



Universidade de Brasília

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E CIÊNCIA DA INFORMAÇÃO
E DOCUMENTAÇÃO.
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO E DOCUMENTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

O Comportamento de busca de informação dos pesquisadores da área de insetos frugívoros da Universidade Federal da Grande Dourados (UFGD) para a elaboração de taxonomia.

Rosenilda Marques da Silva Felipe



Universidade de Brasília – UNB
Universidade para o Desenvolvimento do Estado e da Região do Pantanal - UNIDERP
Faculdade de Estudos Sociais Aplicados – Face
Departamento de Ciência da Informação e Documentação – CID
Pós-Graduação em Ciência da Informação

Rosenilda Marques da Silva Felipe

O Comportamento de busca de informação dos pesquisadores da área de insetos frugívoros da Universidade Federal da Grande Dourados (UFGD) para a elaboração de taxonomia.

Dissertação apresentada ao Curso de Pós-Graduação em Ciência da informação do Departamento de Ciência da Informação e de Documentação do mestrado interinstitucional do convênio UnB/Uniderp, como requisito parcial à obtenção do título de Mestre em ciência da informação. Linha de pesquisa: Arquitetura da Informação

Prof^a. Dra. Sofia Galvão Baptista
Orientadora

Campo Grande/MS

2006



UNIVERSIDADE DE BRASÍLIA

Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação
Departamento de Ciência da Informação e Documentação Programa de Pós-Graduação em Ciência da Informação.

FOLHA DE APROVAÇÃO

Título: O comportamento de busca de informação dos pesquisadores da área de insetos frugívoros da Universidade Federal da Grande Dourados(UFGD) para elaboração de taxonomia.

Autora: Rosenilda Marques da Silva

Área de Concentração: Transferência de Informação

Linha de Pesquisa: Gestão da Informação e do Conhecimento

Dissertação 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 a obtenção do título de **Mestre em Ciência da Informação**.

Dissertação aprovada em: 24 de maio de 2006.

Aprovado por:

Profa. Dra. Sofia Galvão Baptista
Presidente – Orientador (UnB/PPGCINF)

Profª. Drª. Marisa Bräscher Basílio Medeiros
Membro Interno - (UnB/PPGCINF)

Profª. Drª. Lígia Maria Arruda Café
Membro Externo – (UCB)

Profª. Drª. Sueli Angélica do Amaral
Suplente – (UnB/PPGCINF)

DEDICATÓRIA

Ao meu filho Pedro Henrique, que agüentou firme, nove meses dentro da barriga, quase sempre apertado entre uma cadeira e uma mesa de computador, e depois mais cinco meses dividindo a mãe com o computador.

AGRADECIMENTOS

Agradeço primeiramente a Deus, Rei dos Reis, Senhor dos Senhores, por me sustentar, guardar e capacitar.

Em especial, ao meu marido Laércio Felipe, pelo incentivo, apoio, amor, compreensão e companheirismo.

À professora Sofia Galvão Baptista, pela orientação, disposição e atenção, durante a realização deste trabalho.

À professora Marisa Brasher, pela atenção, pelas dicas, pelos questionamentos sempre tão importantes para o melhoramento deste trabalho.

Ao professor Manoel Araécio Uchôa Fernandes, pela ajuda, atenção e disposição em ajudar. Sem sua ajuda seria impossível a realização deste trabalho.

À família Macedo (Manoel, Ângela, Glauce e Mirela) por terem me recebido com tanto carinho em sua casa durante os seis meses que passei em Campo Grande.

À minha sogra Ana Ruiz Felipe, por que esteve sempre pronta para ajudar.

À Sônia, Ângela Canesin, Vanessa e Cristiane pela disposição em contribuir para a realização deste trabalho.

A todas a pessoas que me ajudaram de uma maneira ou de outra, seja orando por mim, ou cuidando do Pedro Henrique, para que eu pudesse concluir este trabalho.

RESUMO

Esta pesquisa procura mostrar o valor da organização da informação num momento onde a informação tornou-se de vital importância para uma sociedade que valoriza o saber e o conhecimento, mas que sempre esbarra em barreiras como a dificuldade de recuperação da informação, dificuldade de transmissão e comunicação do conhecimento e a dificuldade de padronização das informações. Em busca de soluções para estes problemas, procura-se mostrar, que o estudo do comportamento do usuário e o estudo das necessidades de informação, quando aliados a mecanismos de organização da informação, pode contribuir significativamente para a melhora da recuperação da informação. Buscou-se na literatura a identificação de ferramentas que dessem apoio para a organização da informação visando sempre o melhoramento da recuperação da informação, e dessa forma chegou-se a mecanismos como as taxonomias e as ontologias. O produto final desta pesquisa foi a construção de uma taxonomia embasada no comportamento de busca e no estudo das necessidades de informação dos usuários do mestrado de Entomologia da Universidade Federal da Grande Dourados, que estudam a Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais. O estudo do comportamento de busca demonstrou que os usuários encontram dificuldades inerentes á recuperação de informação devida á falta de padronização dos termos utilizados nas publicações da área, bem como, pelo grande número de termos em inglês que são utilizados nas publicações. A taxonomia poderá servir de modelo de organização das informações geradas no mestrado de Entomologia.

Palavras-chave: Organização e Representação da Informação, Taxonomia, Recuperação da Informação, Estudo do Comportamento do Usuário, Estudo da Necessidade de Informação.

ABSTRACT

This research shows the value of the organization of the information at a moment where the information became of vital importance for a society that values knowing and knowledge, but always met barriers such as: the difficulty of information retrieval, difficulty of transmission and communication of the knowledge, and the difficulty of standardization of the information. In search for solutions to these problems, we try to show, that the study of the behavior of the user and the study of the information necessities, when allied to mechanisms of organization of the information, can contribute significantly for the improvement of the information retrieval. The identification of tools searched in literature that always gave support for the organization of the information seeking the improvement of the information retrieval, and in this form got to mechanisms as taxonomy and ontology. The final product of this study was the construction of a taxonomy based in the behavior of search and the study of the necessities of information of the users of the master's degree of Entomology of the University Federal da Grande Dourados, who study the Biodiversity of frugivorous insects, its natural hosts and enemies. The study of the search behavior demonstrated that the users found difficulties inherent to the information retrieval, because of not standardization of the terms used in publications of the area, as well as, for the great number of terms in English, which are used in publications. The taxonomy can be used as model of organization of the information generated in the master's degree of Entomology.

Keywords: Organization and Representation of the Information, Taxonomy, Information Retrieval, Study of Information Behavior, Study of the Information Necessity.

LISTA DE FIGURAS

Figura 1 – Exemplo de Taxonomia	74
Figura 2 – Esquema representativo do ambiente de estudo	89
Figura 3 – Categoria Insetos Frugíveros e suas subordinadas	113
Figura 4 – Categoria Manejo Integrado e suas subordinadas	114
Figura 5 – Categoria Controle natural e suas subordinadas	115
Figura 6 – Categoria Controle Biológico e suas subordinadas	116
Figura 7 – Categoria Infoquímicos e suas subordinadas	117
Figura 8 – Categoria Inventário de espécies e suas subordinadas	118
Figura 9 – Categoria Diversidade e suas subordinadas	119
Figura 10 – Categoria Anastrepha e sua subordinadas	120
Figura 11 – Categoria Análise Faunística subordinada da Anastrepha	120
Figura 12 – Categoria Ceratitis Capitata e suas subordinadas	121
Figura 13 – Análise Faunística subordinada da Ceratitis Capitata	121
Figura 14 – Categoria Neosilba e suas subordinadas	122
Figura 15 – Categoria Análise Faunística subordinada da Neosilba	122
Figura 16 – Categoria Dasiops e suas subordinadas	123
Figura 17 – Categoria Análise Faunística subordinada da Dasiops	123

LISTA DE QUADROS

Quadro 1 – Espiral do conhecimento	25
Quadro 4 – Lista endentada da categoria Manejo Integrado de Pragas	102
Quadro 5 – Lista endentada da categoria Inventário de espécies	104
Quadro 6 – Lista endentada da categoria Diversidade de espécies	105

LISTA DE ABREVIATURAS E SIGLAS

CEUD- Centro Universitário de Dourados.

CSDGM - Content Standard for digital Geospacial Metadata

CWM - Common Warehouse Meta Model

DAML - DARPA Agent Markup Language

DC- The Dublin Coreb

DOM - Document Objeto Model

DTD'S – Document Type Definition

GC – Gestão do Conhecimento.

HTML – Hypertext Markup Language

HTTP - Hipertext Transfer Protocol

IES – Instituições de Ensino Superior.

LOM- Learning Objects Metadata

MARC- Machine Readable Card

OIL- Ontology Inference Layer

OWL – Web Ontology Language

RDF –Resource Description Framework

SGML - Standard Generalized Markup Language

SIC – Sistema de Informação Baseados em Co mputador.

UFMS - Universidade Federal de Mato Grosso do Sul

URL - Universal Resource Locator

W3C - World Wide Web Consortium.

WWW- World Wide Web

XML – extensible Markup Language

XSL - eXtensible Stylesheet Language

UFGD – Universidade Federal da Grande Dourados

SUMÁRIO

LISTA DE FIGURAS	VII
LISTA DE QUADROS	VII
LISTA DE SIGLAS	VIII
1 Introdução.....	12
2 Definição do Problema.....	14
2.1 Necessidade de organização da informação	14
2.2 O Comportamento do usuário de informação	16
2.3 Formulação do problema	17
3 Objetivos	19
3.1 Objetivo geral	19
3.2 Objetivos específicos.....	19
4 Justificativa	20
5 Revisão de Literatura.....	22
5.1 Contextualização da sociedade da informação e o excesso de informação	22
5.1.1 Popularização das tecnologias e telecomunicação e sistemas de informação	27
5.2 Recuperação da informação	30
5.2.1 Dificuldades de recuperação	33
5.2.1.1 Dificuldades na área empresarial e iniciativas	35
5.2.1.2 Dificuldades na área acadêmica e iniciativas	37
5.2.1.3 Dificuldades na internet e iniciativas	39
5.2.1.3.1 Web Semântica	42
5.2.1.3.2 XML – eXtensible Markup Language	45
5.2.1.3.3 DTD – Documento Type Definiton.....	48
5.2.1.3.4 Resource Description Framework.....	49
5.3 Organização das informações	50
5.3.1 Teoria Geral da Terminologia.....	52
5.3.2 Teoria do Conceito.....	55
5.3.3 Estudo do comportamento do usuário.....	57
5.3.4 Estudo das necessidades da informação.....	61
5.3.5 Mecanismos de organização da informação	63
5.3.5.1 Metadados.....	63
5.3.5.2 Dublin Core metadado	66
5.3.5.3 Ontologia	68
5.3.5.4 Taxonomia	74

6 Metodologia	83
6.1 Caracterização da pesquisa	83
6.2 Ambiente de estudo	86
6.2.1 Mestrado em Entomologia e Conservação da Biodiversidade	87
6.3 Definição da população / Amostra da pesquisa	90
6.4 Técnica e instrumento de coleta de dados	90
6.5 Resultados do pré-teste	91
7 Análise dos Dados	93
7.1 Necessidades de informação	93
7.2 Comportamento de busca dos usuários	93
8 Elaboração da Taxonomia	99
8.1 Representação da taxonomia – lista endentada	102
8.2 Representação da taxonomia – árvore hierárquica	113
9 Considerações Finais	124
9.2 Considerações sobre a Taxonomia	126
9.3 Perspectivas para novos estudos	127
Referências.....	129
Apêndice A – Esquema representativo da revisão da literatura	141

1 Introdução

A presente pesquisa tem sua origem a partir de um problema enfrentado pela sociedade da informação, ou seja, o excesso de informação. O excesso de informação é um problema que tem gerado inúmeras dificuldades, e dentre elas, pode-se citar a de recuperação das informações, que está diretamente ligada a outros problemas, como por exemplo, a dificuldade de padronização, de armazenamento adequado, e dificuldade de comunicação das informações.

As dificuldades com relação à recuperação da informação ocorrem em diferentes âmbitos do cotidiano humano, e pesquisadores das diversas áreas do saber têm concentrado esforços no sentido de acabar ou pelo menos minimizar esse problema que vem afligindo e prejudicando o desempenho de atividades pessoais, comerciais e científicas.

Inúmeras iniciativas têm sido tomadas para que a recuperação e a comunicação da informação seja efetivada com maior rapidez e precisão¹, e a Ciência da Informação é uma das ciências que tem se preocupado com essa problemática, principalmente a área de organização da informação, que através de mecanismos como metadados, ontologias e taxonomias, organiza as informações antes delas serem adicionadas em sistemas de informação, melhorando assim a recuperação.

A Entomologia é a área das ciências biológicas que estuda os insetos, e dentro das diferentes linhas de pesquisa, está a que se detém nos insetos frugívoros, e, partindo do pressuposto que existe confusão terminológica nesta área, e a partir da constatação da importância da organização da informação como tema de pesquisa, buscou-se na Ciência da Informação e na Ciência da Computação, subsídios teóricos para a realização desta pesquisa.

¹ Segundo Lancaster, uma alta precisão ocorre quando em uma busca a maioria dos itens recuperados, quando não todos, forem considerados pertinentes.

Neste trabalho, inicialmente será dado enfoque aos referenciais teóricos que embasam a problemática deste estudo.

Posteriormente, será abordado a metodologia e os instrumentos utilizados de forma a desenvolver a pesquisa.

Por fim, pretende-se, embasado no estudo do comportamento dos usuários de informação, ou seja, dos pesquisadores envolvidos com o estudo da Biodiversidade de insetos frugívoros, identificar seus comportamentos de busca e os termos que eles utilizam para realizarem suas pesquisas e elaborar uma taxonomia para organizar o conhecimento gerado por esses pesquisadores, padronizando os termos que poderão ser utilizados na produção científica da área.

Para um melhor entendimento das razões pelas quais foram abordados os assuntos da revisão da literatura, será apresentado no Apêndice A, um esquema representativo do conteúdo de assuntos.

2 Definição do Problema

2.1 Necessidade de organização da informação

Atualmente as informações são encontradas em diversos formatos e em inúmeros canais de divulgação, porém, grande parte dessas informações não é recuperada. É muito comum também, ao se realizar uma busca, recuperar muita informação com pouca relevância².

Uma das razões pelas quais as informações não são encontradas quando procuradas, ocorre pelo fato dos mecanismos utilizados para recuperação, não poderem abarcar tipos e formatos, que não foram antecipadamente previstos ou estruturados.

A organização da informação surge como um meio para estruturar as informações, de modo que, ao serem buscadas, sejam encontradas informações com alta relevância². Existem na literatura, diversas propostas para organização da informação ou padronização de terminologias que visam a otimização da recuperação, dentre elas, pode-se citar as taxonomias.

Na definição de Terra (2004), taxonomia é um sistema utilizado para classificar e facilitar o acesso à informação. Seu objetivo é representar conceitos através de termos; melhorar a comunicação entre especialistas e outros públicos; propor formas de controle da diversificação e oferecer um mapa do processo de conhecimento. É portanto, um vocabulário

² : Segundo Lancaster (1993), é a relação entre um enunciado de necessidade de informação e fontes potenciais de informações. Por exemplo, um periódico é considerado relevante para um enunciado de necessidade de informação, se ele examina o problema ou situação abrangida pelo enunciado.

controlado de uma determinada área do conhecimento e um instrumento que permite alocar, recuperar e comunicar informações dentro de um sistema.

A estruturação das informações através da taxonomia permite a elaboração de linguagens documentárias, que podem ser empregadas na organização de conteúdos informacionais e na sistematização de conceitos, e posteriormente, vir a ser utilizadas em sistemas baseados em computadores, permitindo assim, uma recuperação mais eficiente das informações.

Porém, a elaboração de uma taxonomia envolve especialistas no assunto, pois a organização e padronização que a taxonomia prevê, precisa estar de acordo com o que os usuários dessa informação necessitam. A formulação de taxonomias sem a participação de especialista da área pode gerar problemas, como por exemplo:

- ✓ Utilização de termos na taxonomia que não são os mesmos usados pelos usuários daquelas informações;
- ✓ Estruturação dos conceitos diferente da maneira como os usuários costumam buscar as informações;

Questões citadas acima são reais causadoras de problemas para os mecanismos de recuperação que forem embasados em taxonomias sem a participação dos especialistas.

2.2 O comportamento do usuário de informação

Apesar da existência de diversas ferramentas tecnológicas destinadas à recuperação de informação, elas por si só, não podem garantir, que, por haver uma gama imensa de informações em suas bases de dados, as mesmas não ficarão escondidas ou isoladas dentro dessas bases, porque quando se trata de recuperação de informação, a possibilidade de acesso não garante o sucesso da atividade.

Os usuários muitas vezes, ao tentarem realizar uma busca, utilizam termos que para eles, representam a descrição temática do item buscado, porém, não conseguem recuperar essas informações, porque os termos que lhes pareciam lógicos utilizar na hora da busca, não foram previstos anteriormente, e o mecanismo de recuperação não os compreende.

Freitas (1995), diz que a forma com que a informação é trabalhada deve ser observada, sob o risco de, no momento em que o usuário estiver envolvido em um processo decisório, ser-lhe fornecido apenas ruído³, e afirma ainda, que a eficácia do tratamento das informações depende da forma como é administrada e do bom entendimento de certos conceitos e relações, e que, não é concebível que um importante e “caro” recurso, como é a informação, não seja tratado com um grau de seriedade e competência, assegurando ao usuário, um bom suporte informacional.

O Sucesso do usuário final na recuperação e no acesso à informação depende de vários fatores, incluindo quão bem ele entende a linguagem, a estrutura dos dados na base, e quando ele entende, qual(is) os dados disponíveis satisfazem sua necessidade de informação” Figueiredo (1999, apud Garcia; Silva , 2005).

Por isso, a importância de conhecer o comportamento dos usuários e os termos utilizados por eles, antes de se organizarem as informações a serem armazenadas em sistemas, os quais posteriormente constituir-se-ão fontes de informações.

³ Segundo a Wikipédia, é todo fenômeno aleatório que perturba a transmissão correta das mensagens.

2.3 Formulação do Problema

Os problemas que os pesquisadores da biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais enfrentam com relação à recuperação de informação, são semelhantes aos que ocorrem em outras áreas, ou seja, os mecanismos de recuperação de informação, nem sempre recuperam as informações desejadas.

Abaixo, seguem exemplos de problemas enfrentados pelos pesquisadores (usuários de informação).

Buzzi (2003, p. 34) considera os termos biodiversidade e diversidade como sinônimos, mas quando se faz uma pesquisa no Google acadêmico (<http://scholar.google.com.br/>), este que é um mecanismo que recupera documentos acadêmicos, com as frases: biodiversidade de insetos frugívoros e diversidade de insetos frugívoros, obtém-se as seguintes respostas:

- com biodiversidade de insetos frugívoros = foram recuperados 48 respostas;
- com diversidade de insetos frugívoros = foram recuperadas 68 respostas;

Ou seja, o Scholar Google, não faz controle de sinonímia.

Um outro caso encontrado, foi com relação aos termos e suas variações, por exemplo, na enciclopédia Wikipédia versão em inglês ([http://en.wikipedia.org/wiki/Main Page](http://en.wikipedia.org/wiki/Main_Page)), quando se faz uma busca por frugivorous, ele redireciona para frugívoros e então recupera a informação buscada, ou seja, ele considera um termo como sendo a variação do outro.

No scholar Google (<http://scholar.google.com.br/>), ao se realizar a mesma busca, encontra-se os seguintes resultados:

- ✓ frugívoros insects, recupera-se 1780 respostas.
- ✓ frugivorous insects, recupera-se 2550 respostas.

No Scholar Google, a variação do termo não foi prevista e em 770 documentos a variação foi determinante para a recuperação.

Diante da situação anteriormente exposta, se houvesse uma padronização dos termos utilizados nas publicações sobre insetos frugívoros, ou, se os mecanismos fossem preparados para assumir as variações dos termos, talvez a recuperação fosse mais eficaz.

Diante dos problemas de recuperação enfrentados pelos pesquisadores, conhecendo os conceitos e a importância da organização da informação e do mecanismo taxonomia, e tomando por base o comportamento dos usuários e suas necessidades de informação, este trabalho tem o intuito de responder às seguintes questões:

- 1) Qual a necessidade de informação dos pesquisadores que estudam a Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil Central?
- 2) Qual é o comportamento de busca dos usuários de informação quando estudam a Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil Central?
- 3) Quais as relações conceituais para o desenvolvimento de uma taxonomia para a organização e representação do conhecimento gerado sobre a Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil Central?

3 Objetivos

3.1 Objetivo Geral

Identificar as necessidades de informação e o comportamento de busca dos usuários que estudam a biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil Central, para a elaboração de uma taxonomia.

3.2 Objetivos Específicos

- a) Identificar a necessidade de informação dos usuários que estudam a Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil Central;
- b) Identificar o comportamento de busca dos usuários da pós-graduação em Entomologia e conservação da Biodiversidade, que pesquisam sobre a Biodiversidade de insetos frugívoros seus hospedeiros e inimigos naturais do Brasil Central;
- c) Elaborar uma taxonomia para organização da informação gerada sobre a biodiversidade dos insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil Central.

4 Justificativa

A sociedade da informação representa um novo paradigma técnico-econômico e pode ser encarada como um fenômeno global que vem transformando as atividades sociais e econômicas em todo o mundo.

Segundo os autores que abordam essa área do saber, o principal insumo dessa nova economia é a informação, esta que assumiu um importante papel dentro do cotidiano das pessoas e das corporações e passou a ser produzida e consumida em grande escala, o que na literatura, passou a ser denominado de excesso de informação.

Existe tanta informação cercando o dia-a-dia das pessoas e das corporações, que muitas vezes, ao invés de ajudar, ela acaba atrapalhando o andamento tanto das atividades corporativas, quanto pessoais e científicas.

Parece estarmos diante de um paradoxo: ao mesmo tempo em que se fala em excesso de informação, fala-se em ansiedade por informação, pois nem toda a informação que é armazenada é recuperada adequadamente na hora em que se precisa, e muitas vezes a qualidade da informação recuperada não atende às expectativas dos usuários.

A literatura aponta diversas razões que refletem diretamente na relevância ou na qualidade das informações recuperadas, sendo algumas delas, por exemplo: a falta de organização das informações antes de serem armazenadas em sistemas de recuperação, falta de planejamento dos sistemas de recuperação de informação, e o uso inadequado de tecnologia para armazenamento.

Devido aos problemas que a recuperação da informação vem enfrentando, surgiram diversas técnicas, tais como Processamento de Linguagem Natural, Inteligência Artificial e Web Semântica, mas todas elas precisam levar em consideração a organização das

informações, antes que as mesmas sejam adicionadas a sistemas de informação. Outro fator de tão grande importância quanto à organização, é a análise do comportamento de busca dos usuários.

Na área de organização da informação percebe-se algumas iniciativas, podendo-se encontrar no mercado, ferramentas que se comprometem com os objetivos dessa área, como por exemplo, ferramentas de taxonomia, ontologias e outras.

A taxonomia é um importante mecanismo de organização da informação que pode ser utilizada para sistematizar e organizar informações através da criação de grupos hierárquicos de categorias de conceitos podendo representar conceitos através de termos, e dessa forma, melhorar a comunicação entre os especialistas de uma determinada área.

O estudo do comportamento dos usuários pode funcionar como suporte para que a taxonomia possa abordar termos e conceitos, que realmente sejam do interesse dos usuários.

Levando em consideração os problemas enfrentados pelos pesquisadores que estudam a Biodiversidade de insetos frugívoros e também as abordagens encontradas na literatura á respeito da taxonomia, este trabalho busca através da elaboração de um mecanismo de organização da informação fornecer subsídio para o melhoramento da recuperação da informação.

