

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Maria Aparecida Lourenço Santana

**A INDEXAÇÃO TEMÁTICA DE RECURSOS FUNDAMENTADA POR ESTRUTURA
PROFUNDA E ABORDAGEM OBJETO-RELACIONAMENTO**

Belo Horizonte

2005

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Maria Aparecida Lourenço Santana

**A INDEXAÇÃO TEMÁTICA DE RECURSOS FUNDAMENTADA POR ESTRUTURA
PROFUNDA E ABORDAGEM OBJETO-RELACIONAMENTO**

Dissertação apresentada ao Programa de Pós-Graduação, Curso de Mestrado da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Produção, organização e utilização da informação

Linha de pesquisa: Organização e uso da informação

Orientadora: Prof^a Madalena Martins Lopes Naves

Belo Horizonte

2005

Santana, Maria Aparecida Lourenço Santana

S232i A indexação temática de recursos fundamentada por estrutura profunda e abordagem objeto-relacionamento [manuscrito] / Maria Aparecida Lourenço Santana. – 2005.
165 f, il.

Orientadora: Madalena Martins Lopes Naves
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.
Bibliografia: p. 145-150
Glossário: p. 151-164
Anexos: p. 165

1. Sistemas de recuperação da informação – Teses 2. Ciência da informação – Teses 3. Semântica – Processamento de dados – Teses 4. Análise de assuntos – Teses 5. Linguagens de indexação – Teses I. Título II. Naves, Madalena Martins Lopes III. Universidade Federal de Minas Gerais. Escola de Ciência da Informação.

CDU 025.4

*Dedicado a Bruna, Anita e Almir –
“pedras fundamentais” na minha vida.*

Agradeço por esta dissertação às contribuições de

Professora Madalena Martins Lopes Naves – orientadora do mestrado,
dedicada e paciente,
pelas incansáveis correções e direcionamentos,

Professoras Lídia Alvarenga e Marlene de Oliveira – mestras brilhantes
que não me pouparam trabalhos e nem reconhecimentos,

Professor Marcello Bax – por ter lançado idéias fundamentais,

Todos os professores do mestrado, pelo compartilhamento de saberes,

Professor Márcio Augusto Gonçalves – mestre na UNA,
pelo incentivo na graduação,

Viviany, Lucimary e Goreth– secretárias do PPGCI, pela amizade atenciosa,

Eliane, Nádia e Gisele - profissionais da biblioteca da ECI, pelas inúmeras ajudas,

Moisés – funcionário da ECI, amigo paciente, a quem tantas
vezes atrasei a saída do serviço,

Turma do Curso de Especialização , pelos incentivos e
discussões enriquecedoras,

Cida Shikida, Julianne e Marina – amigas e incentivadoras durante o mestrado,

Elaine e Marcelo – amigos e incentivadores,

Ricardo de Oliveira – amigo e apazível debatedor,

Almir – esposo compreensivo e apoio desse mestrado,

Bruna – filha e companheira de madrugadas de estudo,

Anita – mãe amorosa, primeira pessoa cujo pensamento cativei e
que em mim depositou fé,

A Deus e toda sua Criação, Gigantes que me sustentaram, que me emprestaram os ombros;
ombros sobre os quais subi para empreender mais esta caminhada.

FOLHA DE APROVAÇÃO

SUMÁRIO

LISTA DE FIGURAS	9
LISTA DE QUADROS	10
RESUMO	11
ABSTRACT.....	12

PARTE I - Introdução

1	INTRODUÇÃO.....	14
1.1	Problema de pesquisa	16
1.2	Pressupostos e campo de pesquisa	19
1.3	Justificativa	20
1.4	Objetivos	21

PARTE II - Desenvolvimento

2	REVISÃO DE LITERATURA	24
2.1	Metarrepresentação	24
2.2	Análise de assunto	35
2.3	Linguagens de indexação	41
2.3.1	Controvérsias entre linguagens de indexação	46
2.3.2	Caracterizações das linguagens de indexação	50
2.3.2.1	Processos permutados	51
2.3.2.2	Processos lógicos	53
2.4	Estrutura profunda	63
2.5	Projeto de bases de conhecimento	71
2.5.1	Modelagem orientada ao objeto	72
2.5.2	Esquema de descrição de recursos - RDF	77
2.5.2.1	Modelo RDF	77
2.5.2.2	Sintaxe RDF	79
2.6	Sistemas de informação	83
3	METODOLOGIA	85
3.1	Método de modelagem	87
3.1.1	Modelo de representação de assunto	89
3.2	Método de análise de conteúdo	92
3.3	Estudos preliminares	95

3.4	Etapas da metodologia para o objeto empírico	99
4	A COLETA DE DADOS	102
4.1	Coleta de dados – fase 1	104
4.2	Coleta de dados – fase 2	114
5	ANÁLISE E DISCUSSÃO DOS RESULTADOS	119
 PARTE III - Conclusões e recomendações		
6	CONCLUSÕES E RECOMENDAÇÕES	139
 REFERÊNCIAS		
145		
 BIBLIOGRAFIA		
149		
 GLOSSÁRIO		
151		
 ANEXOS		
1	CD – Exemplos e casos estudados	165

LISTA DE FIGURAS

1	Triângulo semântico (Ogden e Richard).....	33
2	Diagrama de conceitos	34
3	Contexto da análise de assunto	36
4	Relações expressas pela indexação relacional de Farradane.....	58
5	Esquema da estrutura profunda em linguagem de indexação de assunto	70
6	Esquema explicativo da abstração de tipos.....	75
7	Esquema de tripla expressado em grafos	78
8	Exemplo de representação em tripla expressado em grafos	78
9	Esquema explicativo do modelo de representação proposto	90
10	Exemplificação da fatoração aplicada ao modelo proposto	97
11	Parte do código HTML do recurso Ausência_scheme	105
12	Parte do código fonte do recurso 1subject_arqu	107
13	Parte do código fonte de recurso no arquivo 3Subject_metadados.	108
14	Página do Google com a especificação do esquema que descreve o assunto	109
15	Parte do código fonte html que especifica o esquema de descrição do assunto – 4Subject_metaeschema	110
16	Janela normalmente encontrada para acesso a recursos indexados , arq. 7Subject_LCSH	112
17	Página de <i>web site</i> que organiza conteúdos pelas classes da CDU- 9Subject_Keyword.doc.....	113
18	Representação DEPA _m -OR no HiperEditor – Caso 1.....	121
19	Representação DEPA _m -OR no HiperEditor – Caso 2	125
20	Representação DEPA _m -OR no HiperEditor – Caso 3	127
21	Representação DEPA _m -OR no HiperEditor – Caso 4	129
22	Representação DEPA _m -OR no HiperEditor – Caso 5	131
23	Representação DEPA _m -OR no HiperEditor – Caso 6	133
24	Representação DEPA _m -OR no HiperEditor – Caso 7	135

LISTA DE QUADROS

1	Modelo do processo rotacional KWIC	52
2	Modelo do processo de deslocamento do PRECIS	53
3	Exemplo de enumeração de unidades diádicas (R _i , a,b)	57
4	Modelo da representação em dois contextos do PRECIS	60
5	Indicadores de função do POPSI	62
6	Equivalência entre processo de modelagem conceitual orientada ao objeto e estrutura profunda de linguagem de indexação – DEPAm	76
7	Exemplo de sintaxe RDF/XML	79
8	Campos de registro de sistemas de informação – fase 1.....	93
9	Campos de registro de sistemas de informação – fase 2.....	94
10	Fatoração de assunto – etapa 1.....	96
11	Fatoração de assunto – etapa 2	97
12	Combinação DEPAm, OO e RDF/XML	98

RESUMO

Este estudo foi motivado pelo problema da recuperação de recursos eletrônicos, quando são representados sem a hierarquia semântica que descreve seu campo conceitual. Na atividade do indexador percebeu-se a ausência de estruturas de representação semântica com potencial para processamento por máquinas e pessoas. A revisão bibliográfica foi feita a partir da pressuposição de que haveria estruturas semânticas ideais para representar o assunto dos recursos, tanto da área da Ciência da Informação quanto da Ciência da Computação, considerando duas abordagens teóricas diferentes. A primeira, oriunda da teoria geral da classificação de assunto propõe a estrutura profunda DEPAm (Disciplina, Entidade, Propriedade, Ação, modificador). A segunda, originária da modelagem orientada ao objeto, propõe o modelo ER (Entidade-Relacionamento), a representação por tripla e a descrição por esquema e sintaxe RDF/XML (Resource Description Framework / Extended Markup Language). Fundamentado sobre essas teorias propôs-se o modelo de indexação de assunto DEPAm-OR. Através da metodologia envolvendo modelos e análise de conteúdo, fez-se a coleta e a análise dos dados dos recursos eletrônicos disponibilizados por serviços de informação. A implementação da estrutura de Objeto Relacionamento-OR, pôde ocorrer a partir dos *softwares* livres HiperEditor e HiperNavegador. Concluiu-se que o modelo DEPAm-OR, proposto, é uma ferramenta importante. Seu uso foi recomendado pois permitiu a representação dos assuntos dos recursos de forma simples. Também concluiu-se que o modelo acrescentou-lhes grande explicitação semântica, determinada pelas convenções da estrutura profunda.

Palavras-chave: Sistemas de recuperação da informação – estruturas representação do assunto – estrutura profunda – análise orientada ao objeto - análise de assunto – cabeçalhos de assunto – linguagens de indexação.

ABSTRACT

This study was motivated by the problem of electronic resources recovering without the semantic hierarchy describing their conceptual field. The indexer activity was seen as the absence of semantic representation structure construction with some potential to be processed by machines and people. The bibliographical revision has been conducted from the assumption that there would be ideal semantic structures to represent those resources, both in the area of Information Science and in the area of Computer Science, considering two different theoretical approaches. The first, coming from the general theory of subject classification, proposes the deep structure DEPAm (Discipline, Entity, Property, Action, modifier). The second, coming from the object-oriented modeling, proposes the ER (Entity-Relationship) modeling, which involves a triple representation and a description based on scheme and syntax RDF/XML (Resource Description Framework/Extended Markup Language). Upon such theories, the subject indexing model DEPAm-OR was proposed. Through the methodology involving models and content analysis, it was made the data collection and the analysis of the electronic resources made available by information services. The implementation of the Object Relationship-OR structure was became possible though the free softwares, HiperEeditor and HiperNavegador, employment. It has been concluded that the proposed DEPAm-OR model is an important tool. Its usage is recommended since it allows a simple representation of the resource subjects. It also adds them a great semantic explicitness such as it is determined by the deep structure conventions.

Key Words: Information retrieval systems – subject indexing structures – deep structure – object-oriented analysis – subject analysis – subject headings – indexing languages.

PARTE I

INTRODUÇÃO

1 INTRODUÇÃO

Esta dissertação junta-se a outros estudos preocupados com as inovações ocorridas no final do século XX que, sob o escopo das tecnologias da informação, repensaram as estruturas e processos de representação e recuperação da informação. Procura estudar a representação do conhecimento, quando esse processo se põe em conexão com expansões das estruturas computacionais de modelagem de informações. Observa que, com a convergência das tecnologias de microeletrônica, de computadores e de telecomunicações, foram elevados para a casa dos milhões os microcomputadores conectados em rede. Esse fenômeno tem provocado novos estudos sobre o aproveitamento da capacidade dos computadores para as atividades de registro do conhecimento, usando modelos de representação com maior potencial de interconexão.

O tema é muito extenso e, por isso, procura-se cobrir o assunto delimitando-se uma de suas nuances. Os aspectos tratados foram desmembrados em quatro partes, que se subdividem em capítulos. A parte I, composta pelo capítulo 1 e suas subdivisões, procura delimitar a perspectiva de estudo, contendo a identificação do problema de pesquisa, o estabelecimento dos pressupostos teóricos, a justificativa da pesquisa e a definição dos objetivos a serem alcançados. Para auxiliar a leitura, foi incluído um glossário no final da tese com definições de conceitos utilizados no texto.

A parte II está subdividida em três capítulos que contêm o desenvolvimento do estudo. O capítulo 2 é dedicado à revisão de literatura, tópico que explora seis temáticas

interligadas pelo escopo do projeto: metarrepresentação, análise de assunto, linguagens de indexação, estrutura profunda, projeto de bases do conhecimento orientadas ao objeto e sistemas de informação.

O capítulo 3 trata da metodologia de pesquisa. Identifica a pesquisa de *sites* como objeto empírico para aplicação metodológica, formula as etapas do trabalho, elabora a pressuposição e descreve a coleta de dados.

O capítulo 4 trata da análise e discussão dos resultados. Na análise, aplica-se o modelo pressuposto ao conjunto analisado, através das quatro atividades propostas: determinação da atinência, fatoração do assunto, representação DEPAm-OR e descrição por RDF/XML. A discussão dos resultados passa pelas dificuldades e vantagens encontradas no estudo do modelo de indexação temática proposto.

Na parte III, composta pelo capítulo 5, foram descritas as conclusões e recomendações a que se chegou após a pesquisa. Ressaltam-se não só a viabilidade da aplicação, mas também o engajamento das teorias entre áreas e o amplo potencial que o modelo DEPAm-OR representa para domínios do conhecimento em que a precisão da descrição e a semântica explícita de representação e recuperação são fatores essenciais.

Expostos os elementos que compõem este estudo, inicia-se, a seguir, o desenvolvimento dos itens introdutórios.

1.1 Problema de pesquisa

No cenário da representação temática de recursos¹ e da recuperação de informações, o indexador² tem um papel primordial. Como responsável pela criação de registros que descrevem a representação, e considerando-se que esses registros detêm o resultado do processo de análise e de tradução de assuntos, não caberia a esse profissional uma real contribuição para a melhoria das formas de representar o conhecimento? E então, disponibilizando-se as suas competências para os tempos atuais, como utilizar o potencial dos computadores como ferramentas de representação e de recuperação?

As linguagens de indexação, importantes ferramentas de trabalho do indexador, têm sido instrumentos tradicionais de indexação temática. Tais sistemas possuem funcionalidades fundamentais para abordar o problema da representação de recursos no ambiente dos computadores, como, por exemplo, a possibilidade de coordenação de assuntos dos recursos informacionais *a priori*. Dois tipos de instrumento de coordenação são as ferramentas baseadas em linguagens simbólicas e linguagens verbais, que, juntas, formam a classe dos vocabulários controlados.

A representação do assunto dos recursos, utilizando vocabulários controlados, sem se levar em consideração as potencialidades de processamento, cruzamento e construção

¹ O termo “recurso” foi usado em substituição ao termo “documento”, por descrever um conjunto maior de objetos de informação materializados em diversos suportes, dentre os quais o papel e o eletrônico. Essa escolha segue as considerações da Internacional Federation Library Association – IFLA, no documento Functional Requirements for Bibliographic Description, de 1998, e do Anglo-American Cataloging Rules – AACR2, de 2002.

² É chamado de indexador o profissional da informação especializado em definir sobre o que trata o recurso, para fins de criar uma representação por assunto que sirva à localização ou agrupamento daquele recurso numa coleção ou universo maior de recursos. Pode-se afirmar que ele cria registros que descrevem a representação.

de associações novas e de dados pelos computadores, pode ser vista como um problema atual da representação temática de recursos. Para compreender essa questão, pode-se observar o formato descritivo de documentos, o metadados Dublin Core³. Nele há o campo “DC:subject”, ou seja, o assunto do documento. Na prática da representação de recursos, o indexador preenche esse campo com a frase: “Este documento trata de...”. A informação que complementa essa frase é o conteúdo do metadado “DC:subject”. Quando se depara com essa frase, um leitor humano está supostamente habilitado para encontrar sua significação e sentido. No entanto, um programa de computador, ao processar essa frase, não é capaz de construir uma boa rede semântica, mas apenas fazer casamento de padrões entre os termos de busca e os termos presentes na sentença do assunto do documento. Com isso, observa-se que a frase de indexação da forma linear, tal como expressa, não é uma estrutura que coopera semanticamente com programas de computador que fazem busca por casamento de padrões, se considerada especialmente a função de coordenação. A frase de indexação, neste estudo, é o termo preferido para expressar a sentença resultante da atividade do indexador, ou seja, reflete a análise e a expressão sobre o que é um recurso (atinação do recurso).

As estruturas computacionais para representação de conteúdos vêm se desenvolvendo para permitir adição sistemática de semântica aos documentos, como forma de obter resultados de busca mais relevantes. Nesse sentido, também a frase de indexação ou as palavras-chave, construídas sob uma sintaxe desestruturada para processamento

³ “Em uma conferência sobre metadados em 1995, realizada em Dublin, Ohio, um *workshop* discutiu o item ‘semântica’ e concordou com um conjunto principal de temas comuns na maioria dos tipos de documento. Esse

por máquinas, não cooperam com a adição de semântica útil para programas de computadores.

Então, como elaborar a representação de assunto para atingir potenciais de interpretação, inferência e geração de informações, de maneira que esses processos possam ser realizados por pessoas ou por programas de computador, na promoção de melhor recuperação de informações?

A sintaxe, como a forma de descrição, e a semântica, como o significado da descrição, embutidas nas linguagens de indexação, podem ser utilizadas na *web* para descrever o assunto de recursos digitais remotos?⁴

Tais questões refletem as preocupações que instigaram esta pesquisa, este estudo e a busca de respostas. Tomam-se estudos da teoria geral da indexação de assuntos como dispositivo teórico viabilizador de possibilidades de solução e pergunta-se: seriam as estruturas da teoria geral da indexação de assunto adequadas ao processamento e à autogeração de informações por programas de computador?

Que estrutura de representação em sistema de informação computadorizado, também adequada ao padrão de interpretação/inferência humano, serviria para suportar a representação do assunto, intercambiando a interpretação/inferência entre o

conjunto de propriedades se tornou conhecido como iniciativa Dublin Core (DC)” (AHMED *et al.*, 2001).

⁴Recurso digital remoto, conforme a AACR2 (2002), é aquele em meio digital, que não pode ser tocado e está inserido num computador local ou em um nó da rede.

processamento humano e a potencialidade de autogeração de informações por programas de computador?

Para elucidar tais questões, foram definidos alguns pressupostos, expostos na próxima seção.

1.2 Pressupostos e campo de pesquisa

Diante das preocupações expressas no item 1.1, esta pesquisa se constituiu a partir dos pressupostos de que:

- Existem, na teoria geral da classificação de assunto, estruturas que funcionam como dispositivos para representação do assunto e potencializadores de interpretação, inferência e autogeração de informações para pessoas, como também para processamento, cruzamento e autogeração de informações para programas de computador.
- É possível descrever o assunto, utilizando ferramentas computacionais descritivas para a *web*, de maneira a se obter maior potencial semântico, ou seja, potencial de interpretação, inferência e autogeração de informações para pessoas, como também para processamento, cruzamento e autogeração de informações por programas de computador.

Com tais preocupações e pressupostos, este estudo pode ser classificado na linha de pesquisa denominada organização e uso da informação, no campo de estudo da

recuperação da informação e no subcampo linguagens de indexação, delimitado pela abordagem da indexação temática de recursos.

1.3 Justificativa

Esta pesquisa se fundamenta na ontogeneidade da necessidade de representação do conhecimento, na valorização do conhecimento adquirido na ciência da informação sobre o tratamento temático de recursos e na necessidade de confrontação e casamento entre estudos em sistemas de classificação e estruturas computacionais, com o objetivo de representar informações de forma a conjugar esforços interdisciplinares, para atender a novas demandas em sistemas de informação.

Sob o enfoque central do objeto informação, na perspectiva da ciência da informação, o homem, ontologicamente, tem buscado mecanismos de registro do conhecimento de forma que, no seu tempo, e historicamente marcados, esses mecanismos atendam às suas necessidades informacionais, que se acirraram nos tempos atuais, com o uso intensivo de tecnologias de informação computadorizadas.

Além disso, ao retomar temas centrais para a área, como é o caso das linguagens de indexação e os processos de tratamento temático de recursos pelo indexador, este estudo se justifica como de interesse contemporâneo, pois não se atém a descrever tradicionalmente processos e usos, mas os confronta teoricamente com novas ferramentas tecnológicas e novos mecanismos de suporte físico.

Além disso, justifica-se esta pesquisa pela necessidade da área de indexação de informação de aliar seus desenvolvimentos a novas configurações de estruturas computacionais. O casamento da indexação temática com estruturas computacionais orientadas para objetos acena com fortes possibilidades de prover soluções para problemas de recuperação de informações tornados urgentes, em face de altos índices de recuperação de recursos contra baixas medidas de precisão.

Assim, este estudo pretende ser mais uma contribuição para o desenvolvimento do conhecimento científico sobre a representação de informações na área da ciência da informação e teve seu desenvolvimento orientado para a consecução dos objetivos especificados a seguir.

1.4 Objetivos

O desenvolvimento desse projeto objetivou, sob o aspecto geral, propor uma forma de representação do assunto que conjugasse esforços intersectoriais entre a ciência da informação e a informática, a fim de, através da descrição temática dos recursos e da representação por estrutura orientada ao objeto, obter-se maior potencial de interpretação e inferência humano, a fim de clarear a possibilidade de processamento por programas de computador dos campos semânticos expressados.

Buscar atingir esse objetivo amplo, significou perseguir pequenos objetivos, que foram:

- Compreender mecanismos de registro do conhecimento através da identificação de estruturas subjacentes que compõem diversos tipos de esquema de classificação de assunto.
- Evidenciar estruturas de informação e processos de descrição conceitual para fins de se obterem melhores potenciais de representação semântica, de forma a melhorar o atendimento de demandas de usuários.
- Evidenciar um modelo de representação temática amparado por teorias da classificação da ciência da informação e estruturas de descrição da ciência da computação, o que representa uma interseção das duas áreas disciplinares com fortes interfaces para assuntos de recuperação de informações.

O aporte teórico para o desenvolvimento e o alcance dos objetivos menores, ou específicos, baseou-se em estudos de seis temáticas principais: metarrepresentação, análise de assunto, linguagens de indexação, estrutura profunda, projeto orientado para o objeto e sistemas de informação, descritos na subsequente revisão de literatura.

PARTE II

DESENVOLVIMENTO

2 REVISÃO DE LITERATURA

Dentre as temáticas compreendidas como necessárias para a fundamentação e a compreensão teórica desta pesquisa, estudou-se primeiramente a literatura sobre a metarrepresentação; em segundo lugar sobre a análise de assunto; em terceiro, sobre as linguagens de indexação; em quarto, sobre as estruturas profundas; em quinto, sobre a modelagem orientada ao objeto; e, em sexto lugar, localizou-se o contexto de congruência de todos os tópicos nos sistemas de informação.

2.1 Metarrepresentação

A metarrepresentação é um conceito elaborado por Dodebei (2002)⁵ e seu encaixe teórico é explicado através da construção progressiva dos conceitos de conhecimento, conhecimento registrado e representação, chegando-se então à metarrepresentação. Todos esses elementos são vistos como inter-relacionados pelo elemento conceitual.

A questão da representação insere-se no que se percebe como representado e pode ser classificada em diversos níveis. O primeiro nível de representação é atribuído ao que é percebido no mundo real, isto é, ao conjunto de coisas que podem ser observadas, tocadas e sentidas. Sob essa abordagem, os objetos passariam pela primeira etapa de representação, quando são percebidos e assimilados pelas pessoas em suas mentes, fato a que se dá o nome de representação mental.

Assim, qualquer objeto percebido seria associado a uma representação mental pela pessoa. Através de processos cognitivos de associação e reconstrução de significação, a representação mental é transformada em conceitos no cérebro de cada pessoa. Esse sujeito, de posse de suas construções conceituais e um sistema social de linguagem, enuncia os seus conceitos na forma de frases, organizadas e veiculadas num objeto informacional qualquer, como um livro, uma fala gravada, uma música escrita ou gravada, um relatório, um artigo, um mapa, um *software*, etc. Esse conhecimento conceitual e sensorial, quando expresso e materializado num objeto informacional, forma o segundo nível de representação.

Sobre esse objeto, ou melhor, recurso, o indexador pode então trabalhar, buscando a representação do conhecimento contido em relação a toda uma coleção de conhecimentos que ele já tenha registrado ou pretenda registrar. A descrição conceitual feita pelo indexador é o terceiro nível de representação. A adequada representação do conhecimento registrado no terceiro nível permite a recuperação de informações e seu uso por pessoas que busquem aqueles conceitos, expressos pelo autor do recurso e representados pelo indexador.

No momento da representação em terceiro nível, o indexador conta com ferramentas de auxílio para, mais do que representar, metarrepresentar o conhecimento primeiramente representado pelo autor e que será recuperado por usuários do sistema de informação. Assim, compreende-se a metarrepresentação como uma função. Dodebei (2002)

⁵ Vera Lúcia Doyle Dodebei é pesquisadora e professora da UNIRIO.

apresentou esse conceito na literatura brasileira, ligando-o às linguagens documentárias. A autora afirma que metarrepresentação é um termo que expressa a função desempenhada por linguagens de indexação na promoção da representação e da recuperação da informação (DODEBEI, 2002).

Nesse sentido, a metarrepresentação pode ser compreendida como a representação além da representação. Conforme definida pela autora, a metarrepresentação significaria a função de construir representações em terceiro nível e se encaixaria nas atribuições do indexador quando este constrói representações das representações contidas nos recursos, conceitos metarrepresentados em linguagens de indexação, em catálogos, em metadados ou em notações de recursos.

A metarrepresentação foi aqui considerada um conceito importante, porque especifica a diferença entre a simples noção de representar, entendida como a ação de colocar algo em lugar de, e a significação de representação nas linguagens de indexação. Entende-se que o processo de ler um recurso, encontrar sua atenção, extrair-lhe conceitos representativos, acomodar tais conceitos numa hierarquia de classificação, designar-lhe relações de equivalência, contrariedade, associações partitivas e de especialização, preocupar-se com o arranjo de elementos individuais em relação a toda uma coleção ou a todo o conteúdo de um volume é, pois, muito mais do que representá-lo. Tais processos constituem uma função que ultrapassa a representação e, por isso, considerou-se prelevante o termo “metarrepresentação”, a despeito de outras interpretações surgidas na elaboração da pesquisa.

Embora se tenha procurado esclarecer os níveis de representação, o conceito de conhecimento, que convive com a representação, ainda não foi abordado. Neste estudo, tratou-se do termo “conhecimento” como conhecimento registrado, representação e metarrepresentação do conhecimento ou simplesmente representação e metarrepresentação. O termo “conhecimento” refere-se comumente a processos intelectivos individual ou coletivamente situados – o caso coletivo é o das empresas, qualificado usualmente pelo termo “conhecimento organizacional”. Tem-se também o conceito de conhecimento registrado, relacionado a objetos de informação. Com essa acepção, inclui-se o conceito de objeto no contexto da representação do conhecimento.

Compreende-se que o termo “objeto”, ainda pouco encontrado na literatura da área, deve ser lido com algum cuidado semântico. Dahlberg (1978) enunciou uma noção de objeto ao elaborar os conceitos de “objetos gerais” e “objetos individuais”. Nesse caso, o termo “objeto” não se refere ao recurso em si. Em Dahlberg, objetos gerais e individuais podem ser traduzidos como classes de conceitos gerais e individuais. A autora descreve os conceitos individuais como aqueles que ocorrem “toda vez que o objeto é pensado como único, distinto dos demais, constituindo uma unidade inconfundível (coisas, fenômenos, processos, acontecimentos, atributos, etc.)” (DAHLBERG, 1978, p.101). Os objetos gerais seriam aqueles que prescindem de tempo e espaço e que podem ser qualificados como “conceitos gerais” (ex.: UFMG – objeto individual; universidades – objeto geral ou conceito geral).

Ainda sobre as diversas contribuições que referenciam noções de objeto, tema retomado na seção 2.5.1, Buckland (1991) definiu o conceito de informação como

coisa. Para o autor, dados são, na realidade, objetos e, como tais, podem comunicar conhecimento imparcial. Compreenda-se conhecimento imparcial como aquela informação registrada que pode ser verificada e não uma alusão à imparcialidade dos conteúdos do dado em relação ao sujeito autor.

Na história da evolução dos registros humanos, os recursos nem sempre foram compreendidos como objetos, nem comunicavam conhecimento imparcial. Registro e recuperação do conhecimento se deram, muitas vezes, através da poesia e da prosa rítmica, dos mitos e das lendas, formando uma espécie de enciclopédia da era da oralidade. Segundo relatos da evolução nos registros de conhecimento, as escritas cuneiformes e hieroglíficas são antecedentes do alfabeto, só então formando um sistema de registro visual, da nossa era, a era da escrita (Dodebei, 2002). O relato dos registros do conhecimento, e sua evolução no tempo, são fatos que têm chamado a atenção dos estudiosos de ferramentas de representação e recuperação, como pode ser observado no relato de Foskett, ainda em 1973.

Antes de o conhecimento ser registrado (e mesmo ainda hoje em sociedades muito primitivas), eram os próprios indivíduos que constituíam o repositório do conhecimento, a ponte entre gerações sucessivas e entre os que criavam novas informações e os que precisavam utilizá-las. O volume de informações que pode ser assim transmitido é limitado, tendo a sociedade passado a progredir quando informações de vários tipos começaram a ser registradas em formas relativamente permanentes, que podiam servir como substituto da pessoa do 'sábio'. (FOSKETT, 1973, p. 3)

O recurso que contém em si o conhecimento tornado coisa ou objeto é também o que, neste estudo, compreende-se como resultado do ato de representar. No entanto, o conceito de representação tem inúmeras outras acepções além de "ato de colocar algo

no lugar de”. Na ciência da informação, Dodebei (2002) descreve a representação como um fenômeno complexo que abrange três estágios: redução semântica, pluralidade de significados e produção de sentido.

A redução semântica se dá em todo processo de intermediação e é descrita como efeito de qualquer processo de representação. Nesse sentido, os níveis de redução semântica seriam tanto menores quanto mais próxima estiver a linguagem de comunicação empregada pelos atores comunicadores. Contudo, a linguagem em si já representa uma redução de significado (DODEBEI, 2002).

A pluralidade de significados seria provocada pelo fato de a linguagem ser uma multicomposição de fragmentos utilizados como instrumento do pensamento para exprimir conteúdos. Assim, não existiria um significado determinado e único, elaborado num pensamento. Tampouco existiria uma palavra que exprimisse fielmente o pensamento. Em consequência disso, não se poderia transmitir, com exatidão, qualquer pensamento a outra pessoa. Não se conseguiria perceber, na recepção, a informação com os mesmos elementos de significação do momento de sua transmissão (DODEBEI, 2002; MERLEAU-PONTY, 1992; SPIRKIN, 1972; GORSKY, 1966).

Já a produção de sentido se apresenta como um paradoxo na questão da representação, diante da redução semântica do estágio anterior. Tanto a redução semântica como a produção de sentido são questões centrais, uma vez que uma age por sucessivas reduções de significação e a outra promove a rearticulação de

significados. Compreende-se a produção de sentido, justaposta à redução semântica, como uma tensão necessária, pois é na redução de informações que a representação busca obter maior potencial de conhecimento e economia. Na modernidade, a ressignificação para o conhecimento é um processo que requer “um complexo cenário, onde sejam agregadas populações de fontes e canais de informação, de modo a permitir processos seletivos, organizados e econômicos de construção de sentido” (GONZALES DE GOMEZ, 1993, apud DODEBEI, 2002, p. 36).

Outra acepção para representação é proposta por Alvarenga (2001), que considera o conceito a partir da produção de objetos de informações, dividindo-os em representações primárias e secundárias. Tem-se, por um lado, o conceito de representação primária, para definir documentos primários produzidos a partir da mediação do autor e o mundo real. Por outro lado, têm-se as representações que, segundo Alvarenga,

podem ser vistas como de natureza secundária, envolvendo a substituição da informação primária por registros específicos a ela relativos. Nos catálogos dos sistemas e serviços de informação mantidos especialmente no âmbito de bibliotecas, a informação primária textual, sonora ou gráfica, encontra-se representada por novos conjuntos de informações, dando origem a produtos que poderiam ser considerados metainformações ou metadocumentos. (ALVARENGA, 2001, p. 1-2)

Destaca-se, na conceituação de Alvarenga, a visão de que a representação para recuperação, nos moldes do trabalho do indexador, forma produtos do tipo “meta=além de”, quando a autora menciona os termos “metainformações” e “metadocumentos”. O

mesmo sentido é partilhado em Dodebei (2002), que associa o termo “metarrepresentação” à função de representação.

Dodebei descreve a representação a partir da sua funcionalidade nas linguagens de indexação: representar “além de”, ou seja, metarrepresentar. Alvarenga e Dodebei, ao formularem conceitos para a representação nos dias atuais, possuem pontos de vista complementares. Dodebei aborda o conceito de representação sob o ângulo funcional, ligado à ação de metarrepresentação, e Alvarenga o analisa sob o ponto de vista de produto. A idéia de complementaridade dos conceitos surge porque o resultado da função de metarrepresentar dá origem aos produtos metainformações e metadocumentos.

Para Dodebei, a metarrepresentação é a representação da representação, pois a função de representar, utilizando linguagens de indexação, encerra uma economia de símbolos e uma cadeia de reduções semânticas, expressas nas configurações de representações criadas. Articulando-se à mesma idéia, as metainformações seriam as informações das informações e os metadocumentos os documentos dos documentos.

Contudo, tomando-se tanto a função de metarrepresentação quanto o seu produto, metainformação (informação sobre a informação) e metadocumento (documento sobre o documento), percebe-se que a existência destes fundamenta-se no elemento componente e essencial da representação, o conceito. O verbete “conceito” é definido pelo *Minidicionário Aurélio da língua portuguesa*⁶ como “pensamento, idéia”. Na

⁶ FERREIRA, Aurélio B. H. F. *Minidicionário Aurélio da língua portuguesa*. 3.ed. rev. e ampl. Rio de Janeiro: Nova Fronteira, 1993.

filosofia, desde Aristóteles, o conceito (*logos*) já era considerado a essência necessária da realidade.

A noção aristotélica de essência é completada, em Dahlberg, pela noção de verdade e o conjunto de características analisáveis. Na “teoria do conceito”, este é descrito como formado pela “compilação de enunciados verdadeiros sobre determinado objeto, fixada por um símbolo lingüístico” (DAHLBERG, 1978, p.102). Sobre os enunciados verdadeiros, afirma a autora, podem ser feitas análises que revelam atributos, ou seja, as características para análise.

