* livre. O estudo revelou também atividades concorrentes entre os sistemas. Na esfera federal de governo, observa-se que o Senado Federal e a Câmara dos Deputados têm realizado esforços para a conversão para o formato eletrônico do mesmo acervo de leis (leis de hierarquia superior), indicando a possibilidade de que se esteja empreendendo recursos humanos e financeiros de maneira improdutivo.

De acordo com informações obtidas nas entrevistas, verifica-se que existem, em duas casas legislativas, intenções de viabilizar o desenvolvimento de tesouros para serem usados nos sistemas. Outra característica levantada junto aos SIL e que provavelmente representa um obstáculo à integração dos acervos de todas as instituições é o formato de armazenamento dos arquivos. As entrevistas realizadas revelaram que as instituições têm trabalhado isoladamente, seguindo orientações técnicas ou políticas diversas e que, caso haja intenção futura de se integrar os acervos a um sistema mais amplo (supondo-se a criação de um Sistema Nacional de Legislação, por exemplo) os esforços serão majorados.

Quanto aos aspectos de tratamento intelectual dos acervos do NJUR, NJUR-DF e LEGIN, o estudo revelou que a indexação é realizada por humanos, embora haja sistemas que utilizem recursos de indexação automática, como é o caso da Câmara Legislativa do DF, que utiliza indexação textual em sistema de diretórios, e do NJUR que utiliza recursos de indexação textual oferecidos pelo banco de dados. No entanto, o instrumento utilizado como base de indexação intelectual nos três sistemas estudados é o Thesaurus do Senado Federal que, segundo Gonçalves (2004), não possui diretrizes formalmente estabelecidas. Esse fato pode ser indicativo da necessidade de se criar um instrumento padronizado, baseado no esforço cooperativo, e que possa ser utilizado em um âmbito mais amplo. De tal iniciativa poderia surgir, por exemplo, uma Terminologia Legislativa Brasileira, oferecendo insumos não apenas para as pessoas envolvidas em sistemas legislativos, mas também para profissionais do ramo jurídico, como advogados, juizes, entre outros.

Um aspecto importante a ser considerado sobre o controle terminológico, é que os instrumentos de controle tradicionais, como os tesouros, baseiam-se no estabelecimento de restrições rigorosas aos termos, como uma medida destinada à manutenção de certa integridade do vocabulário. Dessa sorte, um conceito composto, como, por exemplo, “tanques de armazenagem de combustível”, indexado pelo documentalista como “*Combustíveis*” e “*Tanques de armazenagem*”, poderá ser pesquisado pelo usuário como “*Armazenagem de combustível*” e “*Tanques*” (Aitchison; Gilchrist, 1979, p. 48). Como esse tipo de problema é passível de ocorrer com muita facilidade, e devido ao fato de que grande parte da informação disponível na *Internet* é indexada automaticamente, com base em algoritmos que analisam a linguagem natural, coloca-se em questão, neste trabalho, a validade de se utilizar as linguagens documentárias com o objetivo central de aplicar restrições ao vocabulário.

Em contrapartida a essa prática, o FORMATER foi sugerido neste trabalho de pesquisa como um instrumento destinado ao estabelecimento de todas as possíveis variações semânticas de determinado termo num dado domínio do conhecimento, ampliando-se, assim, as possibilidades de sucesso nas operações de busca e recuperação. É provável que muitas das restrições de vocabulário tenham sua origem na própria limitação tecnológica. O texto do Unisist, (1975, p. 10), é explícito em relação a esse fato: “*a eventual utilização de equipamento eletrônico de processamento de dados pode exigir: emprego somente de maiúsculas; exclusão de sinais diacríticos; limitação do número de caracteres que um descritor pode assumir*” e previa, inclusive, que a evolução dos equipamentos faria com que essas restrições desaparecessem: “*essas restrições provavelmente desaparecerão num futuro próximo, quando os equipamentos eletrônicos de processamento de dados estiverem mais bem adaptados às necessidades da informação.*”

Com os recursos informáticos atuais (2004) e com a metodologia sugerida neste trabalho, parece possível a manutenção de bancos de termos paralelos, que armazenem as variações do termo de entrada em relação a número, gênero, grafia, entre outras, bem como indicações das suas relações conceituais como sinonímia, homonímia, hiperonímia, hiponímia, entre outras, de modo a fornecer ao usuário e ao especialista em informação maior detalhamento em relação ao conteúdo conceitual do acervo.

Com relação às espécies normativas encontradas nos SIL analisados, são diversas. Obviamente essa diversidade torna o tratamento informatizado mais difícil, tendo em vista que espécies normativas diferentes possuem também estruturas diferentes. Esse fato motivou a proposição, nesta pesquisa, de um formato padronizado de registro documentário para normas legislativas. Tendo em vista questões de interoperabilidade e, ainda, de expansibilidade, optou-se por sugerir a implementação do formato utilizando-se a linguagem XML, que possui destacáveis qualidades com relação à integração de sistemas. Quanto ao aspecto da diversidade de formatos de registros, a linguagem XML favorece a manutenção e a implementação de novos elementos de dados, pela sua natureza extensível. Encontrando motivação nesses fatores é que se formulou a proposta do FORMALEGIS. O FORMALEGIS pode ser considerado uma proposta inicial de formato, que merece discussão pela comunidade legislativa. No entanto, a sua difusão e o seu estabelecimento como uma recomendação no tratamento da informação legislativa poderão trazer grandes benefícios, como a integração nativa entre sistemas legislativos de instituições diferentes dada pela padronização de formatos de armazenamento; a possibilidade de acompanhamento histórico das normas, não somente após sua publicação mas, também, as etapas do processo legislativo que a antecedem; a possibilidade de organização de obras legislativas pertinentes a temáticas específicas como, por exemplo, Legislação Temática sobre Meio Ambiente, Legislação Temática sobre Esportes, entre outras.

O estabelecimento do FORMALEGIS pode ser o primeiro passo para o desenvolvimento de uma ou de várias ontologias legislativas que permitam a realização de inferências por agentes inteligentes, na extração de informação a partir dessas bases ontológicas.

5.1 Sugestões de pesquisa

Em relação aos Sistemas de Informação Legislativa, dois universos não foram cobertos por esta pesquisa: o âmbito dos municípios brasileiros e o das instituições do poder legislativo em outros países. Com relação ao primeiro, julga-se relevante a investigação, com o objetivo de identificar as características dos sistemas de informação legislativa municipais, bem como as características de seus acervos, a fim de se verificar se a metodologia proposta pode ser

estendida ou adaptada às suas particularidades. Com relação ao segundo universo, julga-se importante a obtenção de informações sobre os sistemas daqueles países que possuem maiores avanços com relação à representação e organização da informação legislativa, de modo que se possa estabelecer uma comparação entre as metodologias utilizadas nesses países e as iniciativas brasileiras, incluindo-se o FORMATER e o FORMALEGIS. Uma vez que os formatos sugeridos nesta pesquisa ainda são modelos, sugere-se pesquisas para a sua implementação e validação em situações reais.



Bibliografia

Bibliografia

AITCHISON, J.; GILCHRIST, A. *Manual para a construção de tesouros*. Rio de Janeiro: BNG/Brasilart, 1979. 142 p.

ALVES, I. M. Atividades terminológicas no Brasil. *Terminometro*, número especial nº 3. A Terminologia no Brasil, p. 8-9. União Latina. 1998.

ALVES, M. J. *Sistema LEGIN*. Brasília, Câmara dos Deputados, 30 nov. 2004. Entrevista

AMARAL, M. Montesquieu e o pensamento político. 2003. *O Portal da História – Teoria Política*. Disponível em: <<http://www.arqnet.pt/portal/universal/teoria/montesquieu.html>>. Acesso em: 11 nov. 2003.

ANDRETA, C. A. Legislação como assunto: uma proposta de extensão para a Classificação Decimal de Direito. In: PASSOS, E. (org). *Informação Jurídica: teoria e prática*. Brasília: Thesaurus, 2004, 237 p. pág. 153-172.

ARNITZ, R.; PICHTI, H. *Introducción a la Terminología*. Madrid: Fund G Sanchez Ruiperez. 1995. 382 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). NBR 13789: *Terminologia -- princípios e métodos -- elaboração e apresentação de normas de terminologia*. Rio de Janeiro: ABNT, 1997a. Folheto. 17 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). NBR 13790: *Terminologia -- princípios e métodos -- harmonização de conceitos e termos*. Rio de Janeiro: ABNT, 1997b. Folheto. 6 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). NBR 13791: *Terminologia -- indicativo de línguas e simbologia*. Rio de Janeiro: ABNT. 1997c. Folheto. 14 p.

ATIENZA, C. A. *Documentação jurídica: introdução à análise e indexação de atos legais*. Rio de Janeiro: Achiamé, 1979.

AUBERT, F. H. O Citrat e a terminologia bilíngüe na USP. *Terminometro*, número especial nº3, A terminologia no Brasil, p. 16-17. União Latina, 1998.

AUGER, P.; ROUSSEAU, L. J. *Méthodologie de la recherche terminologique*. Office de la Langue Français. Série Études, Recherches et Documentation, 1977.

BAEZA-YATES, R.; RIBEIRO NETO, B. *Modern Information Retrieval*. New York: Addison-Wesley, 1999. 513 p.

BARBOSA, M. A. Contribuição ao estudo de aspectos da tipologia de obras lexicográficas. *Ciência da Informação*. Brasília, v. 24, n 3, p. 322-327, set/dez. 1995.

BAX, M. P., CAMPOS, F. Como as Máquinas de Busca Indexam Páginas HTML na Web In: Congresso Brasileiro de Biblioteconomia e Documentação, 2000, Rio Grande do Sul.

