

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

FLÁVIO VIEIRA PONTES

**ORGANIZAÇÃO DO CONHECIMENTO EM BIBLIOTECAS DIGITAIS DE TESES E
DISSERTAÇÕES:
uma abordagem baseada na classificação facetada e taxonomias dinâmicas**

Belo Horizonte

2013

FLÁVIO VIEIRA PONTES

**ORGANIZAÇÃO DO CONHECIMENTO EM BIBLIOTECAS DIGITAIS DE TESES E
DISSERTAÇÕES:
uma abordagem baseada na classificação facetada e taxonomias dinâmicas**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da UFMG como requisito para obtenção do Grau de Doutor em Ciência da Informação.

Linha de Pesquisa: Organização e Uso da Informação

Orientadora: Prof^a Dra. Gercina Ângela Borém de Oliveira Lima

BELO HORIZONTE

2013

Pontes, Flávio Vieira.

P814o

Organização do conhecimento em bibliotecas digitais de teses e dissertações [manuscrito] : uma abordagem baseada na classificação facetada e taxonomias dinâmicas / Flávio Vieira Pontes. – 2013.
233 f. : il., enc.

Orientadora: Gercina Ângela Borém de Oliveira Lima.
Tese (doutorado) – Universidade Federal de Minas Gerais,
Escola de Ciência da Informação.

Referências: f. 210-222

Apêndices: f. 223-233

1. Ciência da informação – Teses. 2. Organização da informação – Teses. 3. Sistemas de recuperação da informação – Teses. 4. Bibliotecas digitais – Teses. 5. Classificação facetada – Teses. I. Título. II. Lima, Gercina Ângela Borém de Oliveira. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 02:004



UFMG

Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

FOLHA DE APROVAÇÃO


"ORGANIZAÇÃO DO CONHECIMENTO EM BIBLIOTECAS DIGITAIS DE TESES E DISSERTAÇÕES: UMA ABORDAGEM BASEADA NA CLASSIFICAÇÃO FACETADA E TAXONOMIAS DINÂMICAS"

Flávio Vieira Pontes

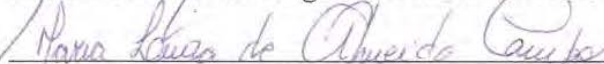
Tese submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como parte dos requisitos à obtenção do título de "**Doutor em Ciência da Informação**", Linha de Pesquisa: "**Organização e Uso da Informação - OUI**".

Tese aprovada em: 04 de abril de 2013.

Por:




Prof. Dra. Gercina Ângela Borém de Oliveira Lima - ECI/UFMG (Orientadora)



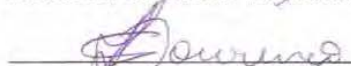
Prof. Dra. Maria Luíza de Almeida Campos - UFF



Prof. Dr. Alcione de Paiva Oliveira - UFV

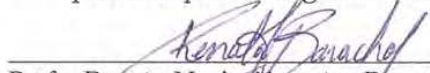


Prof. Dr. Maurício Barcellos Almeida - ECI/UFMG



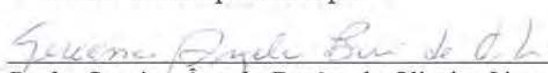
Prof. Dra. Cíntia de Azevedo Lourenço - ECI/UFMG

Aprovada pelo Colegiado do PPGCI



Prof. Renata Maria Abrantes Baracho Porto
Coordenadora

Versão final Aprovada por



Prof. Gercina Ângela Borém de Oliveira Lima
Orientadora



UFMG

Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

ATA DA DEFESA DE TESE DE **FLÁVIO VIEIRA PONTES**, matrícula: 2009651736

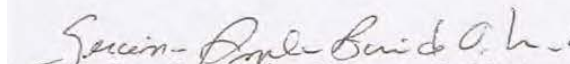
Às 14:00 horas do dia 04 de abril de 2013, reuniu-se na Escola de Ciência da Informação da UFMG a Comissão Examinadora aprovada *ad referendum* pela Coordenadora do Programa de Pós-Graduação em Ciência da Informação em 01/03/2013, para julgar, em exame final, o trabalho intitulado **Organização do conhecimento em bibliotecas digitais de teses e dissertações: uma abordagem baseada na classificação facetada e taxonomias dinâmicas**, requisito final para obtenção do Grau de DOUTOR em CIÊNCIA DA INFORMAÇÃO, Área de Concentração: Produção, Organização e Utilização da Informação, Linha de Pesquisa: Organização e Uso da Informação - OUI. Abrindo a sessão, a Presidente da Comissão, Profa. Dra. Gercina Ângela Borém de Oliveira Lima, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

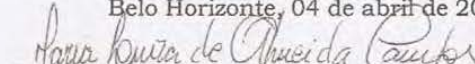
Profa. Dra. Gercina Ângela Borém de Oliveira Lima - Orientadora	APROVADO
Profa. Dra. Maria Luíza de Almeida Campos	APROVADO
Prof. Dr. Alcione de Paiva Oliveira	APROVADO
Prof. Dr. Maurício Barcellos Almeida	APROVADO
Profa. Dra. Cíntia de Azevedo Lourenço	APROVADO

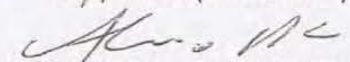
Pelas indicações, o candidato foi considerado APROVADO.

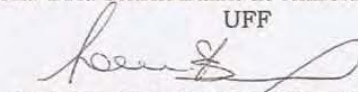
O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.


Belo Horizonte, 04 de abril de 2013

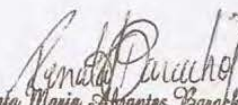

Profa. Dra. Gercina Ângela Borém de Oliveira Lima
ECI/UFMG (Orientadora)


Profa. Dra. Maria Luíza de Almeida Campos
UFF


Prof. Dr. Alcione de Paiva Oliveira
UFV


Prof. Dr. Maurício Barcellos Almeida
ECI/UFMG


Profa. Dra. Cíntia de Azevedo Lourenço
ECI/UFMG


Prof(a). Renata Maria Abrantes Paracho Pontes
Coordenadora do Programa de Pós-Graduação
em Ciência da Informação

Obs: Este documento não terá validade sem a assinatura e carimbo da Coordenadora.

DEDICATÓRIA

A meus pais.

AGRADECIMENTOS

“Aqueles que passam por nós, não vão sós, não nos deixam sós. Deixam um pouco de si, levam um pouco de nós.” Antoine de Saint-Exupéry

“Todos os dias eu lembro a mim mesmo de que minha vida interior e exterior são baseadas no trabalho de outras pessoas, vivas e mortas, e que devo me esforçar para retribuir na mesma medida tudo o que tenho recebido e ainda estou recebendo.” Albert Einstein

Redigir um texto de agradecimento rivaliza, em grau de dificuldade, com a própria atividade de pesquisa e redação da tese. Isto porque a lista corre sempre o risco de ser injusta. Basta uma pequena traição da memória para que se deixe de incluir pessoas cuja contribuição tenha sido fundamental.

Ciente do risco, inicio por agradecer à Universidade Federal de Viçosa, pela oportunidade de aperfeiçoamento profissional, e ao CNPq, instituições sem cujo apoio este trabalho não teria sido possível.

A todos os professores e colegas do Programa de Pós-Graduação em Ciência da Informação da UFMG, um profundo reconhecimento. Em especial a Benildes Maculan, pela parceria e contribuição para este trabalho de pesquisa, valeu Bê!

Aos meus pais, Almiro e Naná, pela minha criação e apoio durante toda a minha vida, por me enxergarem melhor do que sou. Espero poder retribuir por tudo que recebi. Aos demais e queridos familiares, por estarmos juntos neste caminho. A Milene, pelo apoio silencioso e dedicação, por estar ao meu lado nessa etapa da minha vida.

À minha orientadora, professora Gercina, pela disponibilidade, orientação e apoio incondicional, sempre que necessário.

Aos professores Maurício Almeida, Cíntia Lourenço, Célia Dias, Maria Luiza Campos, Alcione de Paiva e Eduardo Wense, que me honraram ao aceitar o convite para compor a Banca de Defesa desta Tese.

Mais uma vez, a todos os meus sinceros agradecimentos.

[...] existem duas condições necessárias e constituintes da pergunta, sem as quais não pode existir a pergunta: certo saber (pré-saber) e um não saber. Um certo saber torna a pergunta possível e o não saber já ciente, ao qual aquilo que se pergunta já é conhecido de uma forma geral e indeterminada, mas ainda não em forma específica e determinada[...] de um tal não-saber ciente surge o querer saber e deste nasce a pergunta[...] tematicamente na pergunta já está o não saber e o querer saber.