5 Revisão de Literatura

5.1 Contextualização da sociedade da informação e o excesso de informação

O termo sociedade da informação passou a ser muito utilizado nos últimos anos em substituição ao conceito chamado de “sociedade pós-industrial” e representa para muitos, um novo paradigma técnico-econômico. Pode-se dizer que é um fenômeno global; possui um elevado poder transformador nas atividades sociais e econômicas; e exige capacidade de transformar informação em conhecimento com velocidade e segurança.

Diversos autores discorrem sobre essa nova sociedade e visualizam características diversas. Segundo Castells (2000), as características do novo paradigma são:

- 4) **A informação é sua matéria-prima:** As tecnologias se desenvolvem para permitir ao homem atuar sobre a informação propriamente dita.
- 5) **Os efeitos das novas tecnologias têm alta penetrabilidade:** A informação é parte integrante de todas as atividades humanas, e todas elas podem ser afetadas pelas novas tecnologias.
- 6) **Predomínio da lógica de redes:** Graças às novas tecnologias, essa lógica é característica de todo tipo de relação.
- 7) **Flexibilidade:** A tecnologia favorece processos reversíveis permitindo modificações e reorganizações de componentes e permite capacidade de reconfiguração.

8) **Crescente convergência de tecnologias:** O desenvolvimento tecnológico tem permitido uma grande interligação de áreas do conhecimento. As diversas áreas do saber tornam-se interdisciplinares.

Em todas as características definidas por Castells (2000), percebe-se a forte influência da tecnologia, e nesse aspecto, existem diversos discursos e acalorados debates sobre os benefícios e prejuízos dessa influência.

Porém, este trabalho não visa abordar a sociedade da informação em seus aspectos mais profundos. A abordagem será apenas no sentido de contextualizar mudanças paradigmáticas, que trouxeram como consequência a valorização da informação, que passou a ser considerada como o principal insumo dessa nova economia, e trouxe consigo, problemas que serão abordados mais a frente.

Ao se observar as diferentes abordagens sobre as características da nova sociedade da informação, percebe-se a inegável influência da tecnologia, porém, existem diversos pontos de vista sobre a valorização das tecnologias no processo de transformação da nova sociedade. Alguns autores defendem o determinismo tecnológico seguindo a lógica de que as mudanças resultam das tecnologias; outros defendem o evolucionismo, que diz ser a sociedade da informação apenas uma etapa de desenvolvimento.

A nova economia norteadada pela necessidade de criar, armazenar, recuperar e gerenciar informação, aliada à evolução tecnológica criou uma nova fase cultural, chamada por Weil (2000) de cultura informática.

Esta cultura é fortemente influenciada pela valorização das informações, o que pode estar causando a produção acelerada das novas informações, e a tecnologia, surgida ou não devido às mudanças paradigmáticas, tem contribuído grandemente para com o aumento e a disseminação das informações.

A explosão informacional está forçando as pessoas a terem que aprender a gerenciar as informações que as cercam, e inúmeros são os problemas e desafios enfrentados, tanto a nível psicológico, nas pessoas, quanto a nível organizacional, e nesse sentido, uma organização ou corporação pode ser uma empresa, uma universidade ou uma organização religiosa.

A internet é a grande responsável por essa explosão informacional, pois segundo uma pesquisa do Nefcraft (2005), em dezembro de 2004, havia aproximadamente 56 milhões de websites; já em janeiro de 2005, 58 milhões; e em março do mesmo ano, foram encontrados mais de 60 milhões de websites fornecendo conteúdo informativo para todo o mundo.

Os alunos da School of Information Management and Systems da University Of California At Berkeley fizeram um estudo sobre a quantidade de informações que são criadas a cada ano, bem como o fluxo e canais de informações. O projeto chamado de “How much Information?” realizou um estudo em 1999 e detectou que havia entre 1 e 2 hexabytes de informação armazenadas naquele ano. Outro estudo fora feito em 2003, com base nos dados do ano de 2002, e a quantidade alcançou os 5 hexabytes de informação armazenadas nos quatro meios físicos: impresso, filmes, magnéticos e ópticos, e visto e ouvido em quatro canais que são: Televisão, Rádio, Telefone e Internet.

Segundo University of California at Berkely (2003), 98% das informações foram armazenados em meio magnético, principalmente em disco rígido. Já o fluxo de informações através dos canais como TV, Rádio, Telefone e Internet, produziram quase 18 hexabytes de novas informações em 2002.

Devido a essa alta produção de informação, conceitos que se preocupam com o armazenamento, a disseminação, a transmissão e a recuperação das informações, tem ganhado especial atenção, como é o caso da gestão da informação e do conhecimento.

É muito comum, na nova sociedade da informação também conhecida como sociedade do conhecimento, algumas abordagens a respeito do conhecimento e dos tipos de conhecimento existentes.

Nonaka & Takeuchi (1997), afirmam que existem dois tipos de conhecimento, o tácito e o explícito. O conhecimento tácito é aquele individual que cada pessoa adquire, que envolve suas próprias experiências, suas habilidades e suas crenças. O conhecimento explícito é o conhecimento empacotado, ou seja, é o conhecimento modelado, pronto para ser repassado.

Alguns desafios enfrentados pelas organizações (seja ela com fins lucrativos ou não) são: capturar o conhecimento tácito (individual), armazenar, modelar e disponibilizar esse conhecimento (explícito), de maneira que esse conhecimento explícito possa trazer novos conhecimentos tácitos, e assim por diante, o que Nonaka & Takeuchi chamam de espiral do conhecimento, que é composta por 4 estados: Socialização, Externalização, Internalização e Combinação.

QUADRO 1 – Espiral do Conhecimento de Nonaka & Takeuchi



Fonte: Santos (2001)

Santos (2001) comenta sobre cada estado da espiral de Nonaka e Takeuchi, onde:

- a socialização é o compartilhamento do conhecimento tácito, através da imitação, observação ou prática (tácito para tácito);
- a externalização é a conversão do conhecimento tácito em explícito e a sua comunicação a outras pessoas (tácito para explícito);
- a combinação é a padronização do conhecimento; é juntá-la em um manual ou guia (explícito para explícito);
- internalização ocorre quando novos conhecimentos explícitos são compartilhados e as pessoas começam a internalizá-las de maneira a torná-los úteis em sua atividade (explícito para explícito);

A espiral do conhecimento de Nonaka e Takeuchi pode ser aplicada nas diversas áreas do saber, para a produção de novos conhecimentos, pois se o conhecimento fica escondido pode ser que ocorra um fenômeno muito comum, chamado de redescoberta, esse fenômeno pode ser encarado como um retrocesso para a ciência, pois pode haver tempo desperdiçado para a descoberta de algo que já fora descoberto antes.

Dessa forma a gestão do conhecimento (GC) não deve ser encarada apenas como um processo que leva à vantagem competitiva, pois a sua preocupação ultrapassa o principal valor apregoado pela vantagem competitiva que é o próprio lucro.

Como exemplo disso, pode-se citar as Instituições de Ensino Superior (IES), que segundo Maccari (2003) constituem uma organização que lida com o conhecimento, gera conhecimento e centra sua razão de existir, no fato de serem responsáveis pela disponibilização dos mesmos, porém, o desempenho das IES, parece não corresponder à natureza de seu negócio.

Segundo a pesquisa realizada por Maccari em 2003, nas cinco melhores IES do país (dados do guia do estudante de 2001), não existem procedimentos e atividade formais

que suportem diretamente a GC nas instituições pesquisadas, o que existe, é uma visão voltada para os aspectos ligados à infra-estrutura para a GC, e uma preocupação com a obediência a parâmetros estabelecidos pela CAPES.

Com a popularização das tecnologias e telecomunicações, a existência de inúmeras ferramentas que se propõem a apoiar os conceitos da gestão do conhecimento, e a conscientização da importância deste conceito em todos os âmbitos do saber, talvez as IES possam ultrapassar barreiras que limitam as possibilidades e alternativas de enriquecimento e amplificação das funções da GC.

5.1.1 Popularização das tecnologias e telecomunicação e sistemas de informação

A sociedade da informação está assistindo a um crescimento tecnológico jamais visto antes, computadores com processamento em altíssimas velocidades, grande capacidade de memória e armazenamento, aumento na velocidade e eficiência na telecomunicação, técnicas de reconhecimento de voz e muitos outros.

Todo esse crescimento tecnológico tem permitido a criação de inúmeras ferramentas que tratam as questões decorrentes da valorização da informação.

Existem diversas tecnologias e sistemas de informação que visam ajudar nos processos de armazenamento, disseminação, transmissão e recuperação das informações, porém a tecnologia é também uma das causadoras da alta produção de informação.

É perceptível a existência de uma grande quantidade de sistemas, com o propósito de gerenciar as informações, porém, muitos deles sem a preocupação para com a organização das informações, antes de serem armazenadas nas bases de dados, logo, há muita informação

armazenada sem nenhum critério de organização, e isso em longo prazo, traz o desgaste da não recuperação adequada.

Como os conceitos de sistemas de informação são grandemente utilizados nessa nova sociedade, é importante definir o que significa sistema de informação, e para essa definição, primeiramente será abordado o conceito de sistema.

A palavra sistema envolve um complexo de idéias. É possível se pensar em sistema, olhando para o sistema solar ou mesmo, para o sistema funcional do corpo humano.

Com vistas a estabelecer a relação de sistemas de informação com o conceito de sistema de modo geral, será aqui apresentada uma abordagem baseada nos princípios do pensamento complexo de Edgar Morin (1996) e da visão sistêmica de Bertalanffy (1997);

Todo sistema faz parte de um sistema maior, com o qual mantém relações, numa contribuição para seu funcionamento, assim como dele recebendo elementos para a execução de suas próprias funções. Pode-se denomina-lo como sendo sistema ou subsistema, mas o nome não deve prejudicar o entendimento do que ele realmente é, pois o que devemos considerar, para sua caracterização, são seus objetivos e suas funções. (Melo 2002, p. 15).

Num sistema de produção, por exemplo, os elementos físicos de entrada e os considerados como recurso são claramente distintos, onde matérias-primas são consumidas e equipamentos são utilizados.

Melo (2002) afirma que as informações de entrada num sistema de informação têm características muito peculiares, diante do que ocorre com as matérias-primas de entrada de um sistema de produção. No sistema de informação, ela não é consumida, pois tem uma natureza lógica; produz uma saída, mas continua disponível da mesma forma, como entrou no sistema.

Um sistema de informação pode ser todo e qualquer sistema que recebe dados, processa-os e devolve dados ou informações como saída. O grande objetivo dos sistemas de

informação é conseguir fazer com que a informação de saída desse sistema, seja satisfatória para o usuário ou para outros sistemas.

Para Stair (2002), um sistema de informação constitui um tipo especial de sistema, que pode ser definido como um conjunto de elementos ou componentes inter-relacionados que coletam (entrada), manipulam (processamento) e disseminam (saída) os dados ou a informação, e fornecem um mecanismo de feedback para atender a um objetivo.

Para Turban (2003), um sistema de informação é um sistema que coleta, processa, armazena e dissemina informações com um propósito específico, e como qualquer outro sistema, um sistema de informação abrange entrada (dados), e saídas (relatórios, cálculos), e processa essas entradas e geram saídas que são enviadas para usuários ou até mesmo para outros sistemas.

Um sistema de informação não necessariamente tem que ser baseado em computador, ele pode ser manual ou computadorizado, porém, normalmente, quando se fala em sistema de informação, esse assunto é abordado à luz dos conceitos tecnológicos.

Os sistemas de informações manuais podem ser gráficos, tabelas, curvas de tendências, e outros, e podem ajudar os analistas de investimentos, por exemplo, a tomar decisões e monitorar preços de produtos.

Segundo Stair (2002), alguns investidores ganham milhões de dólares usando os sistemas de informações manuais de análise de ações.

Os sistemas de informações baseados em computador (SIC), são compostos por hardware, software, banco de dados, telecomunicação e pessoas.

Segundo Turban (2003), os sistemas de informações baseados em computadores, usam computadores e tecnologia de comunicação para executar suas tarefas, e os seus principais componentes são:

- Hardware (monitores, placas, processadores, etc);

- Software (programas);
- Banco de dados (Conjunto organizado de arquivos que se relacionam);
- Rede (Um conjunto de maquinas conectadas que compartilham recursos);
- Procedimentos (estratégias, políticas, métodos e regras para utilizar o sistema de informação);
- Pessoas (componente mais importante dos sistemas de informação; inclui aqueles que trabalham com o sistema de informação ou usam sua saída);

O gerenciamento eficiente das informações depende da eficácia dos sistemas de informações que se propõe a fazer esse papel, de tecnologia e de pessoas que saibam como lidar com a questão.

5.2 Recuperação da informação

O termo *Information Retrieval* (Recuperação da informação) foi criado por Calvin Moores no ano de 1950, e trata dos aspectos intelectuais e da descrição das informações bem como de qualquer sistema ou equipamento utilizado para realizar a recuperação da informação.

A recuperação da informação consiste no resgate de documentos correspondentes, a partir de uma demanda feita por usuários, conforme expõe Ferneda (2003).

Para outros, recuperação da informação é definida como um sistema, onde um usuário define alguns termos de busca e a partir de então, alguns documentos são selecionados e devolvidos ao usuário.

A partir da década de 60, houve um grande avanço na recuperação da informação, onde Maron e Kuhns (1960) lançaram os modelos probabilísticos para a recuperação da informação. Posteriormente, através do projeto SMART foram produzidos inúmeras publicações científicas, e modelos de recuperação de informação.

Segundo Ferneda (2003), nos primeiros sistemas de recuperação da informação, a busca era feita através de contagem e frequência das palavras do texto, e na retirada das palavras que tinham pouca relevância. Acreditava-se que métodos puramente estatísticos seriam suficientes para resolver as questões que envolviam a recuperação da informação, porém, no decorrer do tempo e através das pesquisas, foi notado que métodos baseados em análise semântica seriam muito mais precisos.

Desde 1983, alguns autores começaram a apresentar idéias do uso de linguagem natural e lógica Fuzzy, que direcionaram as pesquisas em inteligência artificial (IA).

A inteligência artificial surgiu de maneira natural devida às pesquisas matemáticas, na tentativa de um aprofundamento para uso da semântica, no tratamento necessário à recuperação da informação, porém, os modelos estatísticos continuaram sendo aperfeiçoados e foram criados diversos modelos novos, porque a quantidade de informações estava aumentando cada dia mais, e tanto a Inteligência Artificial, como a Linguagem Natural, cresceu em uso.

As aplicações de técnicas da Inteligência Artificial (IA), na recuperação da informação, se dão através de pesquisadores ligados à ciência da computação, que verificam juntamente com a ciência da informação, as aplicabilidades da IA a campos fora da ciência da computação. Tais pesquisadores acabaram desenvolvendo alguns protótipos, considerando

resultados práticos, mas voltaram às pesquisas em sua área de origem, sem consolidar avanços significativos na recuperação da informação.

A essência da recuperação da informação é buscar documentos relevantes para uma determinada pesquisa que expresse a necessidade de informação de um usuário através de um sistema de recuperação. Inúmeros trabalhos têm sido feitos no sentido de se conseguir uma infra-estrutura que permita recuperar dados, com índice satisfatório de relevância.

Segundo Ferneda (2003), há quem faça confusão entre banco de dados e sistema de recuperação de informação. Assim vale a pena esclarecer que nos bancos de dados o que são recuperados, são dados que satisfazem a uma determinada expressão, já nos sistemas de recuperação de informação o que se busca recuperar são documentos sobre um determinado assunto.

Os sistemas de recuperação de informação tratam de objetos lingüísticos, ou seja, textos, e traz para seu arcabouço teórico, todos os problemas que o tratamento de linguagem natural possui.

Os bancos de dados têm uma estrutura semântica bem mais definida, e existe um esforço para se conseguir uma padronização dos sistemas de recuperação de informação, semelhante ao que ocorre nos bancos de dados. A tentativa de se padronizar um sistema de recuperação de informação ou pelo menos de se conseguir uma aproximação aos padrões que os bancos de dados possuem, têm sido possível, ao se instituir vocabulários controlados, thesauros, taxonomias, ontologias, etc.

5.2.1 Dificuldades de recuperação

Com o advento da internet, os sistemas de recuperação de informação têm ganhado especial atenção, pois a quantidade de documentos na rede correspondente a um mesmo assunto é assustadoramente grande, e a recuperação desses documentos de maneira precisa e rápida tem sido o desafio dos pesquisadores dessa disciplina interdisciplinar que é a recuperação da informação.

Burke (1999 apud Ferneda, 2003) diz que o ambiente digital que vem configurando as últimas décadas, como os acervos de objetos digitais, textos, imagens, sons, vídeos, páginas Web e diversos objetos digitais, requerem diferentes tipos de tratamento e representação para uma eficiente recuperação da informação.

Nesse sentido, a representação e a organização da informação tem muito a contribuir, pois a representação procura descrever ou identificar cada documento através de seu conteúdo, e isso pode ser feito através de um processo chamado de indexação.

Durante o processo de indexação, os conteúdos dos documentos são analisados, e depois, através de uma linguagem adequada, são traduzidos, e então podem ser representados, como por exemplo, por um thesauros, taxonomias ou ontologias, que são mecanismos que identificam um documento e seus pontos de acessos para posterior recuperação.

Segundo Ferneda (2003), a análise de um documento pode envolver uma interpretação do conteúdo desse documento com a intenção de agregar assuntos que não estão diretamente explicitados em sua expressão textual, mas que podem ser extraídos por um indexador humano.

Uma outra maneira de se indexar é através de métodos automáticos, onde sistemas baseados em computadores, com algoritmos precisos podem fazer o trabalho de indexação.

Esta é uma maneira prática, tendo em vista que os softwares e computadores hoje estão bastante acessíveis.

O grande problema da recuperação da informação consiste na expressão de busca que o usuário deve formular no momento da busca por um documento, pois a criada pelo usuário, deve coincidir com a palavra que foi utilizada na representação do documento, o que se torna extremamente complicado.

Um dos modelos criados para tentar minimizar os problemas anteriormente citados é o baseado em processamento de linguagem natural.

O processamento de linguagem natural surge como uma possível solução, porque leva em consideração objetos lingüísticos em uma tentativa de simular o processamento das expressões humana utilizadas na recuperação da informação. Porém, Faloutsos e Oard, citado por Fernald 2003, dizem que o desenvolvimento de sistemas de recuperação de informação que podem entender os documentos exige técnicas computacionais de alta complexidade, por isso, as técnicas de processamento de linguagem natural são muitas vezes usadas na melhoria de algumas tarefas, como por exemplo, indexação automática.

Apesar do processamento de linguagem natural poder ser utilizado para indexação de documentos, isto não elimina os outros métodos de indexação, ele simplesmente é um mecanismo complementar.

A dificuldade com a recuperação da informação não é algo novo, porém, na nova sociedade da informação se agravaram devido ao excesso de informação que circulam em todos os âmbitos da sociedade. Ao se fazer uma análise nos âmbitos empresarias, científico e na internet é possível perceber que a recuperação tem grandes desafios, mas também conta com inúmeras iniciativas que já estão sendo tomadas na tentativa de minimizar esses problemas.

5.2.1.1 Dificuldades na área empresarial e iniciativa.

Existem dois tipos de organizações ou corporações: as que visam lucros e as que não visam lucro, como é o caso dos grupos sociais, grupo religioso, universidades e outras. Nas que não visam lucro, este não representa sua meta primária (STAIR, 2002).

As corporações da nova economia que visam o lucro estão inseridas num mercado instável e competitivo, cercadas por ameaças e oportunidades e de acordo com Porter (1985 apud TURBAN 2003), existem cinco forças básicas que influenciam a competitividade das empresas:

- A rivalidade entre os concorrentes no mercado;
- A ameaça da entrada de novos concorrentes;
- O poder de negociação dos compradores;
- O poder de negociação dos fornecedores; e
- A ameaça da entrada de novos produtos ou serviços.

Na nova economia, a concorrência acirrada exige tomada de decisões rápidas e precisas, mas há tanta informação armazenada de maneira desorganizada dentro das empresas, que em vez de ajudar, muitas vezes acabam atrapalhando o desempenho das atividades corporativas. Por isso, as empresas estão tendo que investir pesadamente no gerenciamento das informações e do conhecimento que se encontram circulando dentro de seu ambiente organizacional, para se manterem no mercado.

Uma das iniciativas tomadas pelas empresas, no intuito de driblar a concorrência é o investimento em sistemas de informação para transformar dados em informações relevantes para a empresa.

Esses sistemas podem ser manuais ou computadorizados, porém precisam processar transações de maneira rápida e eficiente; armazenar grande quantidade de dados; comunicarem-se rapidamente com outros sistemas de informações (com pessoas e com outras máquinas); recuperarem informações de qualidade para fornecer suporte a tomadas de decisões.

O conceito de gestão do conhecimento deve permear os objetivos desses sistemas de informações, e tem o intuito de acumular, criar, armazenar e compartilhar conhecimento de maneira que esse conhecimento seja utilizado para criar novos conhecimentos em uma corporação.

O excesso de informação é um desafio comum tanto à gestão do conhecimento quanto à recuperação da informação. Há quem diga que o grande desafio para as corporações, é como gerenciar as informações existentes, e não propriamente como obter novas informações ou conhecimento.

A gestão do conhecimento pode contar com o apoio das tecnologias, e nesse aspecto, existem diversas ferramentas tecnológicas que visam ajudar a gestão da informação e dar apoio à implantação de sistemas de gestão do conhecimento, como exemplo, pode-se citar o GED (gerenciador eletrônico de documentos), Workflow (Fluxo de trabalho), Produtos de Groupware, Data warehouse (armazéns de dados), Data Mining (mineração de dados), Sistemas Inteligentes e muitos outros.

As ferramentas de gestão do conhecimento são tecnologias amplamente definidas, que permitem a captura, o gerenciamento, o compartilhamento, e a recuperação do conhecimento, porém, mesmo de posse de muita tecnologia, grandes podem ser os problemas gerados tanto pela resistência das pessoas envolvidas em um processo que envolve mudanças, quanto pelo excesso de informações desestruturadas e armazenadas nos sistemas de informações corporativos.

5.2.1.2 Dificuldades na área acadêmica e iniciativa.

No meio acadêmico, a sociedade da informação trouxe a valorização do conhecimento e do saber, o que tem causado uma conscientização sobre a educação jamais vista antes, e isso gerou uma forte tendência à ampliação do número de formação acadêmica, e conseqüentemente, o desenvolvimento e um crescimento científico, este que tem causado um aumento no número de publicações em proporções assustadoras.

O aumento das publicações ocorreu tanto em papel quanto em meios eletrônicos, porém, neste, a desestruturação é mais visível, havendo diversos formatos e padrões, o que acaba se tornando uma barreira à recuperação dessas informações, surgindo discussões a respeito de padronização de produções, e de tecnologias e de comunicação, para que os repositórios de dados possam se comunicar e facilitar a recuperação.

Costa (2005) diz que é possível acompanhar na literatura contemporânea, os acalorados debates sobre formatos, padrões e protocolos de comunicação no ambiente de pesquisa científica, e que o acesso aberto aos resultados de pesquisa causa efeitos nas universidades, nos editores e nas agências governamentais de fomento à pesquisa científica.

Existe muita informação acadêmica sendo produzida e as dificuldades de acesso afetam as comunidades científicas de todas as áreas. Na tentativa de amenizar esse problema e devido aos avanços tecnológicos e das telecomunicações iniciativas como “open archives”, “preprints” e “e-prints” têm sido tomadas.

O conceito “Open archives” é o conceito de arquivo aberto, ou seja, um determinado trabalho à disposição de todos.

Brody e Harnad (2004 apud COSTA 2005), chamam a atenção para o acesso livre (open access), afirmando, que esse tipo de acesso a resultados de pesquisas acelera um

aumento considerável no impacto das pesquisas, e com isso, aumenta a produtividade, o progresso e as suas recompensas.

Algumas comunidades científicas já aderiram ao conceito de Open archive e estão se estruturando. Costa (2005) relata o desejo de algumas comunidades acadêmicas tornarem disponíveis seus resultados de pesquisa, e cita como exemplo as ciências exatas e naturais que criaram a base de física, matemática, computação e biologia quantitativa chamada de arXiv: <http://arxiv.org>, as ciências da vida (PubMed central: <http://www.pubmedcentral.nih.gov>) e outras.