BERNERS-LEE, T. *Information Management: A proposal*. CERN: 1989. Disponível em: <<http://www.w3.org/History/1989/proposal.html>>. Acesso em: 07 mai. 2003.

BERNERS-LEE, T. *Semantic Web road map*. 1998b. Disponível em: <<http://www.w3.org/DesignIssues/Semantic.html>>. Acesso em: 09 mai. 2003.

BERNERS-LEE, T. *Uniform Resource Identifiers (URI): Generic Syntax*. 1998a. Disponível em: <<http://www.ietf.org/rfc/rfc2396.txt>>. Acesso em: 06 mai. 2003.

BERNERS-LEE, T. *Web Architecture from 50,000 feet*. 1999. Disponível em: <<http://www.w3.org/DesignIssues/Architecture.html>>. Acesso em: 06 mai. 2003.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. *Scientific American.com*. 2001. Disponível em: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&pageNumber=2&carID=2>>. Acesso em: 06 mai. 2003.

BESSÉ, B. Terminological Definitions. In: WRIGHT, S. E.; BUDIN, G. (org) *Handbook of Terminology Management*. Vol. I. Basic aspects of Terminology Management. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1997. 370 p.

BLUM, R. O. A Internet e os Tribunais. *Internet Legal*, 2001. Disponível em: <<http://www.internetlegal.com.br/artigos/>>. Acesso em: 25 ago. 2003.

BORGES, R. C. P. *Interlegis – Rede de Comunicação e Participação Legislativa*. 2000. Disponível em: <<http://www.iadb.org/leg/Seminar/Documents/interlegis%20Peres%20Nascimento%20Port.pdf>>. Acesso em: 31 ago. 2003.

BOSAK, J.; BRAY, T. XML and the Second-Generation Web. *Scientific American*, New York. May, 1999. Disponível em: <<http://www.sciam.com/article.cfm?articleID=0008C786-91DB-1CD6-B4A8809EC588EEDF>>. Acesso em: 12 mai. 2003.

BOWMAN, J. K. *Search Engine optimization and web site promotion*. 2004. Disponível em: <<http://spider-food.net/>>. Acesso em: 16 dez. 2004.

- BRASIL. *Constituição da República Federativa do Brasil*: promulgada em 5 de outubro de 1988. Organização dos textos, notas remissivas e índices por Juarez de Oliveira. 4. ed. São Paulo: Saraiva, 1990. 168 p.
- BRASIL. Decreto nº 4.176, de 28 de março de 2002. *Diário Oficial da União*, Poder Executivo, Brasília, DF, 1 abr. 2002.
- BRASIL. Lei Complementar nº 95, de 26 de fevereiro de 1998. *Diário Oficial da União*, Poder Executivo, Brasília, DF, 27 fev. 1998.
- BRICKLEY, D. *Semantic Web history: nodes and arcs 1989-1999. The www proposal and RDF*. 2001. Disponível em: <<http://www.w3.org/1999/11/11-WWWProposal>>. Acesso em: 09 mai. 2003.
- BRUNINGHAUS, Stefanie; ASHLEY, Kevin D. Toward adding knowledge to learning algorithms for indexing legal cases. In: *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, p. 9-17, Oslo: Norway, June, 14-18, 1999. 220 p.
- CABRÉ, M.T. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antártida/Empúries, 1993. 529 p. p. 123-224 .
- CÂMARA DOS DEPUTADOS. *Sistema de Legislação Brasileira – manual de entrada de dados e recuperação de informações*. Brasília: Câmara dos Deputados, 2000.
- CAMPOS, C. F. *Fundamentos de Terminologia*. Ouro Preto: Imprensa Universitária/ UFOP, 1992. 115 p.
- CARNEIRO, M. V. Diretrizes para uma política de indexação. Belo Horizonte: *Revista da Escola de Biblioteconomia*. UFMG, 14(2):221-241 set. 1985.
- CARVALHO, D. de Q. *Classificação decimal de direito*. - 4. ed. rev. e atual. Brasília: Presidência da República, 2002. Disponível em: <<http://legislacao.planalto.gov.br/cddir/cddir.nsf>>. Acesso em: 27 dez. 2004.
- CAVALCANTI, A. M. B. *Proposta de microestrutura para dicionário terminológico bilingüe português-íngles para tradutores*. 2000. 88 p. Dissertação (Mestrado em Linguística) – Instituto de Letras, Departamento de Linguística, Línguas Clássicas e Vernácula – LIV, Universidade de Brasília, Brasília.
- CAVALCANTI, C. R. *Indexação e tesouro: metodologia e técnicas*. Brasília: Associação dos Bibliotecários do Distrito Federal. 1978. 87 p.

CAVALCANTI, C. R., Indexação. In: *Estudos avançados em biblioteconomia e Ciência da Informação*, volume I. Brasília: ABDF, 1982.

CENDÓN, B. V. Ferramentas de busca na Web. *Ciência da Informação*, Brasília, v. 30, n. 1, p. 39-49, jan./abr. 2001.

CHAUMIER, J. *Les techniques documentaires*. Paris, França: Presses Universitaires de France. 1994. 127 p.

CICILIATI JR, F. *Sistema S/APL*. Brasília, Programa Interlegis, 01 dez. 2004. Entrevista.

CINTRA, A. M. *Para entender as linguagens documentárias*. São Paulo: Polis/APB, 1994. 72 p .

CLAY, B. *Search Engine Chart*. Disponível em:
<<http://www.bruceclay.com/searchenginechart.pdf>>. Acesso em: 19 set. 2004.

COLAB. *Projeto de Lei*. Disponível em: <<http://genesis.interlegis.gov.br/tecnologia/XML-Legal/Modelos/ProjetoDeLei>>. Acesso em 01 dez. 2004.

COMITÊ GESTOR DA INTERNET NO BRASIL. *Indicadores - Crescimento da Internet*. Disponível em: <<http://www.cg.org.br/indicadores/brasil-mundo.htm#mundo>>. Acesso em: 22 ago. 2003.

COSERIU, E. *Teoria da linguagem e linguística geral; cinco estudos*. Rio de Janeiro: Presença; São Paulo, EDUSP, 1979.

DACONTA, M. *The Semantic Web. A guide to the future of XML, Web Services, and Knowledge Management*. Indianapolis, Indiana: Wiley Publishing Inc. 2003.

DAHLBERG, I. Fundamentos teórico-conceituais da classificação. *Revista de Biblioteconomia de Brasília*, v. 6, n. 1, jan/jun. 1978b, p. 9-21.

DAHLBERG, I. *Ontical structures and universal classification*. Bangalore : Sarada Ranganthan Endowment, 1978 64p.

DAML. About the DAML Language. Publicado em 03 mai. 2003. Disponível em: <<http://www.daml.org/about.html> >. Acesso em: 12 nov. 2004.

DAY, M. Metada Formats. In: *DESIRE Information Gateways Handbook*. 2000. Disponível em: <<http://www.desire.org/handbook/2-3.html>>. Acesso em: 16 dez. 2004.

DEMPSEY, L.; HEERY, R. *Metadata: a current view of practice and issues*. Bath, United Kingdom: UKOLN – The UK Office for Library and Information Networking, University of Bath, 2000. Disponível em: < <http://www.ukoln.ac.uk/metadata/publications/jdmetadata/>>. Acesso em: 24 dez. 2004.

DICTIONARY.COM. 2004. Disponível em: <<http://www.dictionary.com>>.

DISTRITO FEDERAL (Brasil). Câmara Legislativa. *Elaboração de textos legislativos: fundamentos, modelos e regras práticas*. Brasília: CLDF, 1996. 268 p.

DOGPILE. *About DogPile*. Disponível em: <<http://www.dogpile.com/info.dogpl/search/help/about.htm>>. Acesso em: 05 out. 2004.

DUBUC, R.; LAURISTON, A. Terms and Contexts. In: WRIGHT, S. E.; BUDIN, G. (organizadores) *Handbook of Terminology Management*. Vol. I. Basic aspects of Terminology Management. Amsterdam/Philadelphia: John Benjamins Publishing Company. 1997. 370 p.

ENGEL, G. I. Terminologia e Transferência de Tecnologia. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990.

FAULSTICH, E. A Terminologia no Brasil: histórico e perspectivas II. *Terminometro*, número especial n°3, A terminologia no Brasil, p. 10-12. União Latina. 1998b.

FAULSTICH, E. L. J. A Terminologia na Universidade de Brasília. *Terminometro*, número especial n°3, A terminologia no Brasil, p. 13-15. União Latina. 1998a.

FAULSTICH, E. L. J. Metodologia para projeto terminográfico. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990

FAULSTICH, E. L. J. Socioterminologia: mais que um método de pesquisa, uma disciplina. *Ciência da Informação*. v. 24, n° 3. 1995

FELBER, H. *Terminology Manual*. Paris: Unesco, 1984. 375 p.

FELBER, H.; PICHT, H. *Métodos de terminografía y principios de investigación*. Madrid: Instituto Miguel Cervantes, 1984. 254 p .

FERREIRA, A. B. de H. *Novo Dicionário da Língua Portuguesa*. Rio de Janeiro: Editora Nova Fronteira, 1986.

FERREIRA, A. B. de H: *Dicionário Aurélio Eletrônico – Século XXI*. Versão. 3.0. Rio de Janeiro: Nova Fronteira, 1999. CD-ROM.

FUJITA, M. S.L. Linguagem documentária em odontologia: uma aplicação do sistema de indexação Precis. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990.