As características de um conceito, na realidade, não são somente características, mas categorias, uma vez que podem ser formadas por unidades de níveis muito gerais. Acredita-se que um recurso não é classificado por si, mas pelos conceitos que contém. Em um recurso, inúmeros conceitos articulados formam um todo semântico. E esse todo é, certamente, articulado por conceitos e categorias de conceitos. Alvarenga mencionou a articulação conceitual ao afirmar que “o que na realidade se classifica em uma biblioteca tradicional ou digital não são os documentos, mas os conceitos contidos nesses documentos” (ALVARENGA, 2001, p.7) e que o conceito, além de elemento essencial da metarrepresentação, também se revela invariante, pois, para o processo de representação, a busca pelo elemento de classificação recai irremediavelmente sobre o conceito, independentemente de o recurso que o emana ser textual, sonoro ou em imagem, fixa ou em movimento.

Pensar na articulação conceitual, formando um todo semântico, leva à concepção de que o conceito estaria ligado a redes de padrões de inferência, associações e

relacionamentos (SHERA, 1957). Essa idéia é encontrada em Platão, Aristóteles e, posteriormente, no Triângulo Semântico de Ogden & Richard. Tal triângulo, conforme Fig. 1, apresenta os vértices X, Y e Z, sendo Y a expressão lingüística, Z a coisa ou estado de coisas e X o significado. Umberto Eco (1991, apud DODEBEI, 2002), numa releitura de alguns filósofos, afirma que Z, para Platão, era uma experiência no mundo, uma realidade situada no mundo das idéias, e que, para Aristóteles, era uma substância primeira, em toda a sua concretude. O X, para Platão, era um conceito e, para Aristóteles, uma disposição da mente. Todos esses elementos – X, Y, Z – são conectados por linhas que expressam formas de relacionamento conceitual.

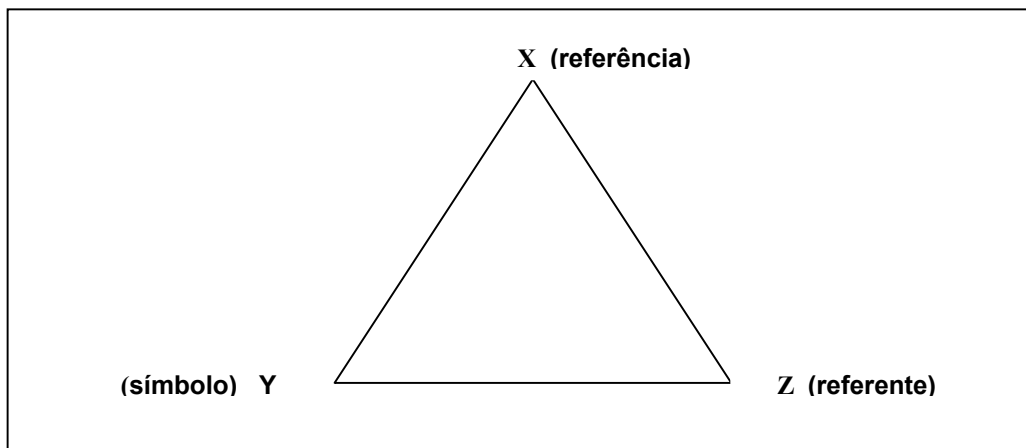


FIGURA 1: *Triângulo semântico (Ogden e Richard)*
 Fonte: DODEBEY, 2002, p. 50.

Vickery (1986) apresentou um diagrama simples que articula esses mesmos vértices do triângulo semântico, chamando o vértice X de conceito, o Y de símbolo e o Z de referente, como mostra a Figura 2.

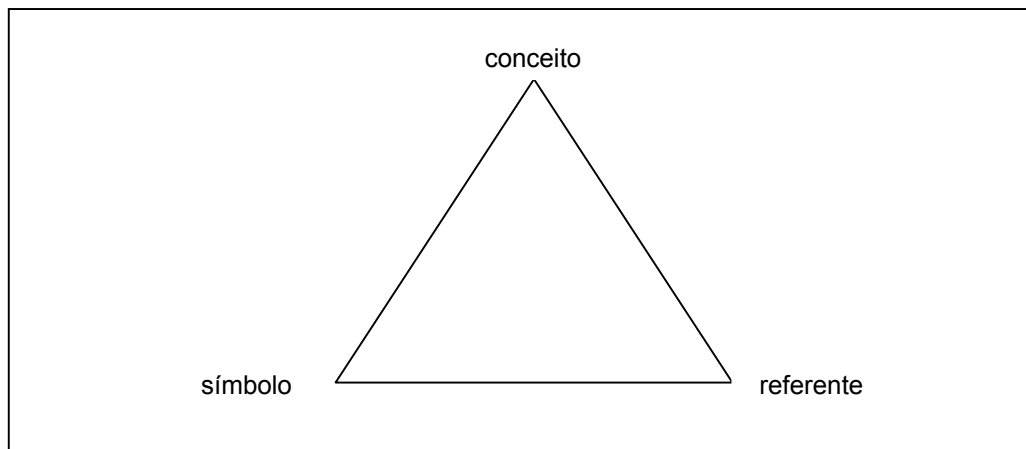


FIGURA 2: *Diagrama de conceitos*
Fonte: Vickery, 1986.

Na metarrepresentação, as interações entre o conceito expresso, o objeto que deriva o conceito e a significação atribuída determinam padrões de inferência, associações e relacionamentos a serem registrados em sistemas de recuperação de informação minimamente da mesma natureza que aqueles visualizados no triângulo semântico. Dahlberg (1978) explicita relações lógicas, hierárquicas, partitivas, de oposição e funcionais entre conceitos.

As relações lógicas entre conceitos, como implicação, intersecção, disjunção e negação, são úteis na construção e no funcionamento das linguagens de indexação. Na relação de implicação, um conceito é considerado como contido em outro mais geral. Na relação de intersecção, ocorre a coincidência entre elementos de dois conceitos. Na relação de disjunção os conceitos se excluem mutuamente e, na negação, um conceito inclui característica cuja negação se encontra em outro.

Na área da Ciência da Informação, muitas das características das representações conceituais que se dão no campo da metarrepresentação, são elaboradas pela teoria e prática da análise de assunto. Os elementos mais importantes apreendidos sobre a análise de assunto estão descritos na seção seguinte.

2.2 Análise de assunto

No contexto da análise de assunto, são descritos o conceito, o processo e alguns pontos de vista considerados importantes.

A análise de assunto, neste estudo, é vista como o processo-base que permite a metarrepresentação (2.1) do conhecimento obtido sobre os recursos (recursos de informação, em papel ou não). Aborda-se esse tema através da sua conceituação e da visão lógica dos seus subprocessos, como também através da elaboração de algumas considerações.

A análise de assunto insere-se num escopo mais amplo, o da indexação temática. Lancaster (1989), numa revisão publicada pelo *Arist – Annual Review of Information Science and Technology*, que cobriu a literatura editada de 1986 a 1989, define a indexação temática e suas interfaces como

a presença, identificação e expressão, da matéria assunto, em textos de documentos, bases de dados, linguagens natural e controlada, requisições de informação e estratégias de busca. Do ponto de vista do usuário, a análise de assunto está ligada ao acesso por assunto. Assim, também são pertinentes os meios pelos quais a informação sobre um assunto pode ser recuperada de várias fontes (LANCASTER, et al., 1989, p. 35).

A conceituação de Lancaster permite estabelecer a indexação temática como um processo que, além de formar metarrepresentações, é complexo e, por isso, envolve diversos atores nos seus subprocessos. Estes estariam ligados aos papéis desempenhados pelo autor, pelo indexador e pelo leitor num sistema de informação. Sobre os sistemas de informação, seguem-se, na seção 2.5, conceituações e revisão de alguns aspectos dos processos envolvidos. A Fig. 3 mostra, através de diagrama de contexto, como interagem esses atores. Para o processo de análise de assunto, o indexador tem o seu papel ressaltado.

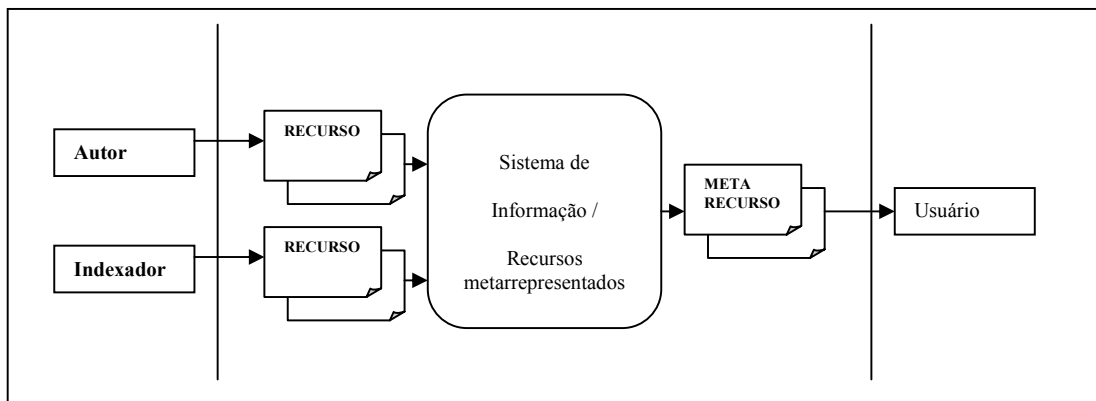


FIGURA 3: Contexto da análise de assunto
Fonte: Elaboração própria.

Na Fig. 3, o autor do recurso é quem fornece a representação primária do conhecimento, com toda a subjetividade das suas percepções. O conteúdo descrito pelo autor pode, inclusive, conter erros, sob o ponto de vista de uma outra referência. No entanto, tais erros não importam para a análise de assunto, pois sua função está ligada ao desempenho da fidelidade entre o conteúdo do recurso e o produto da sua metarrepresentação.

O usuário, por sua vez, está ligado à análise de assunto pelas requisições de pesquisa e acesso a recursos metarrepresentados. É a partir do comportamento de resposta desse ator que se fazem as avaliações de precisão e relevância dos conteúdos da representação secundária, além de outros tipos de controle de qualidade de sistema.

O indexador, no papel central dos subprocessos de um sistema de informação, responde por diversas etapas da análise de assunto. Cesarino e Pinto (1980), citando um esquema elaborado por Lancaster, descrevem a atuação do indexador em dois momentos. O primeiro acontece nos subprocessos de recepção dos documentos, análise do conteúdo, análise conceitual, tradução conceitual com ferramenta de vocabulário controlado, classificação dos documentos, armazenagem e construção de perfis dos documentos através dos termos de indexação. O segundo momento ocorre na recepção dos pedidos de busca, análise de conteúdo de busca, análise conceitual, tradução da análise conceitual em termos de indexação e construção de uma estratégia de busca. Os dois momentos culminam na existência de um procedimento de busca que considera o casamento entre os perfis da metarrepresentação e da busca efetuada pelo usuário.

O primeiro momento, conforme Cesarino e Pinto (1980), resume-se sob o conceito de análise de documentos, dividido em análise objetiva e análise subjetiva. Como processo de análise objetivo, foram mencionadas as descrições dos documentos em termos de suas características físicas, que responderiam à questão: “Qual a aparência deste documento?”. Geralmente, preocupam-se com as descrições do nome da obra (título), autor, edições, editor, data de publicação, número normalizado, especificação do tipo de suporte, etc.

Considerando-se a década de 1950 até esse início de século XXI, cerca de cinquenta anos, as descrições objetivas têm sido fortemente desenvolvidas através de padronizações, que ganharam consenso por todo o mundo, através dos códigos de catalogação, padrões internacionais para todos os tipos de recurso informacional, e formatos de representação intercambiáveis (catalogação compartilhada) , legíveis por máquina.

Já os processos de análise subjetiva são aqueles que “pretendem a descrição do documento em termos de seu conteúdo informativo (assunto abordado), respondendo à questão: ”De que trata este documento?”” (CESARINO; PINTO,1980, p. 33). Segundo as autoras, é na descrição dos conteúdos dos recursos que residem os maiores desafios para a indexação.

A atividade de analisar o assunto de um recurso pode ser dividida em três etapas: compreensão do texto como um todo, identificação dos conceitos e seleção dos conceitos válidos para serem indexados (CESARINO; PINTO, 1980).

Todo o conjunto de subprocessos envolvidos na análise de assunto para a identificação de um documento foi chamado por Langridge (1989) de sumarização. Para o autor, é indiferente o processo de análise de assunto de um livro imenso ou de um pequeno artigo, e o objetivo da análise seria o de tornar possível a seleção de um ou dois documentos de uma coleção, em resposta a uma pergunta particular.

Langridge (1989) define aplicações diferenciadas para a análise de assunto. Primeiramente, tal análise estaria ligada à caracterização do documento como um todo,

a ser recuperado de uma coleção. Em segundo lugar, a análise de assunto estaria ligada à análise exaustiva, para permitir que itens específicos de informação sejam recuperados do conteúdo de um documento, também chamado de índice interno.

Pode-se visualizar os subprocessos da análise de forma objetiva e subjetiva. O que o indexador persegue é a análise objetiva. Para tanto, busca responder a duas perguntas. A primeira é: “O que é o documento?”. Essa pergunta deverá ter como resposta uma das formas de conhecimento, também chamada de disciplina. É, portanto, uma resposta abstrata. A segunda pergunta é: “Sobre o que ele é?”. Para essa questão deve ser descrito o tópico do assunto naquela disciplina. A resposta se formula em termos de um fenômeno ou do que percebemos como existente no mundo (LANGRIDGE, 1989).

A pergunta “sobre o que é um documento” é conceitualmente referenciada em português como “atinência”, “sobrecidade” ou “temacidade”, termos que surgem como tradução do inglês *aboutness*. Atinência, único termo de que se encontrou registro na língua portuguesa, é definido no Houaiss(2001)⁷ como “qualidade ou característica de atinente”. Atinente, por sua vez, significa o que diz respeito a algo ou o que concerne a algo, definição que encontra o mesmo sentido de “sobre o que é” algo.

⁷ HOUAISS, Antônio. VILLAR, Mauro S. *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro, Objetiva, 2001.

Langridge (1989) chamou de escorregadio o conceito de *aboutness*, uma vez que é usado como *intrinsic aboutness* para se referir a elementos intrínsecos ao documento, e *extrinsic aboutness* quando se trata de considerar variáveis externas.

Hjørland atribuiu o uso de *aboutness* a uma “tentativa de se escaparem às dificuldades do conceito de assunto” (1992, p.174). Afirma que, com o uso do termo, os seus adeptos atribuem clareza e objetividade à análise de assunto que, na sua visão, não existem. O autor considera idealista a questão do assunto e que, na determinação do assunto de um documento, nem o autor, nem o leitor e/ou o especialista em informação estariam corretos ou teriam qualquer conhecimento objetivo sobre ele.

Neste estudo, considerou-se que a argumentação de Hjørland é um tanto generalista, sendo necessárias alguns comentários. O autor não considerou a função da análise de assunto em um sistema de informação ou sistema de recuperação da informação. Não observou a economia que a análise de assunto possibilita na representação dos recursos, quando trata da redução semântica, da pluralidade de significados e da construção de sentido. Não constam, ainda, questões de custo e benefício, nem é observada preocupação com a relação com o usuário, que tem, ontologicamente, lançado o homem na busca de mecanismos de representação do conhecimento.

Por outro lado, defendendo o *aboutness*, dentre outros aspectos da análise de assunto, Langridge (1989) afirma que falhas de compreensão do que deve ser objetivamente respondido na análise de assunto levam erroneamente a perguntas subjetivas do tipo: “Para que é este documento?” ou ainda: “A quem este trabalho é endereçado?”. Para o

autor, essas questões admitem mais de uma resposta e vão depender do olhar subjetivo do indexador, enquanto questões sobre a forma do conhecimento e tópico abordado são objetivas e terão somente uma resposta certa. Respostas diferentes seriam atribuídas a erros de análise.

Langridge (1989) contribui não só para a representação de documentos, na qual se admite apenas uma resposta certa, mas também para a representação de informações. Para ele, o processo de análise para recuperação de informações é diferente da recuperação de documentos. Exige-se do indexador uma análise de assunto mais detalhada, com a preocupação de caracterização do documento como um todo, ou seja, caracterização dos tópicos aí contidos.

Parceiras da análise de assunto, e muitas vezes interferindo no seu desenvolvimento, as linguagens de indexação proporcionam o registro sistemático da análise de assunto efetuada. Por sua importância e contexto, pois amparam teoricamente o objeto de estudo desta pesquisa, buscou-se elaborar uma conceituação e algumas características da evolução verificada ao longo dos últimos quase cinquenta anos, desde surgimento do *KWIC*, primeiro sistema de índices rotados, no ano de 1960.

2.3 Linguagens de indexação

Até aqui foi descrita a idéia de que o conhecimento, quando registrado, torna-se objeto informacional. Como objeto, o seu tema pode ser analisado em relação ao total da

coleção em que deverá ser classificado ou em minúcias que denotem os tópicos que contém, formando o produto da análise de assunto. O produto considerado, neste estudo, é essencialmente a frase de indexação: uma ou diversas, conforme diferentes critérios adotados pela instituição ou indexador. Relembrando, a frase de indexação é a frase em linguagem natural, expressa pelo indexador ao final da análise de assunto. Normalmente, é a sentença que completa a afirmação: “Este documento trata de ...”. Contudo, as frases de indexação somente se tornam efetivamente produto de representação secundária após organizadas e representadas, ou seja, através do uso de instrumentos de indexação, num conjunto de teorias e práticas abarcadas pelas linguagens de indexação.

O termo “linguagem de indexação” foi escolhido, neste estudo, por ser o mais utilizado na área, dentre outros termos correlacionados: linguagens de indexação (Melton), linguagens descritivas (Vickery), codificação documentária (Grolier), linguagens de informação (Soergel), vocabulários controlados (Lancaster), assim como linguagens de recuperação da informação, linguagem de descrição da informação, linguagens de representação, linguagens de representação documentária e linguagens documentárias.

Linguagens de indexação são linguagens artificiais formadas por um conjunto de regras que servem para representar abreviadamente o conteúdo de um recurso (RIVIER, 1992). Guinchat e Menou (1994), que as denominam de linguagens documentárias, afirmam que são instrumentos convencionais de uso das unidades de informação para a descrição dos conteúdos dos documentos. Destacam, como instrumentos, as

classificações bibliográficas, os cabeçalhos de assunto, as palavras-chave, as listas de descritores, os tesouros ou, ainda, conjuntos léxicos de uma área utilizados no momento do tratamento intelectual dos recursos ou das pesquisas por recursos.

Os sistemas de classificação, que formam as linguagens de indexação simbólicas, têm sido instrumentos tradicionais de indexação temática. Fornecem, *a priori*, a possibilidade de coordenação de assuntos dos recursos informacionais. Os conceitos que representam os assuntos são previamente organizados, formando grupamentos expressos na forma textual ou na forma conjugada de símbolos e texto. Quando os assuntos são sistematizados na forma textual, alfabeticamente organizados, diz-se tratar-se de uma linguagem verbal, caso da Lista de Cabeçalhos de Assuntos da Library of Congress. Mas, quando o sistema de indexação preconiza símbolos para promover o agrupamento e representação conceitual, trata-se, então, de uma linguagem simbólica, cujo exemplo mais usual é a Classificação Decimal de Dewey. Juntas, as linguagens verbais e simbólicas formam a classe dos vocabulários controlados, ou seja, linguagens de indexação que determinam os termos preferidos para se representarem conceitos.

Guinchat e Menou (1994) definem sistemas de classificação como linguagens artificiais, instrumentalizadas para representar todos os conceitos e objetos de um campo determinado do conhecimento. Para exercerem essa função, as linguagens artificiais, tanto as simbólicas quanto as verbais, apresentam os conceitos de forma ordenada, ou melhor, coordenada. São, por isso, instrumentos de indexação temática, pois possibilitam a representação do recurso informacional por seu assunto.

Os sistemas de classificação fazem parte do conjunto de linguagens de indexação que, na apresentação dos conceitos que organizam, utilizam alguma forma de sistematização. As sistematizações, como agrupamento de conceitos, são denominadas pré e pós-coordenação. Na pré-coordenação, os descritores são, na maioria dos casos, palavras compostas que designam a totalidade de uma noção. Exemplo: “Cultura irrigada de cereais”. São formados através de uma pré-visão dos conteúdos dos recursos e pré-visão das necessidades de representação pelas quais o indexador irá passar. Já na pós-coordenação, “ao contrário, as noções são reduzidas a seus elementos constitutivos mais simples. Nesse caso, pode-se ter, por exemplo, o descritor ‘técnicas de culturas’” (GUINCHAT; MENO, 1994, p.141). Na pós-coordenação a possibilidade de combinação de termos, para formar conceito existente em um recurso, só aparece no momento de recuperar esse recurso. Não há uma lista de combinações de termos previa.

Guinchat e Menou (1994) descrevem a essência funcional do uso de linguagens de indexação que relatam como sendo “instrumentos destinados à realização de operações precisas, em condições precisas, para responder a necessidades precisas”. Contudo, a precisão de seus funcionamentos variam muito em relação aos níveis de controle que implementam, ligados à questão da coordenação de conceitos.

Uma questão controversa inerente às linguagens de indexação é a da linguagem natural como ferramenta de representação e recuperação. A linguagem natural é normalmente considerada como sinônimo de discurso comum. É a linguagem habitualmente usada para falar ou escrever e que ocorre no texto livre, ou seja, no

título, no resumo, num extrato ou no texto integral de uma publicação. Quando a recuperação de informações num sistema é feita pelo casamento do padrão texto livre, presente no conteúdo do recurso, e texto livre da estratégia na busca (expressão que enuncia a necessidade de informação), entende-se que o sistema faz busca em linguagem natural.

Afirma-se que um sistema de recuperação de informação em linguagem natural efetua busca utilizando o casamento de palavras padrão texto livre, quando esse sistema retira palavras da expressão empregada pelo usuário no momento da busca de informações e as usa para checar se os documentos armazenados têm registrada aquela palavra. A expressão “casamento” é usada no sentido de compreender que são recuperados apenas os documentos cujos termos da busca “casam” com os termos aí contidos. Nessa situação de recuperação não ocorre a substituição dos termos dados para pesquisa por termos mais representativos da necessidade do usuário. A esse recurso de busca de informação têm-se atribuído altos índices de recuperação (muitos documentos recuperados) e baixos índices de relevância (poucos documentos que atendem aos interesses da busca de informação).

A linguagem artificial, no âmbito das linguagens de indexação, expressa uma técnica de recuperação diversa da linguagem natural. Cada uma agrupa ao redor de si defensores e críticos que apontam suas vantagens e desvantagens. Aqui, entende-se como linguagens artificiais o conjunto de linguagens de indexação usadas para fazer sumarização, representação e busca de informações com controle de vocabulário. Muitas pesquisas têm sido feitas sobre os aspectos positivos e negativos do uso das

linguagens artificiais e natural, tentando apontar o melhor instrumento de funcionamento da representação e da recuperação.

2.3.1 Controvérsias entre linguagem natural e linguagem artificial

Ainda sem conclusões definitivas, os estudos sobre as controvérsias entre recuperação usando linguagem natural ou linguagem artificial têm baseiado seus procedimentos, principalmente, em índices de revocação e precisão no contexto da recuperação de informação. A revocação é uma medida sobre a efetividade de uma busca, expressada como a taxa resultante do número de recuperação de registros relevantes ou documentos, em resposta a um total de registros ou documentos relevantes em um banco de dados. A precisão é uma medida sobre a efetividade de uma busca, expressada como a taxa resultante dos registros ou documentos relevantes recuperados de um banco de dados em face ao número total recuperado em resposta a uma busca. Observa-se que os conceitos de revocação e precisão mencionam a relevância(REITZ⁸).

A relevância, mencionada nos dois índices revocação e precisão, refere-se ao julgamento feito pelo usuário sobre o grau em que um documento recuperado na busca preenche sua necessidade de informação. Assim, um índice de precisão pode ou não expressar completamente a precisão da recuperação. Contudo, a efetividade das

⁸ REITZ, Joan M. ODLIS - Online dictionary for library and information science. Libraries Unlimited. Acessível em <http://lu.com/index.cfm> recuperado em 31/05/2005.

medidas de revocação e precisão depende da relevância dos resultados da pesquisa, conceito atribuído pelo usuário. Mas a relevância não leva em consideração o estágio atual de conhecimento do usuário. Um documento recuperado pode ser relevante para um dado tópico buscado pelo usuário, mas pode não ser pertinente para este usuário. Tal fato ocorre porque o estágio de conhecimento do usuário pode ser muito fundamental e por isso sua necessidade de informação seria mais básica. Nesse caso, se os documentos relevantes fossem muito especializados ocorreria falta de pertinência em relação à necessidade do usuário, e vice-versa.

Pertinência, na recuperação da informação, seria a extensão na qual um documento recuperado, em resposta a uma estratégia de busca, satisfaz a necessidade de informação, que por sua vez, é baseada no estágio de conhecimento atual do usuário. Pertinência é, por isso, um conceito mais específico do que relevância (REITZ).

Estes indicadores de medidas de recuperação de informação são instrumentos registrados no centro das discussões das linguagens de indexação. Alguns posicionamentos das pesquisas, que apontam a disputa entre tipos de recuperação (linguagem natural ou linguagem artificial), são descritos para demonstrar que, embora tenham evoluído muito nos últimos anos, as linguagens de indexação estão longe de serem mecanismos completamente consolidados.

Carrow e Nugent (1977, apud LOPES, 2002), para quem as buscas com termos de linguagens artificiais produziram uma significativa e melhor recuperação, aconselharam o uso de ambos os métodos: linguagem natural e artificial, uma como complemento da

outra. Henzler (1978), também citado por Lopes, defendeu essa combinação durante a elaboração da estratégia de busca.

Lopes (2002), na sua revisão de literatura, examinou ainda diversos outros autores. Calkins (1980) não conseguiu comprovar a hipótese de que a busca em linguagem natural faria retornar todos os itens recuperáveis. Knapp (1982) elaborou técnicas para melhorar a estratégia de busca em mecanismos de linguagem natural. Já Schroder (1983) acrescentou a necessidade de uso de acrônimos nas estratégias de recuperação. Wagers (1983) afirma existirem, nos resumos, termos muito semelhantes aos de uma linguagem artificial. Stevert e Boyle (1983) consideram as linguagens artificiais como instrumentos de precisão (recuperam o que é certo do ponto de vista do usuário) e não de revocação (recuperam em grande quantidade). Piternick (1984) escreve sobre a necessidade de uso de termos da linguagem natural e da artificial, na montagem da estratégia de busca. Harter (1986) concluiu que as linguagens artificiais são rígidas, inflexíveis e precisas, enquanto a linguagem natural é expressiva, flexível e potencialmente ambígua. Boyce e Mclain (1989) lembram a questão da precisão diante da estrutura de informação fornecida pelo sistema, que pode viabilizar ou não os níveis desejados de precisão na recuperação. E Rowley (1990) lembrou a necessidade de novas interfaces hipermídia e hipertexto para a melhoria do desempenho tanto da linguagem natural quanto das artificiais.

Lopes conclui, após a revisão de literatura, que dada a quantidade de estudos vistos, comprova-se que o uso simultâneo dessas linguagens proporciona melhor desempenho nos resultados alcançados na recuperação de informações.

O fato é que, na década de 1990, viu-se progredir, junto a grande volume de recursos digitais textuais, uma imensa gama de pesquisas e aplicações que utilizam a linguagem natural como ferramenta para a organização e recuperação de informações. Proliferaram as estratégias de armazenamento para recuperação baseadas na extração direta e automática de termos do texto. Nesse esquema de tratamento de informações, a recuperação é iniciada pela técnica de extração direta dos termos apresentados na expressão de busca. O resultado da solicitação é obtido pelo processamento do casamento de padrões: termo presente no documento com termo presente na estratégia de busca.

A despeito da grande utilidade encontrada para as aplicações de construção de índices e processos de recuperação baseados em linguagem natural, o que se tem confirmado, ao longo das pesquisas feitas, é que "linguagens de indexação, sendo linguagens projetadas para finalidades especiais, devem ser caracterizadas com relação a essas finalidades" (SVENONIUS, 1981, p. 90). O que a autora aí explicita é que as linguagens artificiais devem ser orientadas, tanto na construção quanto no uso, para finalidades especiais. Essa afirmação remete a Guinchat e Menou (1994), para quem as linguagens se destinam, idealmente, à realização de operações de descrição precisas, em condições precisas, para responder a necessidades precisas.

Além da destinação à precisão, e abordando algumas das características mais específicas das linguagens de indexação, Lancaster (1993) elabora um quadro explanatório das diversas naturezas dos estudos feitos até aquela data, tópico abordado na próxima seção.

Essas questões são importantes, pois o objeto deste estudo é uma abordagem teórica pertencente ao universo das linguagens artificiais. Portanto, ao modelo estudado e proposto nesta pesquisa aplicar-se-iam caracterizações e discussões sobre as especificidades das linguagens artificiais.

2.3.2 Caracterizações das linguagens de indexação

Lancaster (1993) descreve outras especificidades das linguagens de indexação que contribuem para a delimitação teórica do objeto deste estudo. Para ele, as linguagens de indexação têm-se orientado segundo dois direcionamentos: processos permutados e processos lógicos. Na categoria de processos permutados, considera as funções de alternância, rotação e deslocamento. Já para os processos lógicos, relata as funcionalidades de ordem, contexto, relação e função.

A divisão entre permutados e lógicos apóia-se no fato de que os processos permutados promoveram sistematizações de termos baseadas puramente na ordem alfabética. Já os processos lógicos procuraram ordenar os termos através de diversos outros mecanismos de sistematização.

2.3.2.1 Processos permutados

Os processos permutados são caracterizados pela troca de lugar entre os termos de uma sentença, frase ou título, com a intenção de ressaltar o termo que represente o resultado de uma busca de informações. A definição de permutação mais próxima das linguagens de indexação é da lingüística estruturalista, que a considera como a troca dos elementos de uma seqüência, sejam letras, palavras ou sintagmas (HOUAISS, 2001)⁹. Um sintagma, por sua vez, pode ser compreendido como uma unidade conceitual, correspondendo a conceitos formados por palavras simples ou até frases inteiras.

O primeiro índice que utilizou processo permutado totalmente mecanizado foi o sistema KWIC, abreviatura de *Key Word in Context*, de H. P. Luhn, em 1960, nos Estados Unidos, mas pode-se atribuir a Kaiser, Ranganathan e Coates as idéias que, por diferentes caminhos, levaram aos índices permutados e à sua evolução, observada nos processos lógicos descritos adiante.

Um caso de processo permutado, denominado alternância, ocorre no programa *SLIC-Selective Listing in Combination*, desenvolvido por J. R. Sharp (1966, apud LANCASTER, 1993), no qual uma dada seqüência de termos, por processo combinatório, resulta nas coordenações entre todos os termos, obviamente eliminando as combinações redundantes. Ex.: conjunto de termos ABCDE, ordenação alfabética da

⁹ Obra já citada na página 39.

letra A (AB, AC, AD, AE, ABC, ABD, ABE, ABCD, ABCE, ABCDE). Tal sistema gera, inconvenientemente, um grande número de entradas.

O *Excerpta Médica*, periódico da área médica, utiliza no seu índice, como no caso do SLIC, termos alternados, mas, em geral, no máximo a quantidade de entradas que existe na quantidade de termos. Esse processo também é conhecido como *shunting*. Ex.: ABCDE (ABCDE, BACDE, CABDE, DABCE, EABCD). Contudo, tem ainda o inconveniente de gerar um contexto inexistente na relação de termos da frase de indexação inicial.

Já a rotação é um processo idêntico à alternância, exceto pelo fato de o termo de entrada (que permitira a recuperação) ser ressaltado na apresentação do índice. Uma frase com os termos ABCDE, quando submetida à rotação, apresentaria descrições em que o termo sublinhado apareceria numa posição de destaque, como demonstra o método *Key Word in Context – KWIC*, no Quadro 1.

QUADRO 1: Modelo do processo rotacional KWIC

	A	<u>A</u> BCDE
	AB	A <u>B</u> CDE
	ABC	AB <u>C</u> DE
	ABCD	ABCD <u>E</u>

Fonte: Lancaster, 1993, p. 45.

O processo de deslocamento, por sua vez, emprega duas linhas para reduzir a ambigüidade da relação entre termos, como mostra o Quadro 2, exemplificando a permutação realizada pelo *Preserved Context Indexing System-PRECIS*. O processo do

PRECIS será mais bem descrito na seção 2.3.2.2, pois os procedimentos que emprega não se enquadram apenas em processos permutados.

QUADRO 2: *Modelo do processo de deslocamento PRECIS*

<u>A</u>
BCDE
<u>B</u> A
CDE
<u>C</u> BA
DE
<u>D</u> CBA
E
<u>E</u> DCBA

Fonte: Lancaster, 1993, p. 45.

Observa-se que tanto a rotação quanto a alternância e o deslocamento proporcionam alguma medida de preservação do contexto. Geralmente, esses processos pressupõem o emprego de termos de indexação (retirados de sistemas de classificação, listas de cabeçalho de assunto e tesouros) e não de texto livre. Porém, a abordagem de sistemas de indexação preocupados com o contexto de existência dos termos, sua ordem de enunciação, suas relações e funções, é delimitada por Lancaster sob o conceito de processos lógicos.

2.3.2.2 Processos lógicos

Lancaster (1993), na descrição dos processos lógicos, considera as características de preocupação com a ordem de expressão dos conceitos, com a preservação do contexto em que foram enunciados (depois de sofrerem processo de permutação) e com a

expressão das relações e funções, semânticas e pragmáticas, através de estruturas de relacionamento de termos.

Aspectos ligados à ordem de expressão dos conceitos remontam a especificações da lógica de concretude de Kaiser, no início do séc. XX. Segundo esse teórico, os conceitos seriam organizados a partir da idéia do mais concreto ao mais abstrato, seguidos pelas idéias de localização e processos a eles relacionados.