Através dessas iniciativas, o resultado de pesquisas tem se tornado disponível a todos os interessados, antes das pesquisas terem sido terminadas e publicadas.

Os “preprints” são modelos de divulgação de resultados de pesquisa antes de serem publicadas. Esses mecanismos que têm caráter acadêmico e não comercial, demonstram um anseio informacional e um interesse pelo conhecimento, que delineiam uma sociedade que a cada dia mais, valoriza o saber e o conhecimento.

Os e-prints, segundo Café et al (2002), embora sejam referidos na Web como pré-prints, na maioria das vezes são relatórios de pesquisas em versões eletrônicas submetidos para disseminação, comentários e revisão dos pares.

[...] Jordan (1999) amplia o conceito de *pré-prints*, contemplando também a dimensão referida pelos Arquivos Abertos. Para este autor, os *pré-prints* são manuscritos que não foram ainda publicados, mas que podem ter sido revistos e aceitos para publicação, submetidos, mas sem decisão de publicação, ou pretendidos para publicação e estando circulando para comentários. Os *pré-prints* disponíveis na Web podem ser também referidos como *e-prints*. A maioria dos *e-prints* são relatórios de pesquisas em versões eletrônicas submetidos para disseminação, comentários e revisão dos pares; enviados para publicação em periódicos; ou prévias para apresentação em conferências⁴. Semelhantemente, a American Physical Society define *eprint* como um *pré-print* (pré-published article) em forma eletrônica. O conceito é nebuloso, mas é, em geral, suficiente para incluir qualquer trabalho eletrônico, circulado pelo autor, fora do ambiente da publicação tradicional. Isto inclui qualquer informação eletrônica (não necessariamente impressa) fornecida pelo autor. (CAFÉ et al 2002, p. 6)

Diante de tal situação, a recuperação de informação tem um importante papel a cumprir, devendo demonstrar agilidade e eficiência, porque isso é fundamental para o progresso da ciência, portanto, a organização das informações dentro de ambientes acadêmicos também é de extrema importância.

5.2.1.3 Dificuldades na internet e iniciativa.

A Internet que surgiu devido ao aparecimento da tecnologia de sistemas abertos, trouxe mudanças radicais na interação homem-máquina. Os computadores deixaram de ser apenas um equipamento apto a executar tarefas dentro de um ambiente restrito e estático para ser uma janela capaz de se fazer chegar a qualquer parte do mundo em fração de segundos.

O número de computadores ligados à internet aumenta rapidamente, e a recuperação da informação permite que usuários acessem milhares de catálogos on-line, assim como milhares de bancos de dados. A internet funciona como um mecanismo de transporte e a Web (conhecida como World Wide Web ou WWW) é uma aplicação que usa essas funções de transporte.

A Web é um sistema com padrões aceitos mundialmente para armazenar, recuperar, formatar e exibir informações por meio de uma arquitetura baseada em cliente/servidor (arquitetura que prevê uma rede, onde o que armazena o dado é chamado de servidor, e o outro que pode utilizá-lo, é chamado de cliente).

A Web lida com todo tipo de informações digitais, como textos, hipermídia, gráficos e som. A tecnologia por trás da Web foi criada por Timothy Berners-Lee, em 1989,

onde ele propôs uma rede global de documentos hipertextos que permitisse que os físicos pesquisadores trabalhassem em conjunto.

Para Ferneda (2003), a Web é um conjunto de unidades de informação chamadas de páginas, com as seguintes características:

- Um endereçamento chamado de URL (Universal Resource Locator);
- Um protocolo, chamado de http (Hypertext Transfer Protocol), que permite a transferência de um documento (página), armazenada em um servidor ou host, através de sua URL. Após o pedido do usuário, a transferência é executada e o servidor envia a página ao computador do usuário;
- Um padrão para a especificação da estruturação da página Hypertext Transfer Markup Language (HTML) ou eXtensive Markup language(XML), uma linguagem que permite diferentes componentes em uma página Web;

Uma URL é o endereço de um arquivo que pode ser acessado através da internet, por exemplo:

Seja o endereço <http://www.tata.brturbo.com.br/documentos/index.htm>, onde:

[http://](http://www.tata.brturbo.com.br/documentos/index.htm) → protocolo

[www.tata.brturbo.com.br](http://www.tata.brturbo.com.br/documentos/index.htm) → computador host

[documentos](http://www.tata.brturbo.com.br/documentos/index.htm) → caminho dentro do host para encontrar o arquivo

[index.htm](http://www.tata.brturbo.com.br/documentos/index.htm) → nome do arquivo

A URL <http://www.teste.brturbo.com.br/documentos/index.htm> indica que um arquivo chamado de index, que tem a extensão htm(indica que o formato do arquivo é HTML) que pode ser acessado através do protocolo (<http://>), está armazenado no computador

chamado de `www.teste.brturbo`, cujo o domínio é `(.com.br)`, ou seja, a página encontra-se no Brasil dentro do diretório documentos.

A maioria das páginas da Web está escrita em linguagem HTML, a qual possui um conjunto pré-definido de comandos chamados tags (ou marcas), usados para definir componentes relacionados com a aparência e com a funcionalidade das páginas. Uma página HTML pode estar fazendo referência a outras páginas, ou seja, elas podem possuir links e formar uma teia propriamente dita, por isso (Web).

A Web (World Wide Web), nascida juntamente com a internet, é um repositório de informação que cresceu desestruturadamente, e acabou por cunhar o termo “sobrecarga da informação” termo originado do inglês “information overload”, que, na verdade, tem a intenção de traduzir a imensidão de informação passível de ser acessada via Web.

A sobrecarga e a falta de estruturação das informações têm dificultado a busca e a recuperação dessas informações. As ferramentas de busca na Web surgiram com o intuito de ajudar os usuários da informação, mas esses mecanismos têm realizado consultas em base de dados que são desestruturadas, trazendo a recuperação de dados irrelevantes e inadequados.

Um dos fatores que tem causado irrelevância é o fato da busca ser feita através de palavra-chave, ou seja, a análise feita na recuperação é puramente sintática, onde após a busca, é liberada uma lista contendo páginas com a palavra pesquisada, conseqüentemente o usuário gasta muito tempo até selecionar quais dados daquele total que foi recuperado são de fato, informações relevantes.

A internet tem sua base no protocolo HTTP (Hypertext Transfer Protocol) e na linguagem HTML (Hypertext Markup Language) que foram projetados apenas para representação e navegação na rede, mas não permitem adicionar semântica às páginas.

Freitas (2000) diz que as duas características da Internet que dificultam o acesso às informações úteis, específicas e relevantes são: o volume e a falta de definição semântica precisa.

A falta de mecanismos capazes de capturar a semântica do conteúdo das páginas da Web criou uma forte demanda por serviços que se ajustem à classe de serviços estudadas em Inteligência Artificial, e isso passou ser uma válvula de escape para os problemas relacionados à manipulação da informação na Internet.

Devido às dificuldades de recuperação acima mencionadas, algumas iniciativas têm sido tomadas no sentido de melhorar a recuperação das informações, como exemplo, dessas iniciativas pode-se citar a Web semântica, a utilização da XML (eXtensible Markup Language), DTD's (Document type Definition) e RDF (Resource Description Framework), bem como utilização das ontologias e taxonomias.

5.2.1.3.1 Web Semântica

Web Semântica é o nome do projeto do World Wide Web Consortium (W3C), idealizada por Tim Berners-Lee, o inventor dos principais protocolos da Web, para modificar a atual Web.

Segundo Decker et al (2000) & Berners-Lee et al (1999 apud Souza 2004), a Web Semântica pretende embutir inteligência e contexto nos códigos XML utilizados na confecção de páginas Web, para melhorar a interoperabilidade de programas e páginas e possibilitar um uso mais intuitivo pelos usuários.

A interoperabilidade das informações representa um dos desafios para se conseguir a integração de dados, especialmente no contexto Web, onde os recursos são distribuídos e possuem caráter heterogêneo.

A primeira iniciativa que foi tomada com relação à criação de padrões de suporte para diferentes níveis de interoperabilidade, como é o caso do HTTP e HTML, marcos importantes que permitiram a Web tornar-se uma das mais populares realizações computacionais das últimas décadas.

Berners Lee (2000) diz que a idéia da Web semântica é fazer por dados, o que o HTML fez pelos sistemas de informação textual, ou seja, prover flexibilidade suficiente para possibilitar a representação de todos os bancos de dados, criar regras lógicas para ligá-los, e, juntamente com isso, criar valor à informação.

Berners-Lee (2001) anela pela interoperabilidade de programas inteligentes chamados de agentes e a internet, onde possa haver troca de informação entre si, pois isso poderia trazer o benefício da automatização de tarefas dos usuários.

Daconta (2003) aborda o papel da Web semântica e diz que não é destinada somente para Web. Ela pode ser utilizada, por exemplo, em uma intranet corporativa, onde grandes serão os benefícios da mesma, onde poderá, por exemplo, ajudar a encontrar informações rapidamente para tomada de decisão e conseqüentemente possibilitar vantagem competitiva.

Atualmente as páginas da Web são construídas em HTML (Hiptertext Markup Language), derivada do padrão SGML (Standard Generalized Markup Language), que é um padrão que descreve outras linguagens.

A base do SGML é que todos os documentos têm uma estrutura e elementos semânticos que podem ser descritos sem se levar em conta a maneira como serão exibidos.

Todas as linguagens derivadas do SGML podem usufruir de um conjunto de marcas sintáticas (tags) que descrevem dados e comandos de manipulação de documentos, chamada DTD ou Document Type Definition.

O HTML é, na verdade, um conjunto de tags já definidos, que não podem ser alterados ou modificados sem uma nova redefinição do DTD da linguagem e também dos navegadores, para que a linguagem possa ser interpretada, porque cada navegador tem seu interpretador e ele conhece a priori o conjunto de tags do HTML. Pode-se dizer que HTML é um DTD específico do SGML.

Um Navegador ou browser interpreta as marcas (tags) que o documento possui e decide como os dados ali existentes serão exibidos na tela.

Apesar da maioria das páginas da Web terem sido criadas em HTML, essa linguagem tem sido considerada limitada porque não consegue descrever o conteúdo semântico. A consequência dessa limitação foi o surgimento da XML (eXtensible Markup Language), que é uma linguagem recomendada pela W3C (World Wide Web Consortium), semelhante ao HTML, pois ambas surgiram a partir do SGML, logo contêm marcas (tags) que descrevem os documentos, mas enquanto o HTML se preocupa em como os dados serão mostrados ou exibidos, o XML concentra esforços na descrição dos dados em um documento.

A Web semântica possui alguns formalismos que podem ser utilizados para descrever semântica, como exemplo, XML (Extensible Markup Language), RDF (Resource Description Framework), DTD (Document type Definition), Ontologias e Taxonomia.

Grandes são os benefícios, mas grandes também são os desafios da Web semântica, e o primeiro passo para uma recuperação de informação precisa, é a organização das informações, para que os dados ao estarem estruturados possam ter significado.

Para que a Web semântica seja implementada, é necessária tanto padronização de tecnologias e de linguagem quanto uma definição de padrões de dados e ao mesmo tempo, é

preciso definir regras sobre os mesmos, para que novos dados e regras sejam criados automaticamente. Também é preciso que os usuários da internet respeitem essas regras no momento de descrever as informações armazenadas.

Após essas medidas serem tomadas, será possível que os search engines (motores de busca), que hoje buscam páginas HTML através de palavra-chave, ou qualquer outro mecanismo de recuperação, possam montar seus índices a partir de informações confiáveis, baseados em semântica.

A associação do conteúdo de um documento com a sua descrição pode ser feita de maneira manual ou automática.

A seguir serão apresentadas algumas ferramentas que a Web semântica pode utilizar para fazer cumprir seus objetivos, são elas: XML, DTD, RDF, e posteriormente dentro do contexto de organização e representação, serão abordados os conceitos de Ontologias e Taxonomias.

5.2.1.3.2 XML - eXtensible Markup Language

XML ou eXtensible Markup Language é, para alguns autores, uma linguagem considerada muito importante tanto para a internet quanto para grandes intranets, pois prevê uma capacidade de interoperabilidade entre computadores e possui um padrão flexível, aberto e independente de dispositivos.

Harold (2004) define XML como sendo um conjunto de regras para definir marcas semânticas que quebram um documento em partes e identificam essas diferentes partes.

Daconta (2004) diz que XML não é uma linguagem, mas um conjunto de regras sintáticas para criar linguagem de marcação rica em semântica, ou seja, é possível aplicar as regras do XML para criar novas linguagens, chamadas de aplicações XML.

Independente de ser uma linguagem ou não, com o uso da XML, as aplicações podem ser construídas e modificadas com muita rapidez, além disso, ela permite que sejam utilizadas inúmeras formas de visualização dos dados e se preocupa com a descrição dos mesmos.

A XML como já foi citado anteriormente, descreve o significado e a estrutura de um documento e não o formato dos elementos de uma página, podendo-se adicionar formatos nessas páginas. Ela especifica as regras para uma sintaxe de baixo nível, dizendo como as marcas são distinguidas através do contexto.

Na XML os atributos são acoplados aos elementos e assim por diante, sem dizer como essas tags, elementos e atributos são ou qual o seu significado. Ela cria um padrão de elementos que devem ser mostrados sem especificar os nomes dos elementos. Por exemplo, XML diz que tags começam com < e termina com >, entretanto, não diz quais nomes devem ser colocados entre < e > .

Os elementos de um documento XML são especificados através do uso de tags e os tipos de tags podem ser: tags de abertura: <tag>, tag de fechamento: </tag> e tag com auto-fechamento: <tag/>.

Os valores dos dados são especificados entre as tags de abertura e fechamento:

```
<tag> dado </tag>
```

- As tags podem possuir atributo:

```
<tag atrib="valor"/>
```

- As tags podem conter outras tags aninhadas:

```
<tag1> <tag2> dado </tag2> </tag1>
```

Exemplo de um documento XML

```
<?xml version = "1.0"?>
  <clientes>
    <clientes id="1">
      <nome> Joao da Silva</nome>
      <cpf> 657.678.789-00</cpf>
    </cliente>
    <cliente id ="2">
      <nome> Jose dos Remédios</nome>
      <cpf>897.865.476-00</cpf>
    </cliente>
  </clientes>
```

Ao entender essa meta-sintaxe, uma aplicação entende parcialmente todas as linguagens construídas com esta meta-sintaxe. Um browser (navegador de páginas da internet) não precisa conhecer profundamente todas as tags que podem ser usadas por milhares de linguagens de marcação diferentes.

Na linguagem XML não é necessário ficar esperando que um browser se atualize, é possível inventar. Ele descreve estrutura e semântica, mas não o formato, ou seja, as marcas descrevem a estrutura e o significado dos documentos, mas não descrevem o formato dos elementos na página.

Com o uso do XML não é necessário ficar amarrando o seu tipo de dado a listas de tags que não correspondem ao contexto correto. Essa abertura no código traz como

benefício direto o fato de se poder descrever dados com maior significância, e abre caminho para que a tão sonhada semântica seja anexada aos documentos.

Existem duas maneiras para se exibir os dados de um documento XML, utilizando DOM (Document Objeto Model) ou XSL (eXtensible Stylesheet Language).

1) DOM (Document Objeto Model) é uma Interface de programação para documentos HTML e XML, que define a maneira como o documento pode ser acessado e manipulado.

2) XSL (eXtensible Stylesheet Language) é a folha de estilo usada para definir a aparência dos documentos XML, ou seja, é uma linguagem utilizada para transformar XML em HTML. Pode-se dizer que o XSL define como o XML será exibido e reconhecido pelo navegador do usuário.

Diferentes folhas de estilos podem ser aplicadas a um mesmo documento XML, apresentando um dado em diferentes formas, dependendo de cada situação, diferentes formas para diferentes usuários.

5.2.1.3.3 DTD - Document Type Definition

Um DTD ou Document Type Definition é um formato não XML, que pode ser embutido em um documento XML, ou especificado em um arquivo em separado com uma extensão (.dtd). Seu propósito é a construção de blocos válidos para um documento XML, ou seja, definir quais os tipos de elementos que podem existir em um documento, quais atributos esses elementos podem ter, e como as instâncias desses elementos estão hierarquicamente

relacionadas, ou seja, um DTD tem a pretensão de definir o formato e a sintaxe de cada marca que for criado no arquivo XML.

No DTD apenas a estrutura lógica de um documento é descrita, não sendo fornecida informações alguma sobre a semântica da apresentação do documento.

O DTD permite descrever cada tag (marca) e fornecer regras para interpretar cada informação usada em um arquivo XML, e pode ser utilizado, por exemplo, por programas que precisem fazer validações na estrutura do documento, como é o caso de programas que precisam fazer inserções e atualizações de informações em uma base dados.

5.2.1.3.4 Resource Description Framework

RDF ou Resource Description Framework pode ser definido como uma linguagem para representação de informações sobre recursos.

Baptista (2001) define RDF como sendo um modelo para codificar, fazer intercâmbio e reutilizar metadados normalizados. A principal característica do RDF é o suporte para representação uniforme de dados e metadados.

RDF é um modelo recomendado pela W3C, que é desenhado para que se possa, por exemplo, em um site, descobrir os recursos, catalogar o conteúdo do site, escalonar esse conteúdo, etc. Nesse caso, o recurso pode ser uma página HTML, um documento Word ou uma apresentação PowerPoint, ou seja, um recurso é qualquer coisa que possa ser representado por uma URL (Uniform Resource Locator).

Klein (2001) define RDF como sendo um modelo para se representar metadados. O modelo de representação de dados do RDF é uma tripla composta por: Objeto (o próprio

recurso), atributo (propriedades desse objeto) e valor (valor atribuído às propriedades), mas essa tripla, por si só, não consegue atribuir semântica aos dados.

O RDF se comporta da mesma maneira como se comporta o XML, ou seja, não pode sozinho atribuir semântica aos dados, e também reúne algumas características necessárias para prover suporte na interoperabilidade das informações para a Web, porém, não prevê cardinalidade, unicidade, ambigüidade e mecanismos de inferência, características estas que já se acham presentes nas ontologias. O uso de RDF é mais fácil do que as linguagens ontológicas como, por exemplo, DAML, OIL ou OWL, porém, as ontologias são mais avançadas e abrangentes do que o RDF.

5.3 Organização das informações

A diversidade de formatos que circulam nos diversos meios e canais de comunicação tomou conta do cotidiano informacional tanto pessoal quanto empresarial, e muitas informações não são encontradas e/ou recuperadas, quando desejadas.

Os avanços nas telecomunicações têm ajudado grandemente a explosão de informações, pois hoje, é possível haver comunicação com o mundo todo, praticamente em tempo real. Quando se fala em comunicação, não se pode restringir à telefonia para comunicação mútua, pois os serviços de comunicação disponíveis na Internet revolucionaram as comunicações tanto pessoais quanto empresariais, podendo-se ter acesso a uma imensa variedade de dados nos mais diversos formatos em tempo recorde.

Por isso, áreas especializadas vivem momentos de concentração de esforços no sentido de produzir conjuntos de termos comuns para suas áreas, bem como determinar o significado de cada termo, para que possam ser utilizados entre todos os interessados.

Diversos estudiosos têm se preocupado com a organização e representação das informações, e isso tem feito com que os estudos saiam do escopo da documentação, pois, o crescimento assustador do número de informações veiculadas na Internet em diferentes formatos, como documentos, som ou imagem, reforça a importância da organização da informação e do conhecimento, no aspecto da recuperação.

O caráter interdisciplinar do conhecimento faz com que fique difícil uma definição para organização do conhecimento, pois para a Ciência da Computação, a organização e estruturação do conhecimento, tanto na forma implícita como explícita é primordial para que sistemas de informações computacionais possam ser desenvolvidos para atender às necessidades dos usuários.

Para a Ciência da Informação, a organização do conhecimento está atrelada ao desenvolvimento de métodos de classificação que visam facilitar o armazenamento e a recuperação da informação, independentemente desses sistemas serem informatizados ou não.

A intersecção dessas duas áreas com certeza pode trazer importantes contribuições para a recuperação, pois a partir de informações bem estruturadas e ferramentas computacionais eficientes, será possível obter uma recuperação de informação mais eficiente.

Organização do conhecimento é uma atividade cultural interdisciplinar que adiciona valor informacional para coleções contendo conhecimento. Ele fixa pontos de acesso comuns para itens, de maneira que grupos de usuários de sistemas de informações associados sejam melhores servidos (SIGEL, 2000, tradução nossa).

Para Barite (2000 apud Yamaoka 2005) há uma diferenciação entre organização e representação do conhecimento.

Num conceito amplo, a organização do conhecimento pode ser considerada uma disciplina que estuda as leis, os princípios e os procedimentos pelos quais se estruturam o conhecimento especializado em qualquer disciplina, recebendo aportes da informática, da lingüística, da terminologia e da ciência da informação[...] (BARITE, 2000 apud YAMAOKA, 2005, p. 42)

A representação do conhecimento como um ramo da organização do conhecimento que se preocupa com a simbolização conceitual do saber humano, abrangendo a classificação à indexação e o conjunto de aspectos informáticos e lingüísticos. (BARITE, 2000 apud YAMAOKA, 2005, p. 42).

A representação da informação pode ser utilizada, por exemplo, na descrição de um livro ou documentos, onde algumas palavras são utilizadas para representar o conteúdo ou assunto do mesmo, e identificam esse objeto (livro), no sentido de facilitar, quando necessário, a recuperação dessa informação. Essas palavras são representações do conhecimento e seguem alguma corrente teórico–metodológica para serem organizadas de maneira hierárquica. Posteriormente, essa representação pode ser utilizada, tanto para a recuperação da informação quanto para organização dos livros em prateleiras.

5.3.1 Teoria Geral da Terminologia

A Terminologia foi desenvolvida a partir dos estudos do engenheiro austríaco, Eugen Wüster. Inicialmente, o objetivo desse estudo era organizar os termos da Eletrotécnica, e assim, garantir uma comunicação mais precisa. O estudo se expandiu, resultando em uma proposta que se denominada Teoria Geral da Terminologia, que trata dos conceitos de uma língua técnica e seus relacionamentos.

Terminologia é a matéria que se ocupa da designação dos conceitos das línguas de especialidade, esta que segundo Felber (1984 apud Moreira & Oliveira, 2005) pode ser entendida como a totalidade de todos os significados lingüísticos usados em um determinado campo de uma profissão específica.

A contribuição da Terminologia se faz presente no momento em que se pretende entender o significado de um termo e relacioná-lo com outros, dentro de um domínio

específico, permitindo a formulação de relações mais coerentes entre os termos selecionados para compor instrumentos que deseja desenvolver. (MOREIRA; OLIVEIRA; 2005).

A Terminologia pode ser considerada como interdisciplinar, pois é importante em todos os campos da ciência e intimamente ligada a disciplinas como a lingüística, lógica, documentação, classificação, com ontologia e outras, tendo em comum os elementos necessários à organização entre termos e conceitos. A sua não utilização, ou em função de seu uso indevido, os termos de uma determinada especialidade podem perder o significado, tornando-se impossível a comunicação e a troca de informação, podendo assim, haver um prejuízo na evolução da ciência.

Uma das ciências que pode ser grandemente amparada pela Terminologia é a Ciência da Computação, esta que talvez por ser uma ciência nova, não possui ainda, uma padronização terminológica bem definida, e a compreensão da terminologia técnica de uma área é de extrema importância para que haja uma perfeita comunicação entre os especialistas da área.

Para a Terminologia é fundamental a definição do que são “termos” e “conceitos”, por isso, a seguir serão abordadas suas definições.

“Termo” é segundo Sager (1990) a representação lingüística de um conceito, e para a ISO 1087(2000), a designação verbal de um conceito geral em um domínio específico.