FURGERI, Sérgio. *Ensino didático da linguagem XML*. São Paulo: Érica, 2001.

GALINSKI, C.; KROMMER-BENZ, M. Actividades Internacionales en terminología. In: *Actas del Primer Simposio Latinoamericano de Terminología*. Del 4 al 8 de abril de 1988. Caracas-Venezuela: Universidad Simon Bolívar – Union Latina. 1988.

GALINSKI, C.; NEDOBITY, W. La estrategia de Informacion en Termnet. In: *Actas del Primer Simposio Latinoamericano de Terminología*. Del 4 al 8 de abril de 1988. Caracas-Venezuela: Universidad Simon Bolívar – Union Latina. 1988.

GILCHRIST, A. *The Thesaurus in retrieval*. Londres: Aslib, 1971. 184 p.

GLOBAL21. *Blocos Econômicos*. Disponível em: <<http://www.global21.com.br/blocoeconomicos/>>. Acesso em: 31 jul. 2004.

GOMES, G. D. S. *A vinculação de parlamentares a grupos de interesse e seus impactos na reeleição para a Câmara Legislativa do Distrito Federal*. Brasília: UnB, Departamento de Ciência Política, dissertação. 1995. 148 p.

GOMES, H. E. *Classificação, Tesouro e Terminologia – fundamentos comuns*. 2004. Disponível em: <<http://www.conexaorio.com/bit/tertulia/tertul.htm>>. Acesso em: 04 nov. 2004.

GOMES, H. E. *Manual de Elaboração de tesouros monolíngües*. Brasília Prog. Nac. Bibl. Inst. Ens. Superior 1990a 77 p .

GOMES, H.; CAMPOS, M. L. A.; MOTTA, D. F. *Elaboração de Tesouro Documentário – Tutorial*. 2004. Disponível em: <<http://www.conexaorio.com/bit/tesouro/index.htm>>. Acesso em: 05 nov. 2004.

GOMES, Hagar E.; CAMPOS, M. A. Classificação e Terminologia. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990b.

GOMÉZ-PÉREZ, A. Ontological Engineering: a state of the art. *Expert Update*. 1999. Expert Update. Ontono 2(3): 38-43. Disponível em:

<<http://citeseer.ist.psu.edu/cache/papers/cs/22343/http:zSzzSzwww.csc.liv.ac.ukzSz~franzSzExpertUpdatezSzontologies.pdf/ontological-engineering-a-state.pdf>>. Acesso em: 17 jul. 2004.

GONÇALVES, M. N. *Sistema NJUR*. Brasília, Senado Federal, 27 nov. 2004. Entrevista.

GORSKI, D. P.; TAVANTS, P. V. El concepto. In: *Logica*. Traducción direta del reusso por Augusto Vidal Roget. Mexico Ed. Grijalbo 2 ed 1968. 316 p. p. 38-83 .

GUARINO, N.; WELTY, C. Supporting ontological analyses of taxonomic relationships. *Data & Knowledge Engineering*, v. 39, 2001. p. 51-74. Disponível em: <http://www.loa-cnr.it/Papers/dke2001.pdf>. Acesso em: 21 nov. 2004.

GUIMARÃES, J. A. C. A recuperação temática da informação em direito do trabalho no Brasil: propostas para uma linguagem de indexação a partir de estudo crítico-comparativo das linguagens utilizadas em centros especializados na área. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990.

GUINCHAT, C, MENOU, M. *Introdução geral às ciências e técnicas da informação e documentação*. Brasília: IBICT.1994. 540p.

HARMAN, Donna. Automatic Indexing. In: *Challenges in Indexing Electronic Text and Images*. Medford, New Jersey: Asis, 1994. p.247-264

HAROLD, E. R. *XML bible*. Foster City, CA: IDG Books Worldwide, Inc. 1999.

HAROLD, E. R.; MEANS, W. S. *XML in a Nutshell*, 2nd Edition. Gravenstein Highway North, Sebastopol, CA: O'Reilly & Associates Inc. 2002.

HOESCHL, H. C. *Inteligência Jurídica: Novas tendências tecno-jurídicas mundiais*, 2003. Disponível em: <<http://conjur.uol.com.br/textos/16534/>>. Acesso em: 01 set. 2003.

HOUAISS, A. *HOUAISS – Dicionário Eletrônico da Língua Portuguesa*, 2004. Disponível em: <<http://houaiss.uol.com.br/busca.jhtm>>. Acesso em: 06 nov. 2004.

HYVÖNEN, E. The Semantic *Web* – The new Internet of Meanings. In: HYVÖNEN, Eero (editor). *Semantic Web Kick-Off in Finland: vision, technologies, research and applications*. Helsinki, Finland: HIIT Publications. 2002. Disponível em: <<http://www.cs.helsinki.fi/u/cahyvone/stes/semanticweb/kick-off/index.html>>. Acesso em: 26 mai. 2003.

IBGE. *Brasil já tem mais de 180 milhões de habitantes*. Disponível em: <http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=207&id_pagina=1>. Acesso em: 16 dez. 2004

IBICT. *Diretrizes para a elaboração de tesouros monolíngües*. Brasília: IBICT, 1984.

INTEGRATED RESOURCE MANAGEMENT. *Search Engine Optimization: guidelines, tips*. Publicado em 02 set. 2004. Disponível em: <<http://www.integratedresourcemanagement.com/searchengineneeds.html>>. Acesso em: 20 set. 2004.

INTERLEGIS. *Comunidade Virtual do Poder Legislativo*, 2003b. Disponível em: <http://www.interlegis.gov.br/conheca_interlegis/20020219201116/view>. Acesso em: 31 ago. 2003.

INTERLEGIS. *Conheça o Interlegis*, 2003a. Disponível em: <http://www.interlegis.gov.br/conheca_interlegis>. Acesso em: 31 ago. 2003.

INTERLEGIS. *Novos softwares disponíveis para a Comunidade Legislativa*. (2004a). Descrição. 2004a. Disponível em: <http://www.interlegis.gov.br/produtos_servicos/20040315120320/view>. Acesso em: 21 dez. 2004.

INTERLEGIS. *Padronização de Documentos Legislativos*. 2002. Disponível em: http://www.interlegis.gov.br/produtos_servicos/20020109121259. Acesso em 03 dez. 2004.

ISO. *Documentation – Guidelines for the establishment and development of monolingual thesauri*: ISO 2788:1974.

ISO. *Guide for the preparation of classified vocabularies (example of method)*. ISO/R 919-1969

ISO. *International unification of concepts and terms*. ISO/R 860-1968.

ISO. *Naming principles*. ISO/R 704-1968.

ISO. *Terminologia – princípios e métodos: ISO WD 704.1:1996*. 46 p.

ISO. *Terminology-Vocabulary. Terminologie-Vocabulaire: ISO 1087: 1990 (E/F)* Genève ISO 1990 17 p.

KARTOO. *Sistema de ajuda do site*. Disponível em: <<http://www.kartoo.net/a/en/aide01.html>>. Acesso em: 05 out. 2004.

KRIEGER, M. da G. FINATTO, M.J.B. Introdução à Terminologia: teoria e prática. São Paulo: Contexto, 2004.

KRIEGER, M. da G. Terminologia Revisitada. *Delta*, v. 16, n. 2, 2000. p. 209-228

KRIEGER, M. G. Projeto Termisul: formação e cooperação. *Terminometro*, número especial nº3, A terminologia no Brasil, p. 22-24. União Latina. 1998.

KURNIAWAN, Budi. *Java para a web com servlets, JSP e EJB*. Rio de Janeiro: Editora Ciência Moderna, 2002.

LANCASTER, F. W. *Construção e uso de tesouro: curso condensado*. Brasília: IBICT, 1987. 114 p.

LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos Livros, 1993. 347 p.

LASSILA, O. *Introduction to RDF metadata*. 13/11/1997. Disponível em: <<http://www.w3.org/TR/NOTE-rdf-simple-intro>>. Acesso em: 15 mai. 2003.

LIMA, F. *Modelagem Semântica de Aplicações na WWW*. 2003. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil.

LIMA, J. A. O.; SOUZA, M. A. M. *Legislação Brasileira na Internet*. Senado Federal - Prodasen/DSL, 2001. Disponível em: <www.interlegis.gov.br/produtos_servicos/20020109121259/20020109130716/view>. Acesso 27 nov. 2004.

LIMA, M. D. V. *Sistema NJUR-DF*. Brasília, Câmara Legislativa do DF, 27 nov. 2004. Entrevista.

LOPES, R. R. V. Uso de ferramentas automatizadas para expandir procedimentos de indexação e classificação através do uso de bancos de termos. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990.

MAEDCHE, A. *Development and applications of ontologies*. Institute AIFB, Knowledge Management Group, University of Karlsruhe, Germany. 2000. Disponível no endereço <http://events.aifb.uni-karlsruhe.de/ontologytutorial/fgml2000/tutorial_fgml.pdf>. Acesso em: 08 abr. 2003.

MAMMA. *About Mamma.com*. Disponível em: <<http://www.mamma.com/info/about.html>>. Acesso em: 05 out. 2004.

MARCO, D. *Building and Managing the Metadata Repository: A Full Lifecycle Guide*. New York: John Wiley & Sons, Inc. 2000.

MARTINEZ, S. M. Estructuración conceptual y formalización terminográfica de frases en el subdominio de la oncología. *Estudios de Lingüística Española*. v. 19, 2003. Disponível em: <<http://elies.rediris.es/elies19/cap1321.html>>. Acesso em: 17 set. 2003.

MAURITAIN, Jacques. *Elementos de filosofia II : A ordem dos conceitos, logica menor (logica Formal)*. Rio de Janeiro : Livraria AGIR , 1958. 300p.