Depois de Kaiser, surgiu a importante contribuição de Ranganathan, na década de 1930. Esse autor descreveu a idéia da indexação em cadeia, dada pela ordem dos conceitos dispostos na fórmula PMEST. A letra *P* descreve a categoria personalidade, que, segundo Parkhi (1964), abrange personalidade humana, personalidade social, aspecto físico da personalidade humana, refinamento da personalidade da linguagem, o espaço como personalidade, o tempo como personalidade, o estilo como personalidade, o autor e o trabalho na literatura como personalidade. Langridge (1989) simplifica a configuração de Parkhi ao afirmar que uma personalidade é reconhecida pela negativa das categorias MEST – matéria, energia, espaço e tempo, ou pela característica de substância, das categorias de Aristóteles. A letra *M* descreve matéria, aplicada tanto a propriedades físicas ou materiais (ex.: madeira), quanto a propriedades abstratas (ex.: cor, forma) (LANGRIDGE, 1989). A letra *E* descreve energia, ou seja, ação de qualquer tipo. A letra *S*, do inglês *space*, descreve espaço, limitado a espaço geográfico. A letra *T* descreve tempo, relacionada a tempo cronológico. Assim PMEST é o acrônimo dos termos em inglês: *personality, matter, energy, space, time*, e formam categorias de representação de assunto, de característica lógica.

Conforme afirmado no parágrafo anterior, as categorias PMEST foram, então, empregadas para a construção de índices em cadeia. A indexação em cadeia “consiste em se analisar um assunto através de uma seqüência linear e rígida de termos, com a finalidade de construir índices alfabéticos manuais” (RIVIER, 1992, p. 66). Esse procedimento gera índices, ao final do trabalho, com o agrupamento dos termos em torno das categorias PMEST em que foram previamente classificados.

Coates, na década de 1950, influenciado pelo PMEST, elaborou uma lógica na qual ordem era atribuída pela idéia de importância do conceito, do mais importante para o menos importante. Seu princípio de ordem (coisa, parte, material, ação, agente), principiava em coisas, depois parte da coisa, depois material da coisa, depois ação realizada ou sofrida e finalmente o agente passivo da ação.

Assim, têm-se as principais contribuições sobre a questão da ordem relacionada aos processos lógicos, atribuídas a Kaiser, Ranganathan e Coates. Resta a lógica do contexto e relacionamentos.

Concernentes à preservação do contexto e a relacionamentos entre termos, estudos realizados na década de 1960 se preocuparam com a representação diferenciada entre estruturas sintáticas (contexto de ocorrência) e associações semânticas (expressão de relações e funções) dos conceitos, como os estudos de Gardin e Farradane. Gardin (1965), na elaboração do *SYNTagmatic Organization Language-SYNTOL*, afirma que “é fundamental a distinção entre organização semântica e sintática dos descritores” (p.20). Com tal afirmação, o autor considerou a necessidade de explicitar os

relacionamentos entre conceitos, mas não a necessidade de categorizá-los, pois sua questão de fundo era a associação paradigmática como ocorrência *a priori* e a sintagmática como ocorrência *a posteriori*. O autor explica que o termo “sintagmático” designa a dimensão de análise lingüística como unidade relacional formal.

A análise lingüística a que Gardin se refere foi defendida por Ferdinand de Saussure por volta de 1915, ao elaborar os conceitos de eixo sintagmático e eixo associativo (paradigmático). Para Saussure (1981), uma sentença apresenta, na linearidade das palavras, relações de natureza sintagmática ligadas aos sintagmas expressos. Um exemplo da relação sintagmática numa estrutura de sentença é a frase: “O cachorro morde o homem”. Os termos apresentam uma relação de sujeito, verbo e objeto que, nesse contexto, não pode ser invertida. Do contrário, teríamos: “O homem morde o cachorro”. Por outro lado, cada termo da sentença enunciada remete, por si só, a um conjunto de idéias, ausentes na frase, que formam um eixo semântico vertical: as relações paradigmáticas, formadas por um conjunto infinito de associações, que dependeriam das idéias individuais de cada pessoa, e que, por isso, não seriam caracterizadas nem por quantidade, nem por sentido determinados. Se no eixo sintagmático (horizontal) as relações estão presentes e são finitas, no eixo paradigmático (vertical), ao contrário, as relações estão ausentes e são infinitas, formando uma configuração de nó de rede.

O modelo geral de Gardin, que é sintagmático, descreve pares de termos que podem ser ligados entre si por uma relação específica, selecionada dentre um conjunto ilimitado. A unidade estrutural mínima do seu modelo é composta por dois termos (*a* e

b) em uma relação diádica (R_i), ou seja, entre dois. Assim, qualquer esquema de classificação poderia ser expresso nesse formato, como enumeração aberta de unidades diádicas (R_i, a, b), das quais seria fácil reconstruir estruturas totais, conforme especifica o Quadro 3.

QUADRO 3: Exemplo de enumeração de unidades diádicas (R_i, a, b)

Relação diádica <i>i</i>	Unidade <i>a</i>	Unidade <i>b</i>
Ser autor	Madalena M. L. Naves	Dissertação
Ter título	Dissertação	O papel do orientador: o processo de orientação no curso de pós-graduação em biblioteconomia da UFMG
Ser orientada	Dissertação	Jeannette M. Kremer
Ter publicador	Dissertação	UFMG
Ser publicada	Dissertação	1993

Fonte: Elaboração própria

O SYNTOL é importante porque situa-se num período de transição que passou a agregar, junto à explicitação semântica, também a explicitação sintática. Percebe-se isso quando Gardin afirma que, no seu sistema, “nenhum esforço é feito, e talvez nunca seja necessário, construir explicitação dos relacionamentos por trás das mais detalhadas interpretações de hierarquia” (GARDIN, 1965, p. 22).

Farradane (apud LANCASTER, 1993) contribui para os processos lógicos quando busca explicitar as relações que ocorrem nas hierarquias geralmente vistas nos sistemas de indexação. Contribui também para a evolução das linguagens de indexação quando descreve os tipos de relação presentes nos sistemas de classificação, ou seja, os indexadores relacionais, que descreveriam as relações conceituais numa gradação entre associação e discriminação (LANCASTER, 1993),

como mostra a Fig. 4. As relações de Farradane foram extraídas dos esquemas de desenvolvimento do raciocínio propostos pela psicologia infantil, segundo a qual as relações ocorrem em estágios.

O primeiro estágio associativo é a percepção simples sem referência ao tempo; o segundo é a associação temporária entre idéias; e o terceiro é a associação fixa (permanente) de idéias. Os estágios de discriminação são: coincidência simples (conceitos difíceis de discriminar), não-distinto (conceitos que têm muito em comum) e conceituação distinta (conceitos que podem ser completamente discriminados). (LANCASTER, 1993, p. 58)

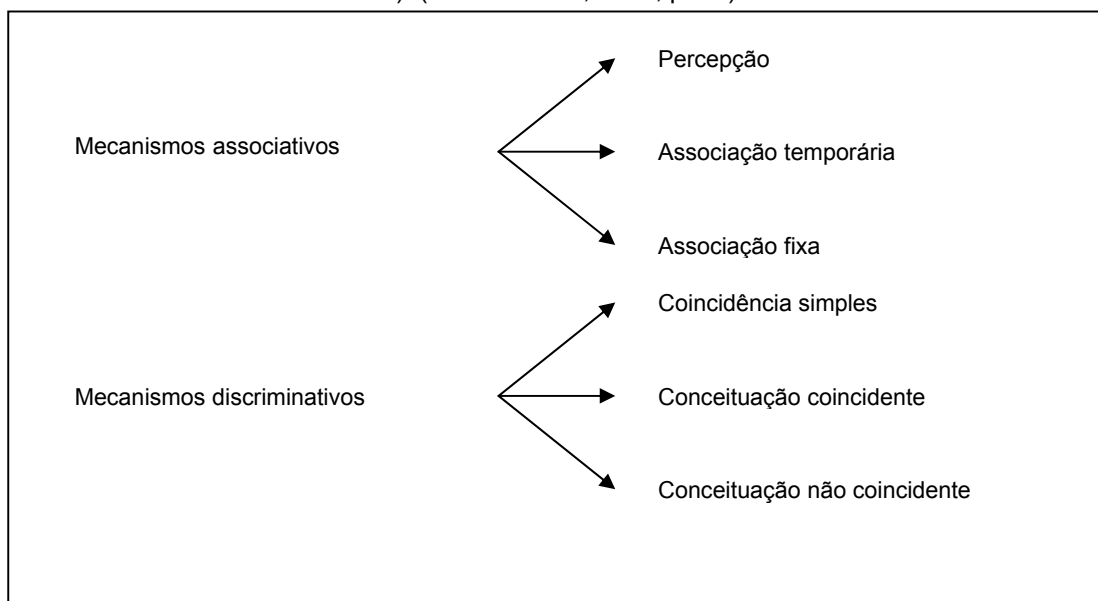


FIGURA 4: *Relações expressas pela indexação relacional de Farradane*
 Fonte: Lancaster, 1993, p. 58, adaptado.

No processo de indexação, enunciados são construídos mediante a reunião de termos “isolados”, usando-se operadores atribuídos a cada mecanismo associativo ou discriminativo. Por exemplo, uma ação é relacionada ao operador “/-“ e, na indexação de um recurso sobre armazenamento de beterrabas, construir-se-ia um analetto pela reunião dos isolados “armazenamento” e “beterrabas” interligados pelo operador relacional de ação “/-“, formando o seguinte exemplo: Beterrabas / - Armazenamento (LANCASTER, 1993, p. 58).

Assim, combinando cada um dos mecanismos discriminativos com cada um dos associativos, Farradane chegou a um conjunto de relações que utilizou na produção de índices permutados, conhecidos como indexadores relacionais. As relações por ele estudadas expressam: 1 – coincidência, 2 – equivalência, 3 – distinção, 4 – auto-atividade, 5 – dimensional, 6 – ação, 7 – associação, 8 – pertencimento e 9 – causação. Coincidência exprime concordância entre padrões (ex.: Balanças /o aferição). Equivalência exprime igualdade de valores, força, peso (ex.: motores a álcool /= motores a gasolina / - combustão). Distinção exprime relação de diferenciação (ex.: xaropes /(açúcar/) solubilidade). Auto-atividade exprime ações de um organismo sobre si mesmo (ex.: aves /* migração). Dimensional exprime propriedades físicas de medidas das matérias (ex.: contêineres /+ padrões). Ação exprime ações temporárias das coisas em geral (ex.: minério de ferro /- fundição). Associação exprime qualquer combinação sintática possível de ser fixada naquele momento, que não seja de pertencimento nem de dependência funcional (ex.: beterrabas /- armazenamento /; lavagem). Pertencimento exprime a relação todo/parte (ex.: unidades escolares /(bibliotecas)). Dependência funcional exprime relações cuja função de um depende da ação do outro (ex.: sistemas /- auditoria /: planejamento).

Outros pesquisadores contribuíram para o estudo das relações e funções nos processos lógicos. É o caso de Selye, citado por Lancaster (1993), que desenvolveu, no *Symbolic Shorthand System* – SSS, um conjunto de indicadores relacionais. O principal indicador trabalhado foi a flecha, que mostrava a direção da ação de uma substância sobre um órgão, fato bem significativo, pois o SSS foi desenvolvido para a área médica.

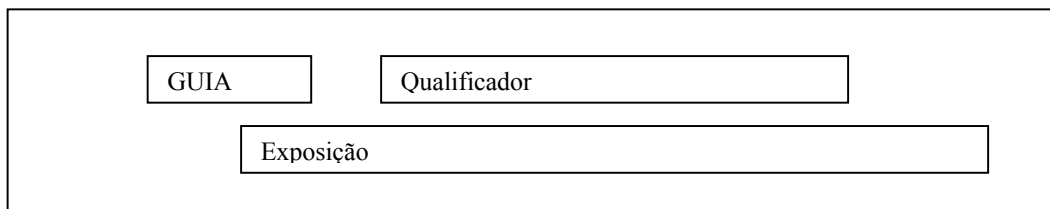
Os estudos e pesquisas que se seguiram, notadamente na década de 1970, incorporaram aspectos de ordem, contexto, relação e função nas elaborações construídas. O PRECIS – *PREserved Context Indexing System*, de Derek Austin, segue essa evolução. Considera que

é metodologicamente fundamentado em duas estruturas: sintática, composta de um esquema de operadores de função que agem como uma espécie de gramática para caracterizar a posição e o significado dos termos no contexto de um cabeçalho de assunto e semântica, que formará o thesaurus do sistema. (FUJITA, 1988, p. 28)

Configura-se o seu desenvolvimento como resultado da evolução dos estudos e circunstâncias reais do final da década de 1960, em que se verificava grande efervescência de idéias sobre a representação de assuntos e a disponibilização tecnológica de computadores para mecanização das elaborações construídas.

Assim, o PRECIS pôde desenvolver sua principal característica, a construção conceitual baseada em preservação do contexto. Cada termo da frase de indexação é essencialmente dependente do termo que o antecede imediatamente. As relações são expressas em duas linhas e mostram, ao mesmo tempo, um contexto mais amplo e um mais restrito, como nas relações hierárquicas, conforme o Quadro 4:

Quadro 4: *Modelo de representação em dois contextos do PRECIS*



Fonte: Lancaster, 1993, p. 55.

Na relação hierárquica, como mostra o Quadro 4, o conceito no alto é sempre um agregador dos conceitos expressos abaixo dele, representando, por isso, um contexto mais amplo daquele campo conceitual. Situação inversa é refletida pelo conceito mais interno. Na ordem em que aparecem, de baixo para cima, os conceitos expressam, sempre, uma especificidade do conceito que os detém, e representam, nesse sentido, um contexto mais restrito.

Para construir as representações pré-coordenadas na forma do Quadro 4, o indexador deveria empregar indicadores de relações entre os termos componentes da frase de indexação. No caso do PRECIS, Derek Austin precisou de 26 operadores de função para pré-coordenar dados dos documentos bibliográficos (LANCASTER, 1993). Esse tipo de linguagem de indexação formava o que ficou conhecido como indexação em *string* ou fileira (RIVIER, 1992).

Na década de 1970, na Índia, é lançado o sistema *Postulated-based Permuted Subject Indexing Language-POPSI*, resultado de experimentos conduzidos pelo Documentation Research and Training Centre – DRTC, no Indian Statistical Institute, iniciados em 1966. Segundo Fujita (1988), o POPSI é um sistema inteiramente baseado em princípios classificatórios, que utiliza cabeçalhos de classificação como termos de entrada na produção de índices. Além dos termos de entrada, há ainda o seqüenciamento dos termos através de ligação em cadeia sob uma ordem preestabelecida.

Bhattacharyya (1979) explica alguns fundamentos e metodologias aplicados no POPSI, como, por exemplo, o princípio de agrupamento de termos por categorias de estrutura

elementar. Destacam-se, entre os postulados e o trabalho conceitual da linguagem POPSI, os princípios de definição do assunto, o agrupamento, a ordenação e a classificação conceitual, baseados em categorias de estrutura elementar denominadas (D)disciplina, (E)entidade, (P)propriedade e (A)ação – Depa, além da compreensão das categorias, entidades-base e entidades-núcleo. As categorias de tempo, espaço, ambiente e forma aparecem como modificadores das categorias elementares Depa.

Seguem-se, no agrupamento dos termos, a classificação organizativa (estruturas sintáticas) e a classificação associativa (estruturas semânticas) do POPSI. Nas estruturas sintáticas, a definição e a indicação dos relacionamentos entre termos são essenciais. O meio utilizado para descrever os relacionamentos foi o postulado de seqüência de categorias de elementos. Alguns símbolos foram instituídos para identificar os relacionamentos, conforme o Quadro 5.

QUADRO 5: *Indicadores de função do POPSI*

1	“	aspas	referência
2	&	e comercial	relação entre assuntos
3	‘	aspas simples	tempo
4	.	ponto final	espaço
5	:	dois pontos	ação
6	;	ponto-e-vírgula	propriedade
7	,	vírgula	núcleo
8	-	travessão	modificador
9	>	sinal de maior	parte - porção

Fonte: Elaboração própria

Esses indicadores, interpolados entre os conceitos, formam a indexação em cadeia, que tornou a linguagem POPSI estruturada para processamento por computador e permitiu a geração automática de índices permutados.

Os sistemas POPSI e PRECIS centralizaram, nos seus mecanismos internos, todos os requisitos de processos lógicos descritos por Lancaster (1993): atribuição de ordem dos termos, manutenção do contexto de existência, uso de indicadores de relação e função no registro e permutação dos termos.

Embora os estudos anglo-saxões tenham sempre tido uma grande repercussão sobre o estado da arte nas pesquisas sobre linguagens de indexação, fundamentos e metodologia, foi nas experiências ocorridas na Índia, particularmente orientadas por S. R. Ranganathan, que se encontraram fundamentos teóricos mais adequados para o embasamento do modelo aqui proposto, conforme descreve a próxima seção, que aborda a estrutura profunda e sua aplicação nas linguagens de indexação.

2.4 Estrutura profunda

Considera-se necessário, antes de qualquer incursão teórica, compreender o conceito de estrutura. No dicionário Houaiss¹⁰, o vernáculo “estrutura” denota dez diferentes definições de senso comum e quatorze definições aplicadas a disciplinas diferenciadas.

¹⁰ Obra já citada na página 39.

Mais ainda, como locução adjetivada, o vernáculo aparece outras dezesseis vezes, em alguns casos com mais de uma significação.

Um outro caminho a percorrer seria através da compreensão do elemento interpositivo *stru* e sua evolução histórica. Segundo o dicionário Houaiss, o elemento *stru* é derivado do verbo em latim *stru* e significa empilhar, reunir, ajuntar, amontoar, criar, construir, erguer em cultismos. Dentre as diversas interposições que esse elemento constrói a partir do século XVIII, registrou-se o seu uso em termos derivados como desestruturar, estruturar, estruturado, estrutura, reestruturar e subestrutura.

No contexto das linguagens de indexação, a palavra “estrutura” denota diversos sentidos. Dois exemplos primários são a estrutura textual, um dos elementos ferramentais da atividade de leitura de documentos, e a estrutura de sistematização dos termos no vocabulário controlado. Outros casos de uso do conceito de estrutura poderiam ser elaborados, mas, neste estudo, faz-se necessária a compreensão de duas conceituações específicas de estrutura, uma advinda da lingüística e outra da informática. Essa última é abordada nos processos de construção de sistemas digitais e será retomada na próxima seção. Já a conceituação advinda da lingüística é a que ora se faz necessária. Nessa disciplina, estrutura é definida como “rede de associações que se constroem a partir de correlações e oposições entre elementos lingüísticos” (HOUAISS, 2001). Nessa conceituação, as noções de associação, correlação e elementos lingüísticos são as mesmas tratadas, até aqui, nas linguagens de indexação.

Contudo, interessa conhecer não apenas a definição simples de estrutura, mas a forma adjetivada “estrutura profunda”. Encontrou-se que esse termo tem o mesmo sentido de estrutura subjacente, proveniente da área de gramática generativa, e é definido como “representação da frase em nível abstrato, na qual se estabelecem as relações semânticas básicas entre os itens lexicais, cuja ordem linear pode ser modificada com aplicação das transformações que forem necessárias, para derivar a estrutura superficial, mantendo as relações semânticas iniciais” (HOUAISS, 2001).

Grandes teóricos, lingüistas estruturalistas, desenvolveram a temática das estruturas profundas e superficiais, na década de 1960. Com isso, a ciência da informação teve a oportunidade de instrumentalizar alguns conjuntos de conhecimentos em estrutura profunda, no escopo das linguagens de indexação.

Partiu de Ranganathan a teoria de que todas as linguagens de indexação eram estruturas de superfície de uma estrutura profunda. Bhattacharyya (1979) elaborou essa idéia na teoria geral da indexação de assunto, que opera sobre uma linguagem de indexação de assuntos (LIA). Nas palavras de Bhattacharyya:

Uma LIA é uma linguagem artificial desenvolvida baseada em estruturas semânticas intrínsecas, elementos artificialmente postulados, e com estruturas sintáticas de proposição de assuntos (...). A estrutura de uma LIA específica deve ser suposta como sendo a estrutura de superfície de uma estrutura profunda de LIA's. (BHATTACHARYYA, 1979, p. 24)

Na elaboração da concepção da estrutura profunda aplicada às linguagens de indexação de assuntos, Bhattacharyya (1979) demonstra como tais estruturas estiveram presentes no princípio de localização de Cutter, princípios de proposição de

assuntos de Dewey, princípios de concretude de Kaiser e princípios de proposição de assunto PMEST de Ranganathan.

Provavelmente, os lingüistas estruturalistas, já no início do século XX, concordariam com a existência de categorias de unidades de linguagem articuláveis em qualquer sentença. Porém, nos estudos sobre representação e recuperação de informação, essas categorias gerais de manifestação do assunto só seriam elaboradas na ciência da informação na década de 1960.

Bhattacharyya, orientado por Ranganathan, é quem enuncia a existência da estrutura profunda na linguagem de indexação de assunto, definida por ele como uma estrutura “composta por constituintes elementares e regras para a formulação de expressões admissíveis, que são usadas para sumarizar em formulações indicativas sobre o que é o conteúdo de uma fonte de informação” (BHATTACHARYYA, 1981, p.12). Na sua concepção, as proposições de assunto abrangem três tipos de estrutura: semântica, elementar e sintática.

Bhattacharyya afirma que a estrutura semântica, baseada nas relações de gênero/espécie, todo/parte e relações colaterais entre facetas, é responsável pela compreensão e significação. Já a estrutura elementar seria baseada nas categorias elementares que, na maioria das vezes, são artificialmente postuladas para dar reconhecimento e significação de campo semântico a substantivos. Finalmente, as estruturas sintáticas seriam as estruturas lineares, horizontais, das descrições de

assunto. Portanto, uma frase se configuraria como uma seqüência de ocorrências dos constituintes elementares.

Ao pensar sobre as três estruturas, percebe-se uma expansão da visão tradicional de funcionamento das linguagens em torno dos eixos sintagmático e paradigmático (GARDIN, 1965; SAUSSURE, 1981). Bhattacharyya (1981) separou a estrutura semântica em duas dimensões diferentes na sua teoria. Ele adicionou uma categoria para os substantivos que denominou de elementares.

Por outro lado, na esfera das categorias elementares, Bhattacharyya retomou a necessidade de lidar com a ocorrência e explicitação das relações sinônimas, quase sinônimas e antônimas.

Embora as categorias elementares tenham sido definidas por Bhattacharyya, Devadason (1985) foi considerado, neste estudo, um melhor interlocutor da teoria geral da indexação de assuntos. O autor desenvolveu uma tese de doutorado sobre essa teoria e, nos seus textos, as categorias elementares são explicadas como:

- (D) Disciplina – uma categoria elementar que inclui campos de estudos convencionais ou campos criados artificialmente, como física, química, biologia, oceanografia, etc.;
- (E) Entidade – uma categoria elementar que inclui manifestações com correlação percentual ou somente existência conceitual de coisas, como contraste de suas propriedades ou ações realizadas por elas ou nelas.
Ex.: cloro, côco, animal, criança, energia, etc.;

- (P) Propriedade – uma categoria elementar que inclui manifestações denotando os conceitos de “atributo” qualitativo ou quantitativo. Ex.: propriedade, capacidade, eficiência, etc.;
- (A) Ação – uma categoria elementar que inclui manifestações denotando o conceito de ação (“fazendo”). Ações podem manifestar-se como auto-ação (ex.: função, migração) ou como ações externas (seleção, avaliação, etc.);
- (m) Modificador – refere-se a uma idéia usada ou de uso pretendido para qualificação de manifestação sem perturbação do seu todo conceitual. Um modificador (especificador, qualificador, especializador, diferenciador) pode modificar a manifestação de qualquer uma das categorias elementares, tanto quanto uma combinação de duas ou mais manifestações de duas ou mais categorias elementares. Modificadores, geralmente, criam relações espécie/tipo.

Uma manifestação usada nesse contexto é uma idéia relativa a qualquer uma das categorias elementares (D – disciplina, E – entidade, P – propriedade, A – ação).

Já os modificadores, segundo Devadason (1985), podem modificar qualquer uma das manifestações das categorias elementares, bem como a combinação de duas ou mais categorias elementares. Definem-se basicamente como de dois tipos:

- 1) Aqueles que requerem um sintagma ou palavra auxiliar (palavra de função) para serem inseridos entre um termo e seu modificador,

formando um sintagma complexo. Ex.: esfregação usando um tambor giratório;

- 2) Aqueles que não requerem palavra auxiliar para serem inseridos no meio, mas, automaticamente, formam um termo composto aceitável, denotando espécie/tipo. Ex.: “vermelho”, formando o termo composto “rosa vermelha”; “infecciosa”, formando o termo composto “doença infecciosa”.

Devadason divide os modificadores entre comuns e especiais. Os primeiros denotam forma, tempo, ambiente e lugar. Os especiais identificam termos que atuam sobre disciplinas, entidades, propriedades ou ações.

Tanto as categorias elementares quanto os modificadores têm suas funções traçadas por setas desenhadas na Fig. 5, onde se observam as categorias DEPAm e todos os tipos de relação que cada manifestação elementar pode realizar no contexto da sua elaboração teórica.

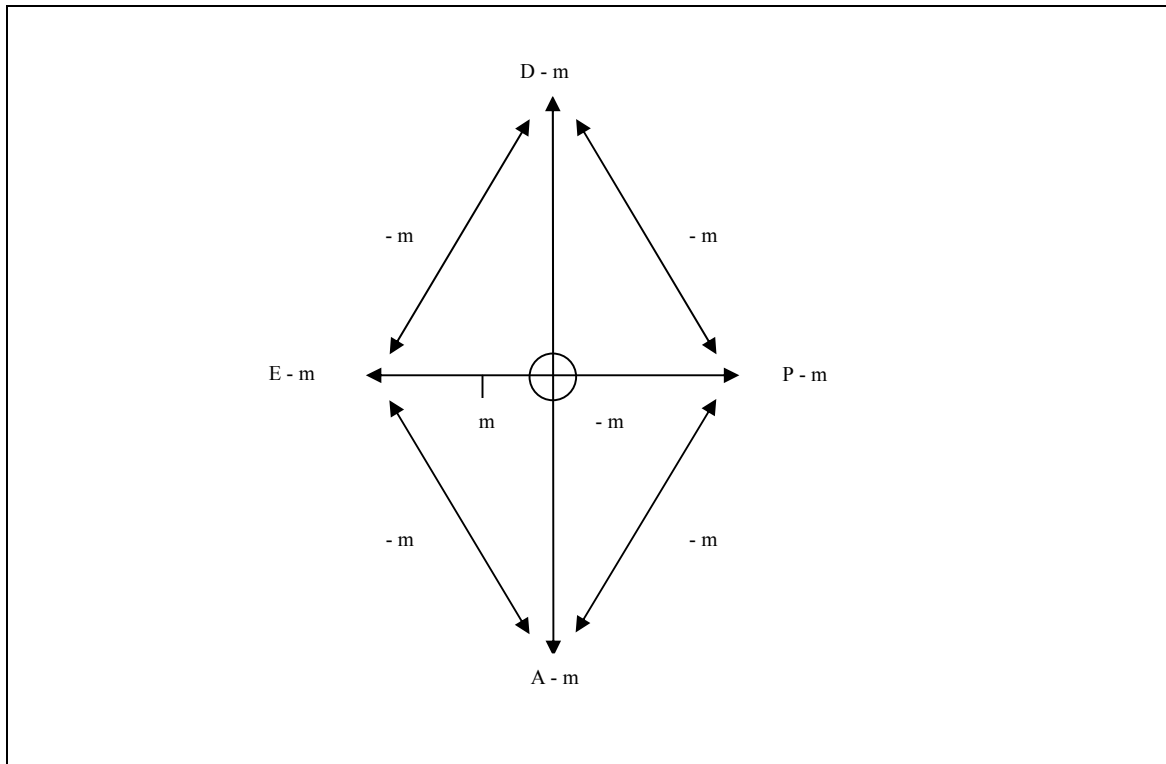


FIGURA 5: *Esquema da estrutura profunda de linguagens de indexação de assunto*
 Fonte: Bhattacharyya, 1981, p.12.

Segundo Devadason (1985), a experiência das categorias em execução no sistema POPSI fez surgir o conceito de *classaurus*: sistema híbrido que permite a pré-coordenação com enumeração tradicional dos assuntos e hierarquização de tabelas classificadas alfabeticamente. Permite também a pós-coordenação, através da construção de um tesouro, com as combinações de tópicos, incluindo, além de descritores, a especificação de termos sinônimos e quase sinônimos.

Com a metodologia *classaurus*, Bhattacharyya implementou abordagens sobre estruturas sintáticas, semânticas e categorias elementares, formando um conjunto de especificações teóricas apoiadas na definição de estrutura profunda e de linguagens de indexação de assuntos. *Classaurus* é a designação empregada por Bhattacharyya para

nomear um sistema de indexação híbrido, com característica de sistema de classificação, na montagem de listas de cabeçalho de assunto adicionadas de notação, e composição de tabelas de termos por categorias de disciplina, entidade, propriedade e ação, nas quais constam também anotações do tipo notas de escopo, termos sinônimos, antônimos e referência bibliográfica.

A proposta desta pesquisa, de estudar um modelo para metarrepresentação de assuntos está fundamentada na confiança na capacidade de representação da estrutura profunda definida na teoria geral da indexação de assuntos. As manifestações das categorias elementares DEPA são o suporte teórico das práticas de análise e fatoração, interpolação e extrapolação aplicáveis à descrição de recursos e especificadas na seção 4 deste estudo, que aborda práticas metodológicas.

Na próxima seção, complementa-se a revisão de literatura com a questão das estruturas definidas pela área computacional para implementação de metarrepresentações.

2.5 Projeto de bases de conhecimento

São diversos os processos necessários à construção de ferramentas de metarrepresentação. Já foram abordados os mecanismos de representação mental, passando-se pelas representações primária e secundária. Também já foram relatados os estudos teóricos sobre técnicas e teorias com fins de metarrepresentação: análise de

assunto, linguagens de indexação e estrutura profunda dos conceitos. Nesta seção, a preocupação é abordar a análise para o desenvolvimento operacional de bases de conhecimento. Acredita-se que esse processo seja responsável por promover a convergência de todas as etapas para uma metarrepresentação conceitual.

2.5.1 Modelagem orientada ao objeto

A análise conceitual orientada ao objeto, no desenvolvimento de bases de conhecimento, pode ser considerada como uma das ferramentas de formação de armazenamento e recuperação de informações, com grande interseção entre os princípios de linguagens de indexação e as estruturas de representação computacionais. Como bases de conhecimento entende-se o conjunto de informações sistematizadas por estruturas computadorizadas, armazenadas em dispositivos digitais e recuperadas por processos lógicos informatizados. Observa-se, nas pesquisas computacionais, que estruturas de representação de conceitos do tipo orientadas ao objeto alcançam grande capacidade de operacionalizar os princípios das estruturas profundas da teoria geral da indexação de assuntos. A orientação ao objeto é uma metodologia de análise e estruturação de fenômenos do mundo real em objetos de informação estruturáveis e implementáveis em *software*, para criação de sistemas de gestão de informações.

Segundo Neelameghan (1992), defensor da idéia de semelhança e de interseção entre as duas abordagens – a análise conceitual da teoria geral da indexação de assunto e a

análise orientada ao objeto computacional –, o processo de conceitualizar a realidade, tal como feito no desenvolvimento de esquemas de classificação para um assunto específico, é muito similar à análise orientada ao objeto. A conceitualização estaria preocupada, basicamente, com elementos a serem representados não *per se*, mas por categorias, ou seja, elementos organizados sob conceitos que são verdadeiras abstrações gerais. Para ele, um objeto é o construto básico de uma análise orientada ao objeto. Sob esse aspecto, o objeto corresponderia ou estaria em correspondência com uma entidade física ou conceitual no mundo real.

Um objeto é uma abstração de um conjunto de coisas do mundo real, de forma que:

- todas as coisas do mundo real do conjunto – as instâncias, tenham as mesmas características,
- todas as instâncias estejam sujeitas a, e em conformidade com as mesmas normas. (SHLAER; MELLOR, 1990, p. 16)

A afirmação de Shlaer e Mellor sobre o que é um objeto gera uma compreensão que aproxima o sentido de objeto ao sentido de entidade, conceito visto na teoria geral da indexação de assuntos. Seguindo essa mesma orientação, Heaney (1995) afirmou que a “modelagem orientada ao objeto agrupa sob um conceito simples [orientação ao objeto] a ‘entidade’ e ‘o que ela pode fazer’” (p. 138).

Heaney (1995), que se dedicou a explorar a orientação ao objeto para aplicação na ciência da informação, concluiu que a modelagem orientada ao objeto parece, em muitos aspectos, ideal para implementações da área. Na sua concepção, a modelagem orientada ao objeto não é um sistema de base de dados, e não estaria, por isso, diretamente ligada a programação de computadores ou linguagens de programação. Seria uma ferramenta de modelagem conceitual para orientar o pensamento sobre os

objetos. Para Heaney (1995), qualquer coisa é um objeto, possui atributos e um conhecido conjunto de operações em que pode participar. Os objetos também possuem as características de poderem ser associados, agregados juntos e/ou agrupados em classes.

No parágrafo anterior, para se abordarem objetos, foi necessário mencionar classes e atributos. A interação entre os termos entidade, objeto, classe e atributo deve ser compreendida a partir de um dos aspectos do paradigma da orientação ao objeto, que é o processo de abstração (PARSAYE et al., 1989).

A abstração de tipos de dados é uma das características mais importantes da orientação ao objeto. Relata-se que uma classe define ambos, a estrutura e o comportamento de tipos abstratos de dados. Tipos abstratos de dados formam objetos, quando estes, pertencentes a uma classe, são chamados instâncias daquela classe (PARSAYE et al., 1989). O paralelo da classe, na estrutura profunda de linguagens de indexação, é a categoria.

Já os objetos podem ser compreendidos como quase sinônimos de entidade. Heaney (1995) afirma que a diferença entre objetos e entidade, nas atividades de modelagem, é puramente de ênfase. Quando a ênfase da modelagem é dada ao relacionamento do objeto, diz-se tratar-se de uma modelagem entidade-relacionamento, que se poderia facilmente compreender como modelagem objeto-relacionamento. Já quando a ênfase é dada ao objeto do mundo real, por si só, suas propriedades e as funções que realiza, desligados da forma de implementação a ser conduzida posteriormente, denominar-se-

ia modelagem orientada ao objeto, que pode ser compreendida também sob o conceito de modelagem orientada à entidade.