(Zilles, Urbano. Fundamentos ontológicos do conhecimento)

RESUMO

As bibliotecas digitais, em especial aquelas cujo acervo é composto de teses e dissertações, carecem de mecanismos mais adequados para exploração do acervo. Os usuários precisam elaborar expressões de consulta, sendo que a inabilidade em expressar precisa e claramente a informação de que se necessita, além de problemas relacionados à ambiguidade da linguagem natural, gera impacto negativo nos resultados. Esta pesquisa utilizou um modelo baseado na Teoria da Classificação Facetada para representação e organização do conhecimento em uma biblioteca digital de teses e dissertações, além de taxonomias dinâmicas, com o objetivo de compor o mecanismo de recuperação da informação. Foram desenvolvidos novos componentes de recuperação e acesso na forma de um protótipo computacional, que foi objeto de um estudo comparativo. O estudo comparativo, desenvolvido na forma de um experimento, buscou determinar a contribuição dos mecanismos de representação e organização do conhecimento, além do novo mecanismo de recuperação da informação, na facilitação da descoberta da informação, assim como no aumento da satisfação dos usuários da biblioteca digital. O mecanismo de busca dinâmica facetada, implementado no protótipo TDF-Bíblia, possibilitou melhorias na experiência do usuário relacionada à busca e à exploração do acervo da biblioteca digital. Essas melhorias puderam ser observadas, principalmente, em três cenários de uso: seleção pragmática de documentos, exploração e análise do acervo. No contexto da avaliação, que foi realizada na forma de um experimento, os resultados obtidos apontaram para melhorias relacionadas à recuperação da informação e à satisfação dos usuários da biblioteca. O mecanismo de busca dinâmica facetada fornece suporte a um novo paradigma de acesso, permitindo a exploração guiada do acervo e fazendo uma ponte entre os processos de busca e a navegação. Além disso, o uso desta abordagem se mostrou intuitivo e facilmente utilizado pelos usuários.

Palavras-chave: Bibliotecas digitais, representação da informação, organização da informação, teoria da classificação facetada, taxonomias dinâmicas, recuperação da informação.

ABSTRACT

Digital libraries, particularly those whose collection consists of theses and dissertations, lack mechanisms best suited for its exploration. Users need to prepare query expressions, but the inability to express precisely and clearly the needed information, in addition to problems related to natural language ambiguity, generates negative impact on results. This research used a model based on Faceted Classification Theory for knowledge representation and organization in a theses and dissertations digital library, besides dynamic taxonomies, to compose the information retrieval mechanism. New retrieval components were developed as a computational prototype, which was object of a comparative study. This comparative study, developed as an experiment, sought to determine the contribution of the mechanisms of knowledge representation and organization, besides the new mechanism of information retrieval, in facilitating the information discovery, as well as increasing digital library users' satisfaction. The mechanism of dynamic faceted search, implemented in the TDF-Bíblia prototype, made it possible improvements in the user's experience related to search and exploration of digital library collection. These improvements could be observed mainly in three usage scenarios: pragmatic documents selection, exploration and analysis of the collection. In the context of the evaluation, conducted as an experiment, the results showed improvements related to information retrieval and library users' satisfaction. The mechanism of dynamic faceted search provides support for a new access paradigm, enabling guided exploration of the collection, and making a bridge between searching and browsing processes. Furthermore, using this approach proved to be intuitive and easily used by users.

Keywords: Digital libraries, information representation, information organization, faceted classification theory, dynamic taxonomies, information retrieval.

LISTA DE FIGURAS

FIGURA 1 - O processo de comunicação	16
FIGURA 2 - Modelo de arquitetura da Biblioteca Digital Alexandria	50
FIGURA 3 - Caso de uso genérico de uma biblioteca digital	51
FIGURA 4 - O Modelo Tríplice de Interação.....	52
FIGURA 5 - O Modelo de Referência DELOS: componentes e atores	54
FIGURA 6 - O Modelo de Referência DELOS: principais conceitos	54
FIGURA 7 - Conceitos do Modelo 5S e sua correspondência com o Modelo DELOS.....	60
FIGURA 8 - FRBR: entidades e relacionamentos do Grupo 1	64
FIGURA 9 - FRBR: entidades e relacionamentos dos grupos 1 e 2	64
FIGURA 10 - FRBR: entidades e relacionamentos dos Grupos 1, 2 e 3.....	65
FIGURA 11 - Modelo de dados do DSpace.....	75
FIGURA 12 - O processo de ingresso de conteúdos no DSpace.....	80
FIGURA 13 - Etapas do fluxo de submissão de conteúdo no DSpace	81
FIGURA 14 - Taxonomia faceta para organização de informações turísticas.....	110
FIGURA 15 - Extensão rasa (a esquerda) e profunda (a direita) do conceito B.	114
FIGURA 16 - Taxonomia dinâmicas.....	118
FIGURA 17 - Cálculo da extensão profunda do conceito C.	119
FIGURA 18 - Conjunto relacionado de conceitos para C ($CR(C)$).	119
FIGURA 19 - Taxonomia reduzida para o foco definido a partir de C.....	120
FIGURA 20 - Componentes de uma biblioteca digital	131
FIGURA 21 - Modelo Tríplice de Interação	132
FIGURA 22 - Etapas metodológicas.....	153
FIGURA 23 - Arquitetura lógica.....	156
FIGURA 24 - Modalidade de busca pelos metadados.	168
FIGURA 25 - Modalidade de busca pela taxonomia facetada.....	168
FIGURA 26 - Tela inicial do protótipo TDF-Bíblia.....	186
FIGURA 27 - Exemplo da Busca Dinâmica-Facetada.....	188
FIGURA 28 - Resultado parcial da Busca Dinâmica Facetada.	189
FIGURA 29 - Integração da Busca Dinâmica Facetada com a busca textual.	190
FIGURA 30 - Gráficos com as frequências das pontuações obtidas nas tarefas de busca e recuperação da informação.	194
FIGURA 31 - Gráficos com as frequências do grau de satisfação com a tarefa de encontrar informação na biblioteca digital.	199
FIGURA 32 - Gráficos com as frequências do grau de satisfação com o mecanismo de busca da biblioteca digital.	200

LISTA DE TABELAS

TABELA 1 - Desempenho nas tarefas de busca e recuperação da informação.	197
TABELA 2 - Satisfação dos usuários.....	202

LISTA DE QUADROS

QUADRO 1 - Comparativo entre os elementos dos padrões Dublin Core, EDT-MS e MTD-BR.....	69
QUADRO 2 - Exemplos de níveis de suporte a preservação.....	76
QUADRO 3 - Exemplos de objetos no modelo de dados do DSpace.....	77
QUADRO 4 - Objetos e ações no sistema de autorização do DSpace.....	79
QUADRO 5 - Ações possíveis para cada etapa do processo de submissão no DSpace.....	81
QUADRO 7 - Relacionamentos paradigmáticos e sintagmáticos.....	92
QUADRO 6 - Relacionamentos presentes em um esquema facetado.....	92
QUADRO 8 - Princípios para a criação de classificações facetadas.....	102
QUADRO 9 - Comparação entre os modelos de avaliação.....	139
QUADRO 10 - Principais etapas da pesquisa.....	151
QUADRO 11 - Configuração da avaliação baseada no Modelo de Avaliação Computacional.....	170
QUADRO 12 - Categorização dos respondentes quanto ao grau de contribuição relacionada ao preenchimento do Questionário Q1.....	175
QUADRO 13 - Categorização das necessidades de informação relatadas pelos respondentes do Questionário Q1.....	177
QUADRO 14 - Questões apresentadas aos participantes.....	179
QUADRO 15 - Atividades do experimento e previsão de tempo.....	180
QUADRO 16 - Respostas obtidas para as questões abertas do Questionário de Satisfação (Q2).....	203

SUMÁRIO

1	INTRODUÇÃO	15
1.1	TEMA.....	16
1.2	PROBLEMA E JUSTIFICATIVA	19
1.3	OBJETIVOS.....	26
1.4	ESTRUTURA DA TESE	26
2	FUNDAMENTOS HISTÓRICOS, TEÓRICOS E METODOLÓGICOS	28
2.1	BIBLIOTECAS DIGITAIS.....	29
2.1.1	Bibliotecas convencionais e digitais – definições e convergências	33
2.1.2	Perspectivas para as bibliotecas.....	37
2.1.3	A pesquisa na área de bibliotecas digitais	40
2.1.4	Iniciativas brasileiras relacionadas a bibliotecas digitais	46
2.1.5	Modelos teóricos para bibliotecas digitais.....	49
2.1.6	Metadados: modelos e padrões.....	61
2.1.7	Sistemas gerenciadores de biblioteca digital	70
2.2	ORGANIZAÇÃO DO CONHECIMENTO E A TEORIA DA CLASSIFICAÇÃO FACETADA.....	84
2.2.1	Princípios e metodologia para a construção de classificações facetadas	101
2.2.2	A representação temática usando o padrão de metadados Dublin Core	105
2.3	CONSTRUÇÃO DE UMA TAXONOMIA FACETADA	107
2.4	FORMALIZAÇÃO DE UMA TAXONOMIA FACETADA.....	109
2.4.1	Modelo de implementação – Taxonomias Dinâmicas.....	112
2.5	AVALIAÇÃO DE BIBLIOTECAS DIGITAIS	123
2.5.1	Avaliação de usabilidade em bibliotecas digitais	141
2.6	TRABALHOS RELACIONADOS E CONSIDERAÇÕES	145
3	METODOLOGIA.....	149
3.1	ETAPAS METODOLÓGICAS.....	150