MEDEIROS, M. B. B. Terminologia brasileira em Ciência da Informação: uma análise. In: *Actas del Primer Simposio Latinoamericano de Terminología*. Del 4 al 8 de abril de 1988. Caracas-Venezuela: Universidad Simon Bolívar – Union Latina. 1988.

MILLER, E.; KOIVUNEN, M. W3C Semantic Web Activity. In HYVÖNEN, Eero (editor) *Semantic Web Kick-Off in Finland*. Vision, technologies, research and applications. Helsinki, Finland: HIIT Publications. 2002. Disponível em: [<http://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/kick-off/index.html>]. Acesso em: 26 mai. 2003.

MOLE – Text Analysis Group. *Boolean Retrieval*. 1999a. Disponível em: <<http://isp.imm.dtu.dk/thor/projects/multimedia/textmining/node2.html>>. Acesso em: 09 set. 2003.

MOLE – Text Analysis Group. *Probabilistic Retrieval*. 1999c. Disponível em: <<http://isp.imm.dtu.dk/thor/projects/multimedia/textmining/node2.html>>. Acesso em: 09 set. 2003.

MOLE – Text Analysis Group. *Vector Space Model*. 1999b. Disponível em: <<http://isp.imm.dtu.dk/thor/projects/multimedia/textmining/node2.html>>. Acesso em: 09 set. 2003.

MOURA, A. M. de C. *A web semântica: fundamentos e tecnologias*. Instituto Militar de Engenharia – Departamento de Engenharia de Sistemas. 2002. Disponível em: <http://ipanema.ime.eb.br/~anamoura/public/tutorial_sbbd2002.zip>. Acesso em: 16 mai. 2003.

NAFRÍA, I. *El futuro de Internet tiene nombre: la web semántica*. 23/05/2001. Disponível no site Baquia.com: <http://www.baquia.com/com/20010523/art00008.html>. Acesso em: 07 nov. 2003.

NAKAYAMA, H. *Análise Temática da Informação*. Nota de aula da disciplina Análise Temática do curso de Doutorado, Departamento de Ciência da Informação e Documentação, Universidade de Brasília 29/03/2001.

NAKAYAMA, H. *Análise Temática da Informação*. Nota de aula da disciplina Análise Temática do curso de Doutorado, Departamento de Ciência da Informação e Documentação, Universidade de Brasília 19/04/2001.

NAKAYAMA, H. Terminologia: tradução e adaptação. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990 a.

NAKAYAMA, H. Tipologia das obras terminológicas. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990 b.

NAKAYAMA, Haruka. *Tradução e adaptação de tesouros – um estudo de caso: Tesouro SPINES*. 1985. 139 p. Dissertação (Mestrado em Biblioteconomia e Documentação) – Faculdade de Estudos Sociais Aplicados, Departamento de Biblioteconomia, Universidade de Brasília, Brasília.

NISO Press. *ANSI/NISO Z39.85-2001 – The Dublin Core Metadata Element Set*. Bethesda, Maryland, USA: NISO Press, 2001. Disponível em: <<http://www.niso.org/standards/resources/Z39-85.pdf>>. Acesso em: 24 dez. 2004.

NISO Press. *Understanding Metadata*. Bethesda, Maryland, USA: NISO Press, 2004. Disponível em: <<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>>. Acesso em: 24 dez. 2004.

NOCETTI, M. A.; FIGUEIREDO, R. C. Línguas Naturais e Linguagens Documentárias: traços inerentes e ocorrências de interação. *Revista de Biblioteconomia de Brasília*, v. 6, n. 1, jan/jun. 1978. p. 23-37.

NORTON, Peter. *Introdução à informática*. São Paulo: Makron Books, 1996.

NOY, N.; MCGUINNESS, D. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford University, 2004. Disponível em: [http://protege.stanford.edu/publications/ontology_development/ontology101.pdf]

NUA - INTERNET SURVEYS. Ecommerce 1998-2003. Disponível em: <http://www.nua.com/surveys/analysis/graphs_charts/comparisons/ecommerce_us.html>. Acesso em: 22 ago. 2003. 2003b.

NUA - INTERNET SURVEYS. *How Many Online?*. Disponível em: <http://www.nua.com/surveys/how_many_online/index.html>. Acesso em: 22 ago. 2003. 2003a.

NUTCH. *Sobre*. Disponível em: <<http://www.nutch.org/docs/pt/>>. Acesso em: 19 set. 2004.

OGDEN, C. K.; RICHARDS, I. A. *O significado de significado*. Rio de Janeiro: Zahar Editores, 1972. 349 p.

ONTOKNOWLEDGE. *Welcome to OIL*. Disponível em:
<<http://www.ontoknowledge.org/oil/>>. Acesso em: 12 nov. 2004.

OUELLET, R.; OGBUJI, U. *Introduction to DAML: Part I*. Publicado em 30/01/2002. Disponível em: <<http://www.xml.com/pub/a/2002/01/30/daml1.html>>. Acesso em: 12 nov. 2004.

PAIVA, M. A. L. de. Informatização da Cultura Jurídica. *Revista Travel Net Jurídica*, 2003. Disponível em: <<http://tvtem.globo.com/juridica/>>. Acesso em: 25 ago. 2003.

PALMER, S. B. *The Semantic web: an introduction*. 2001. Disponível em:
<<http://infomesh.net/2001/swintro/>>. Acesso em: 17 mai. 2003.

PASSOS, E. Doing Legal Research in Brazil. 2002. *Law Library Resource Xchange – LLRX.com*. Disponível em: <<http://www.llrx.com/features/brazil2002.htm>>. Acesso em: 01 set. 2003.

PAVEL, S.; NOLET, D. *Précis de Terminologie*. Travaux publics et services gouvernementaux. Canadá. 2001.

PEREIRA, A. A. de C. Representação documentária de informação legislativa. In: PASSOS Edilenice (org). *Informação Jurídica: teoria e prática*. Brasília: Thesaurus, 2004. 237 p. pág. 153-172.

PEREIRA, P. de P. Processo legislativo: a revisão entre as Casas do Congresso Nacional. *Jus Navigandi*, Teresina, a. 4, n. 46, out. 2000. Disponível em:
<<http://www1.jus.com.br/doutrina/texto.asp?id=137>>. Acesso em: 11 nov. 2003.

PICHT, H. Principios de Investigación Terminologica. In: *Métodos de Terminografía y Principios de Investigación Terminologica*. Madrid: Instituto Miguel de Cervantes, 1984.

PIEIDADE, M. A. R. *Introdução à teoria da classificação*. Rio de Janeiro: Interciência, 1977. 190 p.

POWERS, S. *Paractical RDF*. Sebastopol, CA: O'Reilly & Associates Inc. 2003. 350 p.

PRADO JR, C. *Notas introdutórias à lógica dialética*. 3.ed. São Paulo : Brasiliense , 1968. 264p.

PROFUSION. *Help*. Disponível em:
<http://www.profusion.com/util.asp?TID=XSLTRANSID_HELP&cobid=ess&sid={9C745228-F0B1-4F40-A19C-37BF727DBBB1}#XSLTRANSID_DEFAULTPAGE>. Acesso em: 05 out. 2004.

PYTHON. *Quotes about Python*. Disponível em: < <http://www.python.org/Quotes.html> >. Acesso em: 25 nov. 2004.

QUIVEY, M. A. *The Semantic Web The Future of Our Online Experience*. EDL547 New Technology Research Paper. MTL Program – University of Illinois. 22/04/2002. Disponível em: <<http://students.uis.edu/mquiv01s/Semanticweb.html>>. Acesso em: 12 mai. 2003.

REGISTRO.BR. *Domínios Registrados por DPN*. Disponível em: <<http://registro.br/estatisticas.html>>. Acesso em: 22 ago. 2003.

ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. In: MACHADO, U. D. (editor). *Estudos avançados em Biblioteconomia e Ciência da Informação*. Volume I. Brasília: ABDF, 1982.

ROBREDO, J.; CUNHA, M. B. *Documentação de hoje e de amanhã: uma abordagem informatizada da biblioteconomia e dos sistemas de informação*. São Paulo: Global, 1986.

ROCHA, M. P. C. *Estudo sobre o comportamento informacional de parlamentares e assessores legislativos na Câmara Legislativa do Distrito Federal como subsídio à gestão estratégica de informações no processo legislativo*. 2001. 205 f. Dissertação (Mestrado em Ciência da Informação e Documentação, Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2001.

ROCHA, M. P. C. *Sistema NJUR-DF*. Brasília, Câmara Legislativa do DF, 27 nov. 2004. Entrevista.

RONDEAU, Guy. *Introduction à la terminologie – deuxième édition*. Gaëtan Morin Éditeur, 1984.

ROSNER, Fred. *The Life of Moses Maimonides, a Prominent Medieval Physician*. In *Medical History*. Disponível em: <http://www.aecom.yu.edu/home/einsteinquarterly/PDFs/Volume_19/Number_3/19Rosner125.pdf>. Acesso em: 21 jul. 2004.

ROVER, A. J. Sistemas especialistas legais: uma solução inteligente para o Direito. *Internet Legal*, 2001. Disponível em: <<http://www.internetlegal.com.br/artigos/>>. Acesso em: 25 ago. 2003.

SAGER, J. C. *A Practical course in terminology processing*. Philadelphia: J Benjamins, 1990. 254 p.

SÁNCHEZ, M. T. *La fraseología en el lenguaje biomédico: análisis desde las necesidades del traductor*. 1999. Disponível em: < <http://elies.rediris.es/elies6/index.html#indice> >. Acesso em: 22 dez. 2004.