Então, um objeto ou entidade cuja tipologia de dados é abstraída passa a ter definidos os seus comportamentos e tipos de atributo, mais comumente conhecidos como campos de registro. A Fig. 6 apresenta a organização dos conceitos ora expostos.

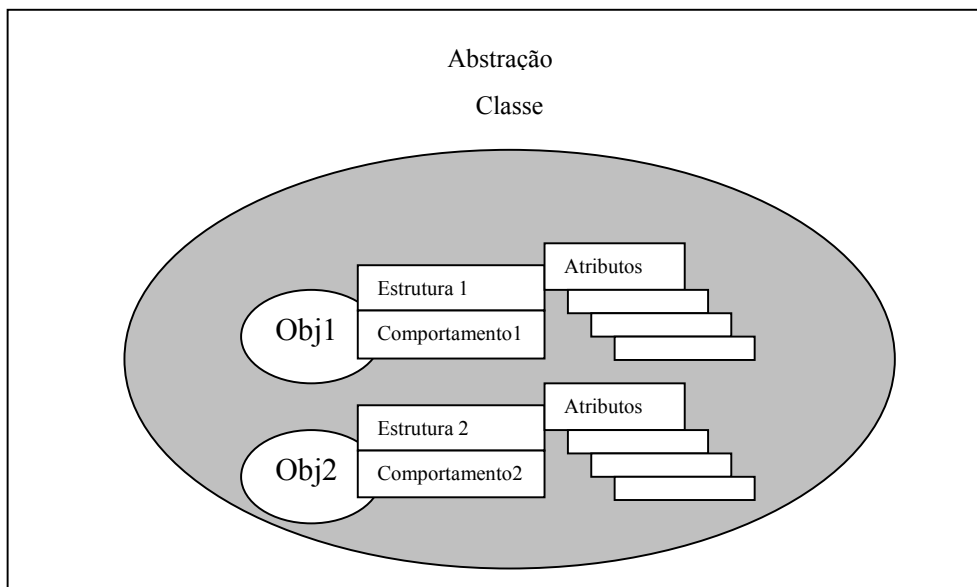


FIGURA 6: Esquema explicativo da abstração de tipos
Fonte: Elaboração própria

Pode-se observar que o esquema DEPAm tem consonância de estrutura entre as manifestações das categorias entidade, propriedade, ação e a descrição dos objetos, atributos e comportamentos, conforme a comparação construída pelo Quadro 6. A estrutura profunda, como extrato do processo de conceitualização da teoria geral da indexação de assuntos, é uma abordagem ampla, preocupada com a universalidade dos conhecimentos representados e, por isso, envolve a categoria disciplina. Na

modelagem orientada ao objeto, a manifestação de disciplina poderia ser compreendida pela idéia de domínio. Um domínio remete ao contexto de ocorrência das classes de objetos. A idéia que expressa é mais de especialização e, portanto, não se aplica à delimitação ideal de áreas do conhecimento.

QUADRO 6: *Equivalência entre processos de modelagem conceitual orientada ao objeto e estrutura profunda de linguagem de indexação - DEPAm*

Equivalência de processos de conceitualização	
Orientação ao objeto	DEPAm
Classe	Categorias
Objetos	Entidades
Atributos	Propriedades
Comportamentos	Ações

Fonte: Elaboração própria

Até aqui, explora-se a equivalência entre os dois processos de conceitualização (Quadro 6), um proveniente da teoria geral da indexação de assunto e o outro da análise orientada ao objeto. Restam abordar os padrões de descrição de recursos, ou seja, a forma como se aplicam à implementação técnica das estruturas profundas. Assim, abordam-se, a partir deste ponto, as estruturas para descrição de metadados, que são padrões para implementação de estruturas profundas em *software*, definidas por esquemas de descrição de recursos – RDF (*Resource Description Framework*). Com isso, pretendeu-se estabelecer, através da revisão de literatura, uma ligação pragmática entre a teoria das estruturas profundas e o modelo proposto nesta pesquisa.

2.5.2 Esquema de descrição de recursos – RDF

O esquema de descrição de recursos RDF, sigla de *Resource Description Framework*, é basicamente composto de duas partes: um modelo para elaboração de metadados e um caso de linguagem, que especifica uma sintaxe, ou seja, ampara a expressão dessa linguagem.

2.5.2.1 Modelo RDF

O modelo RDF opera através do conceito de tripla. Uma tripla é uma expressão simples composta de três elementos: um recurso com a propriedade e o valor. Existem duas visões de tripla, como explicam Ahmed *et al.* (2001). Na primeira, as triplas são recursos com propriedades e, na segunda, são relacionamentos entre recursos. Ao definir uma tripla no esquema RDF, conforme a primeira visão, os autores explicam que:

Na especificação RDF o nome, parte do par nome/valor, é considerado como uma propriedade, e o assunto do metadado é considerado como um recurso. Uma tripla então se torna a combinação de três partes – um recurso com a propriedade e valor. (AHMED et al., 2001, 113)

Na verdade, esclarecem os autores, o padrão RDF foi influenciado por diferentes origens: mundo da gestão do conhecimento, bases de dados, programação orientada ao objeto, dentre outras. As duas visões expressas são a habilidade do esquema de refletir tanto propriedades de recursos quanto seus relacionamentos.

A melhor forma de expressar teoricamente o funcionamento da tripla no esquema RDF é através de diagramas de nós e arcos, chamados de grafos diretamente etiquetados. Essa forma de representação é adequada principalmente quando as propriedades e/ou relações a serem expressas se tornam muito complexas. As Figs. 7 e 8 mostram como é o esquema simplificado de tripla, expresso em grafos.

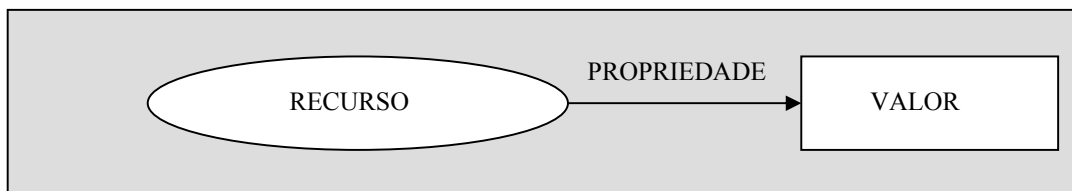


FIGURA 7: *Esquema de tripla expresso em grafos*
 Fonte: Ahmed et al., 2001, p. 115, adaptado.

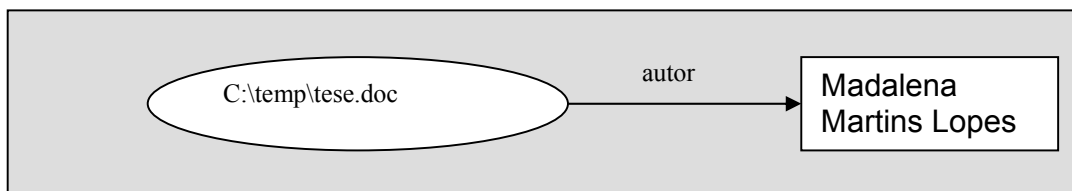


FIGURA 8: *Exemplo de representação em tripla expresso em grafos*
 Fonte: Ahmed et al., 2001, p. 115, adaptado.

Convencionalmente, o formato oval exprime um recurso, o arco (na verdade uma seta), o nome de uma propriedade ou relacionamento, e o quadrado exprime uma expressão literal. Não raro, o arco liga dois recursos. Nesses casos, desenha-se um arco entre dois formatos ovais.

2.5.2.2 Sintaxe RDF

Uma sintaxe para o RDF pode ser construída através da representação em XML – *Extended Markup Language*. Aqui, como no item 2.5.2.1, sobre o modelo RDF, não se objetiva aprofundar em minúcias essa sintaxe, mas apenas dar uma noção de como é construída uma representação conceitual utilizando-se o modelo RDF/XML. A melhor alternativa para se explicar uma sintaxe é a construção de exemplos. O Quadro 7 tem esse objetivo. É um exemplo da sintaxe de representação em XML/RDF.

QUADRO 7: *Exemplo de sintaxe RDF/XML*

<code><rdf:RDF></code>	1
<code><rdf:Description</code>	2
<code> xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"</code>	3
<code> xmlns:dc="http://purl.org/metadata/dublin_core#"</code>	4
<code> rdf:about="http://www.ePolitix.com/Articles/0000005a4787.htm"></code>	5
<code> <dc:Creator>Craig Hoy</dc:Creator></code>	6
<code> </rdf:Description></code>	7
<code></rdf:RDF></code>	8

Fonte: Ahmed et al., 2001, p. 116, adaptado.

A coluna existente no final de cada linha, com um número, não faz parte da sintaxe. Foi colocada no Quadro 7 apenas para orientar de maneira mais didática a explicação dos comandos RDF/XML de cada linha. Também no RDF/XML a tripla é denominada declaração. No exemplo, vamos encontrar linhas de declaração formadas por partes que expressam o recurso, sua propriedade e o valor atribuído à propriedade.

Para compreensão geral, são necessárias mais algumas considerações. A linguagem de marcação XML permite criar etiquetas para os dados. Esse procedimento é diferente das etiquetas de dados do HTML – *HiperText Markup Language*, que traz

predeterminadas as etiquetas para descrição dos dados. As padronizações de etiquetas e conteúdos, na XML, ficam a cargo dos espaços de nomes, cujo termo de referência mais conhecido é originalmente o inglês *namespace*, um dos atributos do item elemento documento. Contudo, as regras da XML permitem ainda o uso de etiquetas de conteúdo, sem qualquer forma de padronização.

Dois outros conceitos importantes nessa sintaxe são os de item "documento de informação" e item elemento de informação. O primeiro "corresponde à entidade documento, mais tipicamente a um número de elemento e outros itens de informação. Em todo documento XML, até no mais simples documento XML, o número de itens de informações pode ser grande" (AHMED et al., 2001, p. 81). Já o item elemento de informação contém propriedades do documento, como as especificações de espaços de nomes, locais, listas de itens menores do documento, atributos diversos, URI – *Uniform Resource Identifier* (identificador de recurso uniforme), hierarquia de pai ou filho.

Após algumas explicações sobre conceitos gerais que envolvem o modelo RDF/XML, pode-se prosseguir na compreensão da sintaxe descrita pelo Quadro 7:

- A primeira linha introduz um conjunto de informações sobre um item elemento documento `<rdf:Description>`. Para esse documento são declarados itens elementos de informação.
- A primeira informação está na linha 2, na qual a etiqueta `xmlns:rdf` (indica o espaço de nome (*namespace*) que padroniza a propriedade `rdf`, usada na primeira linha). Nesse caso, o valor do atributo é o expresso pelas aspas

“<http://www.w3.org/1999/02/22-rdf-syntax-ns#>”. O conteúdo das aspas está referenciando uma URI do *World Wide Web Consortium* – W3CON.

- A terceira linha, exatamente da mesma forma que a segunda, especifica outro *namespace*, o Dublin Core.
- A quarta linha está utilizando uma padronização de etiqueta do W3CON, pois a especificação inicia-se por *rdf:about*. Segundo Ahmed et al. (2001), as declarações começam com a referência ao recurso que a declaração define como *about*. O valor da propriedade *about* é definido pelas aspas. É uma URI: “<http://www.ePolitix.com/Articles/0000005a4787.htm>”.
- A quinta linha define um atributo do recurso especificado na quarta linha. O atributo criador (*creator*) é declarado como uma etiqueta de atributo definida pelo espaço de nomes Dublin Core *dc:Creator*. Tal fato pode ser compreendido porque a parte da etiqueta *dc* é uma abreviação de Dublin Core e está definida no endereço do espaço de nomes *dc*.
- Finalmente, a sexta linha faz o fechamento do conjunto que descreveu as informações pertinentes a um documento `</rdf:Description>`.

Concluindo esta seção, o modelo e a sintaxe RDF/XML foram especialmente considerados neste estudo porque potencializam a metarrepresentação baseada na teoria geral da indexação de assunto, e mais do que isso, formam o casamento entre as estruturas profundas de linguagem, as estruturas de análise orientada ao objeto e o desenvolvimento prático de todo esse conjunto teórico.

Acredita-se que a descrição de um elemento documental e suas propriedades funcionem de forma natural. O fundamento teórico de operacionalização das representações aproximou-se muito mais do processo cognitivo de metarrepresentação feito pelo indexador, pois, se tomarmos uma frase de indexação na estrutura DEPAm, poderemos implementar sua a descrição num sistema de armazenamento e recuperação digital, construindo a representação em máquina, na mesma estrutura em que foi concebida. Isso posto, pode-se afirmar que os mecanismos de relações semânticas baseados em estrutura profunda seriam passíveis de implementação e, ao menos potencialmente, permitiriam maior qualidade da representação pela explicitação das relações entre conceitos.

Com isso, converge-se teoricamente para a verificabilidade dos pressupostos de que existem na teoria geral da classificação de assunto estruturas que funcionam como dispositivos para representação do assunto e potencializadores de interpretação, inferência e autogeração de informações para pessoas, como também para processamento, cruzamento e autogeração de informações por programas de computador e que é possível descrever o assunto utilizando ferramentas computacionais descritivas para a *web*, de maneira a obter maior potencial semântico, ou seja, potencial de interpretação, inferência e autogeração de informações para pessoas, bem como para processamento, cruzamento e autogeração de informações por programas de computador.

É necessário, porém, completar a visão do estudo prático de todas essas elaborações teóricas com a inserção do contexto de seu funcionamento: os sistemas de informação.

Por isso, a próxima e última seção da revisão de literatura aborda os sistemas de informação numa visão de construto social, ou seja, local em que se realizam, de fato, as implementações para atender às necessidades dos usuários de ferramentas de armazenamento e recuperação.

2.6 Sistemas de informação

Ao conceito sistemas de informação pode-se atribuir muitos sentidos. A definição do vocábulo, por si só, já condiciona essa ambigüidade. No dicionário Houaiss (2001), são definidos como formados por conjunto de pessoas, de procedimentos e de equipamentos projetados, construídos, operados e mantidos com a finalidade de coletar, registrar, processar, armazenar, recuperar e exibir informação, podendo servir-se de diferentes tecnologias. Já Araújo (1994) os conceitua como mecanismos que objetivam a realização de processos de comunicação.

O conteúdo de revisão de literatura deste estudo (metarrepresentação, análise de assunto, linguagens de indexação, estrutura profunda e projeto orientado ao objeto) são partes integrantes de um todo: os sistemas de informação. São os sistemas de informação, com todos os seus atores, que contêm, na prática, cada um dos papéis aqui elaborados.

Cada tópico da revisão de literatura representa uma subdivisão na esfera do conceito de sistema de informação. Se um sistema de informação é constituído por pessoas,

procedimentos e equipamentos, foram mencionadas as pessoas que nele interagem, (os autores de recursos e o indexador), os procedimentos ou construções de frases de indexação e seu desdobramento em estrutura profunda, inseridas no contexto das linguagens de indexação, e, finalmente, os equipamentos, na breve menção aos computadores como instrumentos processadores de armazenamento, representação e recuperação de informações.

O foco deste estudo se concentra no ciclo de funcionamento dos sistemas de informação, em que se têm a entrada, o processamento e a saída, podendo-se afirmar que os procedimentos acima mencionados ocorrem na subdivisão de entrada. Araújo (1994) aponta como atribuições de um subsistema de entrada tópicos relacionados à seleção/aquisição, descrição, representação, organização de arquivos e armazenamento. Sob esses aspectos, o estudo de estruturas profundas operacionalizadas através da orientação ao objeto e da implementação em estruturas RDF/XML é uma das abordagens da representação de recursos no subsistema de entrada.

São muitos os processos subentendidos quando se pensa em estudar sistemas de informação. Podem ser estudos voltados para processos de entrada, processamento ou saída. Procurou-se, de maneira simplificada, reduzir a ambigüidade dessa compreensão, delimitando o foco teórico nos processos de representação de recursos como uma atribuição da subdivisão de entrada.

Abordados os sistemas de informação num contexto de convergência entre os conteúdos expostos na revisão de literatura, que podem ser considerados como fundamentais para a compreensão do local de articulação dos fenômenos de representação, põe-se a tratar, no próximo capítulo, das especificações metodológicas desta pesquisa.

3 METODOLOGIA

A pesquisa desenvolvida é, segundo Ander-Egg (1973, apud MARCONI; LAKATOS, 1982), do tipo pesquisa básica ou fundamental. Sua característica é a procura do progresso científico e a ampliação de conhecimentos teóricos, sem a preocupação de usá-los na prática. Não se visou a atender interesse de outrem, apenas procurou-se seguir os caminhos que as questões de acesso a recursos vieram traçando ao longo do século XX e início do século XXI. A escolha do tema ocorreu como evolução de outras temáticas relacionadas, como as estruturas de dados para representar conhecimento estruturado.

No caso do objeto aqui estudado, os recursos digitais acessíveis pela *web*, percebia-se uma tentativa de estruturação de contextos de representação de recursos que não chegavam a atingir o nível do assunto. Com isso, o indexador não tem conseguido representar o assunto com semântica processada por computador, pois até que evoluam os programas para interpretação de sintagmas nominais, verbais,

predicações e suficiente mapeamento conceitual, a descrição da semântica de um recurso será tarefa humana.

Não se deve pensar que a estruturação do assunto não havia sido pensada. Pelo contrário, fazendo-se um estudo retrospectivo, percebeu-se que, ao longo das décadas de 1960 e 1970, muitas iniciativas importantes foram tomadas, incluindo a estruturação da atenção do recurso construída em estrutura profunda, por Bhattacharyya, na teoria geral da indexação de assunto.

Outro desafio que também persistia era a existência de estruturas de representação que pudessem ser lidas tanto por computadores quanto pelo leitor humano, de uma forma mais natural que as estruturas de tabelas relacionadas, disponibilizadas pelos sistemas de processamento computacional existentes.

No entanto, aproximadamente entre os anos de 1997 e 2000 novas estruturas computacionais foram construídas e percebeu-se a sua adequação para atender à demanda das estruturas de representação de assunto em estrutura profunda.

A compreensão dessa realidade despertou a iniciativa de se estudar o fenômeno da representação de assunto em face das novas oportunidades de estruturação da sentença para seu processamento híbrido entre pessoas e máquinas.

Esta pesquisa se desenvolveu de abril de 2003 a fevereiro de 2005 e foi delimitada pelo objeto empírico estudado que foram recursos eletrônicos disponibilizados na *web*.

Embora se tenha pretendido coletar recursos em língua portuguesa, essa escolha não se mostrou viável, por falta de disponibilidade de recursos acessíveis cuja característica estivesse de acordo com as linhas traçadas para o trabalho. Assim, optou-se por fazer as buscas na segunda língua mais comum aos brasileiros, o inglês.

Uma questão central neste projeto foi o fato de a revisão de literatura ter fornecido subsídios teóricos, através das idéias resgatadas, para se desenvolver analogia entre as teorias pesquisadas e as estruturas de representação do conhecimento, enquanto objetos reais. Na elaboração metodológica, percebe-se a recorrência a dois métodos: a modelagem e a análise de conteúdo, descritos a seguir.

3.1 Método de modelagem

Sobre a modelagem como metodologia, relata-se que, com o subsídio teórico resultante da revisão de literatura, foi possível construir um modelo de representação temática de recursos. A partir da construção desse modelo, todo o trabalho passou a ser direcionado para o seu teste e sua avaliação. Portanto, pode-se afirmar que o modelo desenvolvido passou a orientar a aplicação metodológica, tanto na fase de coleta de dados, quanto nas fases de análise dos dados e enunciação dos resultados obtidos.

Acerca dos modelos, um dos significados para o vocábulo é o de “algo eminentemente digno de imitação, exemplar ou ideal” (KAPLAN, 1975, p. 265). Também se menciona

que os modelos evocam traços da elaboração de ferramentas reais e, assim, realizam a vocação de fazer sistematizações que servem aos propósitos da ciência.

Segundo Kaplan (1975), os modelos promovem a construção de um contexto significativo, no qual “descobertas específicas podem ser encaradas como pormenores relevantes” (p. 275). Durante a ordenação dos dados, ao passar de uma observação a outra, um modelo promove a noção de perseguição de uma idéia, enquanto se espera que algo aconteça, pois “quando menos, são conscientes, explícitos e definidos” (KAPLAN, 1975, p. 275).

Contudo, se por um lado os modelos simplificam e esquematizam intrincadas elaborações cognitivas, por outro sofrem o peso de algumas deficiências. Segundo Kaplan (1975), são perigosas pelo menos seis deficiências. O autor cita, em primeiro lugar, a ênfase demasiada em símbolos, ou seja, a crença inconsciente no poder dos símbolos. Em segundo lugar, a ênfase demasiada na forma, que pode imprimir fechamento prematuro de idéias e esconder imperfeição de conhecimento em relação ao objeto. Em terceiro lugar, a supersimplificação, lembrando que o erro não é a simplicidade, mas o fato de esta poder esconder simplificações erradas, nos lugares errados, provocando a caminhada em direção errada. Em quarto lugar, o autor cita a ênfase no rigor, ressaltando que um objeto só empresta a uma pesquisa a exatidão possível tendo em vista os instrumentos existentes à época da sua realização. Em quinto, Kaplan lembra a leitura gráfica, que apresenta o perigo de não se compreender que o modelo é um modelo particular de representação, levando à crença em realidades idealmente existentes. E a última deficiência citada por Kaplan é o realismo

pictórico, que trata da compreensão de modelos como imagem e semelhança do que está sendo modelado, sem considerar que o modelo pode não demonstrar variáveis endógenas (internas) e exógenas (externas).

Embora Kaplan tenha identificado esses seis perigos da metodologia de pesquisa baseada em modelos, é importante frisar que, à época da primeira edição da sua obra, 1969, o autor se mostrava preocupado com o modismo dos modelos nas ciências do comportamento. Algumas vinham adotando a metodologia de elaboração de modelos e, a seu ver, com muito pouca elaboração sobre o pensamento, dedicando-se substancialmente à orientação empírica das ciências do comportamento, pautadas por “submissão servil ao êxito tecnológico das ciências físicas” (MARX apud KAPLAN, 1972, p. 299).

Neste estudo, a modelagem foi orientada pelo conjunto teórico formado pela estrutura profunda – DEPAm (seção 2.4) e a modelagem orientada ao objeto (seção 2.5). Essas teorias levaram ao fornecimento de elementos para a elaboração do modelo de representação DEPAm-OR, descrito na próxima seção.

3.1.1 Modelo de representação de assunto

O modelo proposto, representado na Fig. 9, é, fundamentalmente, a interjunção dos campos teóricos da análise de assunto (seção 2.2) por estrutura profunda (seção 2.4) e a modelagem orientada ao objeto (seção 2.5) da informática. O modelo demonstra

como os conteúdos coletados já detêm algum nível de representação por estrutura profunda. Percebe-se a capacidade de articulação entre os conteúdos coletados (determinações de campos disciplinares, entidades, propriedades e processos) e o modelo de representação de assuntos proposto, com a explicitação das mesmas categorias de disciplina, entidades, propriedades e ações.

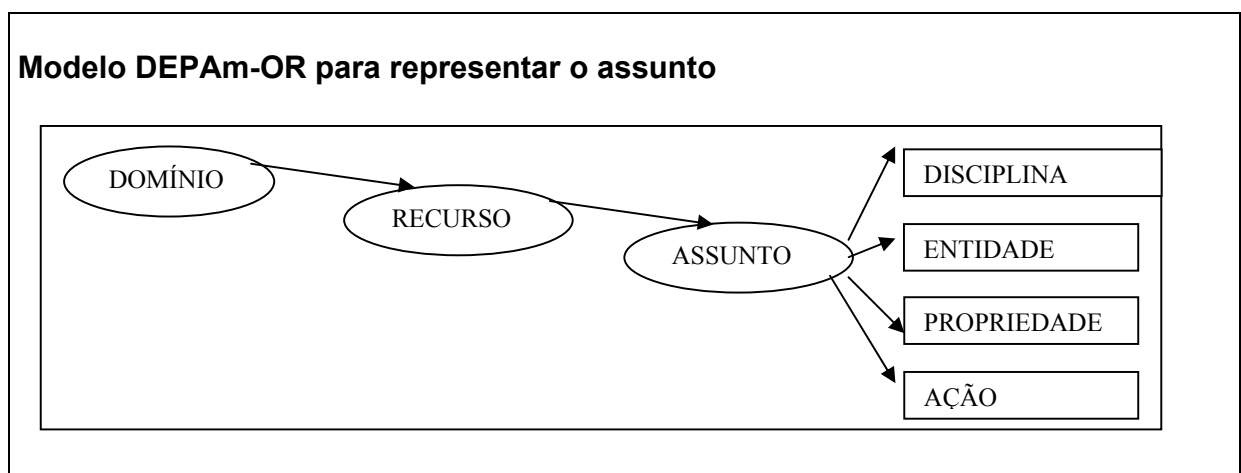


FIGURA 9: Esquema explicativo do modelo de representação proposto – DEPAm-OR
 Fonte: Elaboração própria

No modelo, articulam-se quatro tipos de estruturas de informação. A primeira é o domínio, ou seja, o ambiente que mantém o recurso. O espaço de criação, manutenção e gestão do recurso. A segunda é o próprio recurso, com os atributos que lhe são peculiares, como nome, autor, título, URL, publicador, conteúdo, dentre outros.

A terceira refere-se ao assunto, objeto do problema especificado neste estudo. O assunto de um recurso eletrônico tem recebido diversas formas de tratamento. Observa-se a inadequação dos tratamentos recebidos, tendo em vista o problema de pesquisa que vê, na representação atual do assunto dos recursos, a possibilidade de

incremento de estruturas de representação para a construção de maior semântica numa rede, representação que deve ter o potencial de ser processada tanto por pessoas quanto por máquinas. A quarta estrutura está representada por quatro retângulos, nos quais se lêem disciplina, entidade, propriedade e ação. Esses retângulos foram propostos pela teoria geral da indexação de assunto, seção 2.4, na década de 1970. O que o modelo faz é acrescentar-lhes a integração com as estruturas orientadas ao objeto, adequadas aos novos padrões de desenvolvimento das descrições semânticas disponíveis.

Abordando-se seu aspecto geral, o modelo representa uma estrutura de mapeamento conceitual adequada à representação do assunto, como resultado das atividades do indexador e potencial de interpretação, tanto por pessoas como por programas de computador.

Além disso, ao elaborá-lo, a intenção foi mostrar a disposição para conjunção teórica entre campos disciplinares diferentes e, ainda, que refletisse um mecanismo que se põe como parte de um subsistema de entrada (seção 2.6) inserido no âmbito das linguagens de indexação (seção 2.3) e internamente articulado por elementos da análise de assunto (seção 2.2), estrutura profunda (seção 2.4) e modelagem orientada ao objeto (seção 2.5), formando, por sua natureza, um instrumento de recuperação de informações sob o escopo da Ciência da Informação ou um metadocumento com função de metarrepresentação conceitual (seção 2.1).

Adicionalmente ao modelo proposto, primeira orientação metodológica da pesquisa, recorreu-se ainda ao método de análise de conteúdo.

3.2 Método de análise de conteúdo

Kim (1996) afirma que a análise de conteúdo é, além de objetiva, uma estratégia orientada quantitativamente. Em suas palavras, compreende “estudos quantitativos de recursos ou outras formas de comunicação que examinam frequências / padrões de palavras, frases, conceitos, imagens, temas, caracteres, papéis, etc.” (KIM, 1996, apud POWELL, 1999, p. 96). A análise de conteúdo, neste estudo, foi o método que permitiu examinar padrões de conceitos existentes nas definições de atenção de recursos.

Esse método foi aplicado a recursos indexados por serviços de informação. A condição da busca era que esses recursos fizessem a representação do assunto através do uso de linguagem de indexação. Assim, os sistemas de informação pesquisados forneceram os recursos indexados segundo um esquema de classificação do tipo CDD/CDU/LCC ou, ainda, indexados por descritores de algum tesauro.

Sua aplicação se deu através de observação e extração de conceitos dos recursos pesquisados, de maneira a preencher os requisitos de dados do modelo proposto.

Os elementos de dados da análise de conteúdo foram, na primeira experiência, conforme o Quadro 8, o nome atribuído ao recurso recuperado, a estratégia de busca

aplicada, a URL do recurso (endereço na *web*) e a descrição das características observadas. Usou-se uma codificação para cada exemplo, como E1.1, que significa: Exemplo 1, tópico 1.

QUADRO 8: *Campos de registro de sistema de informação pesquisado – Fase 1*

COLETA DE DADOS (1)

- E1.1-Nome:**
- E1.2-Estratégia:**
- E1.3-URL:**
- E1.4-Características:**

Fonte: Elaboração própria

Em face de dificuldades encontradas na primeira etapa da pesquisa, descritas na seção 4.1, que utilizou os elementos para análise de conteúdo do Quadro 8, foi adotada uma outra sistemática. Os elementos coletados na primeira fase foram nomeados como exemplos de 1 a 13.

Outro conjunto de elementos para orientar o processo de análise de assunto foi elaborado, conforme o Quadro 9. Nessa nova estrutura, para constituir a análise de conteúdo, foram enumerados os elementos conceituais: 1-*site* – endereço do *site*, 2-sistema de informação – nome do *site* ou sistema de informação que disponibilizou os dados, 3-Linguagem de indexação – nome da linguagem de indexação utilizada pelo sistema de informação para classificar os recursos recuperados, 4-Estratégia – indicativo da estratégia de busca implementada para chegar ao recurso, 5-URL disponível – endereço onde se pode acessar via *web* o recurso recuperado, 6-Nome –

indicativo do título do recurso, 7-Classificação – menção à classificação verbal ou simbólica atribuída ao recurso, 8-Observações – descrições das características do recurso, 9-Pesquisador – nome do pesquisador que efetuou a pesquisa.

Para representar a estrutura dos conteúdos a serem descritos, convencionaram-se a numeração e as etiquetas de campos: 1) *site*; 2) nome do sistema de informação; 3) linguagem de indexação; 4) recurso: 4.1) estratégia 4.2) título; 4.3) URL; 4.4) pesquisador; 4.5) observações. Adotou-se a convenção C1 para caso 1, C2 para caso 2 e assim por diante.

QUADRO 9: *Campos de registro de sistema de informação pesquisado – Fase 2*

<p>COLETA DE DADOS (2)</p> <p>C1.1-Site: C1.2-Sistema de informação: C1.3-Linguagem de indexação: C1.4-Recurso C1.4.1-Recurso.Estratégia de busca: C1.4.2-Recurso.URL disponível: C1.4.3-Recurso.Nome: C1.4.4-Recurso.Classificação: C1.4.5-Recurso.Observações: C1.4.6-Recurso.Pesquisador:</p>

Fonte: Elaboração própria

A partir da primeira orientação metodológica traçada, modelo e estrutura de análise de conteúdo, foi elaborado um pré-teste das condições de desenvolvimento da pesquisa.

3.3 Estudos preliminares

Para efetuar as buscas na *web*, era necessário escolher uma ferramenta de busca. A ferramenta TodoBr foi escolhida, por fazer a busca nas páginas em português, atendendo à vocação da pesquisa, desenvolvida nessa língua.

Verificou-se que apenas um *site* utilizava a estrutura CDD para dispor os títulos acessíveis sobre temas religiosos. Esse *site* também permitia a navegação para recuperação, utilizando notações simbólicas hierárquicas. No entanto, os recursos “linkados” não foram encontrados (condição denominada, neste estudo, como problema da disponibilidade).

Um outro *site*, com muitos artigos e leis da área de direito em informática, disponibilizava os conteúdos por categorias funcionais organizacionais e não por sistemas classificatórios (condição denominada, neste estudo, como problema da informalidade).

O recurso escolhido para verificar a plausibilidade da proposição metodológica deu-se através de um exemplo, utilizado com algumas limitações:

- ❖ Um texto, na área de sociologia, retirado da seção anexos da tese de Naves (2000), no qual um dos entrevistados, no caso o EP – Grupo 1, associa ao texto lido termos descritores do assunto. Desses termos, inferiu-se a seguinte frase de indexação: Movimento estudantil como uma trajetória de esquerda no regime militar (ditadura) no Brasil em 1968.

A partir da obtenção do exemplo, elaborou-se a atinência do recurso: Movimento estudantil como trajetória de esquerda no regime militar (ditadura) no Brasil em 1968. (fatoração em DEPA) disciplina, entidade, propriedade e ação. Assim, aplicando-se ao assunto do exemplo, têm-se objetos que expressam as manifestações (DEPA):

D-Sociologia
 E-Estudantes
 P-Trajetoária
 A-Movimento
 m_local-Brasil
 m_tempo-1968

A descrição base da abordagem objeto-relacionamento (seção 2.5) é composta de objeto, relação e valor. Aplicada ao assunto, conforme o metadado Dublin Core, tem-se, como mostra o Quadro 10:

QUADRO 10: *Fatoração de assunto – etapa 1*

Objeto (recurso)	Relação (propriedade)	Conteúdo (valor)
Artigo: 1968, o curto ano de todos os desejos.	Assunto	Movimento estudantil como trajetória de esquerda no regime militar no Brasil em 1968.