“Conceito” segundo Dahlberg (1978, p 102) é a reunião e compilação de enunciados verdadeiros a respeito de determinado objeto. Para o padrão (ISO 1087-1:2000) conceito é uma unidade de conhecimento criada por uma combinação única de características, o que torna o conceito, algo que pode ser capturado.

Para Suonuuti (2001 apud MOREIRA & OLIVEIRA, 2005) o conceito está sempre relacionado a outro conceito, formando assim, um sistema de conceitos que pode variar de muito simples a muito complexo.

Segundo Moreira e Oliveira (2005), o termo estabelece com o conceito uma relação unívoca, e esta relação de unicidade é possível porque o escopo do trabalho da Terminologia abrange apenas línguas artificiais, no qual esta correspondência é garantida pelas normas.

Wüster propôs uma classificação para as relações entre conceitos, diferente da adotada pela ISSO 1087(2000). As duas abordagens oferecem formas diferentes de se estudar a natureza das relações entre os conceitos e ressaltam aspectos distintos, que devem ser considerados, porém, essas diferenças não serão objetos deste estudo neste trabalho.

Os relacionamentos e as definições são de grande importância para a Terminologia e segundo Sager (1990, p.51), o relacionamento indica qual o tipo de ligação que um conceito tem com outro conceito.

A definição é um processo necessário para colocar o termo em sua posição, numa estrutura de conhecimento, e é obtido através de definições pré-existentes, contexto ou consulta a especialistas (SAGER, 1990, p. 51).

[...] a definição pode conseqüentemente se concentrar nas características essenciais que um conceito tem em comum com outros e quais diferenças ele tem com outros conceitos (SAGER, 1990, p.51, tradução nossa).

Como a definição explicita as características de um conceito, ela acaba diferenciando um conceito de outro, podendo assim, identificar equivalência de termos.

A organização da informação calcadas nos alicerces da Terminologia sistematiza conceitos de uma área do conhecimento na perspectiva de se poder representar o conhecimento na tentativa de melhorar a sua recuperação, facilitando assim, a comunicação e a transmissão do mesmo.

5.3.2 Teoria do Conceito

A Teoria do Conceito foi criada por Dahlberg em 1970. Seus princípios dão suporte para a determinação do conceito e suas relações e por isso são utilizados para a elaboração de trabalhos terminológicos como classificação, thesauros e outros.

A teoria do conceito pretende servir como base para análises conceituais de todos os trabalhos terminológicos (ou de representação do conhecimento) e, portanto, fornece o ferramental apropriado para atuar nesse nível. Por meio da Teoria do conceito, pode-se classificar os conceitos e as relações entre eles, obtendo, dessa forma, uma estrutura conceitual (DAHLBERG, 1978, p.101).

Segundo Dahlberg (1978, p. 102), cada enunciado (ou característica) verdadeiro a respeito de um objeto representa um elemento do conceito e o conjunto dos enunciados verdadeiros a sobre aquele objeto é chamado de conceito e pode ser caracterizado como geral ou específico.

Para cada enunciado a respeito de um objeto pode ser realizada uma nova avaliação, gerando assim, novos conjuntos de enunciados verdadeiros, que são chamados de conceitos específicos. Ao se obter um enunciado ou característica bem geral a respeito de um objeto obtém-se o que na teoria do conceito é chamada de categoria.

Segundo Dahlberg (1979) toda vez que conceitos diferentes possuem características semelhantes, existe relacionamento entre esses conceitos, podendo-se estabelecer comparações entre conceitos de maneira a organizá-los. As relações entre conceitos podem ser do tipo:

- hierárquico ou gênero/espécie
- partição (todo e suas partes)
- oposição
- funcional (entre categorias)

- Relação hierárquica: ocorre quando entre dois conceitos, um possui as mesmas características do outro, e mais alguma característica. Um outro tipo de relação hierárquica é a relação de Coordenação, que ocorre entre conceitos específicos do mesmo gênero.

Objeto - árvores

Tipos de objetos - árvores frutíferas /árvores de nozes

Tipos de tipos de objetos - macieiras, pereiras / larajeiras

- Relação Partitiva: ocorre entre o todo e suas partes, ou de um produto e dos elementos que a constituem. Por exemplo:

Todo - árvore

Partes - tronco, galho, folhas, flores, frutos.

- Relação de oposição: Ocorre quando um conceito nega outro conceito, podendo ser das seguintes espécies: Contradição ou Contrariedade.

Contradição: falso - verdadeiro

Contrariedade : branco - preto

- Relação Funcional: Ocorre entre conceitos que denotam ações ou processos e seus complementos.

Doação - presente - casamento

Datilografia - trabalho - conferência

Essas relações entre conceitos podem ocorrer entre os tipos de conceitos: objetos, fenômenos, processos propriedades, relações e entre dimensões.

“Definição” é um item importante dentro da Teoria do Conceito. Segundo Felber (1984 apud Yamaoka, 2005), definição é uma descrição de um conceito por meio de outros conceitos conhecidos, onde na maioria das vezes, será feito em forma de palavras ou termos.

O autor também afirma que a definição é a responsável pela localização de conceito dentro de um sistema de conceitos.

Segundo Campos (2004, p. 30) a área da ciência da informação, apesar de poder contar com teorias bem fundamentadas sobre o conceito e relações conceituais, é fraca em modelos para a elaboração de representações gráficas desses conceitos.

5.3.3 Estudo do comportamento do usuário

Toda informação ou conhecimento armazenado um dia precisa ser recuperado de alguma maneira, caso contrário não teria razão de seu armazenamento, mas o centro das preocupações de quem está diretamente ligado à área de recuperação, é saber, quem serão os usuários das informações armazenadas? ou qual o comportamento de busca desses usuários?

Essas questões têm permeado as discussões sobre recuperação de informação e, atualmente, têm ganhado importância das comunidades científicas que estudam as razões da não recuperação adequada das informações.

A Ciência da Informação tem a preocupação com o usuário da informação, e estuda o comportamento de busca dos usuários, pois acredita que os sistemas de recuperação de informação podem melhorar sua atuação, se forem adequados aos diferentes tipos de usuários de informação.

Segundo Lopes (2002), a evolução histórica da recuperação da informação está embasada nas linhas de desenvolvimento de Lancaster, e afirma que a primeira base de dados americana (National Library of Medicine) indexava suas bases de dados através do uso de tesouros específicos em suas áreas temáticas.

A segunda linha teve seu desenvolvimento no campo do Direito e envolvia a geração de textos completos de leis. Tanto uma como a outra, exige um planejamento acurado da estratégia de busca, visando a uma recuperação de informação de acordo com as necessidades dos usuários.

O conceito de estudo de usuário foi abordado por Ferreira (1995) onde destacou os novos paradigmas e novos usuários de informação. A autora afirma que o usuário de informação tem sido colocado na constrangedora situação de ter que se adaptar aos mecanismos de recuperação, ao invés desses mecanismos se amoldarem às suas características.

Algumas medidas foram tomadas no sentido de minimizar os problemas de interação dos usuários com os sistemas, ou como a literatura aborda, interação “homem-máquina”, e como exemplo dessas medidas, pode-se citar o surgimento de programas com interfaces amigáveis, porém, isso não garante a eficácia da recuperação e não resolve os problemas relacionados às necessidades de informação do usuário.

O sistema pode ter uma boa interface, ter os dados cadastrados, e o usuário por não conhecer a linguagem que deve ser utilizada para a recuperação, não conseguir recuperar o que pretende.

Para que um sistema de informação seja de fato eficiente no quesito recuperação, deve-se fazer um planejamento e criar estratégia de busca, e para tal, é necessário que haja a identificação apropriada dos elementos que descrevem os temas da base de dados desse sistema e levar em conta o usuário e a maneira como ele irá interagir com o sistema. Tudo isso faz parte de uma linha de pesquisa que Garcia (2005) chama de Recuperação interativa da informação.

Segundo Costa (2001 apud Gasque & Costa 2003), várias áreas do conhecimento têm contribuído para estudos do comportamento informacional – information behavior - o que

no Brasil é chamado de estudo de usuários. Ambas as abordagens incluem os estudos de necessidades e uso de informações, com a preocupação de identificar e discutir padrões de comportamento informacional nos diferentes campos do saber.

Uma das razões para se conhecer os hábitos e o comportamento dos usuários da informação deve-se ao fato de se poder construir estratégias de busca adequadas, pois inúmeras são as vezes em que uma informação não é devidamente recuperada, pela falta de estratégia de busca específica para cada tipo de base de dados.

Uma estratégia de busca é uma regra previamente criada dentro dos sistemas de armazenamento de informação que visa permitir a recuperação devida das informações.

Lopes (2002) define estratégia de busca como sendo uma técnica ou conjunto de regras para tornar possível o encontro entre uma pergunta formulada e a informação armazenada em uma base de dados

Spink & Saracevic (1993 apud Lopes 2002, p 60-71) afirma que uma das chaves para a estratégia de busca é a seleção de termos, e sintetiza a questão, fazendo a seguinte pergunta: “Que termo de busca deve ser selecionado para um determinado tema, que represente efetivamente o problema de informação do usuário?”

Sistemas de recuperação de informação que contam com estratégias de busca construídas com base no comportamento do usuário têm mais chances de resultarem em sucesso, pois a recuperação eficiente não depende somente da existência da informação dentro de uma base de dados, ela depende, e muito, de fatores ligados ao próprio usuário, como por exemplo, da compreensão da linguagem ou código utilizado para a organização dessas informações.

Estudos revelam que os usuários de bibliotecas acadêmicas têm maior dificuldade de se aprofundarem em buscas específicas do que os usuários de bibliotecas especializadas,

onde o grau de conhecimento sobre o assunto é maior, e tudo isso influencia no planejamento dos sistemas de recuperação.

O Sucesso do usuário final na recuperação e no acesso à informação depende de vários fatores, incluindo quão bem ele entende a linguagem, a estrutura dos dados na base, e quando ele entende, quais dos dados disponíveis satisfazem sua necessidade de informação. (FIGUEIREDO 1999, apud GARCIA; SILVA 2005).

Ferreira (1997) afirma que qualquer tentativa de padronização de busca por informação deve admitir o indivíduo como centro do fenômeno, e considerar a visão, as necessidades, as opiniões e seus problemas, como elementos significantes e influentes que merecem investigação, quer sejam, para o desenvolvimento de produtos e serviços em ambientes eletrônicos ou não.

Fidel (1986 apud Lopes 2002) relata um estudo sobre sistemas especialistas, demonstrando que a maioria dos sistemas é baseada em análise de texto e não em modelos mentais de busca humana, e por isso não podem processar os critérios relacionados com o pedido de busca e responder questões relativas aos fatores de precisão⁴ e revocação⁵.

Apesar das tecnologias oferecerem oportunidades para um melhor gerenciamento das informações, e permitirem que o usuário seja mais bem atendido, para que isso aconteça é preciso fazer o que Ferreira (1997) sugere, ou seja, que é fazer um planejamento de toda e qualquer atividade de informação de acordo com pesquisas centradas no indivíduo, partindo-se de uma perspectiva cognitiva, e buscando interpretar necessidades de informações tanto intelectuais como sociológicas.

4 Segundo Lancaster, uma alta precisão ocorre quando em uma busca a maioria dos itens recuperados, se não todos, forem considerados pertinentes.

5 Segundo Lancaster, uma alta revocação será aquela em que a maioria dos itens relevantes (pertinentes), se não todos, forem recuperados.

Com a compreensão o comportamento de busca de informação num aspecto mais profundo, torna-se possível capacitar organizações provedoras de informações para servirem melhor às necessidades de seus usuários, o que aumentaria a eficiência específica dos indivíduos em seus aspectos pessoais, sociais e profissionais.

5.3.4 Estudo das necessidades de informação

Partindo do pressuposto de que o ser humano raramente busca informação sem uma razão específica, a perspectiva centrada no usuário e nas suas necessidades de informação ganha importância e se torna base para os sistemas de informações modelados de acordo com o usuário, com a natureza de suas necessidades informacionais, e com seus padrões de comportamento de busca.

Cardoso (2004 apud Garcia, 2005) afirma que expressar uma necessidade de informação é uma tarefa difícil, pois existe uma distância semântica entre a necessidade dos usuários e o que ele expressa na hora de formular uma consulta. Em síntese: é difícil traduzir a real necessidade de informação do usuário, para uma linguagem que o sistema entenda.

Chen e Hernon (1982 apud Ferreira, 1995) dizem que o processo de buscar compreensão do que seja a necessidade de informação deve ser analisado sob a perspectiva da individualidade do sujeito a ser pesquisado.

Esse ponto de vista demonstra que os sistemas de informações podem ser melhores projetados, se procurarem atender às necessidades de informação de seus usuários respeitando sua individualidade. No entanto, como abarcar em um só sistema, características de todos os usuários daquele sistema?

Como uma tentativa de resposta à pergunta acima mencionada, será abordado o conceito de necessidade de informação de Dervin e Nilan (1986) citado por Ferreira (1995), onde afirmam que embora as pessoas tenham suas próprias experiências, subjetivas e únicas, existe também grande similaridade entre situações encontradas por diferentes indivíduos, portanto, necessidade de informação não é conceito subjetivo e relativo somente na mente de um indivíduo.

Pelo contrário, a necessidade de informação representa um conceito intersubjetivo com significados valores, objetivos, etc, e são passíveis de serem compartilhados. Com isso, é possível que seja feita uma generalização de padrões de comportamento de busca e uso de informação, através da ótica do usuário.

O fato da necessidade de informação mudar no tempo e depender do indivíduo que a busca, exige que os sistemas de recuperação informação sejam flexíveis o suficiente para permitirem ao usuário adaptar o processo de busca de informação à sua necessidade. Diante disso, a construção de bases especializadas de dados tem se tornado muito comum.

O estudo do comportamento e das necessidades de informação dos usuários de bases especializadas, com certeza, tem muito a contribuir para com uma melhor organização das informações, pois esse estudo, aliado às diversas tecnologias e ferramentas que apóiam a organização das informações pode trazer uma melhora significativa nos sistemas de recuperação das informações.

A níveis globais, grandes são os desafios da recuperação da informação, e enquanto não se encontra uma solução, o uso de mecanismo para a organização das informações embasadas em comportamento e necessidade dos usuários pode ser utilizado para uma melhora na recuperação das informações, em bases onde há um certo padrão de comportamento de busca.

5.3.5 Mecanismos de organização da informação

Embasados nos conceitos da classificação e da terminologia, surgiram inúmeros sistemas para organização das informações e do conhecimento. Esses sistemas incluem uma variedade de esquemas, que são necessários em todas as áreas do conhecimento humano. Como exemplo, pode-se citar metadados, ontologias, taxonomias e outros.

5.3.5.1 Metadados

O termo metadados surgiu por volta da década de 60, e teve seu uso atrelado aos conceitos de bancos de dados, mas a partir da década de 80, começou a ser utilizado para identificar as informações. Segundo Gilliland-Swetland (2000), a partir da década de 90, o termo metadados passou a ser utilizado pelas comunidades envolvidas com o gerenciamento e a interoperabilidade de dados geoespaciais, e com a administração de projetos de sistemas e manutenção em geral.

Na literatura são encontradas várias definições para metadados, porém, a mais comum diz que são dados sobre os dados, ou seja, é uma maneira de descrever dados, informações ou recursos, onde recurso é qualquer coisa que tenha identificação, como por exemplo: documentos eletrônicos, imagem, serviço ou uma coleção de recursos. Outros autores se referem a metadados como descrição de objetos ou objetos digitais

Para Hilmann (2001), metadados é um conjunto de atributos ou elementos necessários para descrever um recurso em questão. Para Hodge (2001), é a informação estruturada que descreve, explica, localiza ou marca uma informação para que se possa executar uma fácil recuperação.

Na arquitetura de metadado para Web, definida por Berners-Lee (1997), é informação sobre recursos da Web que a máquina pode interpretar sem a interferência humana, e recurso é qualquer coisa, como por exemplo, uma página na Web, um documento ou um arquivo.

Com relação aos tipos de metadados, Hodge (2001) diz que existem os tipos descritivos, administrativos e estruturais desenvolvidos em função dos objetivos que se pretende alcançar.

- Descritivos: descrevem um recurso para poder identificá-lo posteriormente.
- Administrativos: provêm informações para ajudar gerenciar um recurso, como por exemplo, quando e como foi criado o recurso? Quem pode acessá-lo?
- Estruturais: indicam como são compostos os objetos, por exemplo, como são organizadas as páginas de um capítulo.

Para Gilliland-Swetland (2000) os metadados são divididos nos seguintes tipos:

- Administrativo: usado na gestão da informação ou recursos de informação;
- Descritivo: descreve ou identifica informações ou recursos;
- Preservação: gerencia a preservação dos recursos de informação;
- Técnico: gerencia o comportamento dos metadados;
- Uso: administra o nível e tipo de uso de um recurso de informação;

Algumas comunidades que possuem os mesmos interesses têm buscado padronizar alguns vocabulários comuns àquela comunidade, o que se pode chamar de padrão de metadados.

Existem alguns padrões de metadados para finalidades distintas, como por exemplo: DC (The Dublin Core), CSDGM (Content Standard for digital Geospatial Metadata), MARC (Machine Readable Cataloguing), LOM (Learning Objects Metadata), CWM (Common Warehouse Meta Model), e outros. A escolha entre um ou outro, deve ser embasada nos objetivos que se deseja alcançar.

Sendo metadados, dados sobre os dados, uma importante característica é a localização desses metadados, ou seja, onde eles ficam armazenados.

Quanto ao local onde os metadados serão armazenados, Borbinha (2002) diz que pode ser dividido em três categorias: embebidos, associados e os separados.

a) **Metadados Embebidos**, podem ser embebidos ou embutidos nos recursos, como o que ocorre nos formato HTML, onde normalmente são embutidos nos cabeçalho dos documentos.

b) **Metadados Associados**, é inerente ao próprio hipertexto, como o que ocorre no XML, onde é possível definir um recurso, como um grupo de ficheiros ou tags, ligados ao recurso que os descreve.

c) **Metadados Separados**, fica armazenado em um banco de dados ligado aos recursos.

5.3.5.2 Dublin Core Metadados

O padrão de metadados chamado de Dublin Core, é um conjunto de elementos de metadados, que objetiva facilitar a recuperação e descoberta de recursos. É usado por organizações como bibliotecas e agências governamentais, para textos, imagens e outros recursos (METADATA, 2004).

A The Dublin Core Metadata Initiative (DCMI) é uma organização dedicada a promover e difundir a adoção do padrão de interoperabilidade dos padrões de metadados e desenvolver vocabulários especializados para descrição de recursos que permitam o uso de sistemas inteligentes

Segundo Hilmann (2001), o Dublin Core Metadata é formado por um conjunto de 15 elementos, onde cada elemento é opcional ou pode ser repetido. Estes elementos são agrupados em: Conteúdo, Propriedade Intelectual e Características.

Conteúdo:

Título: Um título dado ao recurso;

Assunto: Assunto referente ao conteúdo do recurso;

Descrição: Uma descrição sobre o conteúdo do recurso;

Fonte: Uma referência para um outro recurso que tenha dado origem ao recurso;

Idioma: Idioma do conteúdo intelectual do recurso;

Relação: Uma referência a um outro recurso que se relaciona com o recurso em questão;

Cobertura: A extensão ou cobertura espaço temporal do conteúdo do recurso;

Propriedade Intelectual:

Editor: A instituição responsável pela difusão do recurso;

Criador: Uma entidade principal responsável pela elaboração do conteúdo do recurso;

Direitos: Informações sobre os direitos do recurso e seu uso;

Características:

Contribuinte: Uma entidade responsável pela contribuição do recurso;

Data: Data associada com um evento no ciclo de vida do recurso;

Tipo: A natureza ou gênero do conteúdo do recurso;

Formato: Manifestação física ou digital do recurso;

Identificação: Identificação não ambígua do recurso dentro de um dado contexto;

Borbinha (2000) diz que os elementos do Dublin Core são normalizados, o que ajuda a melhorar a consistência com outras comunidades de metadados, permitindo uma melhor clareza, definição de alcance e consistência interna nas definições dos elementos dos mesmos.

5.3.5.3 Ontologia

A Ontologia vem ganhando bastante atenção e sendo muito utilizada na organização de conteúdos e fontes de dados e faz uma combinação entre termos e relações em um domínio do conhecimento.

A literatura traz diversas definições para ontologias e aborda diversos tipos e muitas propostas de sua utilização em áreas do conhecimento, como: filosofia, computação, ciência da informação e outras.

O termo ontologia tem sua origem do grego “onto”, ser, e “logos”, palavra. A visão Aristotélica influenciou grandemente, tanto a ciência da computação, quanto a ciência da informação.

Alguns problemas têm norteado as pesquisas sobre ontologia, como por exemplo, o reuso, o compartilhamento, aquisição, integração do conhecimento, comunicação de informação entre sistemas, agentes, recuperação de informação e outras.

O objetivo da construção de uma ontologia surge da necessidade de um vocabulário compartilhado para se trocar informações entre membros de uma comunidade.

No artigo de Moreira (2002), é exposta a definição de Guarino, que explica ontologia como sendo um agrupamento das seguintes classes ou níveis:

- uma disciplina Filosófica;
- um sistema conceitual no nível semântico (independente de linguagem);
- um artefato concreto no nível sintático voltado para um propósito específico;
- como uma disciplina filosófica é escrita com “o” maiúsculo;

- quando for empregado no nível conceitual ou semântico, ontologia denotará o conjunto de conceitos e relações comuns a um determinado domínio, e pode ser expresso em linguagens naturais ou formais;
- por último, quando empregada a nível sintático, uma ontologia refere-se a um sistema formado por um vocabulário específico, utilizado para descrever alguma coisa;

Palmer e Frappaolo(2004) entende que uma ontologia é uma rede de relacionamentos, autodescrita, usada para rastrear como os termos ou palavras se relacionam.

Gruber (1994 apud Borst, 1997) define ontologia como sendo uma especificação explícita de uma conceitualização, sendo esta uma interpretação estruturada de uma parte do mundo, onde as pessoas utilizam para pensar e comunicar o mundo.

Este trabalho não tem o intuito de explanar as diferenças conceituais do termo ontologia, mas trará algumas considerações de estudos feitos por pesquisadores nessa área, com o objetivo de situá-la como uma disciplina multidisciplinar e de grande importância dentro da organização da informação.

Moreira (2004) discute algumas das interpretações sobre ontologia e traça um paralelo entre os conceitos de ontologias, são elas:

- ontologia como um sistema conceitual subjacente a uma base de conhecimento;
- ontologia como um tipo especial de base de conhecimento;
- ontologia como um vocabulário usado por uma teoria lógica;
- ontologia como uma especificação de uma conceitualização;

A ontologia da filosofia e a ontologia da ciência da computação são objetos distintos. A filosofia classifica ontologia como sendo uma estrutura conceitual e não chega a representar o conhecimento.

A ontologia da ciência da computação e os tesauros são objetos que operam no mesmo nível (epistemológico), mas possuem propósitos diferentes, onde a primeira é voltada para registro de conceitos de um domínio, com o propósito de inferência automatizada e o segundo, para comunicação entre o usuário e acervos; os tesauros cumprem parte dos objetos que a ciência da computação pretende com o uso de ontologia e por isso são denominadas de ontologias terminológicas.

Dessa forma, a ciência da computação, mais precisamente a sub-área de Inteligência Artificial (IA) tem investido em pesquisas sobre as ontologias, porque esta necessita de um conjunto de vocabulários para representação do conhecimento, uma vez que a IA defende a existência das coisas, se elas puderem ser representadas de alguma maneira, e constata-se isto nas palavras de Gruber.

Para sistemas baseados em Inteligência Artificial o que “existe” é aquilo que pode ser representado. Quando o conhecimento de um domínio é representado em um formalismo declarativo, o conjunto de objetos que podem ser representado é chamado de universo do discurso. Este conjunto de objetos, e a descrição do relacionamento entre eles, são adicionados em um vocabulário de representação, onde os sistemas baseados em conhecimento representam o conhecimento. (GRUBER, 1993, p. 910).