SCIRUS. *About Scirus*. Disponível em: <<http://www.scirus.com/srsapp/aboutus/>>. Acesso em: 05 out. 2004.

SEMANTIC *web primer*. 2002. Disponível em: <http://uwimp.com/eo.htm>. Acesso em: 15 mai. 2003.

SEOCONSULTANTS.COM. *History of Search Engines and Directories - Search Engine History*, 2004. Disponível em: <http://www.seoconsultants.com/search-engines/history/>. Acesso em: 19 jul. 2004.

SEOCONSULTANTS.COM. *Meta Tags - Metadata Elements*. 2004a. Disponível em: <http://www.seoconsultants.com/meta-tags/>. Acesso em 26 dez. 2004.

SERVER WATCH. *December 2004 Netcraft Survey Highlights*. Disponível em: <http://www.serverwatch.com/stats/netcraft/article.php/3444451>. Acesso em: 12 dez. 2004.

SILVA, J. C. G. da; ASSIS, F. S. G. de. *Linguagens de programação: conceitos e avaliação*. São Paulo: McGraw-Hill, 1988. 213 p.

SPITERI, L. F. *The Classification Research Group and the Theory of Integrative Levels*. Faculty of Information Studies. University of Toronto. 1995. Disponível em: <http://www.lis.uiuc.edu/review/summer1995/spiteri.html>. Acesso em: 05 ago. 2004.

STUCKENSCHMIDT, H.; VAN HARMELEN, F. *Information Sharing on the Semantic Web*. 2003. Copiado da *Internet* por meio do aplicativo E-mule <http://www.emule-project.net>. Acesso em: 15 nov. 2004. (manuscrito)

SULLIVAN, D. *Major Search Engines and Directories*. Publicado em 28 abr. 2004. Disponível no *site* Search Engine Watch <http://searchenginewatch.com/links/article.php/2156221>. Acesso em: 13 set. 2004.(B)

SULLIVAN, D. *Score Media Matrix: Search Engine Ratings*. Publicado em 23 jul. 2004. Disponível em: <http://searchenginewatch.com/reports/article.php/2156431>. Acesso em: 20 set. 2004. (A)

SULLIVAN, D. *Who Powers Whom? Search Providers Chart*. Publicado em 23 jul. 2004. Disponível no *site* Search Engine Watch <http://searchenginewatch.com/reports/article.php/2156401>. Acesso em: 20 set. 2004.(C)

SWARTZ, A. *The Semantic Web In Breadth*. 2002. Disponível em: <http://logicerror.com/semanticWeb-long>. Acesso em: 16 mai. 2003.

SZKLAROWSKY, L. F. Distrito Federal: município ou estado e a Lei de Responsabilidade Fiscal. *Coojornal*. Disponível em: <http://www.riototal.com.br/coojornal/szklarowsky004.htm>. Acesso em: 18 nov. 2003.

TÁLAMO, M. F. G. M. *Linguagem documentária*. São Paulo: Associação Paulista de Bibliotecários, 1997. 12 p.

TÁLAMO, M. F. M. Instrumentos de controle terminológico: limites e funções. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990.

TANSLEY, D. *Como criar web pages rápidas e eficientes usando PHP e MySQL*. Rio de Janeiro: Editora Ciência Moderna, 2002.

TAYLOR, C. *An Introduction to metadata*. 2003. Disponível em: <<http://www.library.uq.edu.au/iad/ctmeta4.html>>. Acesso em: 16 dez. 2004.

TCDF. Tribunal de Contas do Distrito Federal. *Portaria nº 290/98*. disponível em: <<http://www.tc.df.gov.br/silegisorig/tcdf/port/presi/por-1998-00290-999.html>>. Acesso em: 20 nov. 2003.

TTT.ORG. *Master List of 12620 Data Categories*. Disponível em: <<http://www.ttt.org/oscar/xlt/webtutorial/mstrlist.htm#top>>. Acesso em: 21.12.04.

UNISIST/UNESCO. *Normas para a elaboração de Thesaurus*. Brasília: Unesco. 1975. 33p.

USCHOLD, M; GRUNINGER, M. *Ontologies: principles, methods and applications*. Edinburg: The University of Edinburg, 1996. Disponível em: <<http://www.aii.ed.ac.uk/project/pub/documents/1996/96-ker-intro-ontologies.ps>>. Acesso em: 15 jul. 2004.

VALENTINE, C; WINNICK, C. *XHTML*. Rio de Janeiro: Campus, 2001.

VAN CAMPENHOUDT, M. *Abrégé de Terminologie Multilingue*, 1997. Disponível em: <<http://www.termisti.refer.org/theoweb1.htm>>. Acesso em: 09.06.01.

VEIGA, C. C. da. A terminologia e a hierarquização de idéias. In: *Anais do II Simpósio Latino Americano de Terminologia e I Encontro Brasileiro de Terminologia Técnico Científica*. 10 a 14/setembro/1990. Brasília: IBICT. 1990

VIVÍSSIMO. *Help*. Disponível em: <<http://vivissimo.com/help.html>>. Acesso em: 05 out. 2004.

W3C *Resource Description Framework (RDF): Concepts and Abstract Syntax*. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>> Acesso em: 21 dez. 2004.

W3C. *HTML 4.01 Specification*. 24/12/99. Disponível em: <http://www.w3.org/TR/REC-html40/cover.html#minitoc>. Acesso em: 10 nov. 2003.

W3C. *HyperText Markup Language (HTML) Home Page*. 11 abr. 2003. Disponível em: <http://www.w3.org/MarkUp/>. Acesso em: 10 nov. 2003.

W3C. *OWL Web Ontology Language - XML Presentation Syntax*. 2003c. Disponível em: <http://www.w3.org/TR/owl-xmlsyntax/#subsec-struct>. Acesso em 27 set. 2004.

W3C. *RDF Vocabulary Description Language 1.0: RDF Schema*. 2004b. Disponível em: http://www.w3.org/TR/rdf-schema/#ch_introduction. Acesso em: 21 dez. 2004.

W3C. *Resource Description Framework*. 2003a. Disponível em: <http://www.w3.org/RDF/>. Acesso em: 05 mai. 2003.

W3C. *Semantic Web*. 2003b. Disponível em: <http://www.w3.org/2001/sw/>. Acesso em: 05 mai. 2003.

W3C. *XML Schema Part 0: Primer Second Edition*. 2004a. Disponível em: <http://www.w3.org/TR/xmlschema-0/>. Acesso em: 30 dez. 2004.

W3SCHOOLS. *Rdf Tutorial*. Disponível em: <http://www.w3schools.com>. Acesso em: 18 nov. 2004.

WALD, Arnaldo. *A elaboração e revisão dos projetos de códigos*. Revista de Direito da Procuradoria Geral do Estado da Guanabara, n. 21, p. 166-189, 1969.

WALL, A. *Keyword Phrases and the Meta Keywords Tag*. 2004a. Disponível em: <http://www.search-marketing.info/search-engine-history/index.htm>. Acesso em: 26 dez. 2004.

WALL, Aaron. *History of Search Engines & Web History*. 2004. Disponível em: <http://www.search-marketing.info/search-engine-history/index.htm>. Acesso em: 19 set. 2004.

WARNER, A J. The role of linguistic analysis in full-text retrieval. In: *Challenges in Indexing Electronic Text and Images*. Medford, New Jersey: Asis, 1994. p.265-275

WILLEY, J. *A history of search engines*. Disponível em: <http://www.wiley.com/legacy/compbooks/sonnenreich/history.html>. Acesso em: 19 set. 2004.

WRIGHT, S. E. Term selection: the initial phase of Terminology Management. In: WRIGHT, S. E.; BUDIN, Gerhard. (organizadores) *Handbook of Terminology Management*. Vol. I. Basic aspects of Terminology Management. Amsterdam/Philadelphia: John Benjamins Publishing Company. 1997. 370 p.

WRIGHT, S. E.; BUDIN, G. (organizadores) *Handbook of Terminology Management*. Vol. I. Basic aspects of Terminology Management. Amsterdam/Philadelphia: John Benjamins Publishing Company. 1997. 370 p.

YAHOO!. *How do I improve the ranking of my web site in the search results?* 2004. Disponível em <<http://help.yahoo.com/help/us/ysearch/ranking/ranking-02.html>>. Acesso em 26 dez. 2004.