Fonte: Elaboração própria

Entretanto, o conteúdo do objeto descrito especifica outra subdivisão, conforme as categorias (DEPA) aplicadas ao assunto. Para demonstrar essas outras relações, tem-se a mesma estrutura descritiva, mantendo o formato de tripla, necessário à modelagem orientada ao objeto implementada por RDF/XML, conforme Quadro 11:

QUADRO 11: *Fatoração de assunto – etapa 2*

Objeto (recurso)	Relação (propriedade)	Conteúdo (valor)
(assunto)	1 Disciplina	Sociologia
Movimento	2 Entidade	Estudantes
estudantil	3 Propriedade	Trajetória
trajetória	4 Propriedade_modificador	Trajetória de esquerda
de	5 Ação	Movimento
esquerda	6 Ação_modificador	Movimento estudantil
no	7 modificador_local	Brasil
regime militar	8 modificador_tempo	1968
no		
Brasil em 1968		

Fonte: Elaboração própria

Aplicando-se a fatoração e demonstrando por arcos e nós (estrutura de grafos), conforme o modelo proposto, tem-se, na Fig. 10:

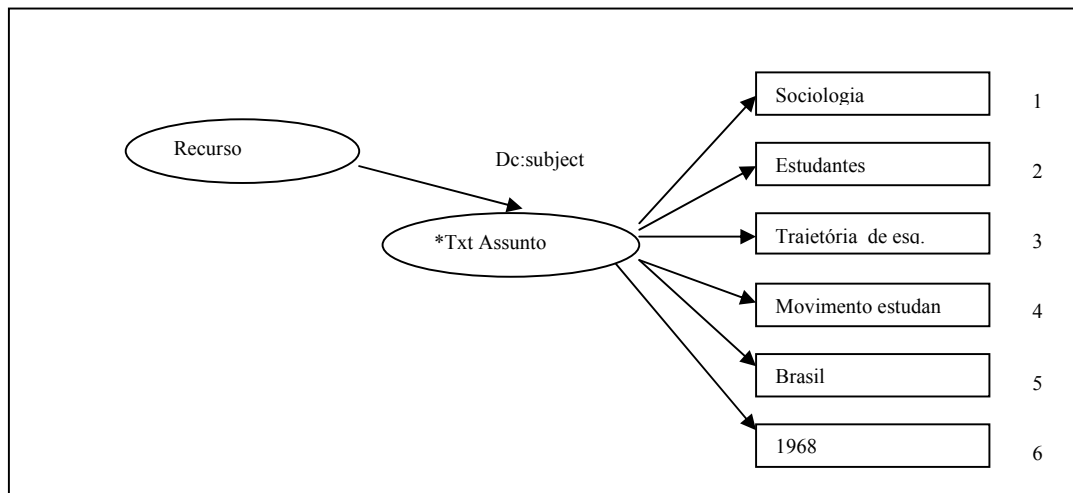


FIGURA 10: *Exemplificação da fatoração aplicada ao modelo proposto*
 Fonte: Elaboração própria

As estruturas computacionais de tripla, por sua vez, são legíveis por *browsers* da *web*. Os conteúdos descritos por essa estrutura (anotação de conteúdo textual) servem para leitura humana e também são processáveis por máquinas de busca, aptas ao processamento de estruturas semânticas dos recursos encontrados. A representação pelo formato RDF/XML ficaria como descrito no Quadro 12.

QUADRO 12: *Demonstração do funcionamento do modelo proposto: combinação de DEPA, OO e RDF/XML*

```
<rdf:RDF>
  <rdf:about = "URL do documento">
    <dc:Creator rdf:resource = "URL do autor do documento">
    <dc:Publisher rdf:resource = "URL do publicador">
    <dc:Subject = "Movimento estudantil como trajetória de esquerda no
    regime militar no Brasil em 1968">
    <dssl:disciplina = "sociologia">
    <dssl:entidade = "estudantes">
    <dssl:propriedade = "trajetória">
    <dssl:propriedade_m = "trajetória estudantil ">
    <dssl:ação = "movimento">
    <dssl:ação_m = "movimento estudantil ">
    <dssl:modificador_local = "Brasil">
    <dssl:modificador_tempo = "1968">
  </rdf:Description>
</rdf:RDF>
```

Fonte: Elaboração própria

O espaço de nome *dssl* é fictício e foi usado, neste exemplo, sem maiores comentários sobre sua definição, mas, num exemplo completo, deveria ter sido descrito em etapa anterior à etiqueta *<rdf:about>*. A definição do espaço de nome envolve o endereço *web* de uma entidade e da lista de conceitos que aquela entidade gerencia, como o vocabulário controlado de um domínio e que o recurso representado estaria utilizando. Esse tipo de linha de codificação não aumenta a compreensão da articulação do

modelo proposto, mas ajuda a aumentar o código RDF/XML, fato que dificultaria sua explicação.

Elaborados o modelo, as estruturas de análise de conteúdo, e verificada *a priori* a viabilidade do prosseguimento dos estudos, tratou-se de especificar, de forma sistematizada, os passos que envolveriam a execução das etapas da metodologia.

3.4 Etapas da metodologia para o objeto empírico

Após as incursões iniciais, relatadas na seção 3.3, foram definidas as etapas para atender aos objetivos do estudo:

- O primeiro objetivo específico mencionou a compreensão de mecanismos de registros do conhecimento. Para esse objetivo, elaborou-se a estratégia de busca de recursos digitais que fossem classificados por diferentes instrumentos, como a Classificação Decimal de Dewey, a Classificação Decimal Universal, a Classificação da Library of Congress e linguagens verbais como tesouros e listas de cabeçalho de assuntos. Na busca por recursos, a frase de indexação deveria estar necessariamente pronta. Como ela não foi encontrada, ficou instituído que uma das etapas da análise de dados seria a construção da frase de indexação.
- O segundo objetivo mencionou o evidenciamento de estruturas e processos de representação conceitual. Foi então efetuada a identificação de estruturas profundas de linguagem (DEPAm) na frase de indexação elaborada para o recurso recuperado, indexado ou não por linguagem de indexação.

- O terceiro e último objetivo mencionou o evidenciamento de um modelo de representação temática amparado por teorias da classificação da Ciência da Informação e estruturas de descrição da Ciência da Computação. Assim, foi implementado um procedimento de identificação de estruturas computacionais para descrição temática do tipo objeto-relacionamento, que se encaixassem no padrão de manifestação estrutura profunda (DEPAm), conforme a modelagem orientada ao objeto, exemplificada na Fig. 8, exemplificação aplicada ao modelo DEPAm-OR proposto.
- Enfim, para o casamento dos objetivos com os pressupostos que vislumbravam a existência de teorias sustentando a classificação e descrição do assunto, tanto na Ciência da Informação quanto na Ciência da Computação, foi elaborado um procedimento analógico de teste do modelo de representação temática proposto. Cuidou-se de estruturá-lo de maneira a conjugar estrutura de linguagem de indexação e estrutura computacional, com potencial de interpretação e inferência por pessoas e autogeração de informações por máquinas, na rede.

O modelo proposto, testado no objeto empírico, foi construído vislumbrando a possibilidade de que programas de computador que processam representação em linguagem XML, propostos pelo W3Con para a *web* semântica, pudessem navegar entre as estruturas DEPAm e inferir o conteúdo representado. Para o processo de inferência, acredita-se que a verificação da hierarquia da categoria do item traduz o que seria o seu teor semântico, levando em consideração as aproximações existentes em

mapeamentos de campo conceitual. Os tópicos especificados anteriormente resultaram nos seguintes passos para execução da metodologia:

1 Coleta de dados

- Seleção de recursos em sistemas de informação com conteúdo classificado por sistema de classificação;

2 Análise dos dados

- Elaboração da atinência, conforme revisão de literatura, item 2.2, determinação do que é um documento;
- Reconstrução da classificação usando estrutura profunda (DEPAm-disciplina, entidade, propriedade, ação, modificadores);
- Formalização gráfica da representação temática, conforme o modelo de representação de assunto primeiramente elaborado;
- Tradução do modelo para a linguagem XML/RDF, implementando-se a representação de maneira legível por pessoas e programas de computador.

A partir daí, a etapa de coleta de dados foi direcionada para alcançar, no mundo real, elementos que pudessem demonstrar, por analogia, a articulação do modelo e sua funcionalidade para a representação de assunto, quando esse assunto proviesse de um domínio existente.

A partir das definições da metodologia, pôde-se então elaborar a coleta de dados.

4 A COLETA DE DADOS

A coleta de dados se deu em duas fases e com procedimentos distintos. A primeira ocupou-se da busca na *Internet* aberta. Essa fase quase não trouxe contribuições, embora tenha sido trabalhosa, conforme descrito na seção 4.1. Usaram-se, para a pesquisa, os recursos disponibilizados por dois motores de busca: o Google e o MetaCrawler. O motor de busca TodoBr foi abandonado, porque sua vocação é a busca em português do Brasil e, após os testes da metodologia de pesquisa, verificou-se que, se mantida a estratégia da limitação lingüística, praticamente não haveria recursos digitais para se trabalhar.

As limitações no desenvolvimento da fase 1 se deram porque os motores de busca não possuem a opção de recuperar recursos a partir da especificação do campo de metadado, como estratégia de busca. Como a intenção inicial da pesquisa era encontrar recursos na *web* que mantivessem registradas as categorias da indexação a eles atribuída, e como os recursos não mostravam essas classificações nos seus corpos textuais, só restou procurar essa informação nos metadados descritivos dos recursos, isto é, nos códigos de linguagem HTML. Nessa fase, a decepção maior foi constatar que, embora já transcorridos dez anos desde a elaboração dos metadados Dublin Core, muito pouco se incorporou à representação dos recursos eletrônicos.

Na segunda etapa, obtiveram-se as melhores contribuições para a temática da pesquisa, embora tenha sido completamente abandonada a busca por recursos na

Internet através da digitação simples da estratégia de busca na caixa de texto do motor de busca.

A segunda fase (seção 4.2) ocupou-se da busca de dados orientada para recursos disponibilizados por instituições especializadas, como o Centro de Documentação e Informação da Usiminas, o Centro de Documentação e Informação de uma grande indústria de aeronaves que, por não ter autorizado a divulgação do seu nome, foi neste estudo denominada Companhia X - Cx, o *site* do Reino Unido BUBL e apenas um recurso disponibilizado na *web*, pela editora Eldritch Press, o único encontrado na primeira fase da coleta de dados.

Essas instituições foram escolhidas por formas e motivos variados. A Usiminas pelo reconhecido trabalho de indexação através da CDU e acesso facilitado por contatos pessoais. A Companhia X foi contatada porque desenvolve correntemente pesquisas sobre representação e recuperação de informações baseadas principalmente num tesouro que já acumula descritores resultantes de vinte anos de trabalhos de indexação. O trabalho de pesquisa da Companhia X já havia sido abordado por esta pesquisadora numa disciplina do Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais (PPGCI/ECI/UFMG). O *site* BUBL foi indicado por um professor da ECI/UFMG, a partir de suas pesquisas sobre o uso dos sistemas de classificação na *web*. E, finalmente, o recurso da Eldritch Press foi obtido pela estratégia de busca na *web*, com o uso de motor de busca, executado na fase 1 da coleta de dados.

A seguir, encontram-se detalhados os procedimentos adotados na coleta de dados, na fase 1 (seção 4.1) e na fase 2 (seção 4.2).

4.1 Coleta de dados – fase 1

Durante as buscas por *sites* de conteúdo indexados por linguagens de indexação, a maioria dos resultados obtidos foi insatisfatória. O objetivo de encontrar conteúdos indexados com os instrumentos CDD/CDU/LCC ou tesauro foi alcançado com dificuldades. Após grande quantidade de buscas infrutíferas, decidiu-se descrever o tipo de recuperação que se obtinha, a fim de ilustrar os problemas encontrados. As seqüências de casos, a seguir, apresentam *sites* recuperados, suas características e algumas observações.

As investigações se deram a partir de estratégias de pesquisa digitadas no motor de busca Google e descritas em cada exemplo incluído neste estudo. Os *links* que o Google retornava eram verificados, um a um, através do resumo apresentado, no qual se buscavam características que indicassem conteúdos indexados por CDD, CDU, LCC ou tesauro. Ao acessar o *link* escolhido, procuravam-se, no texto, indicações sobre indexação temática. Não encontrando, o código-fonte da página era então verificado, à procura de metadados de assunto. O acesso ao código-fonte de uma página se deu pela escolha, no menu de ferramentas, das opções <exibir><código fonte>. Logo, uma janela do *software NotePad* (ou bloco de notas) é aberta e o código fonte em *html* é mostrado.

A seguir relatam-se os 13 casos coletados. Cada caso investigado foi anotado sob quatro aspectos, mencionados na seção 3.2.

E1.1-Nome: Ausência_scheme.doc

E1.2-Estratégia: CDD-CDU-LC-DDC-UDC-LCC-classificação-classification

E1.3-URL: http://ancienthistory.about.com/library/bl/bl_text_aristoteles_categories.html

E1.4-Características: É comum que um recurso disponha somente dos dados: título, palavras-chave e descrição, como no exemplo recuperado nesta URL, Fig. 11. Nessa figura, que é parte do código HTML da página recuperada, aparecem como metadados apenas as variáveis palavras-chave (META NAME="keywords" CONTENT="Aristotle categories") e descrição (META NAME="descrição" CONTENT="The Categories by Aristotle").

```
<TITLE>Aristotle's Categories</TITLE>
<META NAME="keywords" CONTENT="Aristotle categories">
<META NAME="description" CONTENT="The Categories by Aristotle">
<!--GIHEDSTR-->
<META NAME="zgitemplate" CONTENT="7.0">
<meta name="pd" content="Thursday, 05-Aug-2004 22:20:27 GMT">
```

FIGURA 11: Parte do código HTML do recurso Ausência_scheme

Fonte: Elaboração própria

No código-fonte da página não se encontraram metadados referentes a um domínio de nomes que implicasse um sistema de classificação. O uso de um esquema clássico de classificação do conhecimento, com explicitação do assunto em notação ou cabeçalhos de assunto, parece limitado a serviços de informação especializados.

E2.1-Nome: CDD_211204.doc

E2.2-Estratégia: CDD-CDU-LC-DDC-UDC-LCC-classificação-classification

E2.3-URL: <http://deweybrouse.org/>

E2.4-Características: Web site que usa CDD para classificar outros sites, forma um catálogo de sites organizados pelas categorias principais.

Ao se clicar num dos links mostrados, acessava-se outro site (ex.: www.enc.org/) e não necessariamente arquivos com conteúdo indexado por uma linguagem. Os sites acessados, em geral, não traziam referências à categoria de conteúdo que mostravam,

isto é, não era possível, através deles, identificar uma representação de categoria que indicasse sua localização, como área do conhecimento, num todo maior, na Internet.

E3.1-Nome: CDD_Brincar

E3.2-Estratégia: CDD-CDU-LC-DDC-UDC-LCC-classificação-classification

E3.3-URL: <http://thrall.org/dewey/>

E3.4-Características: *Web site* que ajuda a memorizar os códigos de classificação da CDD.

A estratégia de busca digitada na caixa de texto do Google não mostrou ser eficiente. Todas as recuperações pareceram refletir o casamento da estratégia de busca com o conteúdo dos recursos e não com os metadados do assunto. A menos que o assunto estivesse descrito textualmente no recurso, fato não esperado, percebeu-se que o motor de busca não fez a recuperação desejada, ou seja, com conteúdos classificados por CDD, CDU, LCC.

E4.1-Nome: 1subject_arqu

E4.2-Estratégia: subject DDC-UDC-LCC

E4.3-URL: <http://collectionscanada.ca/6/17/s17-208-e.html>

E4.4-Características: No *site* dessa URL encontrou-se um *link* para a seção <publications>. Ao clicar nesse *link* obteve-se acesso a um recurso que descrevia um guia prático para pesquisadores em arquivos. Interessante é que a troca de estratégia de busca retornou um recurso em que as palavras “subject” e “DCC” apareciam no corpo do texto. Contudo, o mais importante foi perceber que a palavra “subject” aparecia entre os *metatags* do código-fonte, conforme mostra a Fig. 12. O texto teve sua cor esmaecida para valorizar a linha que se desejava destacar neste estudo.

A alteração da estratégia de busca para identificar, nos metadados do código-fonte, a representação do assunto classificado por linguagem de indexação, pareceu melhorar os resultados da pesquisa na Internet.

```

<!-- META START -->
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
<meta name="description" lang="eng" content="First-time archives users frequently experience something of a
culture shock as they struggle to adapt to new concepts and procedures. This guide introduces new users to
some of the "customs" of archives.">
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/">
<meta name="dc.title" lang="eng" content="Using archives : a practical guide for researchers">
<meta name="dc.creator" lang="eng" content="Government of Canada, Library and Archives Canada, ">
<meta name="dc.language" scheme="ISO639-2" content="eng">
<meta name="dc.subject" lang="eng" scheme="gcore" content="Archives | Handbooks">
<meta name="searchsubject" lang="eng" content="Research tools">
<meta name="dc.description" lang="eng" content="First-time archives users frequently experience something of
a culture shock as they struggle to adapt to new concepts and procedures. This guide introduces new users to
some of the "customs" of archives."><meta name="sitecode" content="xx">
<meta name="pagelevel" content="1">
<!-- META END -->

```

FIGURA 12: *Parte do código-fonte do recurso 1subject_arqu*
Fonte: Elaboração própria

E5.1-Nome: 2Subject_metadados

E5.2-Estratégia: Dewey Decimal

E5.3-URL: <http://www.sciencedaily.com/releases/2004/11/041124155818.htm>

E5.4-Características: A partir do Google chegou-se ao *site* da URL acima. O conteúdo recuperado mostrou dados sobre o objetivo buscado.

A menção ao nome Dewey, na estratégia de busca, apenas repetiu os insucessos das estratégias anteriores: retornou um recurso sobre o assunto e não um recurso classificado usando o sistema de classificação decimal de Dewey. Ao verificar o código-fonte da mensagem, confirmou-se que não havia, nos metadados, descrição de assunto do recurso. Foram encontrados apenas metadados de descrição e de palavras-chave.

E6.1-Nome: CDD_semantica

E6.2-Estratégia: DDC code

E6.3-URL: <http://tcc.itc.it/people/bentivogli/papers/coling04-ws-WDH.pdf>

E6.4-Características: A partir do Google chegou-se ao *site* da URL acima. O conteúdo recuperado não mostrou dados sobre o objetivo buscado.

Diferentemente do E5, no caso E6 o recurso recuperado não era um artigo de jornal, mas um artigo científico, do qual se esperava melhor representação do assunto nos *metatags*. Porém, nada foi encontrado.

E7.1-Nome: 3Subject_metadados

E7.2-Estratégia: meta name="subject"

E7.3-URL: <http://www.dlib.org/dlib/june97/metadata/06weibel.html>

E7.4-Características: A partir dos *links* apresentados pelo Google, foi acessada a URL <http://mathematik.uni-osnabruek.de/ak-techmik/MathNet/BasicMeta.html>. Dessa URL, então, foi acessado o recurso em questão. Os campos de metadados, verificados através do código-fonte da página, mostraram que o conteúdo recuperado era relevante para a estratégia de busca elaborada. Embora o recurso no campo de assunto apresentasse apenas palavras-chave, no campo descrição, encontrou-se o que seria a atinência do documento, conforme a Fig. 13:

```
<HEAD>
<TITLE>
<META NAME = "DC.title" CONTENT = "The 4th Dublin Core Metadata Workshop Report">
<META NAME = "DC.creator" CONTENT = "Stuart Weibel">
<META NAME = "DC.creator" CONTENT = "Renato Iannella">
<META NAME = "DC.creator" CONTENT = "Warwick Cathro">
<META NAME = "DC.subject" CONTENT = "metadata, conference, workshop, dublin core elements set">
<META NAME = "DC.date" CONTENT = "1997-06-16">
<META NAME = "DC.description" CONTENT = "Summary of the Resultos of the fourth Dublin Core
Workshop">
</HEAD>
```

FIGURA 13: Parte do código-fonte de recurso no arquivo 3Subject_metadados

Fonte: Elaboração própria

A busca pelo conteúdo do metadado descrito na *metatag* do recurso pareceu uma boa estratégia de busca. Elementos relevantes começaram a ser encontrados. Se necessário, seria possível elaborar, a partir da *metatag DC.description*, as etapas restantes do modelo proposto DEPAm-OR. Esse fato é interessante e acena com duas características importantes. A primeira é o registro da análise de assunto no estilo frase de indexação do documento, no metadado *DC.description*. A segunda é o registro de uso de elementos descritivos do conjunto Dublin Core.

E8.1-Nome: 4Subject metaschema

E8.2-Estratégia: meta name="subject"

E8.3-URL: <http://swtafe.vic.edu.au/staff/introduction/guides/webstandards/metadata.asp>

E8.4-Características: Na página de *links* apresentada pelo Google, chamou a atenção o fato de que a linha que descrevia a *metatag subject* identificava o esquema de classificação usado, como mostra a Fig.14.

Pareceu, na pesquisa do E8, que a montagem da estratégia de busca tinha alcançado o ponto-chave para a obtenção dos dados pretendidos, que eram recursos indexados conforme “esquemas” de classificação específicos. O *link* do Google apontava para um recurso indexado segundo o esquema de assuntos da Library of Congress: LCSH- Library of Congress Subject Headings (Fig. 14). O acesso ao *link* mostrado suscitou dúvida quanto à coerência da estratégia, pois o recurso era um documento sobre as políticas de descrição de recursos *web* para o *site*. Por isso, no corpo do texto, havia linhas iguais à linha da estratégia de busca. Por outro lado, ao acessar o código-fonte da página, encontrou-se a especificação do esquema de classificação usado para representar a página, conforme se esperava. A Fig. 15 apresenta uma parte do código que mostra essa ocorrência.

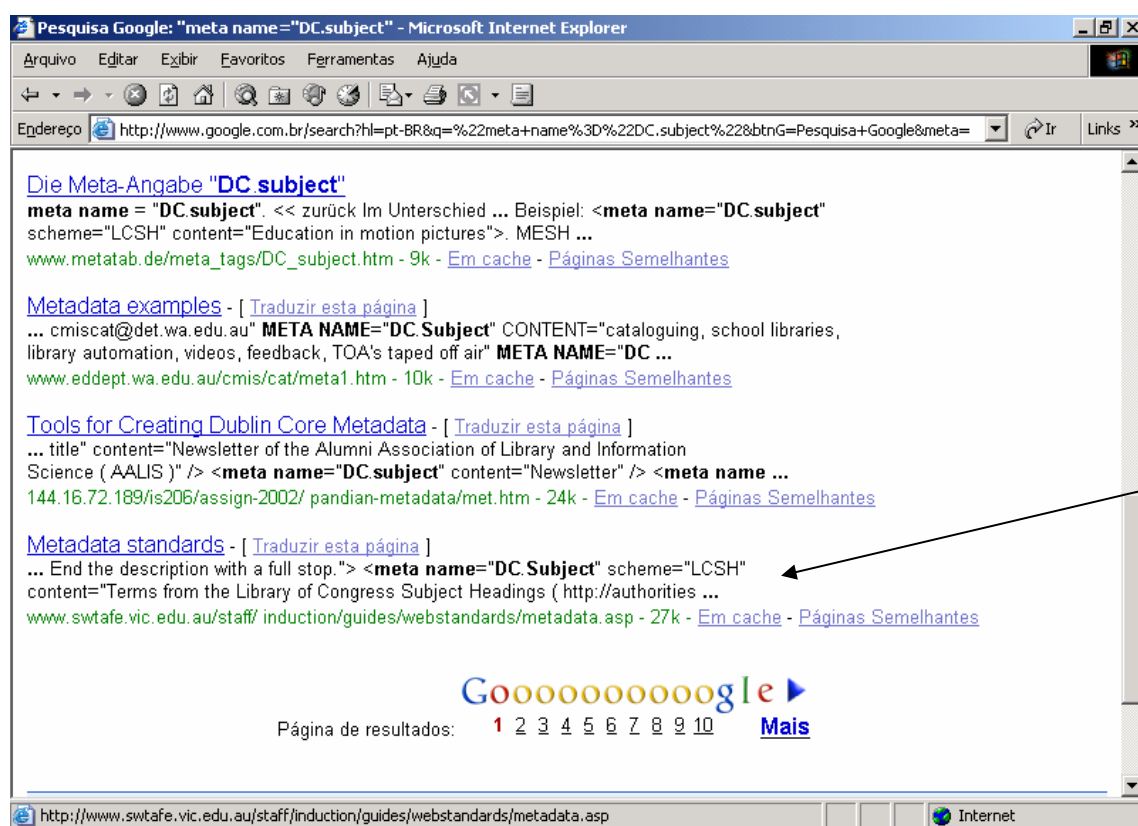


FIGURA 14: Página do Google com a especificação do esquema que descreve o assunto.
Fonte: Elaboração própria

```

<title>Metadata standards</title>
<meta name="Description" content="Metadata standards for South West TAFE">
<meta name="Keywords" content="South West TAFE, metadata, standards; Dublin Core; Edna">
<meta name="DC.Title" content="Metadata standards">
<meta name="DC.Subject" scheme="LCSH" content="Metadata">
<meta name="DC.Identifier" content=
"http://swtafe.vic.edu.au/staff/induction/guides/webstandards/metadada.asp">
<meta name="DC.Date" content="2003-07-07">

```

FIGURA 15: *Parte do código-fonte html que especifica o esquema de descrição do assunto – 4Subject metaschema*

Fonte: Elaboração própria

E9.1-Nome: 5Subject_meta=schema

E9.2-Estratégia: "meta name="DC.Subject" scheme="ERIC" "

E9.3-URL: <http://www.fsu.edu/~gfrost/guide.html>

E9.4-Características: Partindo-se da página do Google, foi acessada a URL mencionada. Novamente, o conteúdo do recurso tratava de assuntos de metadados Dublin Core e, por isso, no corpo do texto apareciam linhas com os termos da estratégia de busca. No entanto, diferentemente do caso E8, ao verificar o código-fonte da página em *html*, encontrou-se, entre diversos elementos descritivos, o metadado `<META NAME=DC.subject.keyword" CONTENT="Dublin Core, document description, cataloging, Internet cataloging, subject headings, subject analysis">`.

Embora no código-fonte dessa URL tenham sido encontradas diversas referências a esquemas de padronização, como, por exemplo, de idioma, de autor, de formato de arquivos para texto e imagem, e de data, para a descrição do assunto, não foi especificado nenhum tipo de esquema de classificação.

Nessa etapa, sentiu-se a necessidade de mudar a ferramenta de busca, pois os recursos recuperados vinham trazendo o conteúdo igual à estratégia de busca, quando o que se buscava eram metadados iguais à estratégia de busca.

Em face desse problema, lembrou-se dos metamotores. Ou seja, havia a possibilidade de se recorrer a motores de busca que buscavam conteúdos, ajudados por outros

motores de busca e que, com isso, incorporavam potenciais de tantos motores de busca quantos teriam sido habilitados a trabalhar quando da sua concepção.

Após algumas tentativas, o metamotor MetaCrawler foi localizado. Os exemplos que se seguem (E10 a E13) foram, então, construídos através de *links* apresentados pelo MetaCrawler.

E10.1-Nome: 6Subject_LCC

E10-Estratégia: scheme="lcc"

E10.3-URL: <http://www.eldritchpress.org/ist/virgin.html>

E10.4-Características: Seguindo-se os *links* apresentados pelo MetaCrawler, verificados um a um, na seqüência, até encontrar algo que fizesse sentido para o tipo de busca empreendido, chegou-se à URL acima. Foi encontrado um recurso que consistia num romance russo de 1877 cuja tradução e publicação em versão *html* ocorreram em 18 de março de 2000: Virgin Soil. O recurso recuperado mostrou, no código-fonte, que estava indexado pela linguagem LCSH e que o recurso-fonte era indexado pelo esquema LCC sob o código PZ3.T936VG. Um outro esquema de classificação notacional também foi usado, o IPL, que, entretanto, foge ao escopo das linguagens em estudo.

Essa recuperação foi a que melhor serviu aos objetivos da pesquisa nessa fase da coleta de dados. Com esse resultado de conteúdos indexados encontrados, reforça-se o tipo de estratégia de pesquisa a ser implementada, ao se usarem os motores de busca de forma a que retornem recursos representados através de uma linguagem de indexação.

E11.1-Nome: 8Subject_LCSH

E11-Estratégia: meta name = "DC.Subject" scheme="LCSH"

E11.3-URL: <http://pictureaustralia.org/members/SLAmap.doc>

E11.4-Características: O *link* indicado pelo MetaCrawler referenciava um recurso indexado pelo LCSH, de nome SLAmap.doc. Ao clicar sobre o *link*, o acesso ao recurso foi barrado por uma caixa de diálogo pedindo nome e senha do usuário. Enfim, não foi possível acessar o recurso indicado pelo MetaCrawler.

O acesso a recursos indexados mostrou-se uma tarefa difícil. Esse tipo de documento quase não se encontra disponível. Os recursos listados por máquinas de busca, geralmente, são protegidos por controle de acesso de usuários, conforme mostra a Fig. 16.

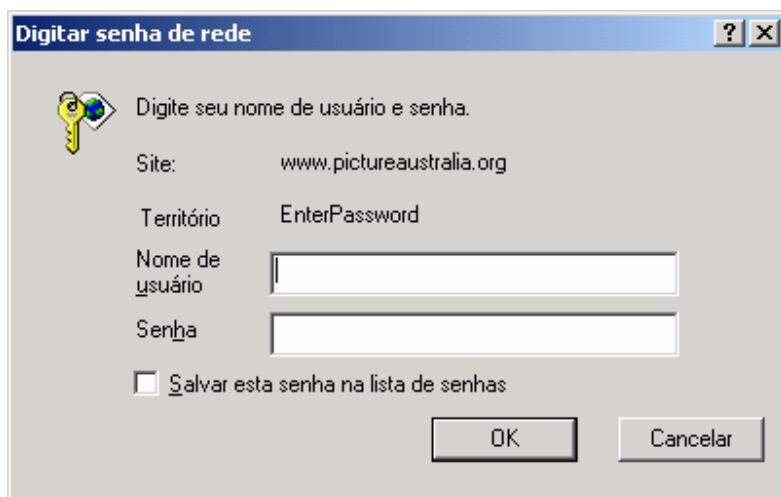


FIGURA 16: Janela normalmente encontrada para acesso a recursos indexados, arq. 7Subject_LCSH
Fonte: Elaboração própria

E12.1-Nome: 8Subject_LCSH

E12-Estratégia: meta name = "DC.Subject" scheme="LCSH"

E12.3-URL: <http://www.oasis-open.org/cover/metaDublin4.html>

E12.4-Características: O recurso apontado pelo *link* do MetaCrawler aparece como conteúdo indexado pelo esquema de classificação LCSH. No entanto, ao se observar o código-fonte do recurso recuperado, não foi encontrada referência para o esquema de classificação de assuntos utilizado. Com isso, concluiu-se que o recurso possuía os dados da estratégia de busca em seu conteúdo e não em seus metadados.

Com essa observação, a revocação obtida pelo acesso à URL do E12 tornou-se desqualificada, em face do objetivo de pesquisa almejado. O que chama a atenção no E12 é que o recurso acessado informa sobre como fazer a descrição de recursos de digitais para acesso via *web*, mas a ele próprio não foram aplicados os princípios de indexação recomendados.

E13.1-Nome: 9Subject_keyword

E13-Estratégia: classification schemes used

E13.3-URL: <http://www.oasis.open.org/cover/classification.html>

E13.4-Características: Ao se trocar a estratégia de busca, foram retornados pelo MetaCrawler sites que classificavam conteúdos de outros sites. Funcionavam como serviço de referência para os sites e não de referências para recursos. Assim, escolheu-se um link que trazia classificação e descrição de páginas da web usando o esquema Classificação Decimal Universal, em www.sosig.ac.uk/. Dentre as classes apresentadas, conforme Fig. 17, escolheu-se aleatoriamente <Geography> e, então, <Demographic Research>. Logo seguiu-se o acesso ao site indexado como conteúdo em pesquisa demográfica, o www.demographic-research.org/. A página mostrada do *Demographic Research on-line journal* mostrou diversos artigos do seu volume onze. Escolheu-se o artigo de número seis, *Is marriage losing its centrality in Italy?* Não se encontrou esquema de classificação de assunto diretamente ligado ao recurso. Ao se observar o código-fonte da página do recurso, foram encontrados os usos mais comuns dos metadados autor, descrição e palavras-chave, sem menção a nenhum tipo de padronização em relação aos elementos Dublin Core.

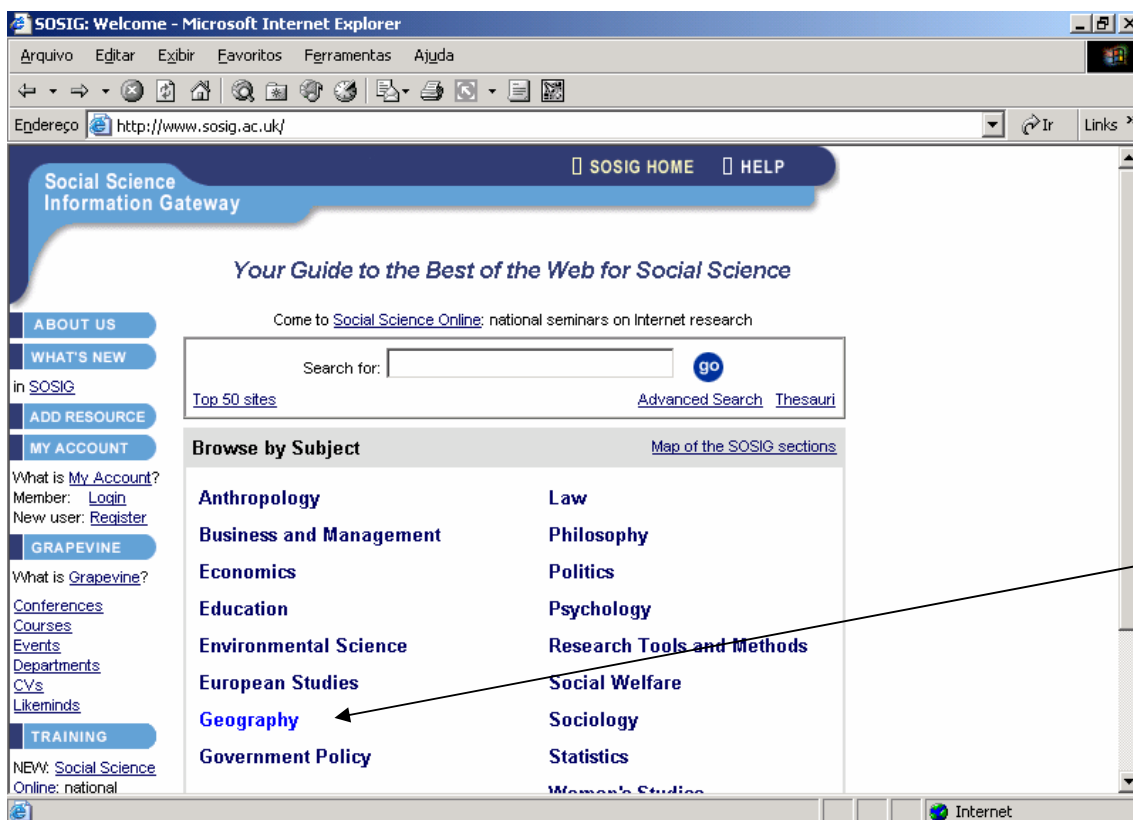


FIGURA 17: Página de web site que organiza conteúdos pelas classes da CDU-9Subject_Keyword.doc

Fonte: Elaboração própria

O fato mais comum encontrado na Web em relação à descrição de recursos recuperáveis pela ferramenta de busca, usando esquemas de classificação do tipo

CDD, CDU, LCC e tesouros, é o uso das classes principais para organizar *sites*. Ao se deparar com os recursos, conteúdos de fato, aqueles acessíveis não sofreram nenhum tipo de elaboração em esquemas de classificação do conhecimento, sobre o assunto de que tratam. Outro fato observado é que os recursos tratados por um esquema de classificação normalmente fazem parte da coleção de uma organização dedicada a esse tipo de trabalho e, em sua maioria, desenvolvem políticas de acesso restrito aos seus recursos.