3.2	DESENVOLVIMENTO DO PROTÓTIPO COMPUTACIONAL	154
3.2.1	Arquitetura e estratégia de implementação	154
3.2.2	O protótipo desenvolvido	166
3.3	AVALIAÇÃO	169
3.3.1	Definição operacional: perguntas a serem respondidas pela avaliação	171
3.3.2	Experimento	173
3.3.3	Codificação dos dados coletados	181
3.3.4	Análise e interpretação dos dados	183
4	RESULTADOS	185
4.1	O PROTÓTIPO TDF-BÍBLIO	185
4.2	AVALIAÇÃO	191
4.2.1	Levantamento Inicial	192
4.2.2	Tarefas de busca e recuperação da informação.....	193
4.2.3	Satisfação do usuário.....	198
5	CONSIDERAÇÕES FINAIS	205
	REFERÊNCIAS	210
	APÊNDICE A.....	223
	APÊNDICE B.....	224
	APÊNDICE C.....	230
	APÊNDICE D.....	232

1 INTRODUÇÃO

O presente trabalho de pesquisa problematiza questões relacionadas às bibliotecas digitais a partir de uma comparação com as bibliotecas convencionais. Colocando à parte o componente tecnológico que potencializa o acesso a partir do momento em que retira as barreiras de tempo e espaço, observa-se que as bibliotecas digitais não proporcionam a mesma eficácia das bibliotecas convencionais no que se refere à transferência da informação. Podemos dizer que bibliotecas digitais são, na maioria dos casos, meros sistemas executores de consultas que devem ser formuladas e objetivadas por seus usuários. Não se observa o componente de interação e integração social e intelectual característico da biblioteca convencional. Na medida em que se retira o componente humano, na maioria das vezes representado no papel do bibliotecário de referência, perde-se muito em termos de capacidade de comunicação, refletindo-se no acesso e transferência da informação.

Diante do exposto, surge a necessidade de tornar as bibliotecas digitais mais eficientes no que diz respeito a sua função de mediação entre os usuários e os itens de seu acervo, tornando o processo de comunicação mais eficaz nestes sistemas. Os caminhos apontados e que serão detalhados neste texto se referem ao uso intensivo de estruturas semânticas de representação e organização do conhecimento. Neste sentido, utiliza-se um modelo baseado na Teoria da Classificação Facetada para representação e organização do conhecimento em uma biblioteca digital de teses e dissertações, além de taxonomias dinâmicas para compor o mecanismo de recuperação da informação. Foram desenvolvidos novos componentes de recuperação e acesso, na forma de um protótipo computacional, que foi objeto de um estudo comparativo. O estudo comparativo, desenvolvido na forma de um experimento, buscou determinar a contribuição dos mecanismos de representação e organização do conhecimento, além dos novos mecanismos de recuperação da informação, na facilitação da descoberta da informação, assim como no aumento da satisfação dos usuários da biblioteca digital.

No tópico a seguir será feita a apresentação do tema deste trabalho, além de apresentada uma breve discussão do contexto no qual ele está inserido, incluindo temas como bibliotecas convencionais e digitais, seus respectivos papéis no processo de comunicação do conhecimento, e ainda problemas e possíveis soluções apontadas por alguns autores.

1.1 Tema

Um tema de pesquisa engloba um conjunto de problemas de pesquisa relacionados e afins. Esta seção se propõe em apresentar o tema geral da presente pesquisa, de modo a, em seguida, delimitar o problema a ser tratado pela mesma.

O objetivo principal e razão de ser das bibliotecas, segundo Mey (1995), é atender às demandas do público (usuários) quanto aos registros do conhecimento, ou seja, quanto ao conteúdo dos itens, suportes físicos do conhecimento. Para alcançar este objetivo, o bibliotecário deve analisar e tratar os itens preparando-os com vista a seu uso. Neste processo de tratamento, os bibliotecários elaboram representações desses itens de forma a facilitar a busca e recuperação. Estas representações abrangem tanto o aspecto físico dos itens quanto o seu conteúdo temático. Os bibliotecários criam representações e instrumentos de acordo com as características da biblioteca, de seu público, e do próprio item, de modo a permitir que o usuário encontre o seu item e que o item encontre o seu usuário.

Lancaster (2004, p.285), ao discorrer sobre os problemas da recuperação da informação, nos alerta que as representações não são perfeitas. Da mesma forma, os pedidos (consultas) raramente são representações perfeitas das necessidades de informação dos usuários e os enunciados de busca podem não ser representações perfeitas dos pedidos. Além disso, o referencial ou conhecimento prévio do usuário pode não coincidir com o referencial de um especialista em informação ou mesmo com o referencial dos autores. Conclui-se que “o problema da recuperação da informação consiste essencialmente em procurar cotejar aproximações de necessidades de informação com aproximações de mensagens” (LANCASTER, 2004). Diante disso, não é de admirar que os resultados nem sempre sejam completamente satisfatórios.

O processo de comunicação em uma biblioteca pode ser representado pelo esquema clássico da comunicação proposto por Shannon e Weaver (1975) e ilustrado na FIGURA 1 abaixo.



FIGURA 1 - O processo de comunicação

Fonte: adaptado de Shannon e Weaver (1975)

Pode-se dizer que as FONTES são os itens que compõem o acervo da biblioteca; TRANSMISSORES são os serviços disponibilizados pela biblioteca; CANAIS são os instrumentos usados pela biblioteca (catálogos, índices, etc.); RECEPTORES e DESTINO são os usuários da biblioteca. Pode-se considerar como FONTE DE RUÍDO tudo que interfere no processo de comunicação e dificulta a recepção ou a compreensão pelo usuário. Como bem observa Mey (1995), assim como as bibliotecas têm mensagens a comunicar ao usuário, este também possui suas próprias mensagens internas quando se dirige à biblioteca, refletindo os seus desejos e demandas por informação. Se o usuário deseja um assunto específico, os canais da biblioteca devem estar preparados para receber sua mensagem. Assim, a FONTE passa a ser o usuário; o TRANSMISSOR, sua forma de expressão; o CANAL, os instrumentos de que dispõe a biblioteca; o RECEPTOR, os serviços da biblioteca; o DESTINO, os itens do acervo. Neste caso a FONTE DE RUÍDO é a dificuldade que o usuário encontra em expressar sua mensagem por meio dos canais disponíveis. Trata-se de uma via de mão dupla (MEY, 1995).

Quando os bibliotecários elaboram suas mensagens, baseadas em representações dos itens, estas mensagens se destinam tanto a informar aos usuários sobre os itens disponíveis, como permitir que os usuários expressem sua mensagem interna, fazendo com que os itens encontrem seus usuários. Portanto, todos os instrumentos e serviços disponibilizados pela biblioteca devem se adequar ao público e aos itens, desempenhando uma função mediadora.

Até alguns anos atrás as bibliotecas convencionais eram a principal fonte de informação para pesquisadores e estudantes. Hoje em dia estas informações se encontram espalhadas em meio digital, principalmente na grande rede de computadores. Bibliotecas digitais já entraram em cena há vários anos. Greenstein e Thorin (2002) apresentam a seguinte definição simplificada de biblioteca digital: "... é uma biblioteca na qual as coleções são armazenadas em formato digital (ao invés dos formatos tradicionais como impressos, microfilme e outras mídias) acessíveis por computadores" (tradução nossa). No capítulo seguinte serão apresentadas e discutidas definições mais completas e consistentes com a importância deste tema.

Quando se passa das bibliotecas convencionais para as bibliotecas digitais, percebe-se que surgem novas possibilidades e também desafios. Mas no que diz respeito ao processo de comunicação apresentado acima, ele não se altera. Uma biblioteca digital tem como objetivo facilitar o acesso de seus usuários ao conteúdo digital, usando para tanto uma infraestrutura tecnológica. O principal papel das bibliotecas digitais continua sendo o de exercer uma função mediadora entre, de um lado os itens (digitais) disponíveis em seu

acervo e de outro as necessidades de informação de seus usuários, fazendo com que a informação possa fluir, com o mínimo de distorções, em ambos os sentidos.