Apêndices

APÊNDICE A — Sites de Casas Legislativas do Brasil

APÊNDICE B — Roteiro de Entrevista Semi-Estruturada

APÊNDICE C — Elementos de Dados Terminológicos do FORMATER

Apêndice A – Sítios de Casas Legislativas do Brasil

Instituição	Estado	URL
Assembléia Legislativa do Acre	AC	www.aleac.ac.gov.br/
Assembléia Legislativa de Alagoas	AL	http://www.ale.al.gov.br/
Assembléia Legislativa do Amazonas	AM	http://www.aleam.gov.br
Assembléia Legislativa do Amapá	AP	www.al.ap.gov.br
Assembléia Legislativa da Bahia	BA	www.bahia.ba.gov.br/assemb/
Assembléia Legislativa do Ceará	CE	http://www.al.ce.gov.br/
Câmara Legislativa do Distrito Federal	DF	http://www.cl.df.gov.br
Assembléia Legislativa do Espírito Santo	ES	http://www.al.es.gov.br
Assembléia Legislativa de Goiás	GO	http://www.assembleia.go.gov.br
Assembléia Legislativa do Maranhão	MA	http://www.al.ma.gov.br/
Assembléia Legislativa de Minas Gerais	MG	http://www.almg.gov.br
Assembléia Legislativa de Mato Grosso do Sul	MS	http://www.al.ms.gov.br
Assembléia Legislativa de Mato Grosso	MT	http://www.al.mt.gov.br
Assembléia Legislativa do Pará	PA	http://www.alepa.pa.gov.br/
Assembléia Legislativa da Paraíba	PB	http://www.al.pb.gov.br
Assembléia Legislativa de Pernambuco	PE	http://www.alepe.pe.gov.br
Assembléia Legislativa do Piauí	PI	http://www.alepi.pi.gov.br
Assembléia Legislativa do Paraná	PR	http://www.pr.gov.br/assembleia
Assembléia Legislativa do Rio de Janeiro	RJ	http://www.alerj.rj.gov.br
Assembléia Legislativa do Rio Grande do Norte	RN	http://www.assembleiam.com.br/
Assembléia Legislativa de Rondônia	RO	http://www.ale.ro.gov.br/
Assembléia Legislativa de Roraima	RR	http://www.al.rr.gov.br
Assembléia Legislativa do Rio Grande do Sul	RS	http://www.al.rs.gov.br
Assembléia Legislativa de Santa Catarina	SC	http://www.alesc.sc.gov.br
Assembléia Legislativa de Sergipe	SE	http://www.al.se.gov.br
Assembléia Legislativa de São Paulo	SP	http://www.al.sp.gov.br
Assembléia Legislativa de Tocantins	TO	http://www.al.to.gov.br

Apêndice B – Roteiro de entrevista semi-estruturada

I – Identificação do entrevistado

Cargo ou área de atuação na Casa Legislativa:

Função em relação ao sistema:

II – Identificação do acervo (assinalar à esquerda dos nomes das normas que o sistema abrange)

<input type="checkbox"/>	Aditamento a Atos Convocatórios	<input type="checkbox"/>	Emenda à Lei Orgânica
<input type="checkbox"/>	Ato Convocatório	<input type="checkbox"/>	Emenda Constitucional
<input type="checkbox"/>	Ato da Mesa Diretora	<input type="checkbox"/>	Emenda Constitucional de Revisão
<input type="checkbox"/>	Ato da Mesa da ANC	<input type="checkbox"/>	Lei Complementar
<input type="checkbox"/>	Ato da Mesa Diretora do Congresso Nacional	<input type="checkbox"/>	Lei Constitucional
<input type="checkbox"/>	Ato das Disposições Constitucionais Transitórias	<input type="checkbox"/>	Lei de Responsabilidade Fiscal
<input type="checkbox"/>	Ato Declaratório	<input type="checkbox"/>	Lei Delegada
<input type="checkbox"/>	Ato do Comando Supremo da Revolução de 1954	<input type="checkbox"/>	Lei Ordinária
<input type="checkbox"/>	Ato do Presidente	<input type="checkbox"/>	Medida Provisória
<input type="checkbox"/>	Ato Institucional	<input type="checkbox"/>	Moção
<input type="checkbox"/>	Constituição Estadual ou Lei Orgânica	<input type="checkbox"/>	Ordem de Serviço
<input type="checkbox"/>	Constituição Federal	<input type="checkbox"/>	Portaria
<input type="checkbox"/>	Constituições Federais anteriores a 1933	<input type="checkbox"/>	Regimento Interno
<input type="checkbox"/>	Decisão Normativa da Presidência	<input type="checkbox"/>	Regimento Interno da ANC
<input type="checkbox"/>	Decreto	<input type="checkbox"/>	Resolução
<input type="checkbox"/>	Decreto do Conselho de Ministros	<input type="checkbox"/>	Resolução da Câmara dos Deputados
<input type="checkbox"/>	Decreto Legislativo	<input type="checkbox"/>	Resolução de Revisão Constitucional
<input type="checkbox"/>	Decreto Legislativo do Congresso	<input type="checkbox"/>	Resolução do Congresso Nacional
<input type="checkbox"/>	Decreto Reservado	<input type="checkbox"/>	Resolução do Senado Federal
<input type="checkbox"/>	Decreto-Lei	<input type="checkbox"/>	Outro (especificar):
<input type="checkbox"/>	Deliberação	<input type="checkbox"/>	Outro (especificar):

Formato do registro documental (assinalar à esquerda dos campos de informações):

<input type="checkbox"/>	Ano	<input type="checkbox"/>	Poder
<input type="checkbox"/>	Assunto	<input type="checkbox"/>	Publicação (ano)
<input type="checkbox"/>	Autor	<input type="checkbox"/>	Republicação (ano)
<input type="checkbox"/>	Catálogo / tema	<input type="checkbox"/>	Texto Integral
<input type="checkbox"/>	Data	<input type="checkbox"/>	Tipo
<input type="checkbox"/>	Ementa	<input type="checkbox"/>	Vide
<input type="checkbox"/>	Indexação	<input type="checkbox"/>	Outro:
<input type="checkbox"/>	Número	<input type="checkbox"/>	Outro:
<input type="checkbox"/>	Número da ADiN	<input type="checkbox"/>	Outro:
<input type="checkbox"/>	Número da Proposição	<input type="checkbox"/>	Outro:
<input type="checkbox"/>	Observação	<input type="checkbox"/>	Outro:
<input type="checkbox"/>	Órgão	<input type="checkbox"/>	Outro:
<input type="checkbox"/>	Origem	<input type="checkbox"/>	Outro:
<input type="checkbox"/>	Palavras-chaves	<input type="checkbox"/>	Outro:

III – Características da indexação do acervo	
Uso de instrumento para controle de vocabulário:	
<input type="checkbox"/> tesouro <input type="checkbox"/> vocabulário controlado <input type="checkbox"/> outros (especifique: _____)	
Aspectos semânticos do controle de vocabulário (padronização das entradas, controle de sinonímia, ambigüidade, polissemia; como são estabelecidos os campos semânticos/matérias legislativas)	
Método de indexação: <input type="checkbox"/> automático <input type="checkbox"/> semi-automático <input type="checkbox"/> manual	
IV – Identificação do sistema	
Nome:	Situação atual <input type="checkbox"/> planejamento <input type="checkbox"/> desenvolvimento <input type="checkbox"/> produção desde ____/____/____
Público alvo:	Quantitativo aproximado de usuários que interagem ou que irão interagir com o sistema:
Objetivos do sistema:	
V – Identificação das características tecnológicas do sistema	
Sistema operacional e plataforma de rede (intranet, internet, lan, man, wan):	Linguagens de programação utilizadas:
Sistema de armazenamento do acervo:	Linguagem de recuperação (operadores booleanos, truncamentos, frase exata, etc)

Caracterização da busca por assunto:

Caracterização do módulo de pesquisa avançada:

O sistema utiliza alguma tecnologia para o estabelecimento de correlações semânticas entre os termos (ex. RDF, RDFS, OWL, DAML ou DAML+OIL) ou que propiciem a realização de inferências sobre as informações, do ponto de vista semântico (ex. agentes inteligentes, mineração de dados, mineração de textos, entre outros)?

Apêndice C – Elementos de dados terminológicos do FORMATER

Tabela 107: elemento Entrada

Nome: Entrada		Código: R61
Rótulo	entrada	
Definição	indicação do termo admitido como descritor	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 108: elemento Categoria Gramatical

Nome: Categoria Gramatical		Código: R62
Rótulo	ctgramatical	
Definição	indicação da categoria ou categorias gramaticais a que pertence o termo	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string)	
Valores padrão	Substantivo Adjetivo Locução Adjetiva—	
Comentário	—	
Atributos	valor—	
Subelementos	—	

Tabela 109: elemento Definição

Nome: Definição		Código: R63
Rótulo	definicao	
Definição	indica a definição do termo em dada língua de especialidade ou domínio	
Ocorrência	obrigatória, no mínimo uma vez	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 110: elemento Idioma

Nome: Idioma		Código: R64
Rótulo	idioma	
Definição	indicação do idioma a que pertence o termo	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string); recomenda-se seguir o padrão internacional adotado para os símbolos de idiomas dado pela norma ISO 639 e pela RFC 3066 , preferencialmente com dois caracteres	
Valores padrão	podem ser indicados quando a lista de idiomas for muito extensa, de acordo com a conveniência do sistema—	
Comentário	—	
Atributos	valor (caso haja a indicação de uma lista de valores padrão)—	
Subelementos	—	

Tabela 111: elemento Equivalente

Nome: Equivalente		Código: R65
Rótulo	equivalente	
Definição	indicação, em outro idioma, do termo a que se refere o registro	
Ocorrência	opcional	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	valor (indicar o código do idioma)—	
Subelementos	Definição, Entrada—	

Tabela 112: elemento Gênero

Nome: Gênero		Código: R66
Rótulo	genero	
Definição	indicação do gênero em que se classifica o termo referido	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string)	
Valores padrão	masculino feminino M m F f	
Comentário	—	
Atributos	valor	
Subelementos	—	

Tabela 113: elemento Variação

Nome: Variação		Código: R67
Rótulo	variacao	
Definição	indica as formas variantes aceitas para o termo	
Ocorrência	opcional	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Variante gráfica, Variante ortográfica, Forma abreviada, Registro técnico, Registro vulgar, Registro geográfico—	

Tabela 114: elemento Fonte

Nome: Fonte		Código: R68
Rótulo	fonte	
Definição	indica a fonte bibliográfica do termo	
Ocorrência	obrigatória, no mínimo uma vez	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Autor da Fonte, Título da Fonte, Edição, Editor, Data de Publicação, Título da Parte, Título do Periódico, ISSN, ISBN, Local, Volume, Fascículo, Série, Paginação, URL, Data de Acesso	