Sobre a fase 1, considerou-se importante e necessário relatar os problemas e os insucessos ocorridos na coleta de dados, a fim de que essas experiências, embora negativas, possam orientar também outros pesquisadores.

Diante das limitações encontradas nessa primeira fase da coleta de dados, passou-se à segunda fase, em que foram adotadas novas estratégias de busca, detalhadas na próxima seção.

4.2 Coleta de dados – fase 2

Uma segunda fase de coleta de dados foi desenvolvida, principalmente em razão das dificuldades encontradas na primeira fase, como falta de permissão de acesso, de recursos do tipo indexado e de descrição sistematizada do assunto. Nessa segunda fase foi feita a busca de dados em *sites* cujos tipos de conteúdo já se conheciam através de experiências, de contatos e estudos ligados à representação e recuperação de informações.

Os itens trabalhados nessa fase da coleta foram descritos na seção 3.2. Sete casos de recuperação de recursos foram estudados. Os dois primeiros usam classificação CDU, os dois segundos usam classificação CDD, os dois seguintes usam um tesauro e, por último, foi estudado o exemplo 10-E10, o único recurso recuperado na fase 1 da coleta e também trabalhado nessa segunda fase.

Os casos de classificação por CDU e por tesauro foram estudados a partir de recursos obtidos por negociação direta e longa (via *e-mail*) com os responsáveis por seu fornecimento. Os questionamentos e a linha de pesquisa do estudo precisaram ser detalhados aos fornecedores dos recursos e o compromisso moral, entre instituições, foi uma das bases da negociação para a obtenção de respostas.

Diferentemente da fase 1, cada caso relatado não tem um parágrafo que o descreve.

As observações feitas, em cada caso coletado, acompanham os demais elementos descritivos do caso.

C1.1-Site: Usiminas – www.usiminas.com.br

C1.2-Sistema de informação: Centro de Documentação e Informação – BDR – Banco de Dados Referências

C1.3-Linguagem de indexação: CDU + INDASSU (vocabulário controlado)

C1.4-Recurso

E1.4.1-Recurso.Estratégia de busca: aço IF

C1.4.2-Recurso.URL disponível: cd anexo arquivo [recursos.html](#)

C1.4.3-Recurso.Nome: Production of UCL IF steel grades at Voest-Alpine Stahl GmbH

C1.4.4-Recurso.classificação: [669.14 :669.784.001.732] [669.14 :539.215.3.001.732]

(669.14) - 6-Ciências aplicadas -> Tecnologia -> 66-Tecnologia química e afins -> 669-Metalurgia ->

669.14 – Ligas de ferro com carbono (exceto ferro fundido). Aço em geral

(669.784) - 6-Ciências aplicadas -> Tecnologia -> 66-Tecnologia química e afins -> 669-Metalurgia ->

669.784 -> Carbono

(539.215) - 5-Matemática e ciências naturais -> 53-Física -> 539-Natureza Física da Matéria -> 539.2-

Propriedades e estruturas de sistemas moleculares -> 539.21-Propriedades de sistemas moleculares ->

539.215-Pulverulência. Pulveribilidade, superfície específica. Fiabilidade. Distribuição coloidal.

Regularidade, homogeneidade. Firmeza, precisão granular.

C1.4.5-Recurso.Observações: “A pesquisa foi Aço IF no Banco de Dados Referências-BDR, o resultado são as referências abaixo, entretanto os nºs de classificação que estão entre [] surgiram da necessidade

de padronização de termos técnicos específicos na siderurgia. Conforme lhe expliquei por telefone possuímos um vocabulário controlado chamado INDASSU que está relacionado com o nº de classificação CDU, onde o *software* utilizado Microsis não recupera pelo texto contido no arquivo em PDF apenas por palavras, título, autor, ano.. ou nº de classificação em CDU”.

C1.4.6-Recurso.Pesquisador: Luciana Rocha Silva – bibliotecária e gerente de informações técnicas do Centro de Documentação e Informações da empresa.

C2.1-Site: Usiminas – www.usiminas.com.br

C2.2-Sistema de informação: Centro de Documentação e Informação – BDR – Banco de Dados Referências

C2.3-Linguagem de indexação: CDU + INDASSU (vocabulário controlado)

C2.4-Recurso

C2.4.1-Recurso.Estratégia de busca: aço IF

C2.4.2-Recurso.URL disponível: cd anexo [recursos.html](#)

C2.4.3-Recurso.Nome: Desenvolvimento de aços livres de intersticiais – IF – via recozimento em caixa para a indústria automobilística.

C2.4.4-Recurso.classificação: [669.14 :539.215.3.001.732] 621.785.3.076 [669.147 :629.33]

(669.14) - 6-Ciências aplicadas -> Tecnologia -> 66-Tecnologia química e afins -> 669-Metalurgia -> 669.14 – Ligas de ferro com carbono (exceto ferro fundido). Aço em geral

(539.215) - 5-Matemática e ciências naturais -> 53-Física -> 539-Natureza Física da Matéria -> 539.2-Propriedades e estruturas de sistemas moleculares -> 539.21-Propriedades de sistemas moleculares -> 539.215-Pulverulência. Pulveribilidade, superfície específica. Fiabilidade. Distribuição coloidal. Regularidade, homogeneidade. Firmeza, precisão granular.

(621.785.3) - 6-Ciências Aplicadas. Medicina.Tecnologia ->62-Engenharia. Tecnologia em geral -> 621-> Engenharia mecânica em geral. Tecnologia nuclear. Engenharia elétrica. Maquinaria 621.785 -> Processo, procedimentos de aquecimento e tratamento térmico. -> 621.785.3-Recozimento e outros processos afins de tratamento térmico.

(629.33) - 6-Ciências Aplicadas. Medicina.Tecnologia ->62-Engenharia. Tecnologia em geral -> 629-Engenharia de veículos de transporte -> 629.3-Engenharia de veículos terrestres (exceto veículos sobre trilhos).

C2.4.5-Recurso.Observações: Idem C1.

C2.4.6-Recurso.Pesquisador: Luciana Rocha Silva – bibliotecária e gerente de informações técnicas do Centro de Documentação e Informações da Usiminas.

C3.1-Site: BUBL - www.bubl.ac.uk

C3.2-Sistema de informação: BUBL Information Service

C3.3-Linguagem de indexação: CDD

C3.4-Recurso

C3.4.1-Recurso.Estratégia de busca: Navegação – Escolha aleatória da classe CDD – Ciências aplicadas

C3.4.2-Recurso.URL disponível: www foresight.org/conferences/MNT6/Papers/Montemagno/index.html e no cd anexo arquivo [recursos.html](#)

C3.4.3-Recurso.Nome: Constructing biological motor powered nanomechanical devices

C3.4.4-Recurso.classificação: 620.5 > 600-Tecnologia(Ciências Aplicadas) >> 620-Engenharia e operações afins >> 620.5-Nanotecnologia >> 3-Nanomedicina >> *link* Recent Articles on Medical Nanomaterials, Nanobiotechnology, or “Nanomedicine”

C3.4.5-Recurso.Observações: O recurso foi acessado inteiramente como recurso eletrônico indexado na *web*. Contudo, ao se ler seu código-fonte, não se encontraram metadados que permitissem a uma ferramenta de busca acessá-lo pela sua classificação CDD, seja pela notação ou pelo cabeçalho de assunto. Ressalta-se, no entanto, a capacidade das ferramentas de busca de encontrar recursos usando parte do texto. Por curiosidade, uma frase do recurso foi copiada e lançada como estratégia de busca no Google: “ATPase biomolecular motor “ Montemagno. Então, mesmo sem conhecer o sistema de

classificação CDD que leva a esse assunto, a ferramenta foi capaz de recuperar o recurso, o primeiro de uma lista de cento e noventa *links* mostrados.

C3.4.6-Recurso.Pesquisador: Maria Aparecida Lourenço Santana.

C4.1-Site: BUBL - www.bubl.ac.uk

C4.2-Sistema de informação: BUBL Information Service

C4.3-Linguagem de indexação: CDD

C4.4-Recurso

C4.4.1-Recurso.Estratégia de busca: Navegação - Escolha aleatória da classe CDD – Ciências aplicadas

C4.4.2-Recurso.URL disponível: www.asep.org/jeponline/issue/JEPDec2004.shtml e no cd anexo arquivo [recursos.html](#)

C4.4.3-Recurso.Nome: Physiological responses to downhill walking in older and younger individuals

C4.4.4-Recurso.classificação: 612.044 > 600–Tecnologia (Ciências aplicadas) -> Ciências Medicas, Medicina: Revistas -> 612-Fisiologia humana -> 612.044-Exercícios e esportes

C4.4.5-Recurso.Observações: Como no C3, indexação CDD BUBL, catálogo de *sites*, não existem codificação nem cabeçalhos de assunto para os assuntos específicos de cada recurso. O sistema CDD foi usado até para descrever as disciplinas a que alguns *sites* estão ligados. Por isso, na etapa seguinte, recursos como esse precisaram de maior cuidado na elaboração das categorias DEPAM. Ressalta-se que no catálogo BUBL ocorrem inúmeras referências sem recursos indexados. Ao se acessar e fazer escolha aleatória de assunto, foi preciso o acesso a mais de dez indicações de *sites* para conseguir algum que tivesse recursos disponíveis.

C4.4.6-Recurso.Pesquisador: Maria Aparecida Lourenço Santana.

C5.1-Site: *Companhia Cx* (Divulgação não autorizada)

C5.2-Sistema de informação: *Sistema de Informação - Slx* (Divulgação não autorizada)

C5.3-Linguagem de indexação: *Linguagem de Indexação - Llx thesaurus, Llx Taxonomy* (Divulgação não autorizada)

C5.4-Recurso

C5.4.1-Recurso.Estratégia de busca: Acessibilidade - Artigos produzidos pelo grupo do pesquisador, autorizados para publicação externa.

C5.4.2-Recurso.URL disponível: www.cs.utexas.edu/users/pclark/papers e no cd anexo [recursos.html](#)

C5.4.3-Recurso.Nome: Exploiting a thesaurus-based semantic net for knowledge-based search.

C5.4.4-Recurso.classificação: > Computing -> Artificial Intelligence -> Knowledge Representation and Reasoning -> Knowledge representation -> Semantic Networks

C5.4.5-Recurso.Observações: “As categorias tesouro *LlxThesaurus* para esse artigo não estão dispostas em uma árvore simples. Inferência é termo específico de raciocínio. Rede semântica é termo genérico. Ontologia é um termo específico de filosofia. Ontologia, como muitos termos da Ciência da Computação, desconfortavelmente cabem no nosso tesouro. Sob bases históricas, ela está subordinada à filosofia. Por outro lado, ela tem também <relacionada a> conexão com *web* semântica. Esse tipo de contradição acontece continuamente – especialmente nas áreas tecnológicas – assim como nosso tesouro cresce ao longo dos anos. *Llx*, a fonte dos nossos termos, é um grafo com múltiplos relacionamentos mais do que uma hierarquia. Nós também algumas vezes fazemos uso de uma taxonomia restrita *Llx Taxonomy*”

C5.4.6-Recurso.Pesquisador: Steve Woods, PH.d.

C6.1-Site: Companhia X - (Divulgação não autorizada)

C6.2-Sistema de informação: *Sistema de Informação - Slx* (Divulgação não autorizada)

C6.3-Linguagem de indexação: *Linguagem de Indexação - Llx thesaurus, Llx Taxonomy* (Divulgação não autorizada)

C6.4-Recurso

C6.4.1-Recurso.Estratégia de busca: Acessibilidade - Artigos produzidos pelo grupo do pesquisador, autorizados para publicação externa.

C6.4.2-Recurso.URL disponível: www.cs.utexas.edu/users/pclark/papers e no cd anexo arquivo [recursos.html](#)

C6.4.3-Recurso.Nome: Building Concept Representations from Reusable Components.

C6.4.4-Recurso.classificação: > Computing ->Artificial Intelligence -> Knowledge Representation and Reasoning -> Knowledge representation -> Semantic Networks

C6.4.5-Recurso.Observações: “Os termos do *Llx Thesaurus* são um vocabulário controlado e nós os usamos através da escolha de palavras-chave. Mas em adição ao *Llx Thesaurus*, nós usamos várias taxonomias para localização de itens dentro de uma hierarquia navegacional. Os termos hierárquicos que eu lhe forneci provêm do *Llx Taxonomy*. Atualmente, ambos os sistemas são vocabulários controlados, mas o *Llx Thesaurus* é um grafo de termos com múltiplos relacionamentos que nós usamos para estabelecer palavras-chave”. Os termos no *Llx Thesaurus* não formam uma árvore, mas no *Llx Taxonomy* formam”.

C6.4.6-Recurso.Pesquisador: Steve Woods, PH.d.

C7.1-Site: Eldritchpress – www.eldritchpress.org

C7.2-Sistema de informação: Eldritch Press

C7.3-Linguagem de indexação: Library of Congress Classification (LCC) e Library of Congress Subject Headings (LCSH)

C7.4-Recurso

C7.4.1-Recurso.Estratégia de busca: Metatag – Usando-se o metacrawler, foi digitada uma parte da metatag pretendida na recuperação “scheme = “LCC”. Dentre os *links* apresentados pela ferramenta de busca, escolheu-se aquele que parecia ter recuperado recurso a partir do casamento de metatag com a estratégia que também era metatag.

C7.4.2-Recurso.URL disponível: www.eldritchpress.org/ist/virgin.htm e cd anexo arquivo [recursos.html](#)

C7.4.3-Recurso.Nome: Virgin Soil

C7.4.4-Recurso.classificação: LCC–PZ3.T936VG / LCSH->Romance, Literatura Russa, -> Literatura do século XIX.

C7.4.5-Recurso.Observações: Esse recurso enumerou 36 elementos Dublin Core na sua representação. O esquema de classificação descrito em primeira ordem foi o IPL.

C7.4.6-Recurso.Pesquisador: Maria Aparecida Lourenço Santana.

Com os dados coletados dos elementos do caso 7, encerra-se a seção de coleta de dados. No próximo tópico (seção 5), descrevem-se as análises de conteúdo empreendidas.

5 ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS

Foram desenvolvidos diversos procedimentos na análise de dados, conforme a metodologia especificada anteriormente. A primeira etapa envolveu a elaboração da atinência, nos casos em que esta não foi explicitada pelo pesquisador. A atinência, estudada na revisão de literatura (seção 2.2), é o elemento de partida para a análise dos dados, em face do modelo de representação proposto. Ela foi determinada com base nos cabeçalhos de assunto ou descritores, título, resumo e, em alguns casos, através da leitura do recurso. Nessa etapa foi necessário traduzir para o português títulos, cabeçalhos de assunto ou descritores em língua estrangeira.

A segunda etapa envolveu a fatoração e a construção dos casos em estrutura profunda – modelo DEPAm, conforme descrito na teoria de Bhattacharyya (2.4). Não havia o objetivo de se alcançar a exaustividade das fatorações possíveis. O fato considerado importante, conforme o objetivo deste estudo, foi a exemplificação do processo de maneira simples. No entanto, alguns recursos pesquisados estavam indexados como polihierarquias. Nesses casos, os exemplos se tornaram um pouco mais complexos.

Na terceira etapa, os casos foram desenhados segundo a estrutura do modelo proposto DEPAm-OR, mostrando como os objetos gerados, na fatoração, configuram-se em objetos relacionados entre si, atendendo à modelagem orientada a objetos.

Na quarta etapa, as representações foram descritas através do esquema-linguagem RDF/XML, a fim de demonstrar adequação e integração dos construtos anteriores,

estrutura profunda e objeto-relacionamento, todos interagindo num mesmo código, que pode ser interpretado por computadores e pessoas.

Objetivando-se uma melhor compreensão, as etapas da análise dos dados foram agrupadas por casos. Assim, junto com o primeiro caso, estão descritas as etapas de um a quatro, formando um todo coerente. Do contrário, teria de ser determinada a atinência dos casos de 1 a 7, depois feita a fatoração dos casos de 1 a 7, e assim sucessivamente. Por se entender que essa última forma de agrupamento dos casos não formaria uma boa coesão semântica, esse procedimento foi descartado.

CASO 1 – Produção de grades de aço UCL IF na Voest_Alpine Stahl GmbH.

Observação: esse recurso foi descrito utilizando notação CDU. Foram utilizadas três notações diferentes que deram origem a uma polihierarquia de dois campos disciplinares.

C1-1ª etapa – Atinência

Atinência polihierárquica (Ciências Naturais e Ciências Aplicadas) –
Produção de grades de aço UCLF IF, ligas de ferro combinado com carbono, na Voest_Alpine Stahl GmbH

C1-2ª etapa – Fatoração DEPA

D-Disciplina > Ciências aplicadas > Tecnologia -> Tecnologia química -> Metalurgia

E-Entidade > Grades

Entidade_modificador > Grades de Aço

E-Entidade > Aço

Entidade_modificador > Aço UCL IF

Entidade_modificador_modificador > Grades de Aço UCL IF

P-Propriedade > Ligas

Propriedade_modificador > Ligas de ferro

E-Entidade > Ferro

P-Propriedade > Combinação

E-Entidade > carbono

A-Ação > Produção

m-Modificador de local > Voest_Alpine Stahl GmbH
D-Disciplina > Ciências naturais. Matemática > Física > Natureza física da matéria
E-Entidade > Sistemas moleculares
P-Propriedade > Pulveribilidade. Friabilidade. Firmeza

C1- 3ª etapa - modelo DEPAm-OR desenhado no HiperEditor

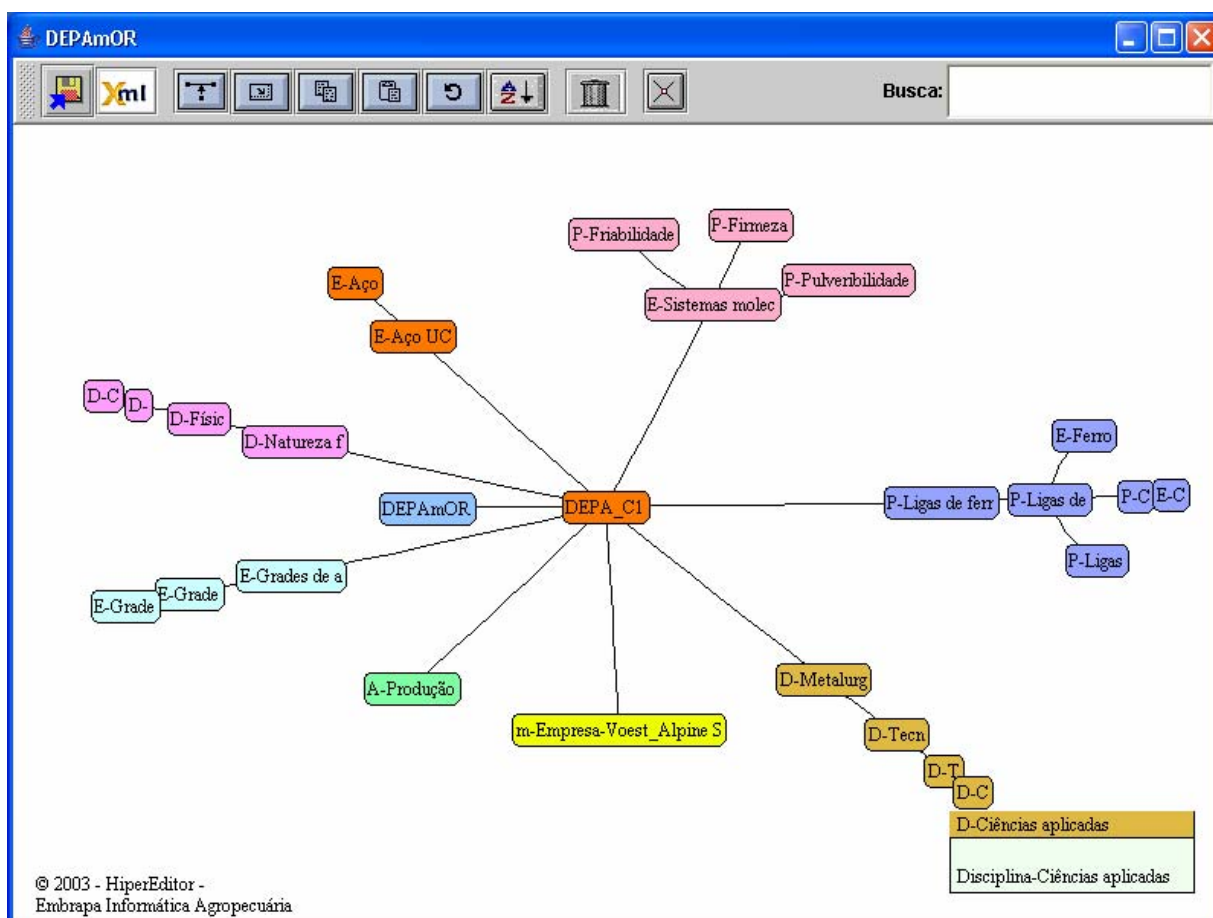


FIGURA 18: Representação do DEPAm-OR no HiperEditor – Caso 1
 Fonte: Elaboração própria

C1- 4ª etapa – Codificação - RDF/XML

```
<rdf:RDF>
  <rdf:about = "recursos.html">
    <dc:subject = "Produção de aço UCLF IF, ligas de ferro combinado com
      carbono, na Voest-Alpine Stahl GmbH">
    <dssl:disciplina = "Ciências aplicadas">
```

```

<dssl:disciplina_m = "Tecnologia">
<dssl:disciplina_m_m = "Tecnologia química">
<dssl:disciplina_m_m_m = "Metalurgia">
<dssl:entidade = "Grades">
<dssl:entidade_m = "Grades de aço">
<dssl:entidade_m_m = "Grades de aço UCL IF">
<dssl:entidade = "Aço">
<dssl:entidade_m = "Aço UCL IF">
<dssl:propriedade = "Ligas">
<dssl:propriedade_m = "Ligas de ferro ">
<dssl:propriedade_m_m = "Ligas de ferro com carbono">
<dssl:entidade = "Ferro">
<dssl:propriedade = "Combinação">
<dssl:entidade = "Carbono">
<dssl:ação = "Produção">
<dssl:modificador_local_empresa = "Voest-Alpine Stahl GmbH">
<dssl:disciplina = "Ciências naturais">
<dssl:disciplina_m = "Matemática">
<dssl:disciplina_m = "Física">
<dssl:disciplina_m_m = "Natureza física da matéria">
<dssl:entidade = "Sistemas moleculares">
<dssl:propriedade = "Pulveribilidade. Friabilidade. Firmeza">
</rdf:Description>
</rdf:RDF>

```

O fato interessante, nesse primeiro caso, é que houve dúvida quanto à ordem de montagem das especificações de cada categoria na ferramenta de navegação hiperbólica¹¹. Por exemplo, a disciplina específica era metalurgia e seu nível mais geral era ciências aplicadas. Então, qual desses níveis deveria estar diretamente ligado ao recurso? O nível mais específico ou o mais geral da fatoração da disciplina?

Nesse caso, C1, optou-se pela representação do recurso partindo-se do nível mais específico do conceito representado, no centro da hipérbole, para atingir o nível mais

¹¹ A navegação hiperbólica, permitida pelos *softwares* livres HiperEditor e HiperNavegador, distribuídos pelo Ministério do Desenvolvimento, Indústria e Comércio, acessíveis para *download* em www.agrolivre.gov.br, permite, no espaço da tela, incluir e navegar na representação conceitual construída. Como se os nós e arcos estivessem dispostos sobre uma superfície circular, a partir do centro, é possível navegar pelos grafos, sendo que o grafo

geral, na extremidade da hipérbole. Assim, quanto mais genérica se tornava a extrapolação da fatoração, mais distante do centro ficava o grafo da sua representação.

Esse estilo de representação, do mais específico para o geral, gerou dúvidas ao se efetuar a descrição. Por exemplo, para disciplina, foi fácil descrever que metalurgia era uma especialidade de questões gerais da física, que, por sua vez, era uma especialidade de ciências sociais aplicadas. Mas, na representação da entidade grade de aço UCL IF, não existe uma representação hierárquica simples, pois a entidade grade é modificada por outra entidade, a substância aço, que, por sua vez, é modificada pelo tipo específico de aço, o UCL IF.

Compreendeu-se, a partir dessa dificuldade, que a fatoração deveria obedecer à naturalidade da expressão da atinência, formando camadas de entidade e propriedade que deveriam ser tratadas isoladamente e não numa hierarquia geral crescente, como se podia fazer com as disciplinas.

CASO 2 – Desenvolvimento de aços livres de intersticiais – IF – via recozimento em caixa para a indústria automobilística

Observação: O recurso foi descrito pela empresa com três notações CDU diferentes. Por isso, sua representação mostra uma polihierarquia de três campos disciplinares diferentes.

C2-1ª etapa – Atinência

Atinência polihierárquica (Ciências naturais e ciências aplicadas)

Desenvolvimento de aços livres de intersticiais, ligas de ferro combinadas com carbono, recozimento em caixa para a indústria automobilística

selecionado promove o redimensionamento de toda a árvore hiperbólica, passando a ocupar o centro e apresentando os grafos antes escondidos pela curva da hipérbole.

C2-2ª etapa – Fatoração DEPAm

D-Disciplina > Ciências aplicadas > Tecnologia. Tecnologia química > Metalurgia

E-Entidade > Ferro

E-Entidade > Aço

Entidade_modificador > Aço livre de intersticiais

P-Propriedade > Ligas

Propriedade_modificador > Ligas de ferro

P-Propriedade > Combinação

E-Entidade > carbono

A-Ação > Desenvolvimento

Ação_modificador > via tratamento térmico

A-Ação > recozimento

m-Modificador local > em caixa

m-Modificador finalidade > Para indústria automobilística

Entidade > Indústria

Entidade_modificador > automobilística

D-Disciplina > Ciências aplicadas > Tecnologia > Engenharia Mecânica > Tecnologia nuclear. Engenharia elétrica > Maquinaria

E-Entidade > Aço

D-Disciplina > Ciências naturais. Matemática > Física > Natureza física da matéria

E-Entidade > Sistemas moleculares

P-Propriedade > Pulveribilidade. Friabilidade. Firmeza

C2- 3ª etapa - modelo DEPAm-OR desenhado no HiperEditor

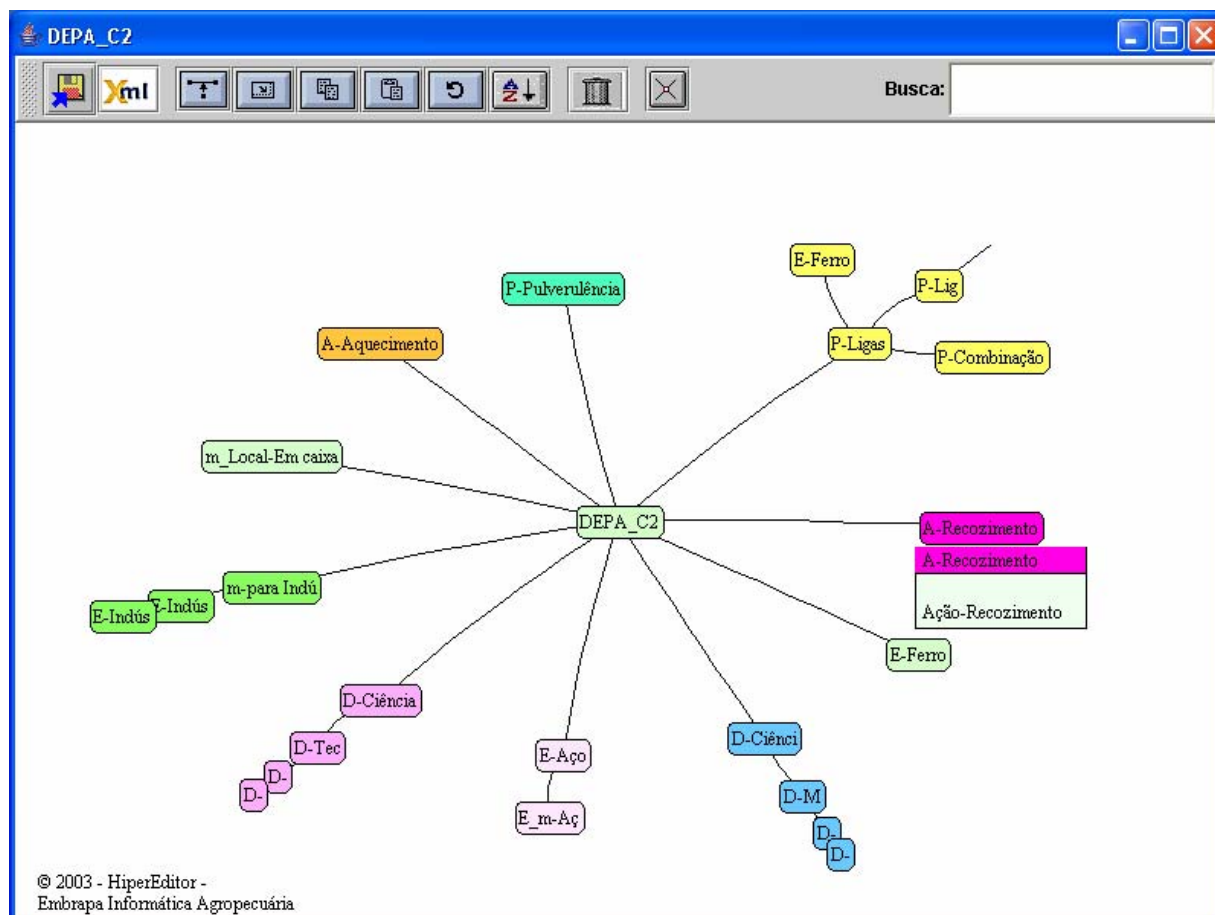


FIGURA 19: Representação DEPAm-OR no HiperEditor – Caso 2
Fonte: Elaboração própria

C2- 4ª etapa – Codificação - RDF/XML

```
<rdf:RDF>
```

```
<rdf:about = "recursos.html">
```

```
<dc:subject = "Desenvolvimento de aços livres de intersticiais, ligas de ferro combinadas com carbono, recozimento em caixa para indústria automobilística">
```

```
<dssl:disciplina = "Ciências aplicadas">
```

```
<dssl:disciplina_m = "Tecnologia">
```

```
<dssl:disciplina_m_m = "Tecnologia química">
```

```
<dssl:disciplina_m_m_m = "Metalurgia">
```

```
<dssl:entidade = "Aço">
```

```
<dssl:entidade_m = "Aço livre de intersticiais">
```

```
<dssl:propriedade = "Ligas">
```

```
<dssl:propriedade_m = "Ligas de ferro ">
```

```

<dssl:propriedade_m_m = "Ligas de ferro com carbono">
<dssl:entidade = "Ferro">
<dssl:propriedade = "Combinação com carbono">
<dssl:entidade = "Carbono">
<dssl:ação = "Desenvolvimento">
<dssl:ação = "Aquecimento.Tratamento térmico">
<dssl:ação = "Recozimento">
<dssl:modificador_local = "em caixa">
<dssl:modificador_finalidade = "para a indústria automobilística">
<dssl:disciplina = "Ciências naturais">
<dssl:disciplina_m = "Matemática">
<dssl:disciplina_m = "Física">
<dssl:disciplina_m_m = "Natureza física da matéria">
<dssl:entidade = "Sistemas moleculares">
<dssl:propriedade = "Pulveribilidade. Friabilidade. Firmeza">
</rdf:Description>
</rdf:RDF>

```

No caso 2, a partir da experiência adquirida com os problemas da descrição do caso 1, elaborou-se a fatoração de maneira mais simples. Se duas disciplinas estavam presentes na atinência, compreendeu-se que deveriam se ligar diretamente ao recurso, e a separação provocada pelas suas especializações deveria ocorrer no nível mais externo da representação em árvore. Assim, fixou-se a forma de representação de geral para específico nas disciplinas e de especialidade para generalidade nas entidades e propriedades. Ou seja, de entidades e propriedades com modificadores (especialidade) para entidades e propriedades sem modificadores (generalidade).

CASO 3 – Construção de motor biológico alimentado por dispositivos nanomecânicos

C3-1^a etapa – Atinência

Atinência (Nanotecnologia) –

Avaliação do desempenho de motor biológico que integra motor de proteínas a motor eletro-nanomecânico para funcionamento híbrido em sistemas orgânicos/inorgânicos.