Outra questão importante que a ciência da computação aborda, é com relação à interoperabilidade de sistemas, onde os agentes fazem um importante papel. Um agente pode ser considerado como um assistente, que utiliza recursos da IA com o intuito de auxiliar um usuário na realização de suas atividades, porém, os agentes precisam de um vocabulário comum, que pode ser provido pelas ontologias.

Através dos agentes, por exemplo, é possível realizar as buscas nos sites da Web e recuperar informações relevantes para o usuário. Um agente é capaz de entender uma ontologia, e de retornar informações mais precisas e confiáveis e podem também ser usados para ajudar na formação da própria base de metadados, alimentando-se assincronamente por

meio de informações encontradas a partir de outros sites, de acordo com informações previamente definidas. (HEND, 2001).

O uso de ontologias tem permitido resultados no desenvolvimento de sistemas de comércio eletrônico, no sentido de normalizar modelos de negócio, o que faz com que sejam rompidas barreiras como, por exemplo, a variedade de sistemas e de configurações utilizadas.

Uma das principais motivações para o uso de ontologias é a possibilidade de compartilhamento e reuso através de diferentes aplicações.

De acordo com Almeida (2003), as ontologias podem ser classificadas de acordo com o grau de formalismo de seus vocabulários, com a estrutura e o assunto da conceitualização, com sua função e com a sua aplicação.

Com relação aos tipos de ontologias, Almeida (2003) utiliza as definições de Van-Heijst Schreiber e Wielinga, responsáveis pela organização de uma biblioteca de ontologias reutilizáveis na área médica. Para eles, as ontologias podem ser classificadas de acordo com suas estrutura e assuntos:

- ontologias terminológicas, que especificam termos utilizados para representar conhecimento;
- ontologias de informações, responsáveis por registros em banco de dados;
- ontologias de modelagem do conhecimento, especificam conceitualizações;
- ontologias de aplicação, possuem definições para modelagem do conhecimento através de uma aplicação;
- ontologias de domínios, expressam conceitualizações para um domínio; ontologia genérica, que expressa conceitualizações genéricas;

- ontologias de representação, explicam as conceituações dos formalismos de representação do conhecimento;

Para Guarino (1996), os tipos de ontologias são baseados em seus conteúdos e se classificam da seguinte maneira:

- ontologias genéricas descrevem conceitos gerais, tais como, espaço, tempo, ação, etc;
- ontologias de domínio expressam conceitos de domínio, que descrevem vocabulários relacionados com um domínio genérico, como por exemplo, medicina, química e biologia;
- ontologias de tarefas expressam conceitos sobre a solução de problemas, independente de domínio, ou seja, descrevem vocabulário relacionado a uma atividade ou tarefa genérica, como, por exemplo, diagnóstico de vendas.
- ontologias de aplicação expressam conceitos dependentes do domínio e de tarefas particulares, normalmente correspondem a papéis desempenhados por entidades de domínio ao se realizar uma certa tarefa;
- ontologias de representação expressam conceitos que fundamentam os formalismos da representação do conhecimento. Um exemplo dessa categoria é a ontologia de frames, utilizada em Ontolíngua.

As ontologias não necessariamente precisam ter sempre a mesma estrutura, e mesmo não havendo consenso sobre os tipos de ontologias e seus componentes, é perceptível a semelhança dos conceitos.

Existem muitas ontologias criadas para determinadas áreas ou para executar certas tarefas, e é possível reutilizá-las, desde que se tome cuidado com a seleção correta, pois nenhuma ontologia pode ser totalmente adequada a todos os indivíduos ou grupos. É possível

construir ontologias a partir de outras ontologias já existentes, o que alguns autores chamam de integração de ontologias e outros, de fusão de ontologias.

Gruber (2001) afirma que uma ontologia conta com componentes como: classe, relações, axiomas e instâncias. As classes são descrição de um conceito que pode ser organizado em uma taxonomia, as relações representam a interação entre conceitos, axiomas representam sentenças verdadeiras, e as instâncias representam os dados.

Com relação às técnicas de construção de ontologias, diversas áreas têm tomado iniciativas nesse sentido, como é o caso da Inteligência Artificial, no desenvolvimento de sistemas baseados em conhecimento, ou as que estudam a Linguagem Natural, ao analisar o significado das palavras, a área de banco de dados que busca interoperabilidade semântica, e a Ciência da Informação, que objetiva melhorar a organização da informação para que haja posteriormente, uma melhor recuperação.

Existem algumas ferramentas tecnológicas que apóiam a construção das ontologias, como por exemplo: CODE4, VOID, ONTOLINGUA, ONTOEDIT, PROTEGÉ, WEBONTO, OILED e outras.

A maioria das linguagens para descrever ontologias é baseada em cálculos de predicados. Como exemplos de linguagens para a criação e o compartilhamento de ontologia, pode-se citar SHOE, XOL, OML RDFS, OIL, DAML+OIL e outras. Todas elas têm o intuito de criar descrição de objetos e seus relacionamentos, e expressar semântica a esses objetos para o caso de uma posterior publicação e compartilhamento dessas ontologias em um sistema de informação e assim, permitir uma recuperação eficiente desse objeto.

De forma geral, as ontologias constituem uma importante ferramenta que suporta especificações e implementações de sistemas computacionais de qualquer complexidade.

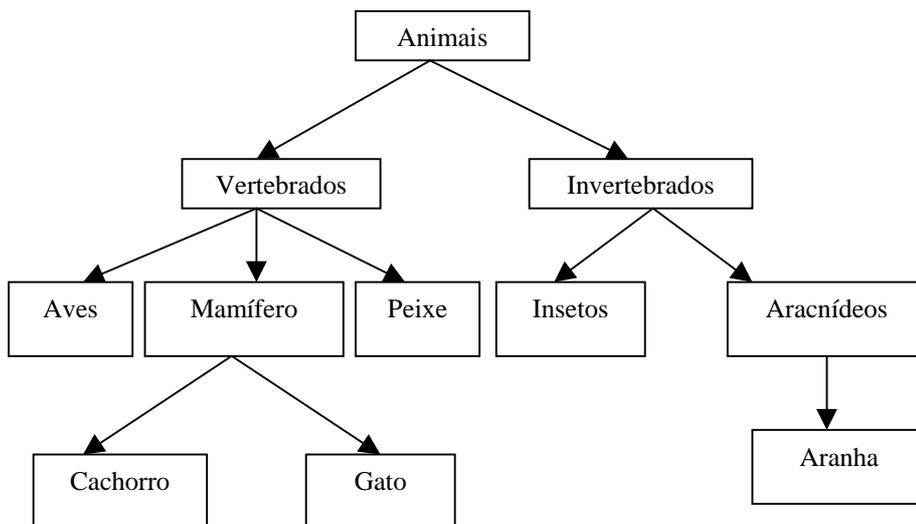
5.3.5.4 Taxonomia

O dinâmico crescimento do número de informações trouxe consigo a necessidade de organização da informação e do conhecimento, e como consequência, a utilização da ferramenta taxonomia.

Taxonomia não é algo novo, pois em 1735, o Sueco Karl Von Linné usou esse conceito na criação de um tipo de classificação que dividia os seres vivos em grupos, de acordo com suas características em comum, respeitando uma hierarquia.

Exemplo da taxonomia de Karl Von Linné:

Figura 1



(Fonte: TERRA, 2004)

As taxonomias são aplicações que podem, por exemplo, ajudar usuários a encontrarem fontes de materiais na internet; leitores a localizam informações em um livro ou jornal; e compradores localizarem produtos ou serviços, etc.

Diversas são as definições e inúmeros os autores que discorrem sobre taxonomia. De um modo geral, todas as definições são diferentes, mas se assemelham em um aspecto; aos vocabulários controlados, classificação e tesouros.

Considerando a abrangência e as especificidades das definições sobre taxonomias, a seguir serão apresentadas algumas encontradas na literatura.

Taxonomia é um sistema utilizado para classificar e facilitar o acesso à informação, seu objetivo é representar conceitos através de termos; melhorar a comunicação entre especialistas e outros públicos; propor formas de controle da diversificação e oferecer um mapa do processo de conhecimento. É portando um vocabulário controlado de uma determinada área do conhecimento e um instrumento que permite alocar, recuperar e comunicar informações dentro de um sistema. (TERRA, 2004)

Segundo Graef (2001), taxonomia é uma estrutura que provê uma maneira de classificar coisas através de uma série de grupos hierárquicos para facilitar sua identificação, estudo ou localização. Para ele, a estrutura taxonômica consiste em duas partes: Estrutura e Aplicações, onde:

- estrutura consiste em categorias ou termos e os seus relacionamentos;
- aplicações são ferramentas de navegação que ajudam usuários encontrarem as informações;

Wason (2000) define taxonomia como sendo:

- um vocabulário controlado de termos ou frases;
- uma classificação ordenada a respeito de uma série de relações naturais;
- um vocabulário que contém apenas um nível constante de uma lista de termos ou frases permitidas;

Thulasi (2001 et al) define Taxonomia como sendo uma maneira sistemática de organizar o conhecimento, que provê uma estrutura hierárquica dos termos do conceito

(categorias e sub-categorias) que ajudam no desenvolvimento de uma língua comum (vocabulário) para a organização e compartilhamento do conhecimento.

Com relação à origem da taxonomia, alguns autores afirmam que ela foi embasada nos conhecimentos de biblioteconomia como vocabulários controlados e classificação, mas a taxonomia pode incluir novos elementos, que a diferencia dos tradicionais tesouros e dos esquemas de classificação (EDOLS, 2001).

Com o intuito de diferenciar os conceitos (taxonomia, ontologia, tesouros e classificação), a seguir serão mostradas algumas características.

Segundo Edols (2000) os elementos que diferenciam a taxonomia dos tesouros dos esquemas de classificação são:

- A taxonomia possui estrutura de suporte, conteúdo e aplicação (ferramentas para navegação);
- A taxonomia frequentemente é criada através da mesclagem e racionalização de diferentes tesouros e índices;
- As taxonomias são personalizadas para refletir linguagem, cultura e objetivos de um recurso em particular;
- As taxonomias são frequentemente criadas usando-se da combinação de esforços humanos e softwares especializados;
- As taxonomias podem se referir a recursos de informação. Nas intranets, elas podem incluir recursos internos como mensagens eletrônicas, memorandos, documentos pessoais bem como informações sobre fontes de livros, partes de livros, relatórios e páginas Web;

Para Conway (2002), a diferença entre a classificação e a taxonomia é mais na teoria do que na prática. Ambos são vocabulários controlados estruturados por relacionamentos lógicos. A diferença entre eles é baseada na complexidade dos relacionamentos expressados, onde a classificação agrupa os termos utilizados, a taxonomia agrupa termos com interatividade.

A ontologia por sua vez, agrupa taxonomias complexas e relacionadas, sustentadas pela descrição e pela definição dos termos e dos relacionamentos. Apesar da sutil diferença, são mecanismos que devem ser utilizados para melhorar a recuperação da informação, da consistência da comunicação, e da criação do conhecimento.

A taxonomia é desenvolvida para prover a uma instituição ou grupo, uma estrutura comum de conceitos e relações entre esses conceitos, para estruturar os elementos léxicos da linguagem, produzindo uma rede semântica comum. A taxonomia permite a elaboração de um vocabulário controlado para recuperar a informação, criar metadados, além de fornecer esquemas que orientam estruturas e leiaute de página da web (CONWAY, 2002, p.105, tradução nossa).

A taxonomia tem grande importância para informações semi-estruturadas, aquelas armazenadas, normalmente, nos servidores de arquivos das corporações, e para informações desestruturadas, aquelas armazenadas nos discos rígidos, e que são de interesse individual. A taxonomia pode permitir que sejam criados links entre os trabalhadores e os conteúdos que são de seus interesses.

Sabe-se que existem vários tipos de taxonomia diferentes, assim como vários métodos para construí-las. A seguir serão citados os três tipos de taxonomias de Conway (2002), que são geralmente elaboradas para ambientes on-line: Taxonomia descritiva, Taxonomias para navegação e Taxonomia para gerenciamento de dados.

- Taxonomia descritiva são vocabulários controlados, ou seja, são construídos embasando-se nos tesouros. São normalmente encontradas nas corporações e designa termos autorizados, termos de entradas ou variantes que permitem aos usuários, durante o processo de busca, utilizarem seus termos preferidos.

A taxonomia descritiva não força os usuários a usarem um único conjunto terminológico, mas, na sua construção, devem ser adicionados diversos tipos de palavras variantes, ortografia variantes, formas variantes, sintáticas variante e dialetos variantes, tudo isso para que o usuário tenha maior liberdade na hora de buscar um assunto.

A taxonomia descritiva quando usada em máquinas de buscas pode melhorar o desempenho das pesquisas, pois retorna informações que foram previamente marcadas como parte do conjunto de termos representantes da informação.

- Taxonomia para navegação é um tipo de taxonomia organizada de maneira diferente da taxonomia descritiva. Criar este tipo de taxonomia envolve determinar agrupamentos apropriados das informações. Ela pretende descobrir informações por meio do comportamento do usuário mediante a utilização de navegadores (browsing). É baseada nos modelos mentais dos trabalhadores e em como a informação está organizada, bem como no comportamento de busca do usuário e não no conteúdo.

- Taxonomia para gerenciamento de dados é uma taxonomia que contém pequeno conjunto de termos controlados, com significância particular enumerada. Ela contém elementos de dados controlados que se referem a atributos específicos. O propósito desse tipo de taxonomia é assegurar facilidade de compartilhamento de dados

Conway (2002) diz que as taxonomias corporativas não devem apenas ser utilizadas para criar lista de termos autorizados em buscas por informação, elas também devem ser utilizadas para criar mapas entre conceitos e conectar funcionários de uma corporação ao conhecimento, em tempo real. Elas podem criar uma rede semântica embasada no negócio da corporação, tornando-se uma importante ferramenta de gerenciamento do capital intelectual da corporação.

Para Conway (2002), o escopo de uma taxonomia corporativa é formado por três fatores que a delimitam:

- a necessidade do negócio da empresa;
- a necessidade de informação que direciona a criação e aquisição do conteúdo;
- comportamento de busca dos usuários da informação;

É possível perceber uma estreita ligação entre os fatores descritos anteriormente, pois embasados na necessidade de negócio da empresa, pode-se determinar o escopo da taxonomia, este que também poderá ser influenciado diretamente pelas necessidades de informação que direcionarão a criação de conteúdos. Da mesma forma, o comportamento de busca de informações dos usuários irá influenciar na maneira como esses conteúdos deverão ser organizados.

As máquinas de busca da web são as principais fontes de pesquisa para muitos usuários, mas trazem uma grande quantidade de documentos, e normalmente, junto com os documentos relevantes, retornam também uma grande quantidade de documentos irrelevantes ou sem qualidade.

Para Pahlevi (2005), a taxonomia pode ser usada para melhorar a precisão das buscas na web. Como exemplo de serviços de buscas baseada em taxonomias, pode-se citar o diretório da web Yahoo (<http://www.yahoo.com>) como um mecanismo que classifica as informações de sites da web através de uma taxonomia e permite aos usuários restringir seu escopo de pesquisa selecionando uma categoria de sua taxonomia.

Segundo Holgate (2004), para se construir uma taxonomia, pode-se escolher uma das maneiras abaixo:

- adquirir uma taxonomia pré-definida;
- construir manualmente uma Taxonomia;
- construir automaticamente uma Taxonomia;
- construir uma Taxonomia híbrida, que é a combinação da

maneira automática, com a maneira manual;

A seguir será detalhada cada uma das maneiras citadas por Holgate (2004).

- Taxonomia pré-definida: conseguida através de empresas que fornecem pré-construção de uma taxonomia. Existem no mercado diversas

empresas que vendem taxonomias pré-definidas. O problema desse tipo é que são criadas categorias de informação não condizentes com o negócio da organização, e normalmente precisam ser revisadas e corrigidas manualmente.

- Construção manual: desenha o entendimento do usuário e do ambiente organizacional. É desenvolvida por um funcionário da organização, que detém vasto conhecimento e é um especialista no conhecimento do ambiente de negócio. Ele, provavelmente, desenvolverá uma boa taxonomia, porém, levará talvez meses para ficar pronta.

- Construção automática: é muito mais rápida, menos cara, e envolve menos especialista do assunto, porém gera como resultado, um conjunto de informações redundantes. Para este tipo de taxonomia existem deferentes softwares que assistem a sua construção automática. Alguns dos maiores pacotes de softwares são: Autonomy, Dialog Infosort, Inxight, Plumtree Coporate Portal, Quiver, Semio e Verity Knowled Organizer.

- Construção Híbrida: provê um meio de construção automática e hierárquica, demanda tempo, e esta embasada em conceitos contidas no próprio corpus do conteúdo. Utiliza a construção automática e posteriormente, pode vir a servir-se da construção manual para o desenvolvimento de uma correta hierarquia.

A escolha de qual tipo de taxonomia será utilizada deve ser embasada em quais problemas a taxonomia terá que resolver, em qual o tipo, e o escopo da informação que deve ser incorporada, no volume de conteúdo e na disponibilidade de especialistas no assunto para evoluírem a taxonomia.

Observa-se na literatura um grande número de autores que discorrem sobre os passos para a criação de uma taxonomia. A seguir serão citados alguns deles.

Segundo Lederman (2005) existem três desafios para a implementação de uma taxonomia: Construção de uma estrutura de vocabulário e relacionamentos, o uso da estrutura de vocabulário para categorizar documentos na medida que são criados e adicioná-los em um repositório e, por último, tornar a busca capaz de usar as categorias definidas.

Para o primeiro desafio, criar e gerenciar um vocabulário estruturado, pois existem muitos softwares com essa finalidade. Uma vez criada uma categoria, alguém deve adicionar os sinônimos e variações desse termo, pois, em uma corporação, alguns termos são utilizados em um determinado departamento e em outro departamento o mesmo termo recebe outro nome, logo, esses termos devem ser relacionados na taxonomia, para que a sua recuperação ocorra com sucesso.

O segundo desafio citado por Lederman (categorização de documentos), também conta com o apoio de softwares, e ocorre quando um documento é criado ou adicionado em um repositório ou rede. Ele deve ser categorizado para que o acesso a esse documento seja feito de qualquer maneira, por várias categorias como, por exemplo, através de qualquer assunto, linguagem, tipo de documento e outros. De qualquer maneira, é importante que alguém confira essa categorização e a modifique ou adicione novas categorias.

O terceiro desafio (capacitar os mecanismos de recuperação da informação a reconhecer as categorias) é uma tarefa árdua, pois é essencial que os softwares tenham a capacidade de mostrar o resultado de busca através de um vocabulário estruturado. Esses softwares devem ter algumas características: mostrar os resultados com alto índice de relevância dentro de cada categoria; permitir que novos termos sejam adicionados, até mesmo, em diferentes linguagens; e permitir combinar categorias com palavras.

Para Conway (2002), o ponto de partida para a construção de uma taxonomia é examinar o comportamento de busca, pois através da pesquisa de usuário é possível descobrir

quais tipos de informações são usados pelos usuários, e os termos mais utilizados para a busca de um determinado assunto.

Uma outra maneira para se descobrir quais os tipos de informação os usuários buscam, é a utilização de registro de problemas, pois a informação pode não ter sido encontrada, devido à maneira como foi categorizada, por isso a importância de se fazer um levantamento de taxonomias existentes, uma vez que o seu reuso pode ajudar a economizar tempo e esforço.

Graef (2001) descreve três passos para a criação de uma taxonomia:

- 1º) criar uma lista de termos (o autor chama de vocabulário), que devem descrever o conteúdo a ser organizado;
- 2º) adicionar relacionamento entre os termos. Além dos relacionamentos, podem ser adicionadas definições e notas;
- 3º) conectar os termos com a fonte de informação. Este terceiro passo é o que o autor chama de etapa Aplicação;

A evolução de uma taxonomia é algo que deve ser prevista, porque depois de uma taxonomia ser criada e após o seu uso, os usuários poderão participar de mudanças, combinação e criação de novos conceitos, e uma reestruturação poderia melhorar a representação da necessidade de busca dos usuários.

6. Metodologia

6.1 Caracterização da pesquisa

A pesquisa é de natureza qualitativa e descritiva, pois tem como objetivo entender a natureza de um fenômeno social.

Esta pesquisa tem como objetivo identificar o comportamento de busca e as necessidades de informação de um grupo de pesquisadores que estudam a Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil Central, no mestrado em Entomologia da UFGD (Universidade Federal da Grande Dourados).

Numa primeira etapa foram realizadas entrevistas com quatro pesquisadores do mestrado em Entomologia e Conservação da Biodiversidade da UFGD, mais precisamente, com os pesquisadores do projeto Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil central.

Com as entrevistas buscou-se, descobrir as necessidades de informação desses usuários, coletar os termos utilizados por eles nas suas buscas por informação e as definições desses termos, e criar as categorias de assuntos que englobasse a área de interesse desses pesquisadores com relação ao projeto anteriormente descrito.

Após as entrevistas, os pesquisadores foram observados ao realizarem buscas por informações em periódicos especializados da área de Entomologia, via internet. Os periódicos que serviram de bases de dados⁶ especializadas para as buscas, foram: Revista Brasileira de

⁶ Para este estudo, a base de dados será o que Date (2000) define como um repositório ou recipiente para coleções de arquivos de dados computadorizados.

Entomologia e a Neotropical Entomology. Ambos são periódicos que fazem parte da Scientific Electronic Library Online – SciELO (Biblioteca virtual que proporciona acesso a uma coleção de periódicos on-line). O intuito da observação foi identificar o comportamento de busca dos usuários.

De posse dos termos comuns utilizados nas buscas por informação e de suas definições (coletadas através das entrevistas e também de uma coletânea de termos técnicos da área de Entomologia) e dos sinônimos desses termos, foi elaborada uma taxonomia para organização das informações ligadas ao estudo da Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil central.

Tendo em vista a diversidade de definições para o termo Taxonomia, a utilizada neste trabalho será a de Graef (2001), que diz que Taxonomia é uma estrutura que provê uma maneira de classificar coisas através de uma série de grupos hierárquicos para facilitar sua identificação, estudo ou localização.

Para Graef (2001), a estrutura taxonômica consiste de duas partes, que o autor chama de “Estruturas” e “Aplicações”. Neste trabalho, somente será desenvolvida a primeira parte (Estrutura), correspondente à criação de uma lista de termos que descreve o conteúdo a ser organizado e a criação dos relacionamentos desses termos.

A segunda parte da estrutura taxonômica de Graef (Aplicações), que corresponde à ligação dos termos com os seus respectivos conteúdos, não está prevista no escopo desta pesquisa e ficará para trabalhos futuros.

Como base para a elaboração da taxonomia foram consideradas a padronização dos termos e definições, advindas da Terminologia, e as categorias de assuntos e os seus relacionamentos, da Teoria do Conceito de Dahlberg.

A elaboração da taxonomia foi realizada manualmente, conforme uma das maneiras definidas por Holgate (2004), e foi do tipo descritiva, conforme definido por Conway; Sliger (2002).

A elaboração da taxonomia foi composta pelas seguintes fases:

- Coleta dos termos utilizados para as buscas referentes aos assuntos relacionados com o estudo da Biodiversidade dos insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil central e suas definições;
- Criação das categorias de assuntos referentes ao estudo da Biodiversidade dos insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil central;
- Identificação dos relacionamentos entre os termos e dos sinônimos e variações;
- Elaboração de um esboço de taxonomia, enviado aos especialistas (pesquisadores) no assunto de insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil central, para possíveis correções;
- Elaboração da taxonomia final para este trabalho (pois uma taxonomia pode, e evolui, sempre que necessário).

6.2 Ambiente de estudo

A instituição objeto de estudo é a Universidade Federal da Grande Dourados (UFGD), que é uma instituição de ensino superior, vinculada ao governo federal.

A UFGD iniciou suas atividades como instituição de ensino superior em 1962, quando ainda se chamava Universidade Federal de Mato Grosso do Sul-UFMS, com sede em Campo Grande-MS.