Tabela 115: elemento Termo Relacionado

Nome: Termo Relacionado		Código: R69
Rótulo	trelacionado	
Definição	indica um ou mais termos relacionados com o termo de entrada	
Ocorrência	recomendado	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Sinônimo, Antônimo, Oposto, Hipônimo, Hiperônimo, Conexo—	

Tabela 116: elemento Contexto

Nome: Contexto		Código: R631
Rótulo	contexto	
Definição	indicação do contexto em que se aceita certa definição formulada para o termo	
Ocorrência	recomendada	
Tipo	tipo complexo*	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Texto, Fonte do Contexto—	

Tabela 117: elemento Fonte da Definição

Nome: Fonte da Definição		Código: R632
Rótulo	dfonte	
Definição	indicação da fonte bibliográfica da definição	
Ocorrência	única e obrigatória	
Tipo	tipo complexo	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 118: elemento Área

Nome: Área		Código: R633
Rótulo	darea	
Definição	indicação da área de conhecimento em que se insere a definição	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 119: elemento URI

Nome: URI		Código: R634
Rótulo	uri	
Definição	indicação do uri da definição	
Ocorrência	única e obrigatória	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 120: elemento Variante gráfica

Nome: Variável Gráfica		Código: R671
Rótulo	vgrafica	
Definição	indicação de variante gráfica	
Ocorrência	opcional	
Tipo	tipo complexo	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 121: elemento Variante ortográfica

Nome: Variável Ortográfica		Código: R672
Rótulo	vortografica	
Definição	indicação de variante ortográfica	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 122: elemento Forma abreviada

Nome: Forma abreviada		Código: R673
Rótulo	vfabreviada	
Definição	indicação de variante abreviada	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 123: elemento Forma extensa

Nome: Forma extensa		Código: R674
Rótulo	vfextensa	
Definição	indicação de variante na forma extensa para termo abreviado	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 124: elemento Registro técnico

Nome: Registro técnico		Código: R675
Rótulo	vrtecnico	
Definição	indicação de variante motivada por registro técnico	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 125: elemento Registro vulgar

Nome: Registro vulgar		Código: R676
Rótulo	vrvulgar	
Definição	indicação de variante motivada por registro vulgar	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 126: elemento Registro geográfico

Nome: Registro geográfico		Código: R677
Rótulo	vrgeografico	
Definição	indicação de variante motivada por registro geográfico	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 131: elemento Conexo

Nome: Conexo		Código: R695
Rótulo	conexo	
Definição	indicação de termo conexo	
Ocorrência	recomendada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 132: elemento Fonte do contexto

Nome: Fonte do contexto		Código: R6311
Rótulo	fcontexto	
Definição	indicação da fonte do contexto	
Ocorrência	obrigatória, quando houver contexto	
Tipo	tipo complexo	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	Autor da Fonte, Título da Fonte, Edição, Local, Editora, Data, Página, Paginação, Período, Volume, Fascículo, ISSN, ISBN	

Tabela 133: elemento Texto do contexto

Nome: Texto do contexto		Código: R6312
Rótulo	tcontexto	
Definição	indicação do texto do contexto	
Ocorrência	obrigatória, quando houver contexto	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 134: elemento Autor da Fonte

Nome: Autor da Fonte		Código: 63111
Rótulo	fautor	
Definição	indicação do autor da fonte	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 135: elemento Título da Fonte

Nome: Título da Fonte		Código: 63112
Rótulo	ftítulo	
Definição	indicação do título da fonte	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 136: elemento Edição

Nome: Edição		Código: 63113
Rótulo	edicao	
Definição	indicação da edição da fonte	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 137: elemento Local

Nome: Local		Código: 63114
Rótulo	local	
Definição	indicação do local de edição da fonte	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 138: elemento Editora

Nome: Editora		Código: 63115
Rótulo	editora	
Definição	indicação da editora da fonte	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 139: elemento Data

Nome: Data		Código: 63116
Rótulo	fdata	
Definição	indicação da data da fonte	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 140: elemento Páginas

Nome: Páginas		Código: 63117
Rótulo	paginas	
Definição	indicação do número de páginas da fonte	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 141: elemento Paginação

Nome: Paginação		Código: 63118
Rótulo	paginacao	
Definição	indicação do intervalo de páginas da fonte	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 142: elemento Título do Periódico

Nome: Título do Periódico		Código: 63119
Rótulo	periodico	
Definição	indicação do título do periódico	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 143: elemento Volume

Nome: Volume		Código: 631110
Rótulo	volume	
Definição	indicação do volume do periódico	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 144: elemento Fascículo

Nome: Fascículo		Código: 631111
Rótulo	fasciculo	
Definição	indicação do fascículo do periódico	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 145: elemento ISSN

Nome: ISSN		Código: 631112
Rótulo	issn	
Definição	indicação do ISSN do periódico	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 146: elemento ISBN

Nome: ISBN		Código: 631113
Rótulo	isbn	
Definição	indicação do ISBN da obra	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 147: elemento Fórmula

Nome: Fórmula		Código: 6711
Rótulo	vgformula	
Definição	indicação de variante do tipo fórmula	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 148: elemento Símbolo

Nome: Símbolo		Código: 6712
Rótulo	vgsimbolo	
Definição	indicação de variante gráfica do tipo símbolo	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 149: elemento Sigla

Nome: Sigla		Código: 6713
Rótulo	vgsigla	
Definição	indicação de variante gráfica do tipo sigla	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 127: elemento Sinônimo

Nome: Sinônimo		Código: R691
Rótulo	sinonimo	
Definição	indicação de termo sinônimo	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 128: elemento Antônimo

Nome: Antônimo		Código: R692
Rótulo	antonimo	
Definição	indicação de termo antônimo	
Ocorrência	opcional	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 129: elemento Hipônimo

Nome: Hipônimo		Código: R693
Rótulo	hiponimo	
Definição	indicação de termo hipônimo	
Ocorrência	recomendada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Tabela 130: elemento Hiperônimo

Nome: Hiperônimo		Código: R694
Rótulo	hiperonimo	
Definição	indicação de termo hiperônimo	
Ocorrência	recomendada	
Tipo	cadeia de caracteres (string)	
Valores padrão	—	
Comentário	—	
Atributos	—	
Subelementos	—	

Anexos

ANEXO A — Categorias de Dados Terminológicos -
ISO 12620

ANEXO B — Perguntas de usuários do SRI/CLDF

Anexo A - Categorias de Dados Terminológicos ISO10620

Cód ISO 12620	ISO 12620 Name	Cód ISO 12620	ISO 12620 Name
A.01	Termo	A.02.03.06	Restrição temporal
A.02	Informação relacionada ao termo	A.02.03.07	Restrição proprietária
A.02.01	Tipo de termo	A.02.04	Formação do termo
A.02.01.01	Entrada principal do termo	A.02.04.01	Origem do termo
A.02.01.02	Sinônimo	A.02.04.02	Etimologia
A.02.01.03*	Quase-sinônimo	A.02.05	Pronúncia
A.02.01.04	Termo científico internacional	A.02.06	Divisão silábica
A.02.01.05	Nome comum	A.02.07	Hifenização
A.02.01.06	Internacionalismo	A.02.08	Morfologia
A.02.01.07	Forma extensa	A.02.08.01	Elemento morfológico
A.02.01.08	Forma abreviada	A.02.08.02	Elemento do termo
A.02.01.08.01*	Abreviação	A.02.08.03	Estrutura do termo
A.02.01.08.02*	Forma curta	A.02.09	Estado do termo
A.02.01.08.03*	Inicialismo (somente iniciais)	A.02.09.01	Autorização normativa
A.02.01.08.04*	Acrônimo	A.02.09.02	Qualificador do planejamento da linguagem
A.02.01.08.05*	Termo truncado	A.02.09.03	Estado administrativo
A.02.01.09	Variante	A.02.09.04	Estado do processo
A.02.01.10	Forma transliterada	A.02.10*	Grau de sinonímia
A.02.01.11	Forma transcrita	A.03	Equivalência
A.02.01.12	Forma romanizada (latinismo)	A.03.01	Grau de equivalência
A.02.01.13	Símbolo	A.03.02	Não equivalente
A.02.01.14	Fórmula	A.03.03*	Direcionalidade
A.02.01.15	Equação	A.03.04	Código de confiabilidade
A.02.01.16	Expressão lógica	A.03.05	Comentário de transferência
A.02.01.17	Categorias de gerenciamento de materiais	A.04	Campo de assunto
A.02.01.17.01	Unidade de manutenção de estoque	A.04.01	Sistema de classificação
A.02.01.17.02	Número da parte	A.04.02	Número de classificação
A.02.01.18	Unidade fraseológica	A.05	Descrição relacionada ao conceito
A.02.01.18.01*	Colocação	A.05.01	Definição
A.02.01.18.02*	Conjunto frasal	A.05.02	Explicação
A.02.01.18.03*	Frase sinônima	A.05.03	Contexto
A.02.01.19	Texto padrão	A.05.04	Exemplo
A.02.02	Gramática	A.05.05	Ilustração não textual
A.02.02.01	Parte do discurso	A.05.05.01	Figura
A.02.02.02	Gênero gramatical	A.05.05.02	Áudio
A.02.02.03	Número gramatical	A.05.05.03	Vídeo
A.02.02.04	Animabilidade	A.05.05.04	Tabela
A.02.02.05*	Nome classe	A.05.05.05	Outro dado binário
A.02.02.06*	Adejtivo classe	A.05.06	Unidade
A.02.02.07	Valência	A.05.07	Extensão
A.02.03	Uso	A.05.08	Característica
A.02.03.01	Nota de uso	A.05.09	Origem do conceito
A.02.03.02	Uso geográfico	A.06	Relação conceitual
A.02.03.03	Registro	A.06.01	Relação genérica
A.02.03.04	Frequência	A.06.02	Relação partitiva
A.02.03.05	Qualificador temporal	A.06.03	Relação seqüencial