C3-2ª etapa – Fatoração DEPAm

D-Disciplina > Ciências aplicadas > Tecnologia > Engenharia e operações afins -
> Nanotecnologia > Nanomedicina

E-Entidade > Motor biológico

P-Propriedade > Integração

Entidade > motor de proteínas

Entidade > motor eletro-nanomecânico

A-Ação > Avaliação

Ação_modificador > do desempenho

m-Modificador - finalidade > Para funcionamento híbrido

Entidade > sistemas orgânicos

Entidade > sistemas inorgânicos

C3- 3ª etapa - modelo DEPAm-OR desenhado no HiperEditor

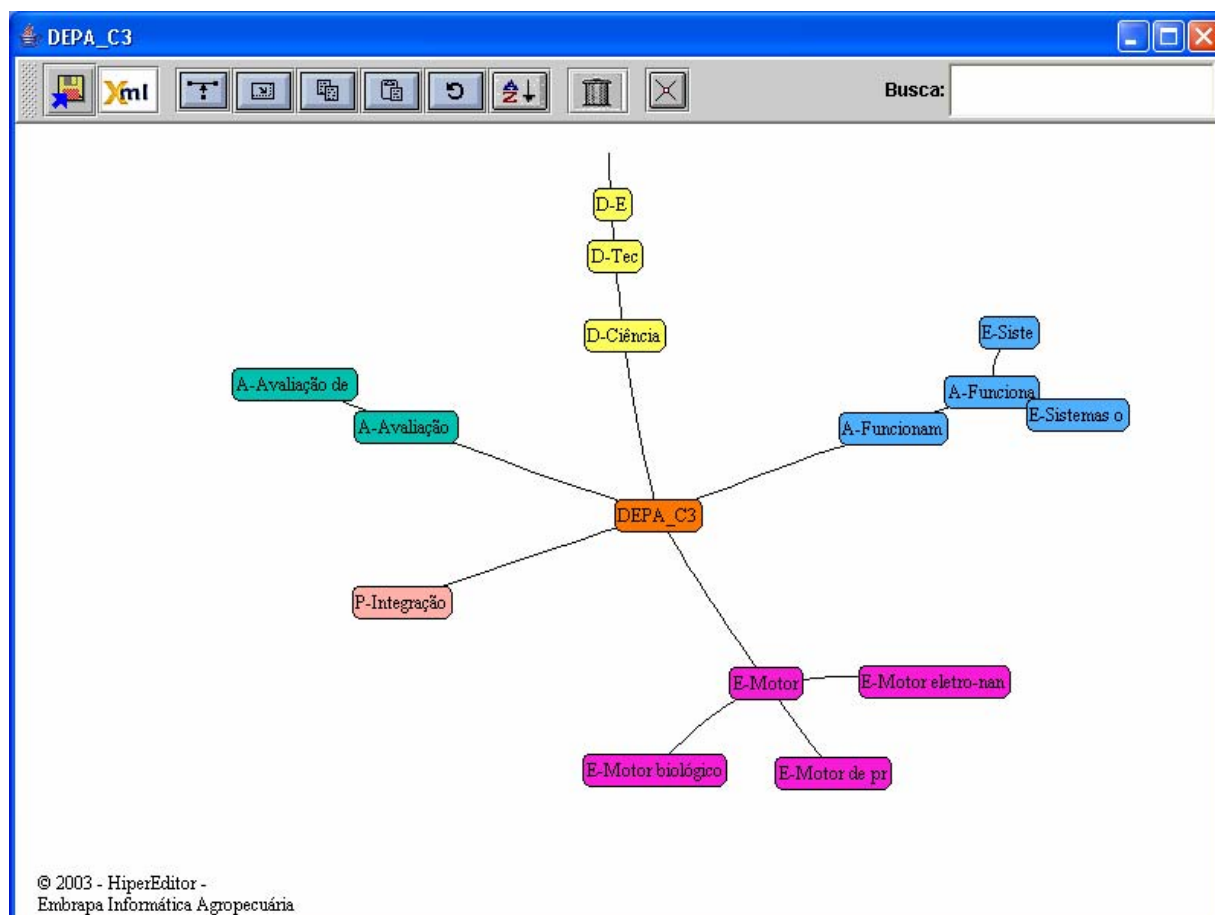


FIGURA 20: Representação DEPAm-OR no HiperEditor – Caso 3
Fonte: Elaboração própria

C3- 4ª etapa – Codificação - RDF/XML

```

<rdf:RDF>
  <rdf:about = "www.foresight.org/conferences/MNT6/Papers/Montemagno/index.html">
    <dc:subject = "Avaliação do desempenho de motor biológico que
      integra motor de proteínas a motor eletro-nanomecânico para
      funcionamento híbrido em sistemas orgânicos / inorgânicos">
    <dssl:disciplina = "Ciências aplicadas">
    <dssl:disciplina_m = "Tecnologia">
    <dssl:disciplina_m_m = "Engenharia e operações afins">
    <dssl:disciplina_m_m_m = "Nanotecnologia">
    <dssl:disciplina_m_m_m_m = "Nanomedicina">
    <dssl:entidade = "Motor biológico">
    <dssl:ação = "integração">
    <dssl:entidade = "Motor de proteínas">
    <dssl:entidade = "Motor eletro-nanomecânico">
    <dssl:ação = "Avaliação">
    <dssl:ação_m = "Avaliação de desempenho">
    <dssl:ação = "Funcionamento híbrido">
    <dssl:entidade = "Sistemas orgânicos">
    <dssl:entidade = "Sistemas inorgânicos">
  </rdf:Description>
</rdf:RDF>

```

A representação do terceiro caso foi simples, dadas as definições elaboradas no caso 2. Esse fato foi animador, pois demonstrou que poucas adequações se fizeram necessárias para a aplicação prática do modelo. As hierarquias atribuídas às disciplinas foram obtidas a partir do caminho de acesso ao recurso na *web*. Não se objetivou conferir se as categorias usadas para classificar os recursos representavam de fato as melhores categorias para o conteúdo apresentado. Apenas se analisou se o modelo proposto era capaz de promover a representação dos recursos, da forma como representavam o conhecimento no ambiente em que este se encontrava.

CASO 4 – Respostas fisiológicas à caminhada descendo montanha em indivíduos mais velhos e mais jovens.

C4-1ª etapa – Atinência

Atenção (Exercícios e Esportes) – Avaliação de respostas fisiológicas à caminhada de descida de montanha em indivíduos mais velhos e mais jovens

C4-2ª etapa – Fatoração DEPAm

D-Disciplina > Ciências aplicadas > Tecnologia > Medicina > Fisiologia Humana > Exercícios e Esportes

E-Entidade > Indivíduos

Entidade_modificador > indivíduos jovens

Entidade_modificador > indivíduos velhos

P-Propriedade > Respostas fisiológicas

A-Ação > Avaliação

Ação > Caminhada

Ação_modificador > Caminhada de descida de montanha

C4- 3ª etapa - modelo DEPAm-OR desenhado no HiperEditor

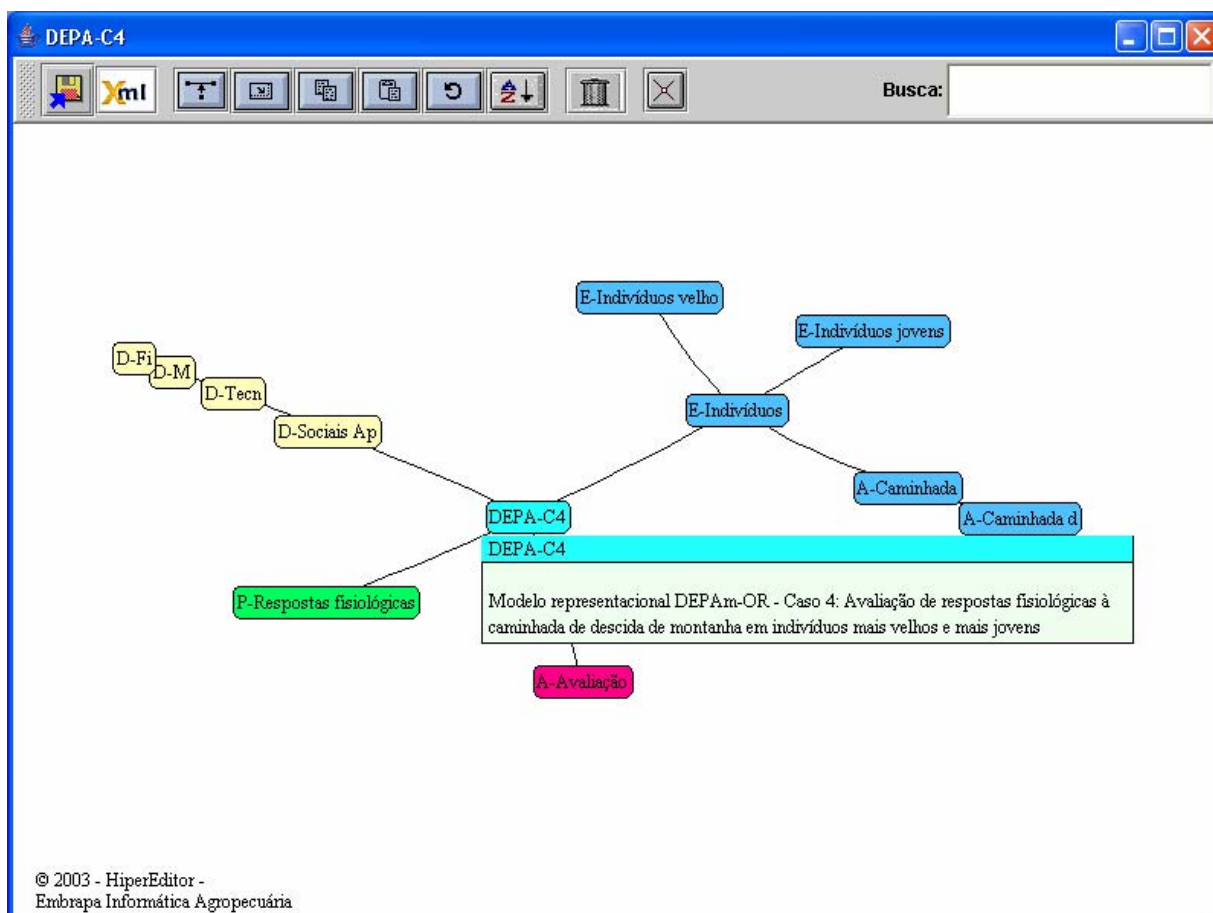


FIGURA 21: Representação DEPAm-OR no HiperEditor – Caso 4
Fonte: Elaboração própria

C4- 4ª etapa – Codificação - RDF/XML

```

<rdf:RDF>
  <rdf:about = "www.asep.org/jeponline/issue/JEPDec2004.shtml">
    <dc:subject = "Avaliação de respostas fisiológicas à caminhada de descida
      de montanha, em indivíduos mais velhos e mais jovens">
    <dssl:disciplina = "Ciências aplicadas">
    <dssl:disciplina_m = "Tecnologia">
    <dssl:disciplina_m_m = "Medicina.
      Fisiologia humana">
    <dssl:entidade = "Indivíduos">
    <dssl:entidade_m = "Indivíduos jovens">
    <dssl:entidade_m = "Indivíduos velhos">
    <dssl:propriedade = "Respostas fisiológicas">
    <dssl:ação = "Caminhada">
    <dssl:ação_m = "Caminhada de descida de montanha">
    <dssl:ação_ação = "Avaliação">
  </rdf:Description>
</rdf:RDF>

```

No caso 4, como no 3, a representação transcorreu sem problemas. A determinação da atinência, item mais subjetivo, foi auxiliada pela leitura quase total do artigo. As etapas de fatoração e representação, na árvore de navegação hiperbólica, transcorreram sem problemas. Muito provavelmente, um indexador iniciante precisará de um pouco de treinamento, a fim de aprender a elaborar as fatorações ou até sentir segurança ao elaborá-las.

CASO 5 – Exploiting a thesaurus-based semantic net for knowledge-based search.

C5-1ª etapa – Atinência

Atinência – Construção de um motor de busca baseado em conhecimentos gerados a partir da rede semântica de conceitos de um tesouro.

C5-2ª etapa – Fatoração DEPAm

D-Disciplina - Ciências naturais. Matemática. Computação. Inteligência artificial. Representação do conhecimento e raciocínio. Representação do conhecimento. Redes semânticas.

E-Entidade - motor de busca

Entidade_m - motor de busca baseado em conhecimento

E-Tesouro

P-Propriedade > Rede semântica de conceitos

A-Ação > Construção

C5- 3ª etapa - modelo DEPAm-OR desenhado no HiperEditor

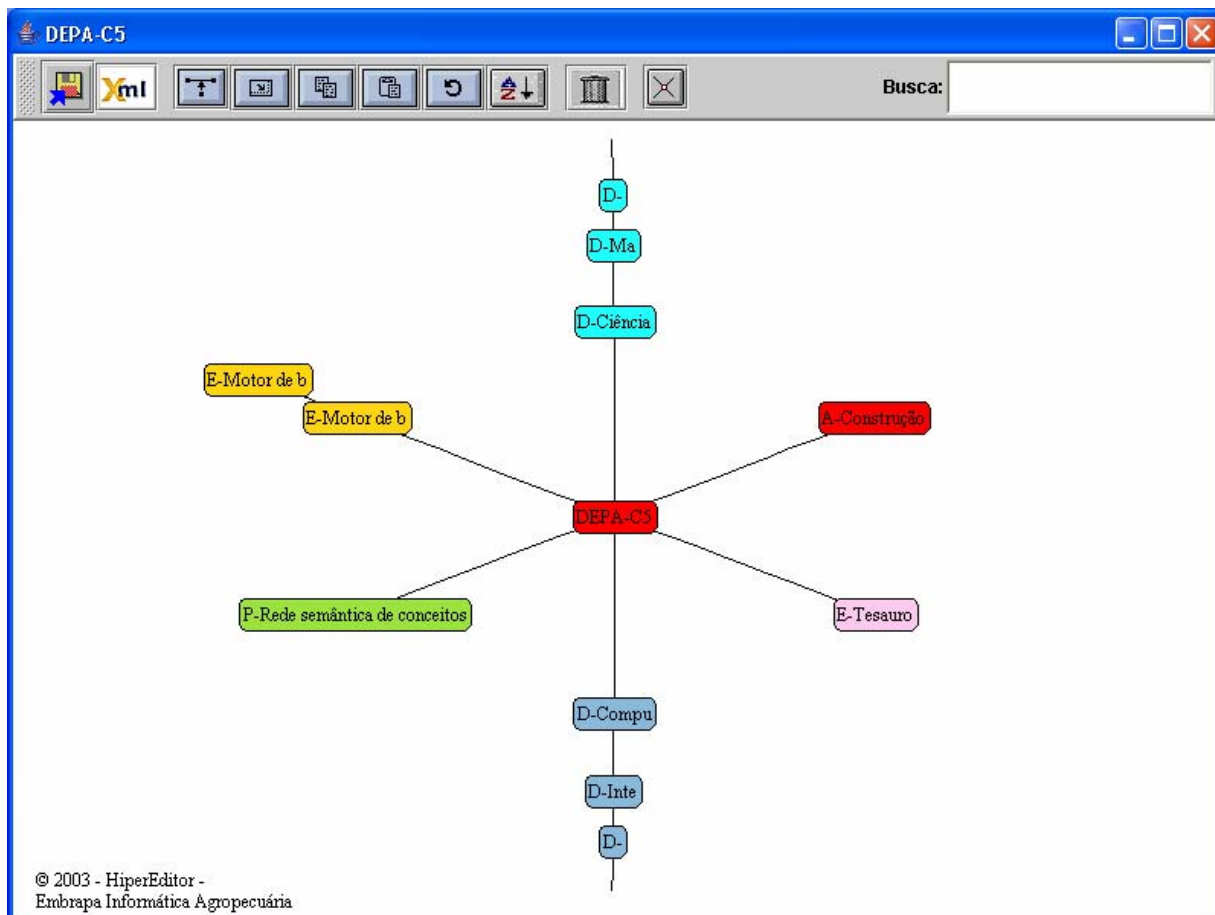


FIGURA 22: Representação DEPAm-OR no HiperEditor – Caso 5
Fonte: Elaboração própria

C5- 4ª etapa – Codificação - RDF/XML

<rdf:RDF>

<rdf:about = "www.cs.utexas.edu/users/pclark/papers">

<dc:subject = "Construção de um motor de busca baseado em conhecimentos gerados a partir da rede semântica de conceitos de um tesouro ">

<dssl:disciplina = "Ciências naturais">

<dssl:disciplina_m = "Computação">

<dssl:disciplina_m_m = "Inteligência artificial">

<dssl:disciplina_m_m_m = "Representação do conhecimento e raciocínio">

<dssl:entidade = "Motor de busca">

<dssl:entidade_m = "motor de busca baseado em conhecimento">

<dssl:entidade = "Tesouro">

```

        <dssl:propriedade = "Rede semântica de conceitos">
        <dssl:ação = "Construção">
    </rdf:Description>
</rdf:RDF>

```

O caso 5, na representação construída, reflete o que o indexador do sistema chamou de taxonomia restrita, pois os termos apresentados pelo vocabulário do tesouro remetiam a diversos campos do conhecimento. Embora a empresa Cx não tenha fornecido outras opções taxonômicas para a representação do recurso, tem-se a convicção de que o modelo proposto poderia fazê-lo, de tantas maneiras quantas fossem necessárias para traduzir adequadamente os conceitos taxonômicos do recurso num domínio conceitual.

Restrições taxonômicas têm sido uma das ferramentas utilizadas para dar a um recurso uma única localização para o seu armazenamento. Por isso, e pela quantidade de opções possíveis para se escolher uma taxonomia principal, o modelo proposto não fica limitado ao arranjo físico dos recursos. Seu compromisso é com o mapeamento conceitual do recurso em relação ao campo semântico em que se localiza.

CASO 6 – Building concept representation from reusable components

C6-1^a etapa – Atinência

Atinência – Construção de sistema de representações conceituais com modelo de composição de componentes reutilizáveis.

C6-2^a etapa – Fatoração DEPAm

D-Disciplina > Ciências naturais. Matemática. Computação. Inteligência artificial. Representação do conhecimento e raciocínio. Representação do conhecimento. Redes semânticas”
E-Entidade > Sistema
Entidade_m > Sistema de representações conceituais
P-Propriedade > Modelo
Propriedade > componentes

Propriedade_m > composição de componentes
 Propriedade_m > Componentes reutilizáveis
A-Ação > Construção

C6- 3ª etapa - modelo DEPAm-OR desenhado no HiperEditor

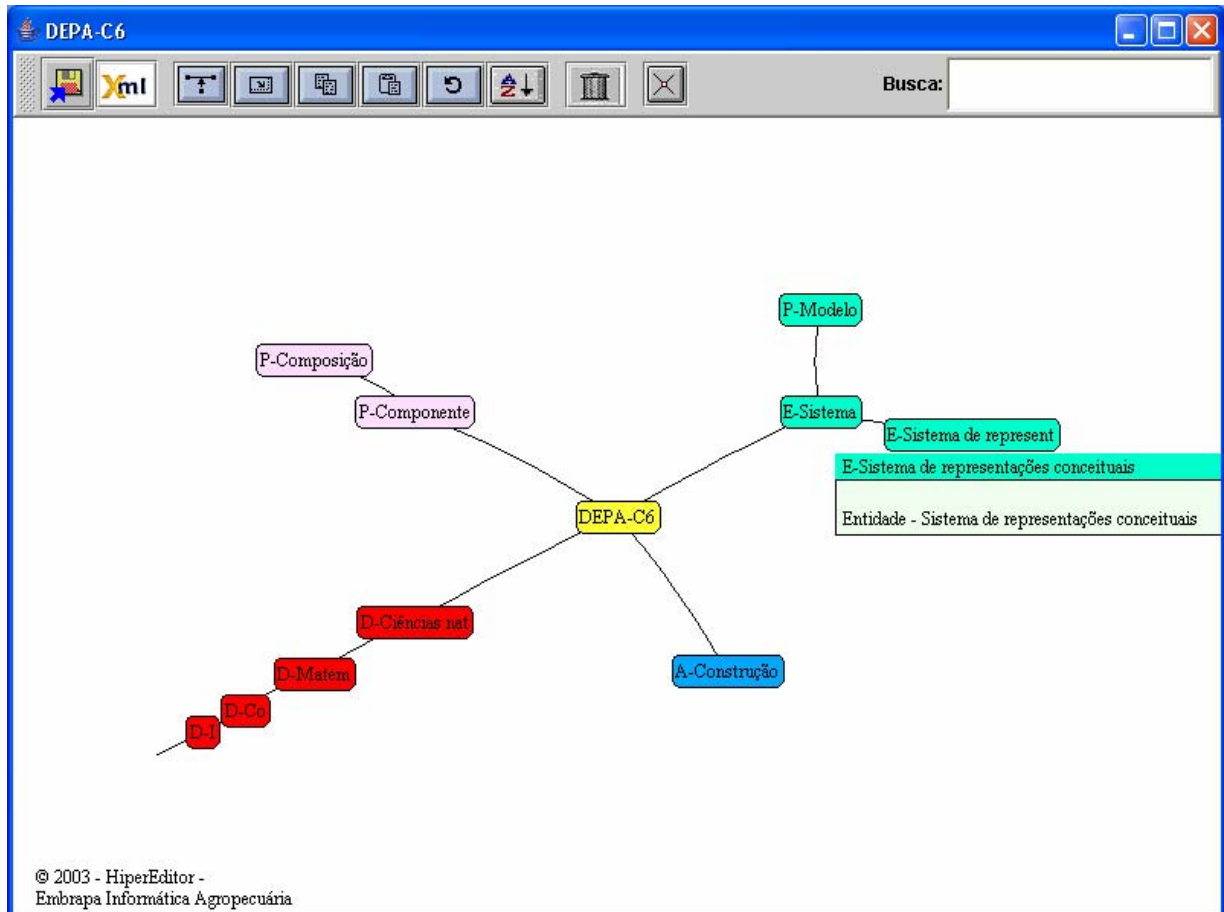


FIGURA 23: Representação DEPAm-OR no HiperEditor – Caso 6
 Fonte: Elaboração própria

C6- 4ª etapa – Codificação - RDF/XML

<rdf:RDF>

<rdf:about = "www.cs.utexas.edu/users/pclark/papers">

<dc:subject = "Construção de sistema de representações conceituais com modelo de composição de componentes reutilizáveis ">

<dssl:disciplina = "Ciências naturais">

<dssl:disciplina_m = "Matemática">

<dssl:disciplina_m_m = "Computação">

<dssl:disciplina_m_m_m = "Inteligência artificial">

```

<dssl:disciplina_m_m_m_m = "Representação do conhecimento e
raciocínio">
<dssl:disciplina_m_m_m_m_m = "Representação do conhecimento">
<dssl:disciplina_m_m_m_m_m_m_m = "Redes semânticas">
<dssl:entidade = "Sistema">
<dssl:entidade_m = "Sistema de representações conceituais">
<dssl:propriedade = "Modelo">
<dssl:propriedade = "Componentes">
<dssl:propriedade_m = "Composição de componentes">
<dssl:ação = "Construção">
</rdf:Description>
</rdf:RDF>

```

Exatamente como no caso 5, o caso 6 provém da Cx, que forneceu a árvore hierárquica do recurso de acordo com uma taxonomia restrita, *Lix Taxonomy*. Em razão disso, a estrutura está apresentada de forma tão simplificada.

CASO 7 – Virgin Soil

C7-1ª etapa – Atinência

Atinência – Romance russo do século XIX

C7-2ª etapa – Fatoração DEPA_m

D-Disciplina > Literatura. Literatura russa

E-Entidade > Romance

Entidade_m > Romance russo

P-Propriedade >

A-Ação >

m-Modificador_tempo: século XIX

C7- 3ª etapa - modelo DEPAm-OR desenhado no HiperEditor

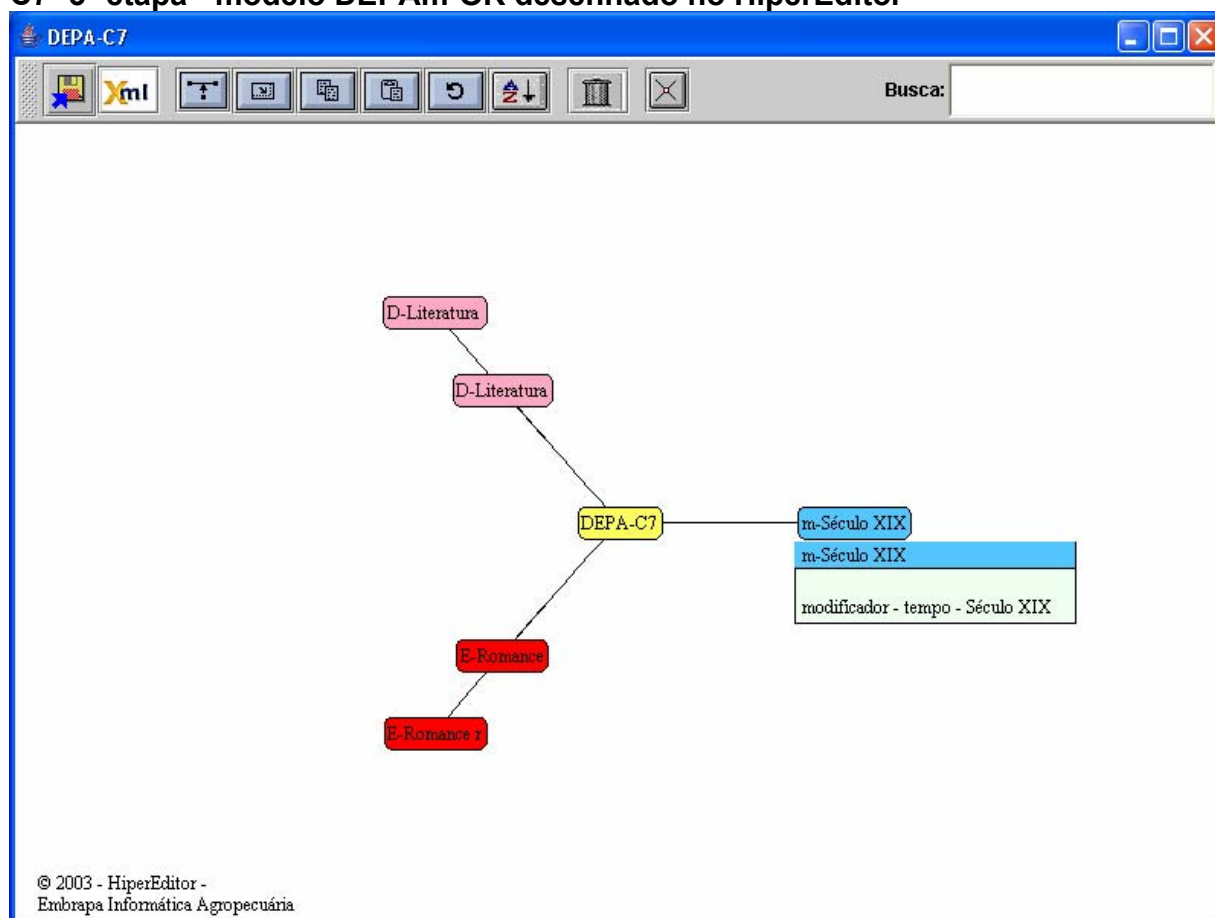


FIGURA 24: Representação DEPAm-OR no HiperEditor – Caso 7
Fonte: Elaboração própria

C7- 4ª etapa – Codificação - RDF/XML

```
<rdf:RDF>
  <rdf:about = "www.eldrithpress.org/ist/virgin.htm">
    <dc:subject = "Romance russo do século XIX ">
    <dssl:disciplina = "Literatura. Literatura russa">
    <dssl:entidade = "Romance">
    <dssl:entidade_m = "Romance russo">
    <dssl:modificador_tempo = "século XIX">
  </rdf:Description>
</rdf:RDF>
```

O caso 7, o único obtido na primeira fase de coleta de dados, por se tratar de um romance, teve suas categorias de classificação especificadas de maneira bem simples

por seu publicador. O fato que chama a atenção, nesse caso, é que a sua hierarquia taxonômica foi obtida a partir dos termos da *LCSH* e estava especificada no código-fonte do recurso. Esse é o único documento recuperado na *web* que permitiria, a uma máquina de busca, inferir sobre o seu conteúdo e mostrá-lo para acesso segundo um campo semântico, em vez do caminho por casamento de termos retirados do conteúdo do documento.

Pode-se afirmar que, dadas as dificuldades encontradas para representar o primeiro e o segundo casos, quando se estabeleceu uma forma lógica adequada para representação em nós na árvore hiperbólica, o restante das representações transcorreu conforme o esperado. Houve exemplos mais simples e outros mais complexos, conforme o uso estivesse ligado a um sistema de informação que representasse um domínio de conhecimento para uma comunidade técnico-científica.

Não houve dificuldades para se elaborar as descrições, pelo formato de descrição de recursos RDF/XML. Estas se revelaram atividades bem simples de serem executadas, a despeito da impressão inicial de que existem letras e codificações em demasia para serem elaboradas e compreendidas pelo indexador, quando este não tem formação em informática de programação de computadores. O formato de tripla, descrito na seção 2.5, mostrou-se eficiente e simples de usar, podendo ser considerado um instrumento a ser facilmente implementado pelo indexador.

As árvores hiperbólicas, construídas através dos *softwares* HiperEditor e HiperNavegador, facilitaram a implementação da representação da faturação como

objetos que se relacionam entre si. Inicialmente, não havia sido prevista essa possibilidade. Contudo, as sucessivas pesquisas permitiram a descoberta desses *softwares* livres, disponibilizados pelo Ministério do Desenvolvimento, Indústria e Comércio em www.agrolivre.gov.br. Esse fato trouxe também a garantia da plausibilidade da conjunção teórica dos dois objetos: estrutura profunda e modelagem orientada ao objeto.

Os nós da árvore hiperbólica implementam, de maneira muito tranqüila, a estrutura DEPAm. Mas a explicitação dos relacionamentos, conforme as possibilidades previstas pelo modelo E-R (entidade relacionamento), não foi garantida. Esse tipo de descrição deveria ter ocorrido nos arcos de ligação entre os nós. Para que não houvesse prejuízo do aspecto teórico, explicitou-se a categoria junto com o nome do atributo, como, por exemplo, no caso 7, em que o objeto representado como *m-século XIX* faz uso desse tipo de construção, pois a letra “m” foi adicionada à descrição *século XIX* para informar o tipo de relação entre o recurso e esse conteúdo, ou seja, uma relação de modificação do tipo tempo.

Após as considerações feitas nesta seção de análise dos dados e discussão dos resultados obtidos, foi possível chegar a certas conclusões e propor algumas recomendações, que se encontram na parte III, a seguir.

PARTE III

CONCLUSÕES E RECOMENDAÇÕES

6 CONCLUSÕES E RECOMENDAÇÕES

A revisão de literatura apoiou os pressupostos definidos, orientou a metodologia e, ao final, possibilitou afirmar a efetividade do modelo proposto.

Os pressupostos inicialmente definidos consideravam a possibilidade de verificação da existência de estruturas teóricas de indexação aplicáveis ao assunto dos recursos, de forma a torná-los objetos adicionados de um maior potencial semântico representado, possíveis de serem interpretados e gerar inferências tanto para pessoas como para programas de computador.

Também era considerada a possibilidade de se encontrar ferramentas adequadas à representação do assunto, da forma teórica enunciada pelo primeiro pressuposto. Isso exigia uma estrutura de representação congruente, que possibilitasse a adição, aos recursos, de descrição semântica possível de ser interpretada e de gerar inferências tanto para pessoas como para programas de computador.

Os pressupostos foram confirmados. O modelo DEPAm-OR mostrou a efetividade do entrelaçamento entre dois campos teóricos, o campo das linguagens de indexação, com a estrutura profunda aplicada ao assunto do recurso, e o campo da modelagem de dados, com a estrutura de entidade-relacionamento aplicada com o esquema e a sintaxe RDF/XML. Foi verificada a plausibilidade do desenvolvimento da indexação do assunto envolvendo o uso do modelo proposto.

Durante as análises realizadas, também foram observadas a preservação da semântica do assunto em estrutura profunda e as possibilidades de interpretação e inferência tanto por pessoas como por programas de computador.

Em relação ao problema da descrição do assunto na forma textual, tal como encontrado nos códigos *html* dos recursos, o modelo DEPAm-OR é uma proposição que orienta o indexador sobre uma nova forma de representá-los. Partindo do princípio das estruturas da linguagem, exposto na revisão de literatura, o modelo DEPAm-OR permite adicionar estrutura semântica à representação do assunto. Casando-se a estruturação semântica DEPAm com os desenvolvimentos de programas de computador para a *web* semântica, torna-se possível a atividade de indexação cooperar para a formação de campos conceituais com uma semântica profunda, ditada pelas estruturas profundas de representação.

A metodologia utilizada apoiou fortemente o desenvolvimento da pesquisa, ao orientar a elaboração de um modelo e alinhar as etapas sucessivas para o teste do modelo. Sobre as críticas que a técnica de modelagem possa provocar, mostra-se que o trabalho de constituir uma ampla revisão de literatura permitiu afirmar que o modelo construído espelha direcionamentos teóricos há muito resolvidos, de ambas as áreas estudadas, não devendo sofrer, assim, por eventuais problemas de concepção teórica.

A coleta de dados empreendida, a despeito do que se acreditava inicialmente, não foi tarefa fácil. O método de análise de conteúdo prevê, na sua essência, a existência e a disponibilidade dos conteúdos a serem analisados. Não existe a criação de dados,

coerentemente com o propósito de pesquisa, como ocorre numa enquete ou qualquer técnica de pesquisa de *survey*. Houve muita dificuldade para se obter os dados. Os relatos dos exemplos, na fase 1 da coleta, não traduzem a essência do trabalho de pesquisa, nem a decepção por não encontrar material de observação.

Porém, a fase 1 da coleta pode ser avaliada como a expressão de um alerta para o campo de estudo das bases teóricas e práticas dos sistemas de classificação. São muito poucas as instituições e iniciativas particulares que se ocuparam de publicar, na *web*, recursos eletrônicos tratados por uma ferramenta de classificação do conhecimento. Cabe à comunidade de pesquisa em Biblioteconomia e Ciência da Informação não só atentar para esse fato, mas também tomar iniciativas que visem à ampliação e disseminação do conhecimento sobre os instrumentos de que dispõe para a prática da representação de recursos.

A análise de dados demonstrou que o método de análise de conteúdo, quando orientado por categorias pertinentes, faz o trabalho de observação transcorrer sem percalços. Ao encontrar recursos eletrônicos disponíveis, para extrair-lhes os conteúdos relevantes, eram gastos poucos minutos.

O modelo pôde ser considerado fundamental para se verificar a viabilidade da melhor representação temática dos recursos estudados por esta pesquisa. Pela sua simplicidade e facilidade de implementação, poderá contribuir consideravelmente para os processos de indexação temática legíveis por pessoas e por computadores.