Em geral, as bibliotecas digitais permitem buscas com base em metadados descritivos (autor, data de publicação, etc.), buscas textuais em campos específicos (título, palavras chave, resumo, etc.), ou ainda no texto completo dos objetos do acervo (*full text*). Desta forma, elas são capazes de responder a consultas como: “liste todos os itens publicados a partir de 1999 em ordem alfabética” ou “liste todos os itens que contenham as palavras ‘análise’ e ‘assunto’”. Mas, em geral, não são capazes de responder a consultas como “liste todos os itens que tratam de análise conceitual”. Especialmente se o termo “análise conceitual” não fizer parte da terminologia dos documentos, não aparecendo nos campos usados para a indexação. É neste ponto que a adição de semântica em uma biblioteca digital pode ser útil. Informação semântica, representada por metadados vinculados a cada item, além de estruturas semânticas de representação do conhecimento do domínio, possibilitam um contexto para as buscas, e pode ser usada para responder a estes tipos de consultas. A flexibilidade e a capacidade adicional de busca e exploração do acervo proporcionada por esta abordagem podem tornar uma biblioteca digital mais útil para a maioria dos usuários.

Estruturas semânticas de representação e organização do conhecimento são objetos antigos de pesquisa da biblioteconomia e da ciência da informação. Sistemas de classificação, listas de cabeçalhos de assunto e tesouros são instrumentos encontrados na maioria das bibliotecas convencionais. Estes instrumentos, em conjunto com a assistência de bibliotecários de referência, possibilitam serviços de valor agregado para os usuários destas bibliotecas. No entanto, quando se passa para o contexto digital, observa-se a ausência deste tipo de serviço. O que se percebe é que as técnicas e instrumentos desenvolvidos ao longo dos anos pela biblioteconomia não vêm sendo amplamente adotados ou adaptados para o meio digital.

Sorgel (2009) indica dois importantes princípios para aplicação de estruturas de representação e organização do conhecimento em bibliotecas digitais: 1) tais estruturas devem ser usadas “por trás dos panos”, ou seja, de forma oculta para o usuário e de modo a assisti-lo, incrementando as buscas e o processamento dos resultados; 2) nos momentos em que for útil para os usuários interagir com tais estruturas, deve-se fornecer visualizações e interações amigáveis de modo a guiar os usuários na interpretação das informações e do seu contexto.

Os usuários geralmente desejam fazer buscas por assuntos, conceitos ou ideias, e estas são as buscas mais difíceis em uma biblioteca digital. Na maioria das vezes o usuário tem de fornecer palavras, mas estas palavras nem sempre expressam de forma

adequada o que se tem em mente. Muitas vezes os usuários não conseguem elaborar uma expressão de busca. As bibliotecas digitais deveriam dar suporte ao usuário neste tipo de busca por significado e este é o principal papel das estruturas semânticas de representação e organização do conhecimento. São exemplos destas estruturas, com grau variável de formalismo: dicionários, glossários, tesouros, listas de cabeçalhos de assunto, mapas conceituais, esquemas de classificação e taxonomias, mapas de tópicos e ontologias.

Tendo sido apresentado o tema do presente trabalho, assim como os principais conceitos envolvidos, incluindo a visão de alguns autores, passaremos agora à discussão das questões que conduzem e justificam a presente pesquisa.

1.2 Problema e justificativa

Um problema de pesquisa conduz a busca de um maior entendimento de questões postas pelo real, ou ainda a busca de soluções para problemas nele existentes, tendo em vista a sua modificação para melhor (LAVILLE e DIONNE, 1999). A presente pesquisa se pode ser enquadrada nesta segunda abordagem.

Laville e Dionne (1999) definem uma problemática como o quadro no qual se situa a percepção de um problema. Nesse sentido, a problemática ou quadro de referência é o conjunto de fatores que fazem com que o pesquisador conscientize-se de um determinado problema, veja-o de um modo ou de outro, imaginando eventuais soluções. O problema e sua solução em vista não passam da ponta de um iceberg, ao passo que a problemática é a importante parte escondida. Uma operação essencial do pesquisador consiste em desvendá-la. É o que se propõe inicialmente fazer nesta seção, na qual serão brevemente apresentadas as principais questões que motivaram o desenvolvimento desta pesquisa, bem como os fatores que justificam a sua relevância. Partimos do pressuposto de que as bibliotecas digitais compartilham dos mesmos princípios e propósitos da biblioteca convencional. O que muda é a tecnologia de suporte, que ao mesmo tempo em que abriu novas possibilidades, derrubando as barreiras de tempo e espaço, trouxe consigo algumas limitações que serão discutidas a seguir.

A partir da década de 1960, muitos estudos e experimentos foram realizados na área de recuperação da informação em ambientes digitais (ELLIS, 1996). Estes estudos, na sua grande maioria, se apoiam em métricas quantitativas como revocação e precisão. Revocação é a proporção entre o número de itens recuperados em uma busca considerados relevantes e o número total de itens relevantes presentes no acervo. Já a precisão é a

proporção entre o número de itens relevantes recuperados e o número total de itens recuperados na busca. Observa-se que ambas as medidas, revocação e precisão, são derivadas do critério de relevância. Relevância pode ser entendida como uma medida da adequação entre um item do acervo e a necessidade de informação do usuário. No entanto, argumenta-se que o critério de relevância se mostra bastante subjetivo (ELLIS, 1996). O fato de um item recuperado ser considerado relevante para uma necessidade de informação declarada por uma pessoa depende do momento e das circunstâncias da sua busca, ou seja, do seu conhecimento prévio e está fortemente relacionado com sua capacidade em objetivar esta necessidade de informação em termos de uma declaração. A incapacidade em expressar precisa e claramente a informação de que se necessita impacta na comunicação entre o usuário e o sistema ou mediador. Freyne e outros (2007) observam que as consultas típicas são vagas e pouco específicas em revelar a verdadeira necessidade de informação do usuário. A grande maioria das consultas possuem em média entre 2 e 3 termos, além disso, os usuários normalmente utilizam termos que são diferentes daqueles utilizados nos documentos que eles procuram.

Uma “necessidade de informação”, como observa Dias (2001), pode ser vista como a carência de conhecimento que impede com que a pessoa dê prosseguimento a uma tarefa qualquer e, como tal, por definição, é de difícil diagnóstico. Além disso, o retorno obtido pode mudar a sua noção sobre sua própria necessidade de informação anteriormente declarada. Uma “necessidade de informação” se revela uma entidade bastante dinâmica, cujo dinamismo e mudança se apresentam durante o próprio processo de interação com o sistema de recuperação da informação. Desta forma, uma possível solução para este problema não é desenvolver sistemas baseados no conceito de relevância, mas sim baseá-los no conceito de “capacidade exploratória”. O usuário necessita de um sistema exploratório eficiente, além de um sistema executor de consultas. Esta capacidade de exploração pode ser encontrada nas bibliotecas tradicionais no papel do bibliotecário de referência, dialogando com o usuário em torno de uma consulta para melhor defini-la, mas é raramente encontrada nos esquemas de recuperação da informação presentes nas bibliotecas digitais e na Internet.

Alguns autores detectam nas bibliotecas digitais a ausência da característica de elemento de integração social e intelectual de suas respectivas comunidades e organizações, elemento este de presença marcante na biblioteca tradicional (DIAS, 2001). Para Morigi e Pavan (2004), verifica-se que, nas bibliotecas convencionais, ocorre um maior contato entre o bibliotecário, sempre presente, e os usuários, sendo marcada a interação face a face e a comunicação oral. Os autores afirmam que o contato é mantido por relações visíveis ao olho humano, de forma que, como organização, é capaz de criar sociabilidade ou,

ao menos, referendar tipos de sociabilidade definidos historicamente. Isto permite que os primeiros possam conhecer mais de perto seus usuários, suas necessidades e suas formas de expressão. Este conhecimento pode ser usado na organização e representação do acervo, assim como na construção de instrumentos mais efetivos de comunicação. A maioria das bibliotecas digitais se restringe a meros sistemas executores de consultas que precisam ser formuladas por seus usuários.