Fonte: adaptado de asdfikasdj açlsdjlfçasjdf asçldkífças dfasdfkj

* termo não aprovado

Cód ISO 12620	ISO 12620 Name	Cód ISO 12620	ISO 12620 Name
A.06.03.01	Relação temporal	A.10.02.02	Responsabilidade
A.06.03.02	Relação espacial	A.10.02.02.01	Gerador
A.06.04	Relação associativa	A.10.02.02.02	Inserido por
A.07	Estruturas conceituais	A.10.02.02.03	Atualizado por
A.07.01	Sistema do conceito	A.10.02.02.04	Verificado por
A.07.02	Posição do conceito	A.10.02.02.05	Aprovado por
A.07.02.01	Conceito genérico amplo	A.10.02.02.06	Usuário
A.07.02.01	Conceito maior partitivo	A.10.02.02.07	Retirado por
A.07.02.02	Conceito genérico superordenado	A.10.02.02.08	Exportado por
A.07.02.02	Conceito partitivo superordenado	A.10.02.02.09	Importado por
A.07.02.03	Conceito genérico subordinado	A.10.02.02.10	Administrador do subconjunto
A.07.02.03	Conceito partitivo subordinado	(A.10.02.03)	Contador de uso
A.07.02.04	Conceito genérico coordenado	A.10.03	Identificador do subconjunto
A.07.02.04	Conceito partitivo coordenado	A.10.03.01	Cliente do subconjunto
A.07.02.05	Conceito relacionado	A.10.03.02	Cliente inicial do subconjunto
(A.07.02.05.01)	Conceito maior relacionado	A.10.03.03	Identificador do projeto
(A.07.02.05.02)	Conceito menor relacionado	A.10.03.04	Identificador inicial do projeto
(A.07.02.06)	Relação seqüencial	A.10.03.05	Identificador do produto
(A.07.02.07)	Relação temporal	A.10.03.06	Identificador da aplicação
(A.07.02.08)	Relação espacial	A.10.03.07	Identificador do ambiente
A.08	Nota	A.10.03.08	Identificador unitário do negócio
A.08(.01)	Nota	A.10.03.09	Identificador de segurança
A.08(.02)	Nota	A.10.04	Autorização
A.08(.03)	Nota	A.10.04.01	Função de autorização
A.09	Linguagem documentária	A.10.04.02	Identificador de autorização
A.09.01	Nome do tesouro	A.10.04.03	Senha de autorização
A.09.02	Descritor do tesouro	A.10.04.04*	Título da função do usuário
A.09.02.01	Termo principal (<i>Top term</i>)	A.10.05	Sugestão do usuário
A.09.02.02	Termo maior	A.10.06	Qualificadores administrativos do termo
A.09.02.03	Termo menor	A.10.06.01	Termo herdado
A.09.02.04	Termo relacionado	A.10.06.02	Chave de classificação
A.09.03	Não descritor	A.10.06.03	Termo de busca
A.09.04	Palavra-chave	A.10.06.03.01	Abreviação normalizada
A.09.05	Cabeçalho de índice	A.10.07	Símbolo do idioma
A.10.01	Transações de gestão da terminologia	A.10.08	Termo estrangeiro
A.10.02	Funções de gestão da terminologia	A.10.09	Seqüência de colocação
A.10.02.01	Data	A.10.10	Tipo de entrada
A.10.02.01.01	Data de origem	A.10.11	Estado de trabalho do elemento
A.10.02.01.02	Data de entrada	A.10.12*	Base de dados alvo (para exportação)
A.10.02.01.03	Data de modificação	A.10.13	Fonte da entrada
A.10.02.01.04	Data de verificação	A.10.14	Identificador do conceito
A.10.02.01.05	Data de aprovação	A.10.15	Identificador da entrada
A.10.02.01.06	Data de retirada	A.10.16	Identificador do registro
A.10.02.01.07	Data de padronização	A.10.17	Identificador do arquivo
A.10.02.01.08	Data de exportação	A.10.18	Referência cruzada
A.10.02.01.09	Data de importação	A.10.18.01	Ver

Cód ISO 12620	ISO 12620 Name
A.10.18.02	Ver também
A.10.18.03	Termo invertido
A.10.18.04	Termo permutado
A.10.18.05	Homógrafo
A.10.18.06	Antônimo
A.10.19	Fonte
A.10.20	Identificador da fonte
A.10.21	Identificador do <i>namespace</i>
A.10.21.01	URL
A.10.21.02	Identificador público formal (FPI)
A.10.22	Entidade de origem
A.10.22.01	Pessoa de origem
A.10.22.02	Instituição de origem
A.10.22.03	Base de dados de origem
(A.10.23)	Idioma fonte
(A.10.24)	Idioma alvo
(A.10.25)	Domínio de <i>expertise</i>

Anexo B – Perguntas de usuários do SRI/CLDF

Câmara Legislativa do Distrito Federal
Coordenadoria de Modernização e Informática
Seção de Administração de Sistemas

Pesquisa no sistema de normas jurídicas

1. Pesquisa por assunto
2. Pesquisa qualificada:
 - a. Por parágrafos
 - b. Combinando palavras e parágrafos
3. Pesquisa por seleção
4. Pesquisa pelo nº da norma
 - a. Pesquisa qualificada

Exemplos de solicitações de Pesquisa:

As pesquisas consideradas para exemplos das solicitações referem-se a Leis, Decretos, Leis Complementares, Resoluções e Decretos Legislativos e foram selecionadas entre as solicitações feitas junto ao Setor de Pesquisa e Recuperação da Informação, nos meses de abril e início de maio de 2003:

1. Pesquisar, na legislação vigente, leis e decretos sobre a organização administrativa do GDF, a partir de 1960, inclusive. Ordenar os documentos em ordem descendente de data. Para minimizar a quantidade de páginas impressas, exibir apenas os espelhos das normas e os parágrafos Identificação, Origem, Fonte, Ementa e Vide.
2. Do que se trata a Leicomplementar nº 13?
3. Em que data foi promulgada a Lei 1.246?
4. Quantas normas de origem do GDF tratam de posto de gasolina?
5. Pesquise o artigo 68 da Lei Orgânica, sua regulamentação, se houver, suas alterações, e se houver ADIN, informe seu nº e sentença ou última ação.
6. Pesquise as alterações da Resolução 19. Informe qual o assunto do artigo 10 e suas alterações.
7. Qual a última tabela de vencimentos dos servidores da CLDF e quando foi publicada?
8. Qual a norma mais recente sobre parcelamento do solo? Ela mais antiga?
9. Qual a data de publicação do Decreto nº 20.658?
10. Qual é a lei do DF que trata sobre contrato de gestão?
11. Do que se trata a Lei nº 2.875?
12. Pesquise sobre as normas que tratam sobre abono de servidor da CLDF.
13. Que leis tratam de igrejas ou templos de qualquer religião, culto ou seita, as quais ainda não foram regulamentadas, no âmbito do DF?
14. Pesquise a lei “simples candango” sobre desoneração e facilitação quanto à tributação para pequenas e micro-empresas.
15. Pesquise a lei do silêncio.
16. Qual o anexo da Lei Complementar nº 336?
17. Pesquise se a Lei 1.048/96 foi regulamentada e qual o Decreto.

18. Pesquise, no DCL nº 30.12.92, a lista de autoridades e ordem do mérito legislativo.
19. Pesquise qual a resolução que trata do auxílio pré-escolar.
20. Pesquise os valores dos tetos pagos pelo INSS para aposentados de 1990 até hoje.
21. Pesquise sobre a lei de autoria do Deputado Renato Rainha que dispõe sobre a alteração de normas de gabarito da quadra 40, no Guará.
22. Pesquise todas as Leis Complementares do DF de janeiro até março de 2003 que foram vetadas e aprovadas na CLDF.
23. Pesquise sobre leis que autorizam a concessão e privatização de estacionamento público.
24. Pesquise sobre a lei de incentivo fiscal para informática.
25. Pesquise se o artigo 92 da Lei 8.112/90 está em vigor com o texto original ou se foi alterado.
26. Pesquise sobre o artigo 69-b da resolução 181: qual a data de publicação no DCL?
27. Pesquise a relação dos cargos de todas as administrações regionais e suas alterações.
28. Que leis tratam sobre cercamento de área verde?
29. Pesquise sobre a legislação que trata sobre ressarcimento para servidor da CLDF.
30. Pesquise sobre a legislação que trata sobre programa Habitacional do DF ou Programa habitacional de baixa renda.
31. Qual o decreto que regulamenta a lei 41?
32. Anexo 11 da decisão do TCDF publicado no DODF de 11.12.97, p. 10253 a 10260.
33. Pesquise sobre o acórdão do TJ sobre o CL I.
34. Pesquise sobre a regulamentação do artigo 2467 e 250, § 3º da Lei Orgânica.
35. Qual a ADIN sobre a Lei n. 1516, de 97.
36. Quais as atribuições das secretarias da CLDF?
37. Pesquise se a Lei 531, de 93 está em vigor e quais alterações sofreu.
38. Lei ou Decreto que dispõe sobre política de educação ambiental no DF.
39. Verificar se existe lei do DF que institui a progressividade do IPTU no âmbito do DF.
40. Pesquise sobre todas as leis de autoria do Deputado Xavier, desde seu primeiro mandato.