Ao analisar os dados, concluiu-se que as representações que utilizam os sistemas de classificação CDD, CDU, LCSH e tesouros deixaram perder elementos supostamente coletados durante a determinação da atinência e úteis à representação pelo modelo DEPAm. Não se encontraram as expressões de ação. Para adequar as representações encontradas ao modelo de estrutura profunda, foi preciso voltar ao recurso e retirar novamente a expressão da ação, a partir do título, do resumo ou de uma leitura rápida de todo o recurso.

A representação do assunto não encontrou problemas quanto às técnicas de fatoração e descrição com o uso da estrutura DEPAm, tópicos contidos no primeiro pressuposto. Dado que as estruturas DEPAm auto-organizam os conceitos e os agrupam em categorias, tanto pessoas quanto programas de computadores estariam potencialmente habilitados a afirmar que um conceito ou subcategoria, dentro de outra categoria, na estrutura DEPAm, pertencem à categoria hierarquicamente superior, o que faz gerar informação nova, por inferência entre níveis hierárquicos.

Também se percebe que a representação do assunto fatorado e modelado por estrutura profunda é passível de ser descrita por modelagem orientada ao objeto, é adequada à representação e à recuperação pelo padrão RDF/XML, como especificado pelo World Wide Web Consortium-W3Con para descrição de conteúdos na *web*.

Amparada pela pesquisa realizada, esta dissertação conclui que o modelo DEPAm-OR é uma ferramenta de uso fortemente desejável para melhorar problemas de estruturação e representação do assunto dos recursos eletrônicos, com a finalidade de

construir campos semânticos, de semântica profunda, na *web*, e operacionalizar facilidades de busca e acesso a recursos eletrônicos a partir de especificações da sua semântica profunda.

O estudo do modelo DEPAm-OR para representação de recursos eletrônicos fundamentados pelos conhecimentos sólidos e validados na Ciência da Informação evidenciou dois aspectos. O primeiro é o fato de existir muito conhecimento a ser aproveitado para auxiliar as pesquisas sobre a representação de informações no ambiente virtual. A Ciência da Informação possui a competência de dezenas de anos de experiência, seja para representar um recurso em relação a uma coleção, seja para representar tópicos dentro de um recurso.

O segundo aspecto evidenciado foi a falta de envolvimento relativo à prática de publicação na *web* usando os instrumentos de representação do conhecimento, notadamente sentida pela não participação da comunidade de Biblioteconomia e Ciência da Informação, percebida pela ausência de conteúdos publicados indexados por vocabulário controlado em língua portuguesa. Percebeu-se que, no ambiente eletrônico, não são seguidas as práticas de representação já instituídas e encontradas nas publicações em papel. Se o objeto empírico dessa dissertação tivesse sido livros impressos, teriam havido milhares de recursos em língua portuguesa indexados e próprios para o desenvolvimento da pesquisa.

Acredita-se desejável a continuidade deste tipo de pesquisa, porque contribui para o desenvolvimento de conhecimento, aplicações e produtos que tratam dos problemas

enfrentados com os grandes volumes de informações publicadas na *web*. Ainda, que se contribui para o desenvolvimento e engajamento da comunidade da Ciência da Informação nas temáticas de recuperação de informação, bem como para outros desenvolvimentos, na área da pesquisa aplicada, ao estudar e disponibilizar um modelo teórico de representação de assunto.

Este estudo evidenciou a necessidade de maiores estudos e conclusões sobre usos dos vocabulários controlados no ambiente virtual, modelos de representação de informação com vocação para semântica profunda, implementação de um ambiente de laboratório para o modelo proposto, tipos de metadados para representação do assunto, relações e inferências no modelo DEPAm-OR, navegação e recuperação no modelo DEPAm-OR, dentre outros.

Enfim, ressaltam-se a interdisciplinaridade deste estudo e a necessidade de a Ciência da Informação, nos moldes da década de 1960 e 1970, abranger, em sua atuação, os conhecimentos da era *web* para a evolução de instrumentos de representação e recuperação baseadas em estruturas semânticas.

REFERÊNCIAS

- AHMED, Kal et al. *Professional XML Meta Data*. Chicago: Wrox, 2001. 568p.
- ANDER-EGG, Ezequiel. *Introducción a las técnicas de investigación social: para trabajadores sociales*. 7.ed. Buenos Aires: Humanitas, 1978, apud MARCONI, Marina de A. & LAKATOS, Eva Maria. *Técnicas de pesquisa*. São Paulo: Atlas, 1982, 205 p.
- ALVARENGA, Lídia. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto de bibliotecas tradicionais e digitais. *Datagramazero*, v. 2, n. 6, dez. 2001.
- ARAÚJO, Vânia M.R.H. de. *Sistemas de recuperação da informação*. Rio de Janeiro: UFRJ, 1994.
- BHATTACHARYYA. G. Some significant results of current classification research in India. *International Forum on Information an Documentation*, v. 6, n. 1, p. 11-18, Jan. 1981.
- _____. POPSI Its fundamentals and procedure based on a general theory of subject indexing languages. *Library Science*, v.16, n.1, p.1-34, Mar. 1979.
- BUCKLAND, Michael K. Information as thing. *JASIS*, v. 42, n. 5, p. 351-360, 1991.
- CESARINO, Maria Augusta da Nóbrega; PINTO, Maria Cristina Mello Ferreira. Análise de assunto. *Revista de Biblioteconomia de Brasília*, v.8, n.1, p.32-43, jan./jun. 1980.
- DAHLBERG, Ingetraut. Teoria do conceito. *Ci. Inf. Brasília*, v.7, n.2, p.101-107, 1978.
- DELSEY, Tom. The library catalogue in a networked environment. In: *Conference on bibliographic control in the new millenium (Library of Congress)*. Canadá, Library of Congres, p.1-9, 2001. Descrito de http://lcweb.loc.gov/catdir/biblcontrol/deslsey_paper.html> recuperado em 25/5/2001.
- CARROW, D.; NUGENT, J. Comparision of free-text and index search abilities in an operating information system. In: *Information management in the 1980's*. 1977, New Yourk. Proceedings... New york: ASIS, 1977.v.14, p. 232-238 apud LOPES, Ilza Leite. Uso das linguagens controlada e natural em base de dados: revisão de literatura. *Ciência da Informação*, v.31, n.1, p. 41-52, jan./abr. 2002.
- DEVADASON, F. J. Computerization of deep structure based indexes. *International Classification*, v.12, n.2, p.87-94, 1985a.
- _____. On line construction of alphabetic classaurus: a vocabulary control and indexing tool. *Information Processing and Management*, v.21, n.1, p.11-26, 1985b.

DODEBEI, Vera Lúcia Doyle. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002. 119p.

ECO, Umberto. *A estrutura ausente: introdução à informação*. São Paulo: Perspectiva, 1991. Apud DODEBEI, Vera Lúcia Doyle. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002. 119p.

ELMASRI, Ramez; NAVATHE, Shamkant B. *Sistemas de banco de dados: fundamentos e aplicações*. Trad. Teresa Cristina Padilha de Souza. Rio de Janeiro: LTC, 2002.

FOSKETT, A. C. *A abordagem temática da informação*. Trad. Antônio Agenor Briquet de Lemos. São Paulo: Polígono, 1973. 437p.

FUJITA, Mariângela. Sistemas de Indexação: PRECIS I: perspectiva histórica e técnica do seu desenvolvimento e aplicação. *Rev. Bras. de Bibl. e Doc.*, v.21, n.1/2, p. 21-45, jan./jun. 1988.

GARDIN, J.C. *SYNTOL*, New Jersey: Rutgers University Press, 1965. 106p.

GONZÁLES DE GOMEZ, Maria Nélide. A representação do conhecimento e o conhecimento da representação: algumas questões epistemológicas. *Ci. Inf.* Brasília, v.22, n.3, p.217-222, set/dez, 1993.

GORSKY, D. P. Lenguaje y conocimiento. In: *Pensamiento y lenguaje*. México, D.F.: Grijalbo, 1966, apud DODEBEI, Vera Lúcia Doyle. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002. 119p.

GUINCHAT, C.; MENOU, M. *Introdução geral às ciências técnicas da informação e documentação*. Trad. Miriam Vieira da Cunha. Brasília: IBICT, 1994. 540 p.

HEANEY, Michael. Object-oriented cataloguing. *Information Technology and Libraries*, p. 135-153, september, 1995.

HJORLAND, Birger. The concept of 'subject' in information science. *Journal of Documentation*, v. 48, n.2, p.172-200, June, 1992.

HUTCHINS, W. J. *Languages of indexing and classification*. London: Peter Peregrinus, 1975.

_____. The concept of "aboutness" in subject indexing. *Aslib Proceedings*, v. 30, n.5, p.172-181, May 1978.

KAPLAN, Abraham. *A conduta na pesquisa: metodologia para as ciências do comportamento*. 2. ed. São Paulo: EPU, Edusp, 1975. 440 p.

KIM, Mary T. Research record. *Journal of Education for Library & Information Science*, n. 37, p. 376-382, 1996 apud POWELL, Ronald R. Recent trends in research: a

methodological essay. *Library & Information Science Research*, v. 21, n.1, p.91-119, 1999.

LANCASTER, F.W.; ELLIEKER, Calvin; CONNELL, Tschera H. Subject analysis. *Annual Review of Information Science and Technology*, v. 24, p. 35-74, 1989.

LANCASTER, F.W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos/Livros. 1993.

LANGRIDGE, D. W. *Subject analysis: principles and procedures*. London: Bowker-Saur, 1989. 146p.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em base de dados: revisão de literatura. *Ciência da Informação*, v.31, n.1, p. 41-52, jan./abr. 2002.

MARCONI, Marina de A. & LAKATOS, Eva Maria. *Técnicas de pesquisa*. São Paulo: Atlas, 1982. 205 p.

MARX, M. H. *Psychological theory*. New York: 1955, apud KAPLAN, Abraham. *A conduta na pesquisa: metodologia para as ciências do comportamento*. 2. ed. São Paulo: EPU, Edusp, 1975. 440 p.

MERLEAU-PONTY, Maurice. *O visível e o invisível*. São Paulo: Perspectiva, 1992, apud DODEBEI, Vera Lúcia Doyle. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002. 119p.

NEELAMEGHAN, A. Application of Ranganathan's general theory of knowledge classification in designing specialized databases. *Libri*, v. 42, n. 3, p. 202-226, 1992.

PARSAYE, Kamran; CHIGNELL, Mark; KHOSHAFIAN, Setrag; WONG, Harry. *Intelligent databases: object-oriented, deductive hypermedia technologies*. New York: Willey, 1989. 479p.

PARKHI, R.S. *Decimal classification and colon classification in perspective*. London: Asia Publishing House, 1964.

POWELL, Ronald R. Recent trends in research: a methodological essay. *Library & Information Science Research*, v. 21, n.1, p. 91-119, 1999.

RIVIER, Alexis. Construção de linguagens de indexação: aspectos teóricos. *Rev. Esc. Bibl. da UFMG*, v.21, n.1, p.56-99, jan./jun. 1992.

SAUSSURE, Ferdinand. *Curso de lingüística geral*. São Paulo: Cultrix, 1981.

SHERA, Jesse H. Pattern, structure, and conceptualization in classification. In: International study conference on classification for information retrieval. *Proceedings*. London: Aslib, 1957, p. 15-27.

SHLAER, Sally; MELLOR, Stephen J. *Análise de sistemas orientada para objetos*. São Paulo: McGraw-Hill, 1990. 178 p.

SPIRKIN, A.G. Origen del lenguaje y su papel en la formación del pensamiento. In: *Les théories de l'action*. Paris: Hachette, 1972, apud DODEBEI, Vera Lúcia Doyle. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002, 119p.

SVENONIUS, Elaine. Directions for research in indexing, classification and cataloguing. *Library Resources & Technical Services*, v. 25, n.1, p. 88-103, 1981.

BIBLIOGRAFIA

AITCHISON, Jean; GILCHRIST, Alan. *Manual para a construção de tesouros*. Rio de Janeiro: BNG/Brasilart, 1979. 141p.

ASSUNÇÃO, Jandira. Preciso em português. *Rev. Esc. Bibl. da UFMG*, v.18, n.2, p.151-371, set. 1989.

BARBOSA, Alice Príncipe. *Teoria e prática dos sistemas de classificação bibliográfica*. Rio de Janeiro: IBBD, 1969.

_____. Classificação facetada. *Ciência da Informação*, v.1, n.2, p.73-81, jul/dez. 1972.

BAX, Marcello Peixoto. Introdução a linguagens de marcas. *Ciência da Informação*, v.30, n.1, p.32-38, 2001.

BERNERS-LEE, Tim. Semantic Web. *Scientificamerican*. Descrito em www.scientificamerican.com/2001/0501issue/0501berners-ml>. Recuperado em 24/9/2001.

DEWEY Decimal Classification and Relative Index: *devised by Melvil Dewey*. Joan S. Mitchell, Winton E. Matthews, Gregory E. New (Eds). 21 ed. New York: Forest Press, 1996. v.1.

CHU, Clara M.; O'BRIEN, Ann. Subject analysis: the critical first stage in indexing. *Journal of Information Science*, v.19, p. 434-454, 1993.

DENTON, William. *How to make a faceted classification and put it on the web*. Nov. 2003. Descrito em <http://www.miskatonic.org/library/facet-web-howto.html>>. Recuperado em 27/5/2004.

ECO, Umberto. *Semiótica e filosofia da linguagem*. São Paulo: Ática, 1991, apud _____. *Como se faz uma tese*. 18. ed. São Paulo: Perspectiva, 2003. 170 p.

FOSKETT, D.J. *The information environment: a world view: studies in honour of professor A. I. Mikkailov*. Amsterdam: FID – Fédération Internationale d'Information et de Documentation, Elsevier, 1990. 197p.

FRANCA, Júnia Lessa. *Manual para normalização de publicações técnico-científicas*. 6. ed. rev. e ampl. Belo Horizonte: UFMG, 2003. 230 p.

FRANKLIN, Rosemary Aud. Re-inventing subject access for the semantic web. *Online Information Review*, v. 27, n.2, p. 94-101, 2003.

FROHMANN, B. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, v. 46, n. 2, p. 81-101, jun. 1990.

FUGMANN, R. An interactive classaurus on the PC. *International Classification*, v.17, n.3/4, p. 133-137, 1990.

FUGMANN, Robert. *Subject analysis and indexing: theoretical foundation and practice advice*. Frankfurt: Index Verlag, 1993. 250p.

GILCHRIST, Alan. Thesauri, taxonomies and ontologies – an etymological note. *Journal of Documentation*, v. 39, n.1, p.7-18, 2003.

GREEN, Rebecca. The design of a relational database for large-scale bibliographic retrieval. *Information Technology and Libraries*, p. 207-221, Dec. 1996.

KOBASHI, Nair Yumiko. A organização e a transferência de informações documentárias: problemas e perspectivas. In: *Organização do conhecimento e sistemas de classificação*. Brasília: IBICT, p. 64-67, 1996.

MARCONI, M. A.; LAKATOS, E.M. *Técnicas de pesquisa*. 2. ed. São Paulo: Atlas, 1982, 205 p.

MOREIRO, José; MARZAL, Miguel Angel; BELTRÁN, Pilar. Desarrollo de un método para la creación automática de mapas conceptuales. V *Encontro Nacional de Pesquisa em Ciência da Informação – Enancib*. Belo Horizonte: Escola de Ciência da Informação da UFMG, 10 a 14 de novembro de 2003, CD.

NAVES, Madalena M. L. *Fatores interferentes no processo de análise de assunto: estudo de caso de indexadores*. 2000. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2000.

SLAVIC, Aida. *A definition of thesauri and classifications as indexing tools*. <<http://dublincore.org/documents/thesauri-definiton/>> armazenado em 21/11/2000, recuperado em 12/5/2003.

SVENONIUS, Elaine. Facet definition: a case study. *International Classification*, v. 5, n. 3, p. 134-141, 1978.

SWIFT, D. F.; WINN, V.; BRAMER, D. “Aboutness” as a strategy for retrieval in the social sciences. *Aslib Proceedings*. v. 30, n. 5, p. 182-187, May. 1978.

TRAVIS, Irene L.; FIDEL, Raya. Subject analysis. *Annual Review of Information Science and Technology*, v.17, p.123-157, 1982.

WORLD WIDE WEB CONSORTIUM. *Semantic web activity: advanced development*. Descrito em <www.w3.org/2000/01/sw>. Recuperado em 25/9/2001.

GLOSSÁRIO

Aboutness – ver Atinência.

Alternação – processo do grupo dos processos permutados, constrói combinações dos elementos da frase, a fim de obter todas as combinações de termos para a composição de índices de termos.

Analeto – proveniente da indexação relacional de Farradane, constitui a combinação da qual participam dois termos isolados, ligados por uma descrição de relação.

Análise – processo de fazer abstrações sobre objetos do mundo real.

Análise conceitual orientada ao objeto – ver Análise orientada ao objeto.

Análise de assunto – atividade realizada pelo indexador, composta por três etapas: a leitura do recurso informacional, a seleção de conceitos e a definição da atinência.

Análise de conteúdo – processo de abstração de informações contidas nos recursos de informação.

Análise objetiva – designa a busca de elementos objetivos durante a análise informacional de conteúdo de recursos, em contraposição com a análise subjetiva.

Análise orientada ao objeto – processo de abstração de informações sobre objetos do mundo real para modelar suas características e funções na interação com outros objetos.

Análise subjetiva – designa a orientação da análise de conteúdo quando esta ocorre em situações específicas, pouco objetivas, e que em outro domínio poderia ser compreendida de outra maneira.

Associação paradigmática – associação a que os conceitos remetem, ausente na linearidade da frase, que não permite definir a quantidade de relações que irão surgir, nem em que sentido ocorreriam.

Associação sintagmática – associação entre elementos da frase ou sentença compreendida na linearidade dos elementos, com quantidade e sentido determinados e que permite aceitar a expressão. Na frase *O cachorro morde o homem* a associação sintagmática determina a impossibilidade da relação inversa *o homem morde o cachorro*, sujeito, verbo e objeto tem limite de sentidos nesta expressão.

Atinência – resultado do processo de análise de assunto, expresso pela frase de indexação, em linguagem natural, elaborada pelo indexador, e que especifica sobre o que trata um recurso.

Atinência polihierárquica – tipo de ocorrência de atinência, em que os conceitos expressados pela sentença remetem o recurso tratado a mais de um assunto. Ver também atinência.

Atributos – propriedades ou características de um recurso, objeto ou entidade de informação.

Bases de conhecimento – conjuntos de informações armazenadas em sistemas computadorizados, para atender às necessidades de usuários, geralmente modeladas para uma necessidade pragmática de organização e recuperação.

Cabeçalhos de assunto – conjunto pré-coordenado de termos simples ou compostos que enumeram os conceitos de uma ou mais áreas do conhecimento, usados para preparar pontos de acesso sistematizados em catálogos de recursos, organizados por assunto.

Cabeçalhos de classificação – ver Cabeçalhos de assunto, não devem ser confundidos com os códigos de classificação – numéricos ou alfanuméricos, atribuídos aos assuntos pelas linguagens simbólicas.

Casamento de padrões – procedimento em que dois termos, um originário da pergunta do usuário e outro de índices de termos de documento, são confrontados para se obterem resultados sobre a relevância do documento para a busca iniciada pelo usuário.

Catálogo compartilhada – refere-se ao procedimento americano de minimizar custos de catalogação através da instituição de redes de bibliotecas que integram seus esforços de catalogação e compartilham as catalogações feitas, a fim de que um item seja catalogado apenas uma vez pela rede.

Classaurus – conceito instituído por Bhattacharyya, no POPSI, significa sistema de indexação híbrido, com características de sistema de classificação (notação e cabeçalhos de assunto) e características de tesouros (tabelas sistematizadas por categorias, termo genérico, termo específico, nota de escopo e referência bibliográfica).

Classe – equivalente a categoria neste projeto, agrupa, conceitualmente, um conjunto de entidades ou objetos.

Classificação associativa – vem do conhecimento; processo que busca explicitar a semântica dos conceitos, ou seja, as relações que estes expressam naquele contexto.

Classificação organizativa – vem das políticas; processo que busca explicitar a sintaxe da linguagem de indexação para fins de manutenção de padronização e sistematização dos conceitos expressos.

Códigos de catalogação – instrumentos de orientação da atividade de catalogação, como o AACR2.

Conceito – menor unidade de informação com sentido por si só.

Conceitualização – atividade de definir conceitos.

Conhecimento – resultado do processo mental de associação de conceitos e construção de sentido, geralmente referido como o final da gradação cognitiva que passa por dado, informação e chega ao conhecimento.

Conhecimento organizacional – conjunto de saberes de uma organização, conhecimento na forma coletiva.

Conhecimento registrado – equivalente a recurso, documento, objeto de informação.

Constituintes elementares – categorias básicas da formulação do assunto: disciplina, entidade, propriedade, ação.

Coordenação – forma de combinar o conjunto dos conceitos para recuperação posterior.

Coordenação *a priori* – forma usualmente encontrada nos sistemas de classificação do tipo Classificação Decimal de Dewey e Classificação Decimal Universal, nos quais as combinações entre conceitos são elaboradas previamente e apresentadas de forma enumerativa, com a intenção de cobrir todas as combinações possíveis.

Coordenação *a posteriori* – forma usualmente encontrada na classificação facetada, na qual os conceitos são enumerados em termos simples ou compostos e a combinação entre eles é construída no momento em que se necessita representar um determinado recurso. Não existe a enumeração das combinações possíveis.

DEPAm – acrônimo de disciplina, entidade, propriedade, ação e modificador; forma a estrutura profunda, temática central da teoria geral da indexação de assuntos de Bhattacharyya.

Elemento documento de informação – item da sintaxe RDF, especifica o recurso informacional.

Domínio – delimitação de área do conhecimento ou de prática.

Eixo associativo – eixo vertical, representativo das associações diversas que um conceito possa vir a sugerir.

Eixo sintagmático – eixo horizontal, representativo das relações de linearidade dos termos expressos na sentença.

Eixo paradigmático – termo usado por Gardin, no Syntol (ver também Eixo associativo).

Elemento de informação – características atribuídas a um recurso (ver também Atributo).

Estratégia de busca – seqüência de termos selecionados para efetuar busca por informações em uma interface de pesquisa.

Estrutura de superfície – estrutura que uma sentença apresenta, individualmente.

Estrutura elementar – o mesmo que estrutura profunda.

Estrutura profunda – conjunto das categorias DEPAm – disciplina, entidade, propriedade, ação, assim compreendidas por representarem em alto nível de abstrações o conjunto das estruturas de superfície encontradas.

Estruturas de representação computacionais – conjuntos de formatos e linguagens que permitem implementar estruturas de informação em computadores.

Etiqueta – identificador colocado entre os sinais de maior e menor, usado para denominar o conteúdo de informação que ocorre nos formatos de anotação HTML e XML, mais conhecida por *tag*.

Extrapolação – procedimento que busca em fontes especializadas a classe conceitual imediatamente superior ou anterior ao conceito pesquisado.

Extrinsic aboutness – definição da atinência para fins de colocação de um recurso junto com outros já indexados.

Fatoração – procedimento de desmembramento conceitual dos termos que compõem uma sentença ou frase de indexação.

Frase de indexação – produto da análise de assunto que expressa em uma sentença, em linguagem natural, sobre o que trata um documento.

HTML – acrônimo de HyperText Markup Language, linguagem de publicação de conteúdos na *web*.

Indexação – Sub-campo disciplinar e atividade da prática profissional do indexador que se sub-divide em indexação temática e indexação descritiva. A indexação temática cuida de atribuir a atinência ao recurso. A indexação descritiva cuida de elaborar registros de conhecimento baseados em normas internacionais de descrição de recursos.

Indexação em cadeia – procedimento em que os conceitos, mesmo submetidos a processos de permutação para geração de índices, mantêm suas relações de contexto, através de indicadores de função.

Indexação em fileira – semelhante à indexação em cadeia, mas elaborada no sistema PRECIS, no qual os operadores de relação garantem a ligação do termo presente ao seu precedente e ao seu subsequente.

Indexação em *string* – ver indexação em fileira.

Indexação temática – processo de indexar recursos pela representação do seu assunto ou tema; diferente de indexação descritiva (ver também Indexação).

Indexador – profissional de informação especializado nas etapas de análise de assunto e tradução conceitual de um recurso para a linguagem de um vocabulário controlado.

Indexação relacional – conceito utilizado para se referir ao sistema de indexação de Farradane, que usa operadores relacionais.

Índices permutados – índices construídos a partir de processos de alternância, rotação e deslocamento de conceitos na frase de indexação (ou sentença).

Informação – dados com potencial de construir sentido pelo receptor.

Instância de classe – mesmo que objetos de uma determinada classe.

Interpolação – processo de adicionar um conceito novo entre dois conceitos já existentes numa lista de termos.

Intrinsic aboutness – definição da atenção para partes de um documento, a fim de representar seu conteúdo informativo de maneira pormenorizada (ver também *Extrinsic aboutness*).

Isolado – termo atribuído por Farradane às unidades conceituais individuais.

Ligação em cadeia – tipo de ligação entre termos que ocorre na indexação em cadeia.

Linguagem artificial – aquela que faz a representação de informações, controlando os termos que servirão para representar conteúdos e para elaborar as estratégias de busca (ver também Vocabulário controlado).

Linguagem documentária – (ver Linguagem de indexação).

Linguagem de indexação – linguagem utilizada para representar e recuperar informações contidas em recursos.

Linguagem natural – linguagem comum usada pelas pessoas para se comunicar.

Linguagens simbólicas – linguagens que permitem a representação conceitual por números e símbolos.

Linguagens verbais – linguagens que elaboram e representam os conceitos na forma verbal.

Lógica de concretude – mesmo que princípio de concretude, de Kaiser, no qual a representação ordena os conceitos na ordem do mais concreto ao mais abstrato, seguidos pelos termos que expressam processos e locais.

Manifestação – ocorrência de qualquer uma das categorias DEPAm.

Metadados – informação sobre a informação, explica o significado de um atributo ou propriedade de um recurso.

Metadados Dublin Core – nome dado ao conjunto de elementos considerados essenciais para descrever um recurso bibliográfico, estabelecidos em Dublin, Ohio, 1995.

Metadocumentos – produtos que informam sobre os conteúdos de recursos indexados – produtos da representação em terceiro nível.

Metainformações – informações contidas nos metadocumentos.

Metarrepresentação – função desempenhada por um metadocumento.

Modificador – manifestação que modifica outra manifestação do tipo disciplina, entidade, propriedade e ação.

Namespace – termo que designa o local virtual onde são armazenados, disponibilizados e validados os conceitos sobre uma área específica.

Objetos – mesmo que entidade, possuem estrutura e função quando em interação com outros objetos.

Operador – termo atribuído por Farradane a cada símbolo usado para representar as relações entre conceitos.

Operadores relacionais – conjunto de nove relações psicológicas; relações que expressam o embricamento do tratamento conceitual com teorias da psicologia, sobre a formação de relações conceituais; os operadores são utilizadas para expressar as relações entre termos de uma sentença.

Padrão texto livre – texto na forma como se encontra no recurso original, isto é, em linguagem natural.

Permutação – processo de trocar a ordem dos elementos da frase de indexação, montando combinações.

Pluralidade de significados – atributo da linguagem natural, provocador de ambigüidade, geralmente resolvido pelo falante, devido ao nível de interações desenvolvidas em determinado contexto.

PMEST – acrônimo de *personality, matter, space, time*, categorias baseadas nos princípios de classificação de assuntos de Ranganathan. Na classificação facetada, as informações de um objeto de informação são categorizadas, pela funcionalidade que apresentam, em personalidade, matéria, espaço ou tempo, criando um índice de indexação facetado.

POPSI – acrônimo de *Postulate-based Permuted Subject Indexing*, sistema computadorizado de indexação de assuntos desenvolvido na Índia, por Bhattacharyya, na década de 1970.

Ponto de acesso – termo atribuído aos conceitos escolhidos para representar um recurso e que tem a função de permitir sua recuperação no processo de busca em catálogos e/ou outro instrumento de indexação.

Pós-coordenação – procedimento em que a combinação dos termos só ocorre no momento da busca, situação diversa à da pré-coordenação.

PRECIS – acrônimo de *Preserved Context Index System*. Sistema de indexação computadorizado, desenvolvido por Derek Austin, na Inglaterra na década de 1970.

Precisão – medida de avaliação de índices de recuperação que mede a exatidão do tema recuperado em face do pedido do usuário. Influenciada por problemas lingüísticos de ambigüidade de conceitos, busca soluções através de formas de descrição com maior especificidade semântica. É normalmente estudada junto com a revocação (ver também revocação).

Pré-coordenação – procedimento em que a combinação dos termos de assunto ocorre antes do momento da indexação, gerando tabelas com as enumerações dos assuntos indexáveis.

Processos permutados – conjunto dos processos de alternância, rotação e deslocamento, que executam mudança de posição dos conceitos na sentença, a fim de conseguir combinações de termos que ocorreriam numa estratégia de busca.

Produção de sentido – relativa à associação de idéias e formulação de significações ordenadas.

RDF – acrônimo de *Resource Description Framework*, descreve modelo e sintaxe de representação conceitual orientada ao objeto.

Recuperação – conjunto das atividades relacionadas à análise da necessidade de informações, à estratégia de busca (elaboração da pergunta ao sistema de informações), aos mecanismos de processamento da pesquisa e de mensuração dos resultados obtidos. Assim, o termo designa tanto as atividades dos processos quanto as atividades disciplinares de pesquisa na área da recuperação de informação.

Recurso – quase sinônimo de documento; foi apresentado pela AACR2, em 1998, como um termo mais abrangente, uma vez que inclui tanto os documentos textuais quanto recursos digitais.

Recursos digitais remotos – recursos digitais armazenados em computadores, que não podem ser tocados por contato manual.

Redução semântica – processos que envolvem o uso de sucessivas linguagens para representar percepções dos sentidos.

Relação colateral – relação que aparece sob um mesmo nível hierárquico em uma organização de termos sistematizada.

Relação gênero-espécie – relação hierárquica, como ocorre em “árvores(gênero)” e “árvores frutíferas(espécie)”.

Relação todo-parte – relação hierárquica, como ocorre “árvore(todo)” e “galhos, flores, frutos, tronco, caule, raiz (partes)”.

Relação antônima – relação associativa que expressa a negação conceitual.

Relação hierárquica – relação conceitual que expressa encadeamento de conceitos, na qual, subentende-se a existência de um elemento raiz ligado a um nível inferior formado por elementos filhos; de outra forma, os elementos numa hierarquia inferior formam sucessivamente relações com elementos superiores até chegarem à relação com o elemento raiz

Relação sinônima – relação entre conceitos que expressa equivalência de termos.

Relevância – medida de avaliação de índices de recuperação, mede a quantidade de documentos recuperados que têm relação com a expressão de busca.

Representação – mesmo que representação do conhecimento, utilizada para significar qualquer processo de descrição de conteúdos.

Representação mental – primeiro nível de representação do conhecimento, no qual o indivíduo representa em sua mente os elementos que percebe no mundo externo.

Representação primária – segundo nível de representação, no qual o autor de um recurso descreve num instrumento de registro físico os conceitos que formou mentalmente.

Representação secundária – terceiro nível de representação, ligado à criação de metadocumentos e metainformações, com a finalidade de metarrepresentação.

Representação temática de recursos – atividade de representar os recursos por assunto ou tema.

Revocação – medida de avaliação de índices de recuperação, que mede a quantidade de documentos recuperados em face da quantidade de itens presentes na base. Medida usualmente discutida junto com a precisão (ver também Precisão).

Rotação – um dos processos permutados, no qual ocorre mudança na forma de ressaltar a ordem dos conceitos de uma frase.

Sintagma – unidade conceitual formada por dois ou mais conceitos que não conduzem ao mesmo sentido de quando estão juntos.

Sistemas de classificação – ordenações conceituais de áreas do conhecimento, seguindo ou não orientações classificatórias filosóficas. Enumeram os conceitos, as combinações possíveis e, muitas vezes, determinam códigos de notação para serem usados em arranjos de recursos. São exemplos de sistemas de classificação a Classificação Decimal de Dewey – CDD, a Classificação Decimal Universal – CDU e a Library of Congress Classification – LCC (Biblioteca do Congresso Americano).

Sistemas de indexação – mais abrangentes do que os sistemas de classificação, incluem ainda sistemas verbais como os tesouros, e híbridos como o PRECIS e o POPSI.

Sistemas de informação – conceito que abarca os muitos processos de organização de informações, como os processos de entrada, processamento e saída.

Sobrecidade – ver atinência.

Sumarização – procedimento no qual o indexador promove a redução semântica sistemática para chegar a níveis de representação elementares.

SYNTOL – acrônimo de *SYNTagmatic Organization Language*. Sistema de indexação elaborado por Gardin, na década de 1960, baseado em lógica formal para representação das relações sintagmáticas.

Tema – mesmo que assunto do documento (ver também Atinência).

Temacidade – (ver Atinência).

Teoria geral da indexação de assunto – teoria elaborada por Bhattacharyya para a análise de assunto, fundamentada em diversos princípios classificatórios de Ranganathan.

Termos de busca – mesmo que termos de pesquisa, ou seja, palavras utilizadas na pesquisa de recursos.

Termos de entrada – termos selecionados pelo indexador para representar o recurso.

Tradução conceitual – processo posterior à análise de assunto, no qual o indexador procura, numa ferramenta de vocabulário controlado, os termos autorizados para representar os conceitos escolhidos para o tema do documento em foco.

Tradução de assuntos – ver tradução conceitual.

Triângulo semântico – designa o Triângulo de Ogden e Richard que, na década de 1920, serviu para demonstrar a articulação entre referência, símbolo e objeto.

Tripla – unidade completa de declaração de conteúdos da modelagem orientada ao objeto, na qual as informações são organizadas na forma de recurso, propriedade, valor.

URI – acrônimo de *Uniform Resource Identifier*, descreve uma localização de recurso na *web*.

Vocabulário controlado – formado pelo conjunto das linguagens artificiais, é usado para obter precisão e qualidade de recuperação de informações em um domínio.

XML – acrônimo de *Extended Markup Language*, utiliza uma sintaxe de linguagem que permite descrever as estruturas conceituais com maior conteúdo semântico.

ANEXOS

CD A indexação temática de recursos fundamentada por estrutura profunda e abordagem objeto-relacionamento: exemplos e casos estudados