Em 1970, foram criados os centros pedagógicos de Aquidauana e Dourados posteriormente, os campus de Coxim, Corumbá, Paranaíba e Três Lagoas.

Em 2006, o campus de Dourados, que até então era chamado de CEUD – Centro universitário de Dourados, foi desmembrado da UFMS e passou a se chamar UFGD – Universidade Federal da Grande Dourados.

A UFGD por seu posicionamento estratégico na região vem participando do ensino e da preservação dos recursos naturais do meio ambiente, especialmente da flora e da fauna do Pantanal, região onde está inserida, e tem motivado estudos e pesquisas ecológicas na instituição.

A UFGD oferece diversos cursos nas modalidades de graduação e pós-graduação Lato Sensu, bem como, tem desempenhado um importante papel no âmbito de pós-graduação Stricto Sensu.

A UFGD está dividida em nove departamentos e um núcleo experimental das ciências agrárias (NCA), dentre eles, está o Departamento de Ciências Biológicas (DCB), que atende aos cursos de Ciências Biológicas, Agronomia, Medicina e Geografia.

6.2.1 Mestrado em Entomologia e Conservação da Biodiversidade

A Entomologia é um ramo da Zoologia que estuda especificamente os insetos nos seus diferentes aspectos, como parte da grande área das Ciências biológicas.

A Entomologia se sub-divide em:

- Entomologia Agrícola: estuda as pragas agrícolas e os inimigos naturais;
- Entomologia Florestal: analisa os insetos associados às florestas naturais e artificiais;
- Entomologia Médica-veterinária: considera os insetos vetores de doenças, parasitóides e/ou vetores de doenças humanas ou outros animais;
- Entomologia Forense: estuda os insetos associados a cadáveres visando elucidar crimes;

Inúmeras são as razões que levam ao estudo dos insetos, uma delas é a própria diversidade, pois são inúmeras as espécies. A quantidade de insetos supera a quantidade de qualquer outro animal.

Uma segunda razão que justifica o estudo dos insetos é devido à produção de alimentos e outros produtos para a população humana, pois dependendo da produção agrícola, ela pode ser muito prejudicada, caso a proliferação de determinados tipos de insetos (pragas) não seja devidamente monitorada e controlada.

O mestrado em Entomologia e conservação da biodiversidade da UFGD é curso ligado diretamente ao departamento de Ciências Biológicas, e tem o objetivo de capacitar recursos humanos dessa área, para pesquisa e docência, por meio de uma formação científica, cultural e tecnológica, que contribua para o desenvolvimento regional e do país.

O programa de pós-graduação em entomologia possui quatro linhas de pesquisas, sendo elas:

Artrópodes, Parasitas e Vetores de importância médico-Veterinária;

Biodiversidade e manejo de insetos no cerrado e pantanal;

Ecologia e Estratégia Reprodutivas de Artrópodes;

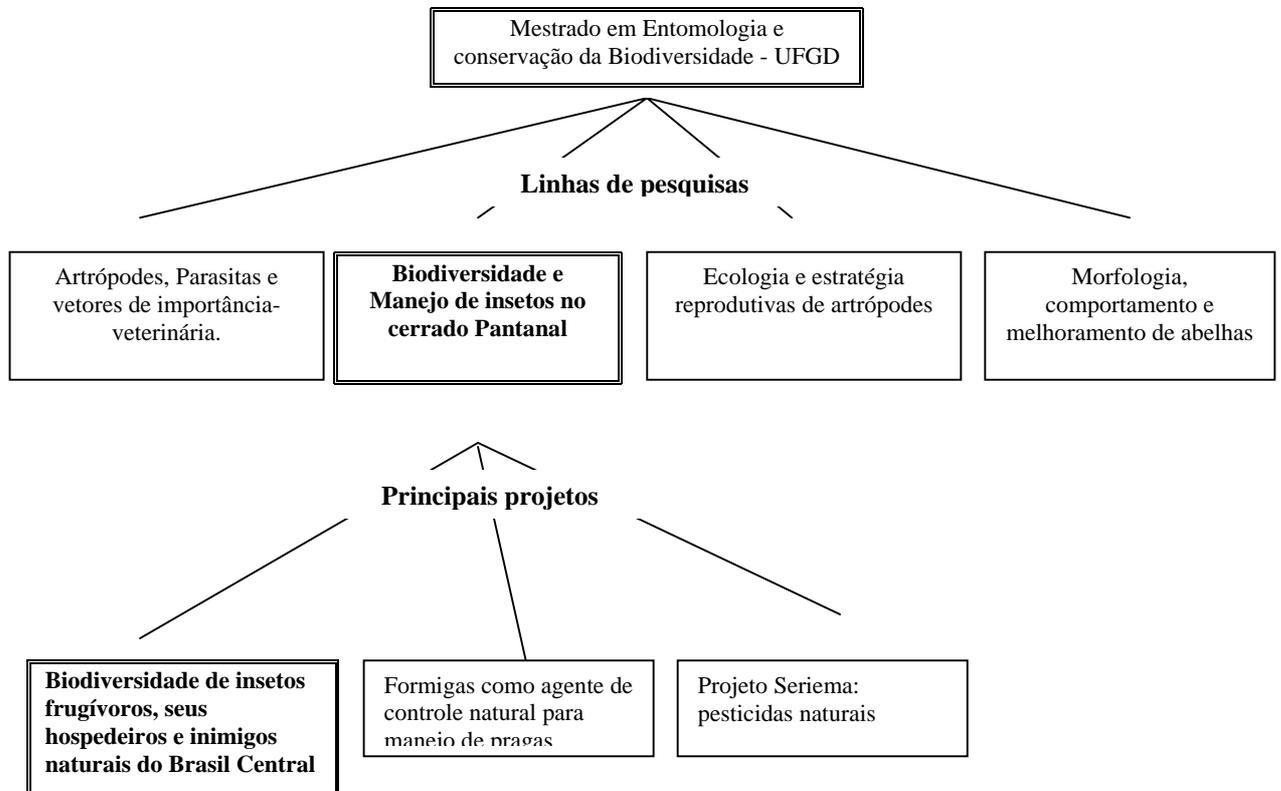
Morfologia, comportamento e Melhoramento de Abelhas;

A linha de pesquisa colaboradora deste trabalho é a biodiversidade e manejo de insetos no cerrado pantanal. Essa linha possui diversos projetos dentre eles o que estuda a **Biodiversidade dos insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil central**, um grande projeto, que vem sendo executado desde 1998.

O projeto tem o objetivo de reunir informações bioecológica sobre as moscas frugívoras (Dípteras: Tephritoidea) seus hospedeiros e inimigos naturais na região oeste do Brasil. Tem como meta, estudar a biologia e o comportamento das espécies de *Anastrepha* com importância econômica, como pragas de frutas cultivadas, viabilizando o emprego de parasitóides e predadores como agentes do biocontrole. Além disso, busca-se evidenciar a existência de feromônios sexuais e/ou deterrentes de ovoposição, identificação, síntese e emprego destes infoquímicos em pomares para o monitoramento e controle populacional das moscas-das-frutas.

Para um melhor entendimento do ambiente de pesquisa, foi elaborado um esquema representativo que se encontra a seguir.

Figura 2 – Esquema representativo do ambiente de estudo



O estudo da Biodiversidade dos insetos frugívoros, seus hospedeiros e inimigos naturais do Brasil Central é um grande projeto, que engloba vários sub-projetos como:

- Biodiversidade de moscas frugívoras (Tephritoidea) capturadas em armadilhas na reserva florestal Picadinha, Município de Dourados –MS;
- Evidenciação de feromônio sexual e/ou deterrente de oviposição em *Anastrepha sororcula* díptera tephritidae;
- Levantamento da diversidade de moscas-das-frutas, Lochaetidae e seus parasitóides em hospedeiros nativos e exóticos no Pantanal sul-mato-grossense;
- Moscas frugívoras, seus hospedeiros e inimigos naturais no pantanal;

6.3 Definição da População /Amostra da pesquisa

De um total de 22 pesquisadores do mestrado em Entomologia, foram selecionados quatro pesquisadores para responderem às questões. A amostra é intencional, pois foram selecionados pesquisadores que têm um grande número de pesquisa e publicações.

6.4 Técnica e instrumento de coleta de dados

A técnica utilizada para coleta de dados foi a entrevista semi-estruturada, seguida de uma observação. As variáveis observadas que fizeram parte do roteiro da observação e da entrevista foram: necessidade de informação dos usuários e comportamento de busca dos usuários de informação.

Os usuários foram observados em um ambiente informatizado, de onde acessaram bases de dados de suas áreas de especialização e realizaram buscas dentro de seus temas de pesquisa.

Durante as fases de entrevista e observação os usuários relatam suas experiências nas buscas por informação quanto a:

- necessidade de informação;
- procedimentos para a localização das informações procuradas;
- assuntos mais procurados;
- termos utilizados na busca da informação

6.5 Resultado do pré-teste

Com a finalidade de verificar a relevância das questões elaboradas para a coleta de dados realizou-se um pré-teste com uma pesquisadora da área de Entomologia que estuda moscas frugívoras. Constatou-se que as questões funcionavam, mas poderiam ser melhoradas no sentido de tentar descobrir mais detalhes sobre a razão da não recuperação das informações, e sobre a terminologia utilizada para as consultas nas bases de dados.

Os tópicos abordados nas entrevistas:

1. Necessidade de informação;
2. Procedimentos de busca utilizados para a localização das informações;
3. Assuntos mais procurados;
4. Termos utilizados na busca por informação;

1) Necessidade de informação: A pesquisadora revela ter necessidade de informações a respeito de: comportamento de atração sexual das moscas do gênero *Anastrepha*; comunicação química entre os machos sexualmente ativos e as fêmeas sexualmente ativas; feromônio em *Anastrepha sororcula*.

Obs. Quanto a esses itens, a pesquisadora revela que tem dificuldade para encontrar informações.

2) Procedimento de busca: a pesquisadora diz que costuma utilizar sites de busca para fazer sua pesquisa, mas na maioria das vezes, usa base de dados especializados e periódicos. Ela revela também, que quando procura por informações em sites de busca, acaba vindo muita informação equivocada, e que em bases especializadas, a confiabilidade é maior.

3) Assuntos mais procurados: a pesquisadora relaciona uma lista de assuntos mais procurados, que são: controle das moscas-das-frutas; utilização de controle biológico em moscas-das-frutas; utilização de feromônio como mecanismo de controle das moscas-das-frutas; monitoramento através de armadilhas e distribuição geográfica das moscas-das-frutas.

4) Termos utilizados na busca de informação: a pesquisadora relaciona uma lista de termos utilizados nas buscas por informação:

- Feromônio de *Anastrepha*;
- Moscas-das-frutas;
- *Anastrepha sororcula*;
- Corte de *Anastrepha sororcula*;
- Acasalamento;
- Cairomônios;
- Feromônios;
- Controle biológico;
- Atrativo alimentar;

Obs1. A pesquisadora afirma que, quando procura somente por *Anastrepha sororcula*, os mecanismos de recuperação trazem muita informação a respeito de sites pessoais, currículos e outros.

Obs2. A pesquisadora comenta, que termos mais gerais são mais fáceis de serem localizados e que os termos mais específicos são mais difíceis.

7 Análise dos Dados

Esta pesquisa que teve por objetivo identificar as necessidades de informação e o comportamento de busca dos usuários de informação que estudam a biodiversidade dos insetos frugívoros, após a coleta e análise dos dados propiciou subsídio informacional para a elaboração de uma taxonomia.

Conforme descrito na metodologia, para se cumprir com o primeiro objetivo - identificar as necessidades de informação dos usuários - utilizou-se a entrevista semi-estruturada como técnica de coleta de dados, e para se cumprir o segundo objetivo, que era identificar o comportamento de busca dos usuários utilizou-se a observação.

Os resultados da análise dos dados foram apresentados de acordo com a ordem dos objetivos.

7.1 Necessidades de informação

As entrevistas foram realizadas no período de 25 de março a 13 de abril de 2006, de acordo com a disponibilidade de cada entrevistado. No total, foram quatro entrevistados, todos pesquisadores que estudavam a biodiversidade dos insetos frugívoros, dois eram docentes e dois mestrandos da pós-graduação em Entomologia.

A realização das entrevistas ocorreu individualmente, com duração média de 100 minutos. Cada entrevistado concedeu uma entrevista, sendo que um dos pesquisadores que

trabalhava com a Biodiversidade de insetos frugívoros desde 1998 foi entrevistado três vezes, em dias diferentes, devido à riqueza de detalhes que ele estava informando.

Em suas entrevistas, os usuários além de relatarem sobre suas necessidades de informação, também informaram sobre:

- os termos que utilizavam para realizarem suas pesquisas;
- as relações entre esses termos e as definições dos mesmos;
- mecanismos e estratégias eram utilizados para as suas buscas por informação;
- as dificuldades encontradas enquanto executavam suas buscas.

Essas informações além de utilizadas para identificar as necessidades de informações dos usuários, contribuíram para com a elaboração da taxonomia.

Os termos relatados como sendo os mais utilizados pelos usuários para as suas buscas por informação foram: insetos frugívoros (*insects frugivorous*), manejo integrado de pragas, controle biológico (ou *biocontrol* ou *biological control*), controle natural (ou *natural control*), inimigos naturais (ou *natural enemies*), parasitóides, hospedeiros, microorganismos patogênicos nemátodos, comunicação química, feromônios, caormônios, moscas-das-frutas (*fruit-fly* ou *fruits-flies*), *Anastrepha*, *Ceratitis capitata*, flutuação populacional, levantamento, análise faunística (*study faunistic*), atrativo alimentar, corte, atrativo alimentar, Tefritídeos (*Tephritidae*), Dipteros, biodiversidade (*biodiversity*) e distribuição geográfica. *Anastrepha obliqua*.

Durante as entrevistas, os usuários descreveram as definições dos termos que foram utilizadas juntamente com as definições adquiridas juntamente a uma coletânea de termos técnicos de Entomologia para criar as relações entre esses conceitos.

Através desse procedimento criou-se a ligação entre os termos controle natural e controle biológico, onde o segundo possui uma relação hierárquica para com o primeiro.

Todas as definições dos termos foram anotadas e utilizadas para a elaboração da taxonomia, e poderão ser conferidas na mesma.

Os usuários disseram que costumam realizar suas buscas em mecanismos de buscas da internet, ou em periódicos especializados, porém, a maioria dos pesquisadores reclamou da eficiência dos mecanismos de buscas da internet. Eles relataram que ao realizarem suas buscas nesses mecanismos, normalmente retornam muita informação irrelevante ou de origem duvidosa.

Os usuários informaram a existência de um grande número de termos em inglês, que são utilizados nas publicações da área. Para alguns usuários, essa é uma das razões pelas quais se torna difícil a recuperação das informações.

Outros usuários relataram que as dificuldades para a recuperação da informação ocorrem porque os termos utilizados nas publicações da área não são comuns, e não existe uma padronização dos mesmos. Salientaram que se houvesse uma certa padronização, certamente a recuperação das informações seria facilitada.

Uma outra dificuldade relatada foi com relação às publicações mais antigas, que são de extrema importância para eles, porém muitas delas não foram ainda digitalizadas e não são encontradas on-line.

Através dos dados coletados com a entrevista, e através da generalização das necessidades do grupo que estuda a biodiversidade de insetos frugívoros, pode-se afirmar que as necessidades de informação desse grupo são:

- Hospedeiros e inimigos naturais dos insetos frugívoros;
- Comportamento dos insetos frugívoros (insect behaviour);
- Estratégias reprodutivas dos insetos;
- Controles natural e biológico dos insetos (biocontrol);
- Levantamento populacional;

- Estudo da população;
- Flutuação populacional;
- Monitoramento dos insetos através de armadilhas McPhail;
- Distribuição geográfica das moscas- das- frutas;
- Controle das moscas- das- frutas;
- Utilização de controle biológico em moscas- das- frutas.

7.2 Comportamento de busca dos usuários

Após as entrevistas terem sido realizadas, os entrevistados foram submetidos individualmente a um ambiente que continha um computador ligado à internet, de onde acessaram os periódicos Revista Brasileira de Entomologia e a Neotropical Entomology e realizaram buscas por informação sobre a biodiversidade de insetos frugívoros.

A principal intenção desse procedimento foi identificar o comportamento de busca desses usuários, pois com a compreensão do mesmo, as organizações provedoras de informação, podem ser capacitadas para melhor servirem às necessidades de seus usuários, aumentando, assim, a eficiência dos indivíduos nos meios pessoais, sociais e profissionais. O ideal seria alcançar o conceito global para uma dada comunidade (FERREIRA, 1995).

Enquanto realizavam suas buscas, os usuários foram observados com relação aos seus comportamentos quanto:

- aos passos utilizados enquanto realizavam suas buscas;
- às estratégias para a realização das buscas;
- às dificuldades encontradas ao realizarem suas buscas.

- Aos termos utilizados nas buscas
- Aos assuntos de seus interesses.

Quanto aos passos utilizados para a realização das buscas, houve uma variação de comportamento, pois para alguns usuários, o primeiro passo foi selecionar o idioma português para depois seguir com a busca. Para outros pesquisadores, quando eles tinham o nome do autor de uma obra, o primeiro passo foi selecionar o autor e então seguiam com a busca.

Observou-se a existência de diferentes estratégias para a realização das buscas dos usuários, mas todos eles demonstraram possuir pouca afinidade com os mecanismos utilizados para este estudo.

Uma das estratégias utilizadas por um usuário foi o truncamento do termo. Como exemplo de truncamento, pode-se citar uma das buscas realizadas num momento em que fora pedido para que buscasse informação sobre mosca-da-fruta. A estratégia foi a seguinte: o usuário desejava informações a respeito de mosca-da-fruta, então ele digitou **fruit*** e executou o comando buscar e como resultado obteve-se uma lista de assuntos que se iniciava com **fruit**, então o usuário selecionou **fruit-fly** e realizou a busca novamente.

Uma outra estratégia, utilizada por outro usuário, foi a seguinte: quando o usuário tentou localizar uma informação e não conseguiu uma recuperação satisfatória, ele foi ao índice alfabético do periódico e tentou localizar um termo que fosse sinônimo ou semelhante ao que ele procurava através de um dos termos que constava no índice de assuntos do periódico (existia um índice alfabético de assuntos) e então utilizou um dos termos que julgou ser o correto e realizou a busca novamente.

Com relação às dificuldades de recuperação, observou-se que a principal razão da não recuperação das informações, deve-se ao fato de existirem muitos termos em inglês nas publicações da área, e também porque os termos utilizados para as publicações são diversificados, ou seja, não existe uma padronização de termos. Observou-se que os usuários

ficavam muito tempo pensando em qual poderia ser o termo sinônimo utilizado para um determinado assunto.

8 Elaboração da Taxonomia

Ferreira (1997) sugere que o planejamento de toda e qualquer atividade de informação seja feito de acordo com pesquisas centradas no indivíduo, partindo-se de uma perspectiva cognitiva, e buscando interpretar necessidades de informações tanto intelectuais como sociológicas.

Para a elaboração da taxonomia, foram utilizadas as informações advindas do estudo das necessidades de informação e do comportamento dos usuários de informação que estudam a biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais.

As categorias são contempladas na Teoria do Conceito e segundo Dahlberg (1978, p. 102) representam um conjunto de características gerais sobre um objeto. Toda vez que conceitos diferentes possuem características semelhantes existe relacionamento entre esses conceitos.

Dessa forma, o primeiro passo para a elaboração da taxonomia, foi criar uma categoria geral de assuntos, que se chamou **insetos frugívoros**. Posteriormente, os assuntos que possuíam características semelhantes foram agrupados, criando-se novas categorias, e então, embasando-se nas relações de gênero e espécie da Teoria do Conceito, foram criadas três categorias, que se relacionam diretamente como a categoria insetos frugívoros, sendo elas: manejo integrado de pragas, inventário de espécies e diversidade de espécies.

Depois de criadas todas as categorias de assuntos, e adicionados os assuntos pertencentes a cada categoria, e os sinônimos/variações dos termos na língua vernácula, foram acrescentados alguns termos equivalentes, em inglês, estes que são remissivos da taxonomia, mas adicionados na mesma, por serem muito utilizados nas publicações da área.

Tanto no pré-teste quanto nas entrevistas foram citadas espécies de moscas-das-frutas que não foram adicionadas na taxonomia porque há uma vasta quantidade de espécies, e como para se referir a uma espécie é necessário primeiramente se referir ao gênero, a taxonomia contemplou somente até o gênero.

Segundo Sager (1990, p.51), definição terminológica é um processo necessário para colocar um termo em sua posição, numa estrutura de conhecimento, e é obtido através de definições pré-existentes, contexto, ou através de consultas a especialistas.

As definições dos termos utilizados na taxonomia foram retiradas a partir do conceito dos próprios usuários e/ou de uma coletânea de termos técnicos da área de entomologia.

Segundo Campos (2004, p. 31), as teorias que têm por objetivo elaborar linguagens documentárias, não possuem um modelo de representação gráfica, mas em forma de lista endentada, com uma notação que, de certa forma, deixa evidente os grupos de termos afins.

O primeiro protótipo da taxonomia foi encaminhado para um dos pesquisadores (especialista) para conferência, e depois da correção de alguns detalhes, a taxonomia foi representada de duas maneiras: em forma de lista endentada, e em forma de árvore hierárquica.

Na taxonomia em forma de lista endentada, poderão ser vistas as categorias de assuntos, os termos que representam os assuntos pertencentes àquela categoria, a hierarquia das categorias e os sinônimos dos termos (quando possuírem), as definições desses termos e as relações entre as categorias. Os termos equivalentes em inglês, estes estão grafados em itálico.

A representação da lista endentada foi feita em três quadros, onde cada categoria diretamente ligada à categoria insetos frugívoros, foi apresentada em um quadro separado.

Na taxonomia em forma de árvore hierárquica foi possível expressar as categorias de assuntos e as relações dessas categorias (subordinações). Essa representação foi dividida em figuras separadas.