Dias (2001) observa que uma série de conhecimentos acumulados sobre como organizar bibliotecas e como tratar informações não está sendo devidamente aproveitada pela instituição correspondente no contexto digital, a biblioteca digital. Aparentemente, as teorias e técnicas desenvolvidas pela biblioteconomia ao longo dos anos não vêm sendo amplamente aplicadas e ou adaptadas para o contexto das bibliotecas digitais. Isso ocorre, principalmente, no que diz respeito ao uso de estruturas de representação e organização do conhecimento. A abordagem de busca textual (*full-text*) predomina. Uma possível explicação para isso pode estar relacionada a certo distanciamento entre os campos teóricos da biblioteconomia e da computação, sendo este último o que mais tem atuado na concepção e desenvolvimento deste tipo de sistema de informação. Hjørland (2000), ao analisar o campo teórico da ciência da informação, argumenta que as abordagens baseadas em tecnologias influenciaram indiretamente a teoria, ou seja, ocorreu a introdução do conceito de “informação” (ciência da informação) em oposição ao conceito mais simples de “documento” (documentação). Segundo o autor, a ambição do campo, como um campo de estudo, é produzir conhecimentos e princípios que possam ser usados por novas tecnologias. No entanto, o que se observa é que o processo de desenvolvimento das novas tecnologias vem ignorando o conhecimento existente. Além disso, a tendência observada é que a ciência da informação tem usado a tecnologia de forma passiva, sem contribuir para o seu desenvolvimento. Este problema, relacionado a não aplicação no contexto digital dos conhecimentos desenvolvidos no campo da biblioteconomia e ciência da informação, também é observado por Milstead (1998) quando afirma:

O grande crescimento e popularidade dos mecanismos de busca, com seus algoritmos primitivos, tem tido alguns efeitos negativos, em minha opinião. Alguns destes mecanismos parece terem sido desenvolvidos por pessoas que vislumbraram uma necessidade, mas que não possuem nem uma vaga ideia de que existe um histórico de pesquisas e desenvolvimento de ferramentas para atender a necessidades similares. Existe pouca evidência de que estes desenvolvedores tenham utilizado *Dialog* ou um catálogo de biblioteca. (MILSTEAD, 1998, tradução nossa)

Não devemos simplesmente criticar os desenvolvedores dos mecanismos de busca utilizados na Internet e nas bibliotecas digitais por não terem considerado as teorias da

biblioteconomia e ciência da informação. Não há dúvida de que os mecanismos de busca são um grande sucesso. Acreditamos que nós, cientistas da informação, é que temos que demonstrar que as estruturas semânticas de representação e organização do conhecimento têm um papel a desempenhar no ambiente digital. Em outras palavras, os mecanismos de busca textual devem ser considerados uma abordagem, entre outras, e as vantagens e desvantagens de diferentes abordagens devem ser demonstradas cientificamente.

Segundo Hjørland e Albrechtsen (1995), todo sistema de informação está destinado a sustentar a produção, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso de conhecimentos e deveria ser concebido para um grupo social concreto e para áreas determinadas. Só há sentido em falar de um conhecimento como informativo ou relevante em relação a um pressuposto conhecido e compartilhado com outros. Estas considerações sugerem a necessidade da formação de um entendimento compartilhado sobre o domínio de conhecimento de modo a facilitar a comunicação no contexto de um sistema de recuperação da informação. Este entendimento compartilhado emerge naturalmente no ambiente de uma biblioteca convencional, apoiado pelos instrumentos e nas interações entre os usuários e os bibliotecários de referência. O problema surge a partir do momento em que a tecnologia, que trouxe ganhos extraordinários em termos do acesso em tempo real e independente de localização, abriu mão, tanto dos instrumentos de representação do conhecimento, como das interações humanas.

Abordagens baseadas em busca textual passaram a ser predominantes no ambiente digital. Com o intuito de oferecer soluções mais sofisticadas para a busca e descoberta de informação, uma abordagem frequentemente adotada é o desenvolvimento de mecanismos que permitem buscas avançadas no acervo. No entanto, como bem observa Cunha (2008), é um mito pensar que o usuário planeje uma estratégia de busca sofisticada. Uma busca geralmente corrobora o princípio do menor esforço; “o usuário deseja que o sistema recupere informação com qualidade e que ele tenha facilidade na utilização”. Este comportamento do usuário se aplica tanto à biblioteca convencional quanto à digital.

Os processos de busca e recuperação de informação, na grande maioria das bibliotecas digitais, assim como ocorre na Internet, se baseiam principalmente nas buscas textuais e na navegação. Isto acontece nas bibliotecas digitais apesar dos processos de tratamento da informação que ocorrem nas mesmas. As buscas textuais se mostram úteis, principalmente quando se trata de grandes volumes de informação, no entanto, elas sofrem da ambiguidade da linguagem natural. Isso negligencia a importância dos metadados e não engaja os usuários no processo de compartilhamento de conhecimento. As buscas retornam uma grande quantidade de resultados, que precisam ser filtrados. Os algoritmos de

ordenação dos resultados (*ranking*) ajudam a encontrar *sites web* relevantes, mas não podem ser facilmente aplicados a documentos em uma biblioteca digital.

Pelo exposto acima, argumenta-se que as bibliotecas digitais, apesar das novas possibilidades trazidas pela tecnologia, apresentam limitações que impactam no processo de comunicação do qual são mediadoras. Quando se compara com as bibliotecas convencionais, observa-se que estas possuem mecanismos mais eficientes de comunicação com seus usuários. Estas limitações estão relacionadas à sua incapacidade de prover mecanismos que possibilitem a formação de um entendimento compartilhado como condição necessária para a eficiência e eficácia do processo de comunicação. As bibliotecas digitais fornecem uma condição necessária, mas não suficiente para a efetiva comunicação, disseminação e compartilhamento do conhecimento. Estas bibliotecas carecem de mecanismos mais eficazes para a exploração do acervo e recuperação da informação. Uma possível solução é o uso de estruturas semânticas de representação e organização do conhecimento. Entre estas estruturas, as classificações facetadas têm sido apontadas como adequadas para a organização do conhecimento em ambientes digitais (GODERT, 1991; INGWERSEN; WORMELL, 1992; KWASNIK, 1999; ELLIS; VASCONCELOS, 2000; LOUIE et al., 2003; ADKISSON, 2005; LA BARRE, 2006; BROUGHTON, 2006). Alguns trabalhos de pesquisa anteriores aplicaram a teoria da classificação faceta para este propósito com relativo sucesso (LIMA, 2004; MACULAN, 2011).

Com o surgimento e consolidação das bibliotecas digitais, um interessante exercício é retomar as cinco leis da biblioteconomia, propagadas por Ranganathan, e verificar se as mesmas continuam válidas diante deste novo contexto: 1) livros são para serem usados; 2) a cada leitor o seu livro; 3) a cada livro o seu leitor; 4) poupe o tempo do leitor; 5) a biblioteca é um organismo em crescimento. Estas leis continuam em perfeita sintonia com o ambiente digital, no entanto, percebe-se a necessidade de atenção especial à terceira lei, ou seja, fazer com que os recursos informacionais estejam bem organizados de forma que “eles encontrem os seus usuários”. Este trabalho de pesquisa se justifica pela lacuna observada no que se refere à aplicação de estruturas de representação e organização do conhecimento na organização de acervos digitais. Tem o potencial de determinar o valor de uma abordagem mais estruturada e teoricamente fundamentada na biblioteconomia para representação e organização do conhecimento em uma biblioteca digital. Mais especificamente, este trabalho tem o potencial de determinar cientificamente as vantagens de uma abordagem teoricamente fundamentada na classificação facetada para a organização do acervo em comparação à abordagem amplamente difundida, baseada em processos automáticos de indexação e busca textual.

Saracevic (2004) levanta uma importante questão relacionada à interação dos usuários com as bibliotecas digitais. A partir da análise dos dados de cerca de 80 avaliações conduzidas e apresentadas na literatura, o autor conclui que os usuários apresentam muitas dificuldades com as bibliotecas digitais:

- Eles geralmente não entendem suas funcionalidades;
- Eles possuem concepções distintas da biblioteca digital que os bibliotecários ou desenvolvedores;
- Eles têm pouca familiaridade com as funcionalidades e possibilidades de interação fornecidas pelas bibliotecas digitais;
- Eles geralmente entram em “becos sem saída” durante a interação com a biblioteca digital.

A percepção dos usuários não é a mesma que as dos bibliotecários e desenvolvedores. Saracevic (2004) ironiza ao afirmar que “os usuários são de Vênus e as bibliotecas digitais são de Marte”. A presente pesquisa pode contribuir para melhorar este quadro, tornando o uso da biblioteca digital mais intuitivo e próximo do modelo mental dos usuários.

As bibliotecas digitais, apesar de todo o avanço tecnológico, se mostram, em determinados aspectos, menos eficientes quando se compara com as bibliotecas convencionais. Existe uma visão compartilhada por diversos autores de que as teorias e técnicas, desenvolvidas ao longo dos anos pela biblioteconomia e a ciência da informação, não vêm sendo sistematicamente aplicadas na concepção destes sistemas. O foco vem sendo colocado exclusivamente na tecnologia, e esta, sozinha, não permite avançar muito.