8.1 Representação da Taxonomia em forma de Lista endentada

Quadro 4 - Categoria Manejo Integrado de pragas

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
<u>INSETOS FRUGÍVOROS</u>	Frugívoros	Insetos que se alimentam de frutas ou vegetais	<i>Insect frugivorous; frugivorous flies;</i>
MANEJO INTEGRADO DE PRAGAS	controle de pragas; controle integrado.	Controle de pragas que envolvem uma ou mais técnicas de controle	<i>integrated pest management; pest control.</i>
Controle natural		Controle de pragas através da própria natureza	<i>natural control.</i>
Controle Biológico	Biocontrol, bio controle;	Controle de pragas através de inimigos naturais	<i>biological control.</i>
Inimigos naturais		Todos os organismos vivos que matam outros organismos vivos	<i>natural enemies.</i>
Predadores	entomófagos;	Inseto que se alimenta de outro inseto	
Parasitóides	endoparasitoides; exoparasitoides;	Parasita interno ou externo que mata o hospedeiro.	<i>parasitoids.</i>
Microorganismos patogênicos	entomopatógenos;	Microorganismos que causam doenças nas frutas	<i>entomopatogênicos microorganisms.</i>
Nemátodos	nemátodos.	Vermes geralmente microscópicos, finos e alongados que podem parasitar as plantas.	
Resistência de plantas a insetos (RPI)		Controle de pragas através da utilização de substâncias retiradas de plantas que possuem resistência natural contra determinados tipos de insetos	

Quadro 4 - Continuação

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
MANEJO INTEGRADO DE PRAGAS Controle natural	(continuação)	(continuação)	(continuação)
Infoquímicos	semioquímicos;	Substâncias químicas retiradas do próprio inseto	<i>infochemicals; semiochemicals</i>
Feromônios	ectohormônio; feromona.	Substância secretada por um animal que altera o comportamento do outro da mesma espécie	
Aleloquímicos		Substância sem valor nutricional, produzida e liberada por um organismo que transmite mensagens químicas entre espécies diferentes e afeta o crescimento, sanidade e comportamento de outra espécie.	
Cairomônios		Composto ou mistura de compostos químicos emitidos por um organismo que induzem uma resposta em um indivíduo de outra espécie, favorecendo o receptor.	
Sinomônio		Substâncias químicas produzidas por uma espécie e recebida por outra diferente, sendo ambas beneficiadas.	
Apneumônio		Composto químico proveniente de hospedeiros não vivos que atraem parasitóides.	
Alomônio		Substância que induz resposta em indivíduos de outras espécies e que favorece o emissor	
Controle Físico		Controle de pragas através de um método físico de supressão populacional.	<i>physical control</i>
Controle Químico		Controle de pragas através de produtos fitosanitários químicos (venenos)	<i>chemical control.</i>
Controle Cultural	tratos culturais	Controle de pragas por meio de limpeza do ambiente, ou através da destruição de resteva (restos de cultura).	<i>cultural control.</i>

Quadro 5 - Categoria Inventário de espécies

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
INVENTÁRIO DE ESPÉCIES	levantamento de espécie.	Estudo para se descobrir quais espécies existem naquela região.	<i>inventory</i>
Monitoramento		Acompanhamento ou avaliação sobre a existência de insetos numa determinada região.	<i>monitoring</i>
Armadilhas		Artifício (equipamento) utilizado para capturar insetos	<i>Trap; trap.</i>
McPhail	frasco caça-mosca	Recipiente de vidro transparente em forma de pêra, as partes da armadilha incluem uma rolha de borracha que sela a parte superior da armadilha e um cabide de arame para prender as armadilhas nos galhos das árvores.	
Amostragem de frutos		Monitoramento através da coleta de amostras de frutos para obter destas as espécies de insetos infestantes	
Questões quarentenárias	barreiras quarentenárias; questões econômicas.	Questões de importância econômica	
Flutuação populacional		Monitoramento através da descoberta de quantos indivíduos de igual espécie ocorrem no decorrer de um período.	<i>population fluctuation.</i>

Quadro 6 - Categoria Diversidade

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
DIVERSIDADE	biodiversidade	Variedade de espécies que compõem a comunidade biológica	<i>biodiversity; species diversity; species richness; diversity of insect.</i>
Tefritóidea	moscas frugívoras.	Superfamília dos dípteros mais evoluídos	<i>Tephritoidea.</i>
Tefritídeos	mosca-da-fruta; moscas-das-frutas.	Família de moscas-das-frutas.	<i>Tephritidae; fruit fly; fruits flies.</i>
Anastrepha.		Gênero de moscas-das-frutas	
Distribuição geográfica		Lugares onde são encontradas as espécies	
Hospedeiros		Recursos dos quais vivem os insetos	
Atrativo alimentar		Cheiro de substância protéica que atraem as moscas	
Acasalamento	Cópula	União de sexos	
Comportamento		função social observável, própria de organismos vivos, que se realiza em contato com o meio ambiente em que o ser vive.	<i>Behavior.</i>
Corte	Cortejamento, comunicação sonora; comunicação química, comunicação visual.	Ato de cortejamento para atrair o inseto.	

Quadro 6 – Continuação

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
DIVERSIDADE Tefritóidea Tefritídeos Anastrepha.	(continuação)	(continuação)	(continuação)
Análise Faunística	estudo da fauna; estudo faunístico; índice faunístico; diversidade faunística.	Relatos através de índices do estudo da fauna de uma determinada região	<i>faunistic study</i>
Frequência		Quantidade de vezes que ocorre uma determinada espécie nas repetidas amostragens	<i>Frequency</i>
Abundância		Porcentagem relativa de ocorrência (numero de indivíduos) de uma determinada espécie em relação a todas as espécies durante o inventário	<i>Abundance</i>
Riqueza de espécies		Número direto de todas as espécies obtidas de um grupo considerado	<i>Species richness</i>
Tabela de similaridade		Tabela que serve para indicar o grau de semelhança na estrutura de duas ou mais comunidades ou ecossistema	<i>Cluster analysis</i>
Densidade		Número de indivíduos de uma determinada espécie por unidade de área	<i>Density</i>
Dominância		Índice que serve para definir o grau de abundancia e frequência de uma determinada espécie.	<i>Dominance</i>
Distribuição espacial		Padrão de distribuição de uma espécie em um ecossistema podendo ser uniforme ou aleatório	<i>Demography</i>

Quadro 6 - Continuação

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
DIVERSIDADE Tefritóidea Tefritídeos	continuação	continuação	(continuação)
Ceratitis capitata.	Mosca do mediterrâneo	Gênero de moscas-das-frutas	
Distribuição geográfica		Lugares onde são encontradas as espécies	
Hospedeiros		Recursos dos quais vivem os insetos	
Atrativo alimentar		Cheiro de substância protéica que atraem as moscas	
Acasalamento	Cópula	União de sexos	
Comportamento		função social observável, própria de organismos vivos, que se realiza em contato com o meio ambiente em que o ser vive.	<i>Behaviour</i>
Corte	Cortejamento, comunicação sonoro; comunicação química, comunicação visual	Ato de cortejamento para atrair o inseto.	

Quadro 6 - Continuação

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
DIVERSIDADE Tefritóidea Tefritídeos Ceratitis capitata.	(continuação)	(continuação)	(continuação)
Análise Faunística	estudo faunístico; índice faunístico; diversidade faunística;	Relatos através de índices do estudo da fauna de uma determinada região	<i>faunistic study</i>
Frequência		Quantidade de vezes que ocorre uma determinada espécie nas repetidas amostragens	<i>Frequency</i>
Abundância		Porcentagem relativa de ocorrência (numero de indivíduos) de uma determinada espécie em relação a todas as espécies durante o inventário	<i>Abundance</i>
Riqueza de espécies		Número direto de todas as espécies obtidas de um grupo considerado	<i>Species richness</i>
Tabela de similaridade		Tabela que serve para indicar o grau de semelhança na estrutura de duas ou mais comunidades ou ecossistema	<i>Cluster analysis</i>
Densidade		Número de indivíduos de uma determinada espécie por unidade de área	<i>Density</i>
Dominância		Índice que serve para definir o grau de abundancia e frequência de uma determinada espécie.	<i>Dominance</i>
Distribuição espacial		Padrão de distribuição de uma espécie em um ecossistema podendo ser uniforme ou aleatório	<i>Demography</i>
Densidade		Número de indivíduos de uma determinada espécie por unidade de área	<i>Density</i>

Quadro 6 - Continuação

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
DIVERSIDADE Tefritóidea	(continuação)	(continuação)	(continuação)
Lonqueídeos	Moscas frugívoras	Família de mosca-da-fruta	
Neosilba.		Gênero da família dos Lonqueídeos	
Distribuição geográfica		Lugares onde são encontradas as espécies	
Hospedeiros		Recursos dos quais vivem os insetos	
Atrativo alimentar		Cheiro de substância protéica que atraem as moscas	
Acasalamento	cópula	União de sexos	
Comportamento		função social observável, própria de organismos vivos, que se realiza em contato com o meio ambiente em que o ser vive.	<i>Behavior</i>
Corte	Cortejamento, comunicação sonora; comunicação química, comunicação visual.	Ato de cortejamento para atrair o inseto.	

Quadro 6 - Continuação

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
DIVERSIDADE Tefritóidea Lonqueídeos Neosilba	continuação	continuação	continuação
Análise Faunística	estudo faunístico; índice faunístico; diversidade faunística;	Relatos através de índices do estudo da fauna de uma determinada região	<i>faunistic study.</i>
Frequência		Quantidade de vezes que ocorre uma determinada espécie nas repetidas amostragens	Frequency
Abundância		Porcentagem relativa de ocorrência (numero de indivíduos) de uma determinada espécie em relação a todas as espécies durante o inventário	<i>Abundance</i>
Riqueza de espécies		Número direto de todas as espécies obtidas de um grupo considerado	<i>Species richness</i>
Tabela de similaridade		Tabela que serve para indicar o grau de semelhança na estrutura de duas ou mais comunidades ou ecossistema	<i>Cluster analysis</i>
Densidade		Número de indivíduos de uma determinada espécie por unidade de área	<i>Density</i>
Dominância		Índice que serve para definir o grau de abundancia e frequência de uma determinada espécie.	<i>Dominance</i>
Distribuição espacial		Padrão de distribuição de uma espécie em um ecossistema podendo ser uniforme ou aleatório	<i>Demography</i>

Quadro 6 - Continuação

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
DIVERSIDADE Tefritóidea Lonqueídeos	(continuação)	(continuação)	(continuação)
Dasiops		Gênero de família dos lonqueídeos	
Distribuição geográfica		Lugares onde são encontradas as espécies	
Hospedeiros		Recursos dos quais vivem os insetos	
Atrativo alimentar		Cheiro de substância protéica que atraem as moscas	
Acasalamento	cópula	União de sexos	
Comportamento		função social observável, própria de organismos vivos, que se realiza em contato com o meio ambiente em que o ser vive.	<i>Behavior</i>
Corte	Cortejamento, comunicação sonora; comunicação química, comunicação visual.	Ato de cortejamento para atrair o inseto.	

Quadro 6 - Continuação

Termos ou frases	Sinônimos	Definição	Termo equivalente em inglês
DIVERSIDADE Tefritóidea Lonqueídeos Dasiops	(continuação)	(continuação)	(continuação)
Análise Faunística	estudo da fauna; estudo faunístico; índice faunístico; diversidade faunística;	Relatos através de índices do estudo da fauna de uma determinada região	<i>faunistic study.</i>
Frequência		Quantidade de vezes que ocorre uma determinada espécie nas repetidas amostragens	Frequency
Abundância		Porcentagem relativa de ocorrência (numero de indivíduos) de uma determinada espécie em relação a todas as espécies durante o inventário	<i>Abundance</i>
Riqueza de espécies		Número direto de todas as espécies obtidas de um grupo considerado	<i>Species richness</i>
Tabela de similaridade		Tabela que serve para indicar o grau de semelhança na estrutura de duas ou mais comunidades ou ecossistema	<i>Cluster analysis</i>
Densidade		Número de indivíduos de uma determinada espécie por unidade de área	<i>Density</i>
Dominância		Índice que serve para definir o grau de abundancia e frequência de uma determinada espécie.	<i>Dominance</i>
Distribuição espacial		Padrão de distribuição de uma espécie em um ecossistema podendo ser uniforme ou aleatório	<i>Demography</i>

8.2 Representação da Taxonomia em forma de Árvore Hierárquica

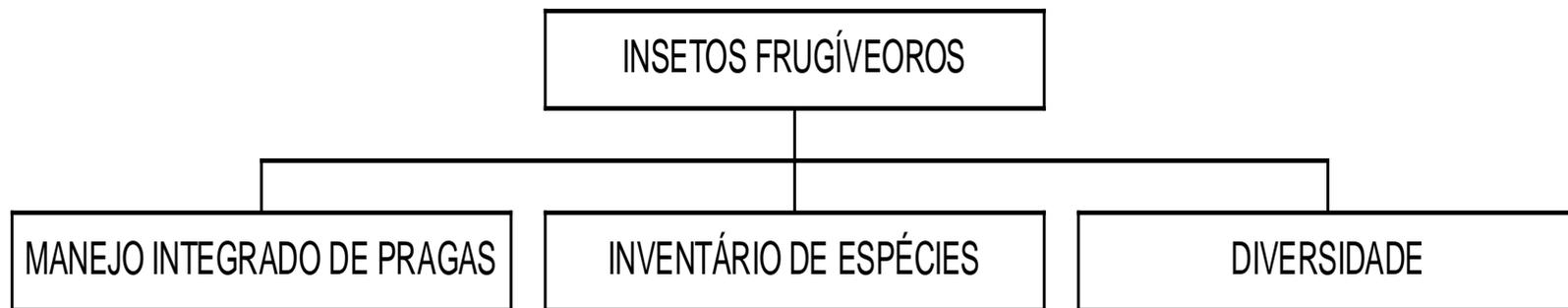


Figura 3 - Categoria Insetos frugívoros e suas subordinadas

Insetos frugívoros/Manejo integrado de pragas



Figura 4- Categoria Manejo Integrado e suas subordinadas.

Insetos frugívoros/Manejo integrado de pragas/controlado natural

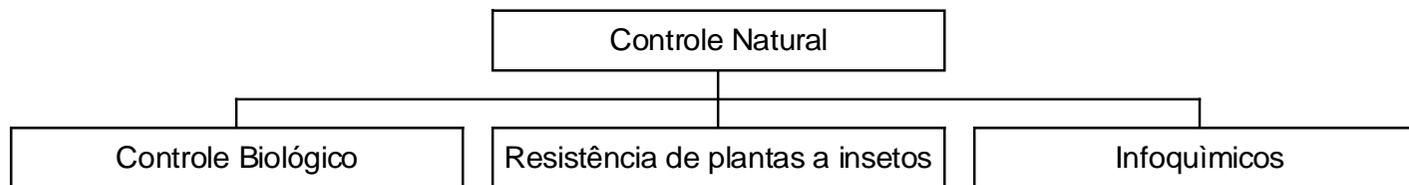


Figura 5 - Categoria controle natural e suas subordinadas.

Insetos frugívoros/Manejo integrado de pragas/Controle natural/controle biológico

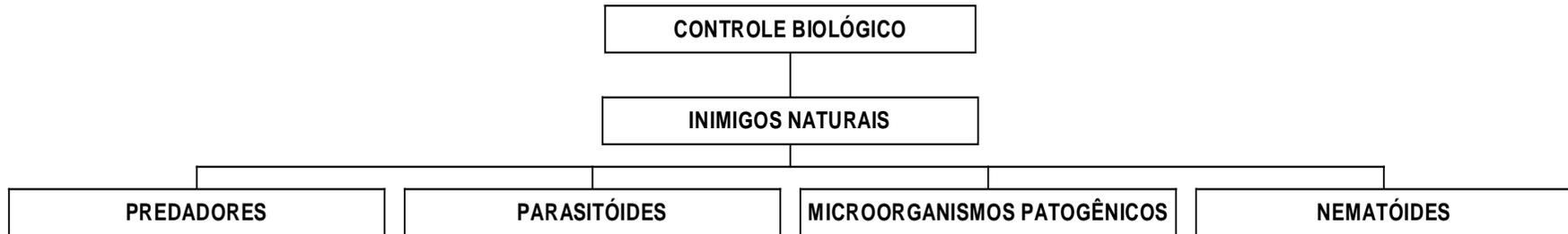


Figura 6 – Categoria Controle biológico e suas subordinadas

Insetos frugívoros/Manejo integrado de pragas/Controle Natural/Infoquímicos

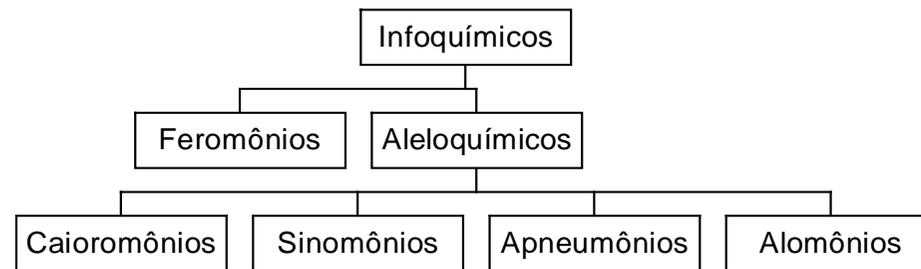


Figura 7 – Categoria Infoquímicos e suas subordinadas



Figura 8 - Categoria Inventário de espécies e suas subordinadas

Insetos frugívoros/Diversidade

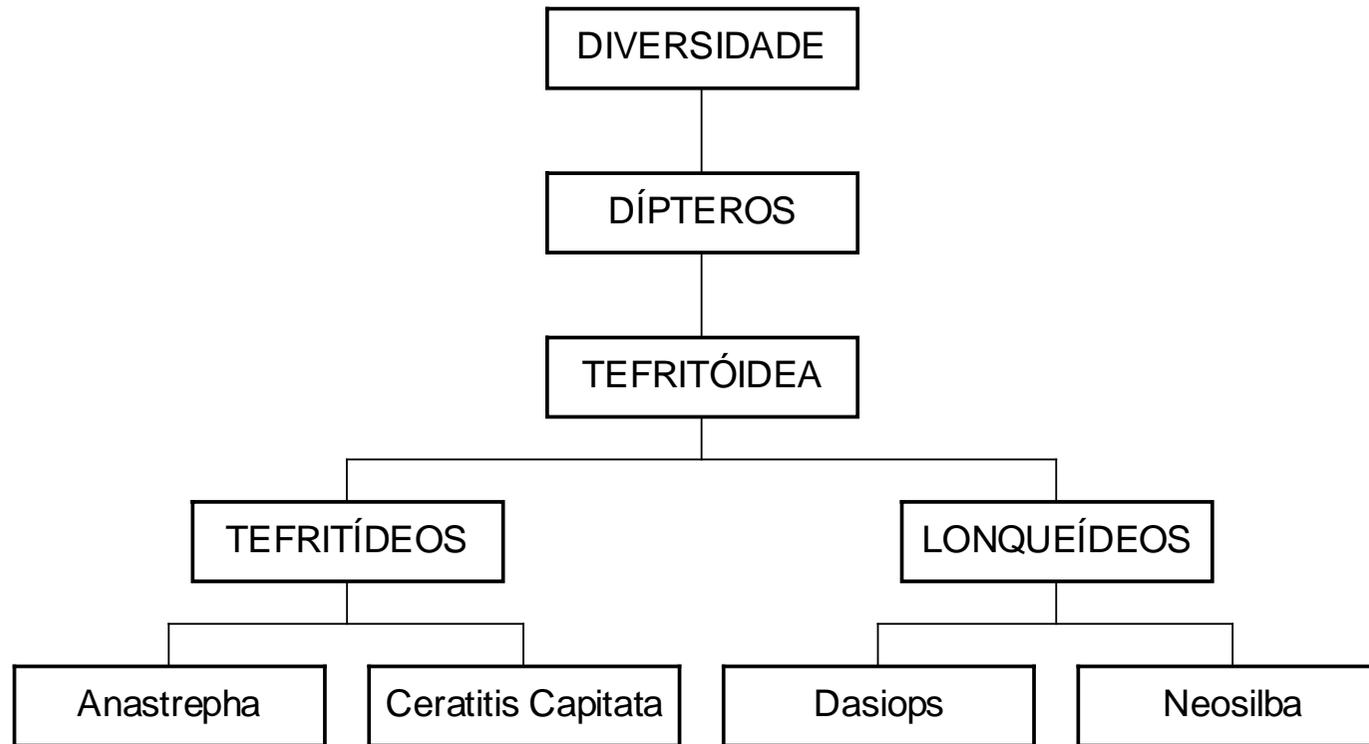


Figura 9 –Categoria Diversidade e suas subordinadas

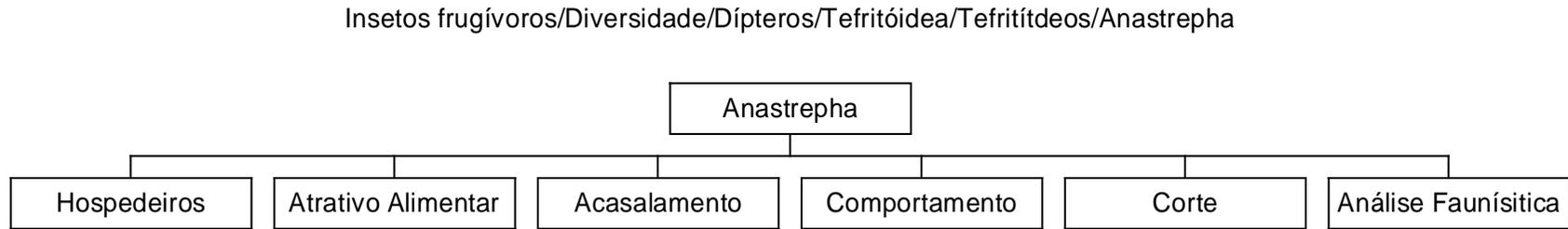


Figura 10 – Categoria Anastrepha e suas subordinadas



Figura 11 – Categoria Análise faunística subordinada da Anastrepha

Insetos frugívoros/Diversidade/Dípteros/Tefritóidea/Tefritídeos/Ceratitís capitata

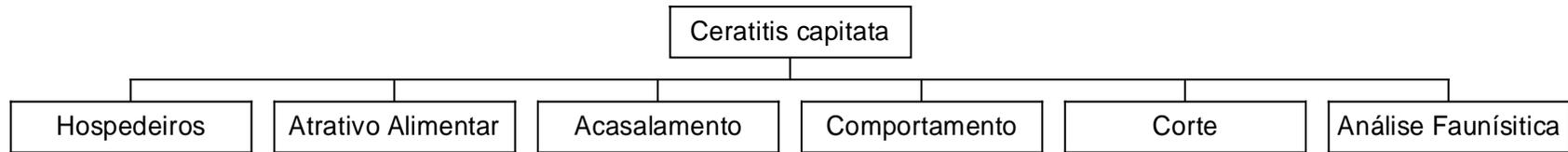


Figura 12 – Categoria *Ceratitís capitata* e suas subordinadas

Insetos frugívoros/Diversidade/Dípteros/Tefritóidea/Tefritídeos/Ceratitís capitata/Análise faunística

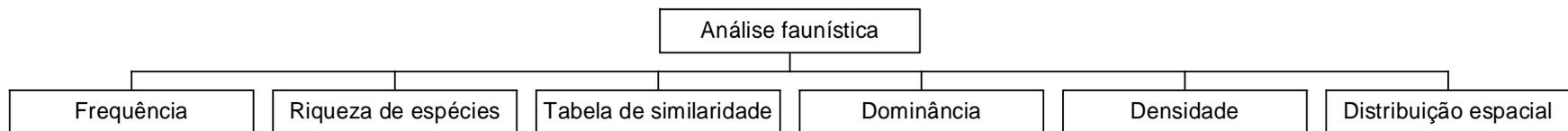


Figura 13 – Análise faunística subordinada da *Ceratitís capitata*

Insetos frugívoros/Diversidade/Dípteros/Tefritóidea/Lonqueídeos/Neosilba



Figura 14 – Categoria Neosilba e suas subordinadas

Insetos frugívoros/Diversidade/Dípteros/Tefritóidea/Lonqueídeos/ Neosilba/análise faunística

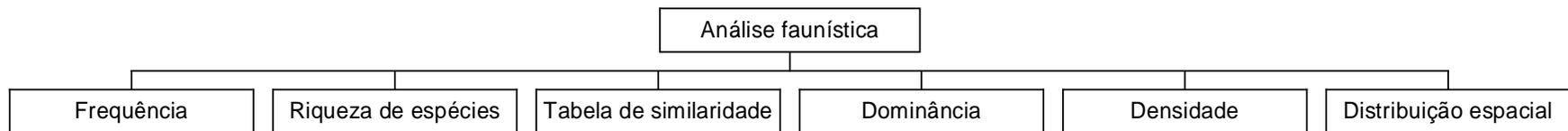


Figura 15 – Categoria Análise faunística subordinada da Neosilba

Insetos frugívoros/Diversidade/Dípteros/Tefritóidea/Lonqueídeos/Dasiops

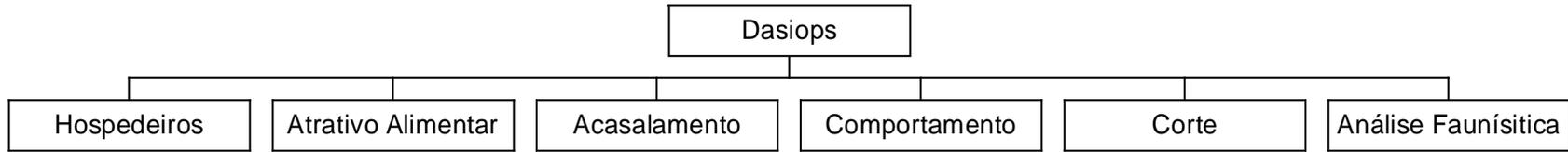


Figura 16 – Categoria Dasiops e suas subordinadas

Insetos frugívoros/Diversidade/Dípteros/Tefritóidea/Lonqueídeos/ Dasiops/análise faunística

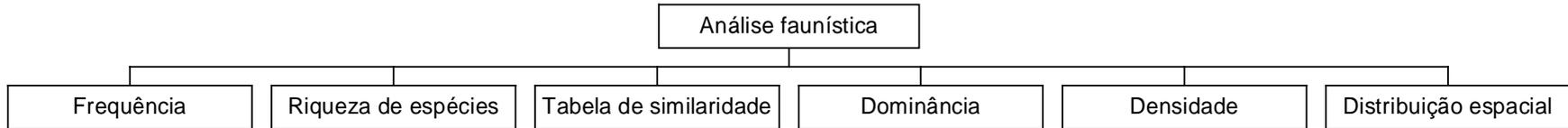


Figura 17 – Categoria Análise faunística subordinada da Dasiops

9 Considerações finais

Numa sociedade que valoriza o conhecimento e a informação, a organização desta tornou-se tão importante quanto a própria informação, pois à medida que houve um aumento no número de informações disponíveis, as dificuldades de recuperação também aumentaram.