O alcance e o impacto das bibliotecas digitais são ainda bastante reduzidos no Brasil, chegando a ser mínima a sua repercussão. Este fato contrasta com a grande necessidade de bibliotecas digitais em um país como o Brasil, especialmente as voltadas para áreas estratégicas como ensino, pesquisa, saúde, empreendedorismo, meio ambiente e cultura. Outro fator que contribui para tal necessidade é a quantidade gigantesca de informação que se torna disponível a cada dia na Internet, informação esta que, sem o devido tratamento, se reduz a informação dispersa, desorganizada, perdida, evidenciando a necessidade de criação de bibliotecas digitais.

Especificamente no caso das bibliotecas digitais de teses e dissertações, um grande esforço vem sendo aplicado pelo IBICT na construção destas bibliotecas nas universidades brasileiras. Os resultados obtidos têm sido bastante satisfatórios no que se refere à construção dos acervos em cada instituição, assim como na interoperabilidade entre estes

acervos de modo a possibilitar buscas integradas. Para tanto, o IBICT atua como provedor de serviços, coletando e agregando os metadados a partir das várias instituições consorciadas e fornecendo o serviço de busca integrada e acesso aos acervos. No que se refere à organização destes acervos, a abordagem utilizada consistiu do desenvolvimento e adoção do padrão de metadados MTD-BR, baseado no Dublin Core, de modo a possibilitar a recuperação baseada em elementos textuais como título, assunto, resumo, etc. As teses e dissertações representam o produto final das atividades de pesquisa em nível de pós-graduação. Estes documentos precisam ser efetivamente acessados e utilizados como insumo em novas pesquisas. A abordagem proposta neste trabalho de pesquisa pode trazer grande contribuição na medida em que pretende melhorar e facilitar a recuperação e consequentemente contribuir para o uso mais efetivo destes acervos no âmbito dos programas de pós-graduação no Brasil.

Pelo exposto, conclui-se que as bibliotecas digitais de teses e dissertações carecem de mecanismos mais adequados para exploração do acervo. Os usuários precisam elaborar expressões de consulta, sendo que a inabilidade em expressar precisa e claramente a informação de que se necessita, além de problemas relacionados à ambiguidade da linguagem natural, gera impacto negativo nos resultados obtidos pelas consultas. Além disso, observa-se a ausência do uso de estruturas semânticas que possam representar adequadamente o conteúdo dos itens do seu acervo e, desta forma, possibilitar formas mais efetivas de exploração do acervo e recuperação da informação. Por essas razões, esta pesquisa tem como proposta a aplicação da teoria da classificação facetada, para a organização do acervo, bem como o uso de taxonomias dinâmicas para o desenvolvimento de novos mecanismos para exploração e recuperação da informação na biblioteca. Com isso, este estudo tem como expectativa responder as seguintes questões: como melhorar e facilitar a interação do usuário, em tarefas de exploração e recuperação da informação em bibliotecas digitais de teses e dissertações? O uso de uma estrutura classificatória facetada, para organização do acervo, e taxonomias dinâmicas para compor um novo mecanismo de exploração e a recuperação da informação, pode melhorar a experiência e a satisfação dos usuários da biblioteca?

Levando-se em consideração a proposta e a expectativa, a seguir são apresentados os objetivos da presente pesquisa.

1.3 Objetivos

Esta pesquisa tem como objetivo facilitar a exploração e a recuperação da informação, por parte dos usuários, em bibliotecas digitais de teses e dissertações, assim como melhorar a experiência e a satisfação dos mesmos. Este objetivo geral será perseguido a partir do uso de mecanismos de representação e organização do acervo de uma biblioteca digital de teses e dissertações, baseados em uma estrutura classificatória facetada, além da implementação de mecanismos de recuperação da informação, baseados em taxonomias dinâmicas. Mais especificamente pretende-se: 1) tornar o uso da biblioteca digital mais intuitivo e próximo do modelo mental dos usuários; 2) contribuir para o uso mais efetivo destes acervos no âmbito dos programas de pós-graduação no Brasil; e 3) avaliar o desempenho de mecanismos baseados na teoria da classificação faceta e taxonomias dinâmicas, implementados em um protótipo de sistema de biblioteca digital, em facilitar a recuperação da informação e na melhoria da satisfação do usuário.

1.4 Estrutura da Tese

A presente tese possui quatro capítulos. No Capítulo 1 foi apresentado o tema da pesquisa, assim como os problemas que a motivaram, incluindo ainda seus objetivos e as justificativas de sua importância.

No Capítulo 2 são apresentados os fundamentos históricos, teóricos e metodológicos que embasaram o trabalho de pesquisa. São apresentadas as origens históricas das bibliotecas digitais, suas bases teóricas, perspectivas de evolução, além das principais linhas de pesquisa na área. Também são apresentadas as principais iniciativas brasileiras nesta área. Em seguida, são apresentados os principais modelos teóricos para bibliotecas digitais, enfatizando-se a importância destes modelos para a pesquisa e desenvolvimento da área. Também são brevemente apresentados os principais modelos e padrões de metadados utilizados para a representação dos acervos, além dos principais sistemas gerenciadores de biblioteca digitais disponíveis. Tendo em vista a proposta desta pesquisa, em utilizar uma estrutura classificatória facetada para organização do acervo em uma biblioteca digital de teses e dissertações, são apresentadas as bases teóricas da classificação facetada, além da discussão de suas vantagens em relação a outros tipos de estruturas classificatórias. Também é apresentada uma abordagem matematicamente

fundamentada para a formalização, representação e manipulação da estrutura facetada, baseada em taxonomias dinâmicas, visando o desenvolvimento de um protótipo computacional. Com a finalidade de subsidiar a elaboração de uma avaliação, são também apresentados os principais modelos e metodologias para avaliação de bibliotecas digitais.

No Capítulo 3 são apresentados os procedimentos e métodos adotados para o desenvolvimento do trabalho de pesquisa, visando atingir os objetivos traçados e apresentados no Capítulo 1, a partir da fundamentação teórica e metodológica apresentada no Capítulo 2.

No Capítulo 4 são apresentados os resultados da pesquisa e no Capítulo 5 são realizadas as considerações finais e apresentadas algumas perspectivas para trabalhos futuros.

2 FUNDAMENTOS HISTÓRICOS, TEÓRICOS E METODOLÓGICOS

Neste capítulo serão detalhadamente apresentados os fundamentos relacionados ao tema e aos objetivos desta pesquisa, e que embasaram o seu desenvolvimento, além de trabalhos já desenvolvidos por outros autores. A seguir é apresentado um breve histórico do surgimento das bibliotecas digitais, assim como alguns conceitos e definições. Também é feita uma breve análise sobre as distinções e pontos de convergência entre as bibliotecas convencionais e as digitais, bem como os relacionamentos e as perspectivas para o futuro dessas bibliotecas. Bibliotecas digitais constituem um campo de pesquisa bastante eclético na medida em que envolve conhecimento em diversas áreas. São brevemente apresentadas algumas das principais áreas de pesquisa envolvendo bibliotecas digitais. Também são apresentadas as principais iniciativas brasileiras na área de bibliotecas digitais, em especial as iniciativas do Instituto Brasileiro de Informação Científica e Tecnologia (IBICT), entre elas a da Biblioteca Digital de Teses e Dissertações (BDTD), de especial interesse para esta pesquisa.

A partir do entendimento de que os modelos teóricos são de grande importância para a consolidação e desenvolvimento da pesquisa em qualquer área, são apresentados dois modelos teóricos desenvolvidos para a área das bibliotecas digitais. Em seguida, são apresentados alguns modelos e padrões de metadados para bibliotecas digitais.

Os sistemas gerenciadores de bibliotecas digitais são de grande importância, visto que incorporam a implementação tecnológica dos serviços e funcionalidades que podem ser aplicados e disponibilizados em uma biblioteca digital. Eles representam o resultado da pesquisa na área de bibliotecas digitais, sendo objeto de constantes avanços. São também apresentados alguns sistemas gerenciadores de bibliotecas digitais, em especial o DSpace¹, que é amplamente utilizado no Brasil e que é adotado como plataforma para desenvolvimento de um protótipo computacional no contexto da presente pesquisa.

Tendo em vista a proposta desta pesquisa, utilizar uma estrutura classificatória facetada para organização do acervo em uma biblioteca digital de teses e dissertações, são apresentadas as bases teóricas da classificação facetada, além de discutidas suas vantagens em relação a outros tipos de estruturas classificatórias. Também é apresentada uma abordagem matematicamente fundamentada para a formalização, representação e manipulação da estrutura facetada na forma de taxonomias dinâmicas, visando o desenvolvimento de um protótipo computacional. Com a finalidade de subsidiar a

¹ Ver em: <http://www.dspace.org>

elaboração de uma avaliação, são também apresentados os principais modelos e metodologias para avaliação de bibliotecas digitais.

Ao final, são apresentados alguns trabalhos relacionados com esta pesquisa, bem como discutidas as relações existentes entre os mesmos, além de algumas considerações acerca de alguns tópicos da fundamentação teórica e metodológica apresentada.

2.1 Bibliotecas digitais

As bibliotecas digitais percorreram um longo caminho de evolução e, neste sentido, é importante sabermos de onde viemos para que possamos entender melhor onde estamos e quais os possíveis caminhos para o futuro.

A ideia de um repositório que registre todo o conhecimento humano se mostra recorrente na literatura. Parece ser um sonho obsessivamente renovado ao longo dos tempos, como aponta Sayão (2008). Ele aparece tanto na ficção literária como em iniciativas pioneiras de mentes criativas como Paul Outlet, H. G. Wells, Vannevar Bush e Ted Nelson. A tecnologia hoje disponível e, em especial a Internet, parecem nos colocar cada vez mais próximos deste sonho. No entanto é reconhecido ser necessário imprimir certo nível de estrutura e organização para que a Internet possa vir a cumprir este papel. Este é o objetivo das pesquisas em torno das tecnologias semânticas, interoperabilidade e recursos distribuídos, que se espera que possam permitir a criação de uma biblioteca universal, sem que para isso seja necessário que todas as informações estejam reunidas em um único lugar.

As primeiras iniciativas concretas no sentido de se construir uma biblioteca universal datam de 1895 quando Paul Outlet e Henri La Fontaine criaram, em Bruxelas, o Instituto Internacional de Bibliografia, cujo objetivo era registrar em fichas a produção mundial de impressos: o Repertório Bibliográfico Universal (CUNHA, 2008). Outlet trabalhou durante toda a sua vida perseguindo o sonho de reunir todo o conhecimento humano e classificá-lo de acordo com o sistema desenvolvido por ele e seu amigo Henri La Fontaine – a Classificação Decimal Universal (CDU). Na sua trajetória, Outlet trabalhou para solucionar os problemas técnicos, teóricos e organizacionais com o objetivo de tornar todo o conhecimento registrado disponível para aqueles que necessitam dele, para ele um problema crucial para a sociedade. Esta iniciativa pioneira fracassou, como aponta Sayão (2008), devido às dificuldades de manutenção de uma colaboração homogênea entre os

diversos países e por problemas técnicos relacionados ao processamento de informações de países e culturas heterogêneas.

Já no século seguinte, H. G. Wells (1866-1946) sugeriu a criação de uma enciclopédia universal, que conteria a memória planetária completa (CUNHA, 2008). Trata-se das primeiras ideias no sentido de construção de um repositório do conhecimento universal apoiado no uso da tecnologia ainda recente do microfilme (SAYÃO 2008). H. G. Wells, autor de clássicos da ficção científica como *A Guerra dos Mundos*, delineou, ainda em 1937, o que seria uma *Permanent World Encyclopaedia*, “[...] um repositório onde conhecimentos e ideias são recebidos, ordenados, sumarizados, classificados, analisados e comparados” (WELLS, 1938, p. 49 citado por SAYÃO, 2008). Este repositório, que integraria toda a inteligência do mundo, teria sua base de conhecimento apoiada na tecnologia de microfilmes, na época, ainda em sua infância. Na proposta de Wells, os microfilmes poderiam ser duplicados e enviados para qualquer lugar, possibilitando que estudantes e pesquisadores pudessem ter acesso. A “memória universal” proposta por Wells seria replicada e distribuída como uma forma de proteção à fragilidade dos registros humanos expostos à violência e à destruição provocadas pela frequência cada vez maior das guerras, vide o exemplo clássico da Biblioteca de Alexandria, destruída por diversas vezes em períodos de guerra. Como bem aponta Sayão (2008), Wells com sua mente visionária tocava em alguns dos principais desafios que estão sendo equacionados hoje pela área de bibliotecas digitais: integração de informações, universalidade e democratização do acesso, fontes de informação distribuída, preservação, além de aplicações importantes para pesquisa e ensino. “Substituindo ‘microfilme’ por ‘arquivos digitais’ verificamos o quão eram exatas as suas utopias” (SAYÃO, 2008).

Alguns anos depois, em 1945, Vannevar Bush – assessor científico do governo norte-americano – descreveu sua visão de um mecanismo para uso individual que seria uma espécie de biblioteca particular mecanizada, o MEMEX (BUSH, 1945). A concepção do MEMEX partiu da constatação de que a mente humana opera por associação, e isto torna indexar a informação de forma alfabética ou numérica, ineficiente. O pensamento é mantido em uma teia de conhecimento no cérebro; seria ideal encontrar uma forma de se fazer algo análogo de forma automatizada. O artigo sugere um aparelho que armazena publicações, livros e anotações, conjuntamente com fotos, como um suplemento à memória humana, e com clareza e permanência superior. O aparelho seria uma espécie de mesa de trabalho, com telas para projeção, teclado, botões e alavancas: o conteúdo seria armazenado em microfilme em um canto da mesa. Este conteúdo poderia ser rapidamente recuperado, sendo indexado por meio de códigos e mnemônicos para fácil acesso. O folheio entre as páginas seria feito por meio de uma alavanca que avançaria ou retrocederia dentro da

publicação selecionada. Existiria um botão que convenientemente levaria à página inicial do repositório. A maior parte do conteúdo para o MEMEX existiria para aquisição: livros, periódicos e jornais. Além disso, existiria uma plaqueta transparente onde poderiam ser colocadas anotações, imagens e memorandos, criados pelo usuário, para serem microfilmados e armazenados. Documentos poderiam ser ligados em pares visualmente, usando códigos para fazer as ligações. Outra alavanca navegaria entre estas ligações. Um caminho poderia ser criado seguindo estas ligações; a este caminho poderiam ser acrescentadas notas e comentários. Este caminho poderia ser armazenado, impresso e copiado para outro MEMEX, permitindo que outros pudessem se aproveitar do conhecimento registrado. Enciclopédias pré-relacionadas apareceriam, assim como bases de conhecimento em todas as áreas científicas. Pesquisadores e interessados teriam acesso a estas trilhas de conhecimento, e trabalhariam agregando e criando novos caminhos. A proposta de Bush é incrivelmente incisiva nos seus pontos principais; parece-nos especialmente relevante a seção onde o autor afirma que o problema da recuperação da informação tem relação com as técnicas usadas para indexar o conhecimento – a ordem alfabética ou numérica é muito pouco intuitiva. Bush previu, já em 1945, um sistema semelhante ao hipertexto, capaz de armazenar a literatura técnico-científica necessária ao exercício profissional do pesquisador moderno, na medida em que permitiria estabelecer comunicação entre os pares – hoje realizada por meio da Internet. Outra característica interessante da proposta de Bush é a possibilidade do usuário criar e manter anotações, além de trilhas de documentos relacionados. Desta forma, o usuário poderia ter um papel mais ativo no processo de organização da informação, coisa que nos dias de hoje ainda não ocorre no contexto das bibliotecas digitais.

De acordo com Sayão (2008), o MEMEX foi uma primeira inspiração para o hipertexto; porém, ao contrário do que muitos afirmam, a máquina de Bush não estabelecia a ideia de hipertexto da maneira e no nível de granularidade que hoje conhecemos. Isto foi estabelecido nos anos de 1960 por outro visionário: Theodor Holm Nelson, mais conhecido como Ted Nelson. Nelson foi quem cunhou os termos hipertexto e hiperímídia, cujos conceitos em sua essência vão muito além do seu uso atual no contexto da Internet. Dentro do projeto Xanadu, o objetivo de Nelson era estabelecer um sistema de bases de dados onde os escritores poderiam publicar diretamente os seus textos vinculando-os a outros documentos, constituindo desta forma, uma biblioteca eletrônica universal *on-line* de documentos interconectados (SAYÃO, 2008). Nelson se preocupou com a questão dos direitos autorais. Dentro da sua proposta, o autor da informação consultada recebia uma determinada quantia automaticamente sempre que alguém acessasse a sua obra. O projeto de Nelson já apontava para a publicação diretamente no sistema hipertextual, sem a

intermediação das editoras convencionais, algo que está sendo perseguido ainda nos dias de hoje, vide as iniciativas em torno dos repositórios institucionais de acesso livre (LEITE, 2009). Apesar do projeto Xanadu nunca ter passado de um protótipo, fato que marcou negativamente a carreira de Ted Nelson, de acordo com Sayão (2008), Nelson modelou muitos dos conceitos fundamentais aos sistemas de hipermídia, incluindo a própria Web, apesar de todo o reconhecimento ser dedicado a Tim Berners-Lee.