As questões que nortearam esta pesquisa são ligadas às dificuldades de recuperação da informação, esta que é uma das preocupações da Ciência da informação, portanto, buscou-se na mesma, subsídios para tentar minimizar esses problemas.

A pesquisa foi desenvolvida junto a um grupo de pessoas que estudam a biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais, do mestrado de Entomologia da Universidade Federal da Grande Dourados.

Qualquer tentativa de descrever padrões de busca de informação deve admitir o indivíduo como o centro do fenômeno e considerar a visão, necessidades, opiniões e danos desse indivíduo como elementos significantes e influentes que merecem investigação (FERREIRA, 1995).

Desta forma, através desta pesquisa buscou-se responder a três questões que foram fundamentais para a elaboração da taxonomia.

Para a primeira questão, buscou-se identificar quais as necessidades de informação dos usuários que estudam a Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais, através de entrevistas.

Para a segunda questão, procurou-se identificar o comportamento de busca de informação desses usuários, através de observação, e, para responder à terceira questão, buscou-se descobrir os termos utilizados nas buscas por informação e as relações desses

termos, utilizando-se tanto a entrevista, como a observação, para que posteriormente, fosse elaborada uma Taxonomia para organização das informações.

Após a coleta e análise dos dados, identificou-se que os usuários que estudam a biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais possuem dificuldades no momento da recuperação da informação, principalmente devido à grande quantidade de termos em inglês utilizados na produção científica da área de Entomologia.

O comportamento de busca dos usuários é diferenciado: cada um possui uma estratégia diferente para a recuperação das informações, sendo que os que possuem mais experiência com os termos utilizados na indexação dos assuntos, realizam suas buscas com mais facilidade do que os que não conhecem os termos.

Os usuários de informação que estudam a biodiversidade de insetos frugívoros quando realizam suas buscas com termos que não são conhecidos, recuperam informações que não são satisfatórias e acabam perdendo muito tempo à procura de termos sinônimos para a realização de novas buscas.

Os mecanismos mais utilizados pelos usuários de informação, que estudam a biodiversidade de insetos frugívoros, são os de busca da internet, mas os usuários acham que esses mecanismos retornam muita informação irrelevante e de origem duvidosa.

Apesar de servir-se dos mecanismos de busca da internet, a biblioteca convencional continua sendo a fonte segura de informação para os usuários que estudam a Biodiversidade de insetos frugívoros.

9.1 Considerações sobre a Taxonomia

A elaboração da Taxonomia foi realizada manualmente, conforme uma das maneiras citadas por Holgate (2004).

Como não havia nenhuma taxonomia pré-definida, algumas informações pertinentes ao estudo da biodiversidade de insetos frugívoros podem ter ficado fora da taxonomia devido ao curto espaço de tempo que houve para o seu desenvolvimento, mas como a taxonomia é um mecanismo dinâmico (à medida que é utilizada, pode ser evoluída) e passível de alteração, novos conceitos ou mudanças de conceitos poderão ser adicionados.

A taxonomia elaborada poderá ser utilizada para organizar as informações sobre a Biodiversidade de insetos frugívoros, seus hospedeiros e inimigos naturais, padronizando-se os termos para as publicações da área. Também poderá ser utilizada para organização das informações antes de serem adicionadas em sistemas baseados em computador, pois segundo Lederman (2005), é essencial que os softwares tenham a capacidade de mostrar os resultados das buscas através de um vocabulário estruturado, devendo-se prever a inserção de novos termos e a combinação de categorias de termo para que as informações recuperadas sejam relevantes.

9.2 Perspectivas para novos estudos

No decorrer deste estudo surgiram algumas conclusões que servirão para futuras pesquisas, tais como:

a) continuidade a este estudo, pois a taxonomia segundo Graef (2001) é composta pelas seguintes fases:

- Criação de uma lista de termos (o que o autor chama de vocabulário); estes termos devem descrever o conteúdo a ser organizado;
- Adicionar relacionamento entre os termos, além de definições e notas;
- Conectar os termos com a fonte de informação. Este terceiro passo é o que o autor chama de etapa Aplicação.

Este estudo contemplou somente as duas primeiras fases da taxonomia de Graef (2001), ficando a terceira fase (conexão dos termos com uma fonte de informação) para trabalhos futuros.

b) Analisar se a utilização da taxonomia pelo grupo de pesquisadores que estudam a biodiversidade de insetos frugívoros promoveu mudanças na terminologia utilizada nas suas publicações.

c) Para elaborar a taxonomia, considerou-se o embasamento teórico tanto na teoria do conceito quanto na terminologia, e segundo Campos (2004), a Ciência da informação apesar de possuir teorias bem fundamentadas a respeito dos conceitos e das relações conceituais, é pobre em modelos que auxiliem a sua representação gráfica. Durante este estudo buscou-se uma maneira de se representar graficamente a taxonomia, porém não foi

encontrado um modelo que a representasse em sua totalidade. Para trabalhos futuros, poderá ser estudado como representar graficamente os conceitos e as suas relações.

REFERÊNCIAS

ALMEIDA, M.B. **Roteiro para a construção de uma ontologia bibliográfica através de uma ferramenta automatizada. Perspectivas em ciência da Informação**, Belo Horizonte, v.8 n.2. julho/dezembro de 2003. Disponível em < www.eci.ufmg.br/mba/text/art_ontobibliog_sub1_WEB.pdf>. Acessado em março de 2005.

ALVARENGA, Lídia Moreira; OLIVEIRA, Alcione Paiva. **O nível do conhecimento e os instrumentos de representação: tesauro e ontologias**. DataGramaZero – Revista de ciência da informação, Brasília, v. 5, n.6, Dez. 2004. Disponível em < www.dgzero.org/dez04/Art_01.htm >. Acessado em março de 2005.

ANWAR, MOHAMMAD; DING, WEI. The Semantic Web – **Semantics for data on the Web**. Disponível em < <http://pine.lc.cl.uh.Edu/ding/6340/theSemanticWeb.pdf>>. Acessado em 21/03/05.

BAPTISTA A. A; MACHADO, A. B. **Um gato preto num quarto escuro – falando sobre metadados**. Revista de biblioteconomia de Brasília, v. 25, n.1, p.77-90 jan/jun 2001. Disponível em < www.unb.br/fa/cid/rbb/25012001/alice.pdf>. Acessado em 20 de abril de 2005.

BERNERS-LEE, Tim e outros. **Semantic Web Development Proposal**. Fevereiro 2000. Disponível em < www.org/2001/01/sw/DevelopmentProposal>. Acessado em 04 de março de 2005.

BERNERS-LEE, Tim. **Metadata Architecture. W3C**, janeiro de 1997. Disponível em <www.w3c.org/designissues/metadata.html >. Acessado em 16 de maio de 2005.

BERNERS-LEE. **“What the Semantic Web Represent.”** Disponível em <<http://www.w3.org/designIssue/RDFnot.html>>. Acessado em 16 de maio de 2005.

BERNERS-LEE, Tim. **Network Working Group.** Disponível em <<http://ietf.org/rfc/rfc2396.txt> >. Acessado em 17 de maio de 2005.

BORBINHA, José, L. **Elementos do Núcleo de Metadados “Dublin Core”, Versão 1.1: Descrição de Referência.** Biblioteca Nacional de Portugal em 18 de Maio de 2000. Disponível em <<http://purl.pt/201/1/#rfc2413>>. Acessado em 16 de maio de 2005.

BORBINHA, José; Freire, Nuno. **Relevância para Bibliotecas e Organizações Relacionadas.** Biblioteca Nacional de Portugal, 2002. Disponível em <<http://metadados.bn.pt> >. Acessado em 16 de maio de 2005.

BORST, W. N. **Construct of engeneering ontologies for Knowledge Sharing and Reuse. Thesis.** 1997. Disponível em <<http://utwent.nl/find/1392> >. Acessado em 28 de junho de 2005.

BUZZI, Z, J. **Coletânea de Termos Técnicos de Entomologia.** Curitiba: Editora da UFPR, 2003.

CAFÉ, L; ARELLANO, M. A. M; BARBOZA, E. F; MELO, B; LAGE, M. B; MENDES, E. **Arquivos Abertos: Inovação Para A Comunicação Científica Na Rede.** XXV Congresso

Brasileiro de Ciências da Comunicação. Salvador, 2002. Disponível em <www2.metodista.br/unesco/JBCC/jbcc183.htm>. Acesso em 11/03/2006.

CAMPOS, A.T. **Linguagens documentárias**. Revista de Biblioteconomia de Brasília, v. 14, n.1, 1986.

CAMPOS, M. L. **Perspectivas para o estudo da área de representação da informação**. Ciência da Informação, Brasília, v.25, n.2, 1995. Disponível em <<http://www.ibict.br/cionline/viewarticle.php?id=482&layout=html>>. Acessado em dezembro de 2005.

CAMPOS, M. L. **Modelização de domínios de conhecimento: uma investigação de princípios fundamentais**. Ciência da informação, Brasília, vol.33, n.1, 2004. Disponível em<<http://www.ibict.br/cionline/viewarticle.php?id=77&layout=html>>. Acessado em abril de 2005.

CASTELL, Manuel. **A era da informação: economia, sociedade e cultura**. In: A sociedade em rede. São Paulo: Paz e Terra, 2000. v. 1.

CONWAY, Susan; SLIGAR, Char. **Unlocking Knowledge Assets**. Redmond: Microsoft Press, 2002. Disponível em <www.microsoft.com/mspress/books/sampchap/5516.asp>. Acessado em setembro de 2005.

COSTA, Sely M. S. **A comunicação científica nos dias atuais: impacto de uma “filosofia aberta”**. Disponível em <http://www.reacao.com.br/programa_sbpc57ra/sbpccontrole/textos/selycosta.htm>. Acessado em outubro de 2005.

DACONTA, M.C., Obrst, L., SMITH, T. **The Semantic Web: A Guide to the Future of XML, Web Services, and knowledge Management**, Wiley Publish, Inc. 2003.

DAHLBERG, Ingetraut. **Knowledge Organization**, agosto de 2000. Disponível em <<http://index.bonn.iz-soz.de/~sigel/ISKO/ko-def.dahlberg.html>>. Acessado em 01 de junho de 2005.

DAHLBERG, Ingetraut. **Teoria do Conceito**. Ciência da informação. Rio de Janeiro, v.7, n.2, p. 101-107, 1978.

DAHLBERG, Ingetraut. **TEORIA DA CLASSIFICAÇÃO, ONTEM E HOJE**. Anais. Brasília, IBICT/ABDF, 1979. v. 1, p. 352-370. Disponível em <http://www.conexaorio.com/bit/dahlbergteoria/dahlberg_teoriam.htm>. Acessado em abril de 2006.

DAVENPORT, T. H; PRUSAK, L. **Ecologia da Informação**. Traduzido por Bernadete Siqueira Abraão. São Paulo: Futura, 1998.

DATE, C. J. **Introdução a sistemas de Banco de dados**. Tradução [da 7ª edição Americana]. Vanderberg Dantas de Souza, Rio de Janeiro: Campus, 2000.

DCH/CEUD, Comissão. **Breve historio do Campus de Dourados**, dezembro de 2000. Disponível em < www.ceud.UFGD.br >. Acessado em 20 de agosto de 2005.

EDOLS, LIZ. **Taxonomies are what?.** Free print, n 97. outubro de 2001. Disponível em <ww.freepint.com/issues/0410010.htm#future>. Acessado em outubro de 2005.

FERNEDA, EDBERTO. **Recuperação de Informação:** Análise sobre a contribuição da ciência da computação para a ciência da informação. Tese de doutorado em ciência da comunicação/USP, 2003. Disponível em < [www.teses.usp.br /teses/disponiveis/27/27143/tde15032004130230/publico/Tese.pdf](http://www.teses.usp.br/teses/disponiveis/27/27143/tde15032004130230/publico/Tese.pdf)>. Acessado em agosto de 2005.

FERREIRA, Sueli M. S. P. **Novos paradigmas e novos usuários de informação.** Ciência da informação, Brasília, v. 25, n. 2, 1995. Disponível em < www.ibict.br/cionline/include/getdoc.php?id=819&article=481&mode=pdf>. Acessado em janeiro de 2006.

FERREIRA, Sueli. M. S. P. Ferreira. **Design de biblioteca virtual centrado no usuário: a abordagem do sense making para estudo de necessidades e procedimentos de busca e uso da informação.** Ciência da informação, v.26, n.2, 1997. Disponível em < http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0100-19651997000200014&lng=en&nrm=iso&tlng=pt>. Acessado em novembro de 2005.

FREITAS, F.L.G. de. **Ontologias e a Web Semântica.** Capturada em 16 de março de 2005. Disponível em < gpia.ucpel.tche.br/~Ipalazzo/Ontolog/Ontologia_web_semantica.pdf>. Acessado em julho de 2005.

FREITAS, F.L.G. de; BITTENCOURT, G. **Cognitive. Multi-Agent System for Integrated Information Retrieval and Extraction over the Internet**. Proceedings of the International Joint Conference IBERAMIA ' 2000 (Ibero-American Artificial Intelligence Conference) SBIA '2000 (Brazilian Artificial Intelligence Symposium), Lecture Notes in Artificial Intelligence, Springer Verlag, Vol 1052 (ISBN 3-540-41276), pp. 310-319, Atibaia, São Paulo, Brazil, November 19-22, 2000.

FREITAS, H; KLADIS, A. C. **Da informação à política informacional das organizações: um quadro conceitual**. São Paulo: RAP, v.29, n.3, junho-setembro de 1995.

GARCIA, R. M; SILVA, H.C. **O comportamento do usuário final na recuperação temática da informação: um estudo com pós-graduandos na UNESP de Marília**. *Revista de Ciência da Informação - DataGramZero*, v.6. n.3. Junho de 2005. Disponível em < www.dgzero.org/jun05/Art_02.htm >. Acessado em Agosto de 2005.

GARVIN, DAVID, in: Harvard Busines Review. **Gestão do conhecimento**. Rio de Janeiro: Campus, 2001.

GASQUE, Kelley. G. D; COSTA, Sely. M. S. **Comportamento dos professores da educação básica na busca da informação para formação continuada**. *Ciência da informação*, Brasília, v. 32, n. 3 , 2003.

GILLILAND-SWETLAND, Anne J. Setting the Stage. In: **Introduction to Metadata: Pathways to digital information**, 2000. Disponível em < <http://www.getty.edu>

/research/conducting_research/standards/intrometadata/2_articles/index.html>. Acessado em agosto de 2005.

GOODALL, George. **Business Taxonomies and Bibliographic objectives**, London, 2003. Disponível em <www.deregulo.com/facertation/pdfs/businesstaxonomies_goodall.pdf>. Acessado em 19 de julho de 2005

GRAEF, J. L. **Managing taxonomies strategically**. Montague Institute Review . 2001. disponível em <http://www.montague.com/abstracts/taxonomy3.html>>. Acessado em 17/03/2006.

GRUBER, T, R. Toward. **Principles for the Design of Ontologies Used for Knowledge Sharing**. In: International Journal Human-Computer Studies 43, p. 907-928. Padova, Italy,

GUARINO, Nicola. **Understanding, Building, and Using Ontologies**. LADSEB-CNR, Padova, Italy, outubro de 1996.

HILLMAN, Diane. **Using Dublin core – Documents**. Disponível em <<http://www.dublincore.org/documents/2001/04/12>>. Acessada em 19 de maio de 2005.

HODGE, Gail. **Metadata Made Simpler**. Bethesda, Niso Press 2001. disponível em <www.niso.org/news/metadata_simpler.pdf>. Acessado em 20 de maio de 2005.

HOLGATE, Lisle. **Creating and Using Taxonomies to Enhance Enterprise Search**. H. W. Wilson Web, 2004.

INGWERSEN, P. **Information Retrieval Interaction**. London: Taylor Graham, 1992. Disponível em <http://www.db.dk/pi/iri/files/Ingwersen_iri.pdf> Acessado em 28 de maio de 2005.

LANCASTER, F.W. **Indexação e Resumos: teoria e prática**; tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 1993. 347 p.

LEDERMAN, Paula. **Implementing a Taxonomy Solution**. AIIM E-DOC Magazine, março/abril de 2005. Disponível em < www.edocmagazine.com/archives_articles.asp?ID=29656>. Acessado em 26 de junho de 2005.

LOPES, L. Ilze. **Estratégia de busca na recuperação da informação: revisão da literatura**. Ciência da informação, v. 31, n. 2, pág 60-71. maio/agosto de 2002. Disponível em < <http://www.ibict.br/cienciadainformacao/viewarticle.php?id=191> >. Acessado em setembro de 2005.

MACCARI, E. A; Rodrigues. L.C. **Gestão do conhecimento em instituições de ensino superior**. R. Negócio, V.8, n.1, p.79-94, abril/ junho de 2003.

MELO, Ivo Soares. **Administração de sistemas de informação**. São Paulo: Thompson, 1999.

METADATA **Reference Guide**. Massachusetts: Massachusetts Institute of Technology, 2004. Disponível em <http://libraries.mit.edu/guides/subjects/metadatas/standards.html>>. Acessado em 23 de maio de 2005.

MOREIRA, A. ALVARENGA, L. OLIVEIRA, A.P. **O nível do conhecimento e os instrumentos de representação: tesouros e ontologias**. Datagramazero v.5. n.6 dez/2004. Disponível em <http://www.dgzero.org/dez04/Art_01.htm>. Acessado em julho de 2005.

MOREIRA, Alexandra. **Uso de uma Ontologia em Sistemas de informação computacionais**, Perspect. Ciênc. Inf., Belo Horizonte, v.7, n.1, p. 49-60, jan/jun. 2002. Disponível em <www.eci.ufmg.br/pcionline/viewarticle.php?id=371>. Acessado em agosto de 2005.

MOREIRA, Alexandra; OLIVEIRA, Alcione. P. **Contribuição da terminologia na modelagem de sistemas computacionais**. Ciência da informação, v.6, n.5, outubro de 2005. Disponível em <http://www.dgzero.org/out05/Art_01.htm>. Acessado em Março de 2006.

NEHMY, R. M. Q; PAIM, Isis. **A desconstrução do conceito de “qualidade da informação”**. Ciência da informação, Brasília, v.27, n.2, p. 36-45, janeiro a abril de 1998.

NETCRAFT. **Web Server Survey**, 2005. Disponível em: <http://news.netcraft.com/archives/web_server_survey.html>. Acessado em 13 de junho de 2005.

NONAKA, I; TAKEUCHI, H. **Criação de conhecimento da Empresa**. Rio de Janeiro: Campus, 1997.

PAHLEVI, Said Mirza;KITAGAWA, Hiroyuki. **Converging Taxonomy context for topic-Focused Web Search**. Maryland, Journal of the American Society, v.56, n.2, Janeiro de

PALMER, Natanael; FRAPPAOLO, Carl. **The Value of Categorization**. The H.W. Wilson Company, Julho/Agosto, 2004. Pathways to Digital Information. Getty Research Institute, 2000 Disponível <http://gett.edu/research/conducting_research/standards_intrametadata>. Acessado em 17 de novembro de 2005.

SAGER, J. C. **A practical course in terminology processing**. Amsterdam/Philadelphia: J. Benjamins, 1990. 254 p.

SIGEL, Alexander. **The “ Knowledge Organization on Internet” Mini- FAQ**, ago. 2000. Disponível em < <http://index.bonn.iz-soz.de/~sigel/ISKO/wiss~org.faq.html>>. Acessado em setembro de 2005.

SANTOS, A, R. **Gestão do conhecimento: Uma experiência para o sucesso empresarial**. Curitiba: Universitária Champagnat. 2001.

SARACEVIC, Tefko. **Ciência da informação: origem, evolução e relações**. Perspectivas em ciência da informação. Belo Horizonte, v.1, n.1, p. 41-62. jan/jun. 1996. Disponível em <www.eci.ufmg.br/pcionline/include/getdoc.php?id=393&article=5&mode=pdf>. Acessado em junho de 2005.

SOUZA, R. ROCHA.; ALVARENGA, L. **A Web Semântica e suas contribuições para a ciência da informação**. Ciência da informação, Brasília, v. 33, n.1, 2004.

STAIR, R.M. **Princípios de sistemas de informação**. Rio de Janeiro: Livros Técnicos Científicos - LTC, 2002.

TARAPANOFF, Kira. Introdução. In: Tarapanoff, Kira, Org. **Inteligência Organizacional competitiva**, Brasília: Editora da Universidade de Brasília, 2001.

TAYLOR, Kit Sims. **The Brief Reign of the Knowledge Worker: Informations Tecnology and Tecnological Unemployment**. In: International confrence on the social impact of information technologies. St. Louis, 1998. Disponível em <<http://online.bcc.ctc.edu/econ/kst/briefReign/Brwebversion.htm>> acessado em de maio de 2005.

TERRA, Schoueri; Michely Vogel. **Taxonomia: Elemento fundamental para a Gestão do conhecimento**. Terra Fórum, 2004. Disponível em www.terraforum.com.br>. Acessado em agosto de 2004.

Thulasi K; Lohrii Kaini Mahemei; Rajashekar.T.B. **Approaches to Taxonomy Development: Some Experiences in the Context of an Academic Institute Information Portal. Indian Institute of Science**. 2001. disponível em <<http://eprints.iisc.ernet.in/archive/00002853/01/icim-paper1.pdf>>. Acessado em março de 2006.

TURBAN, Efraim. **Administração de sistemas de informação**. Rio de Janeiro:Campus 2003.

UNIVERSITY OF CALIFORNIA AT BERKELEY, How much Information? 2003, 2003. Disponível em <<http://sims.berkeley.edu/research/projetc/how-much-info-2003/index.htm>> acessado em 14 de junho de 2005.

WASON, Thomas D. **Dr. Tom's Taxonomy guide: Description, use and Selections**. IMS Global Learning Consortium, Inc., 2000. Disponível em <http://home.minspring.com/TDW/drtomtaxonomiesguidecas.html>>. Acessado em 19 de junho de 2005.

WEIL, Pierre. **A normose informacional**, Revista Ciência da Informação, Brasília, v.29, n.2, maio/agosto de 2000.

WIKIPÉDIA . **A enciclopédia livre**. Disponível em <<http://pt.wikipedia.org/wiki/>>. Acessado em abril de 2006.

WOODLEY, Mary S. **DCMI Glossary. Dublin Core Metadata Initiative**, 2003. Disponível em < <http://dublincore.org/documents/usageguide/glossary.shtml>>. Acessado em 17 de maio de 2005.

YAMAOKA, Eloy Juniti. **Taxonomia e Metadados como elementos estruturantes da Organização e Representação do Conhecimento numa Empresa**. 2005. Dissertação de Mestrado. Universidade de Brasília, Brasília, 2005.

Apêndice A –Esquema representativo da revisão da literatura