De acordo com Sayão (2008), as primeiras aplicações comerciais da computação no contexto das bibliotecas ocorreram no início da década de 1950, quando a IBM desenvolveu o uso de cartões perfurados para dar suporte às operações de processos técnicos da biblioteca. Posteriormente, com o desenvolvimento do *Machine Readable Cataloguing* (MARC), formato legível por computador para representação e intercâmbio de dados bibliográficos, as aplicações da computação no contexto das bibliotecas foram bastante potencializadas. Cabe destacar que tais aplicações eram voltadas exclusivamente para o processamento e intercâmbio dos registros bibliográficos (metadados), e não para a representação digital e manipulação dos conteúdos dos documentos.

De acordo com Sayão (2008) não se pode identificar quando surgiu a primeira biblioteca digital, mas o autor argumenta que o conceito surgiu no início da década de 1980, sendo que a área de estudo de bibliotecas digitais só se configurou como campo de pesquisa a partir de 1990. A biblioteca digital, como é percebida hoje, está muito vinculada ao surgimento da Web, visto que a rede é que definiu as condições tecnológicas e ambientais para a sua concretização enquanto constructo tecnológico e social. Quando se debruça sobre a linha de evolução técnica das bibliotecas digitais fica claro que as suas bases teóricas e práticas são fortemente ligadas à área da computação denominada recuperação da informação. Esta área abrigou pesquisas que antecederam o surgimento da Web, e que foram iniciadas a partir da década de 1950. Sayão (2008) afirma que entre os principais pesquisadores desta área estão Calvin Mooers – que em 1951 cunhou o termo “*information retrieval*” – e Gerald Salton (1927-1995) que é considerado o pai da moderna recuperação da informação, tendo sido um pioneiro no desenvolvimento de técnicas de indexação automática e sistemas de busca cujos conhecimentos as atuais bibliotecas digitais não podem prescindir.

A emergência e o desenvolvimento das bibliotecas digitais foram impulsionados por dois fatores principais: as novas possibilidades abertas pelo rápido desenvolvimento da tecnologia da informação; e o antigo desejo das pessoas, principalmente dos acadêmicos e cientistas, em compartilhar com maior eficiência informações, materiais bibliográficos, bases de dados científicos e resultados de pesquisas (LI, 2009). Sayão (2008) aponta para o fato de, durante a década de 1990, a área de bibliotecas digitais ter se transformado, de um

objeto de preocupação limitado a poucas pessoas da área de Ciência da Computação e de profissionais de Biblioteconomia, em um polo de intensa atração de interesses e de financiamentos, transformando-se numa área altamente institucionalizada.

O grande interesse criado em torno da área gerou uma grande expectativa em torno das potencialidades das bibliotecas digitais não somente como uma visão revigorada da biblioteca convencional, mas como um recurso estratégico dentro de contextos altamente institucionalizados, como governo, cidadania, negócios e pesquisa científica.

De acordo com o exposto acima, é interessante destacar que as origens teóricas e práticas das bibliotecas digitais advêm da computação (recuperação da informação) e não da biblioteconomia. Este fato tem grandes implicações nos sistemas atualmente usados para a implementação de bibliotecas digitais. Em geral estes sistemas não incorporam as técnicas desenvolvidas e aplicadas pelos bibliotecários. Esta constatação tem grande importância na justificativa e direcionamento deste trabalho de pesquisa.

2.1.1 Bibliotecas convencionais e digitais – definições e convergências

Segundo Cunha (2008) a biblioteca convencional ou tradicional é aquela em que a maioria dos itens do seu acervo é constituída de documentos em papel. Ela existe desde a invenção da escrita, tendo sido potencializada pelo advento da imprensa de tipos móveis, em 1440. Cabe destacar que a biblioteca convencional passou por um processo de modernização desde o seu surgimento. Seu catálogo, que inicialmente tinha como suporte o livro, passou para fichas em papel e posteriormente foi informatizado, podendo ser consultado hoje em dia de forma *on-line* para a maioria das bibliotecas.

A biblioteca digital combina a estrutura e a coleta da informação, tradicionalmente usada por bibliotecas e arquivos, com o uso da representação digital tornada possível pela tecnologia da informação (CUNHA, 2008). A informação em formato digital pode ser eficientemente armazenada, preservada e recuperada, permitindo a transposição de barreiras relacionadas a tempo e espaço. Como bem destaca Cunha (2008), a semelhança do que ocorre na biblioteca convencional, a biblioteca digital também deve incluir os princípios consagrados de como a informação é organizada. Neste sentido não se pode considerar como uma biblioteca digital a Internet ou mesmo um *web site* que disponibilize conteúdos digitais selecionados e disponibilizados por indivíduos. O objetivo maior da biblioteca digital é consistente com aquele da biblioteca convencional, isto é, organizar, distribuir e preservar os recursos informacionais. Zafalon (2008) também

compartilha da ideia de que, em sua essência, as bibliotecas convencionais e digitais desempenham o mesmo papel:

... uma coleção de materiais para uso, após serem ordenados e catalogados, em qualquer suporte (as tecnologias de época assim se colocam: desde as tábulas de argila aos registros de computador), com uma política sistemática de formação do acervo (resultando em intenções, conscientes ou inconscientes, particulares ou institucionais) e preocupações voltadas tanto à preservação e ordem quanto ao acesso às suas coleções, para que se possa tirar proveito e oportunidade na condição de usuário, podendo também ter a função de ferramenta social ou de depósito para a memória da humanidade. (ZAFALON, 2008)

Toutain (2006, p.16) nos apresenta a seguinte definição para biblioteca digital:

Biblioteca que tem como base informacional conteúdos em texto completo em formatos digitais – livros, periódicos, teses, imagens, vídeos e outros, que estão armazenados e disponíveis para acesso, segundo processos padronizados, em servidores próprios ou distribuídos e acessados via rede de computadores em outras bibliotecas ou redes de bibliotecas da mesma natureza. (TOUTAIN, 2006, p.16).

Na definição, a autora também destaca a necessidade de “processos padronizados” de aquisição, organização e disponibilização dos conteúdos em uma biblioteca digital. Também aparece na definição a questão da posse do acervo, que no caso das bibliotecas digitais não precisa pertencer à instituição hospedeira.

Fugindo da classificação tradicional das bibliotecas em convencionais e digitais, adotada pela maioria dos autores (CUNHA, 2008; SAYÃO, 2008; TOUTAIN, 2006), que toma como base a ambientação física e as tecnologias adotadas ou seu suporte, Zafalon (2008) propõe o reescalonamento das bibliotecas, de acordo com a forma do acesso, em bibliotecas *in-loco*, em tempo real e híbridas. Tal classificação toma como base a forma do acesso, podendo ser presencial e local (bibliotecas *in-loco*), com acesso remoto *on-line* (biblioteca em tempo real), ou ainda combinações das duas primeiras (bibliotecas híbridas).

Para Sayão (2008), o conceito de uma biblioteca digital meramente equivalente a uma coleção de objetos digitalizados assistida por uma ferramenta de gestão da informação é limitada e inadequada. Neste sentido, o autor sugere a ideia de biblioteca digital como um ambiente distribuído que integra coleções, serviços e pessoas na sustentação do ciclo de vida completo de criação, disseminação, uso, e preservação do conhecimento. Coerente com esta mesma visão, a *Digital Library Federation*² (DLF) fornece uma definição operacional:

² Ver em: <http://www.diglib.org/about/dldefinition.htm>

Bibliotecas digitais são organizações que disponibilizam os recursos, incluindo pessoal especializado, para selecionar, estruturar, oferecer acesso intelectual, interpretar, distribuir, preservar a integridade e assegurar a persistência ao longo do tempo de coleções de trabalhos digitais, de forma que eles estejam prontamente e economicamente disponíveis para uso de uma comunidade definida ou um conjunto de comunidades (tradução nossa).

É interessante notar que apesar desta definição ser amplamente adotada pelas comunidades vinculadas à área de Ciência da Informação, como destaca Sayão (2008), ela revela apenas uma das muitas faces do que é discutido e entendido como biblioteca digital. Harter (1997) já apontava para a impossibilidade de uma definição de consenso por vários motivos, entre eles o fato de o termo “biblioteca digital” ser usado para denotar muitas coisas – de coleções pessoais até a Internet inteira. Na maioria das vezes, essas coisas só têm em comum a manipulação de recursos informacionais em formato digital. Contribui para a dificuldade de definição o fato de um grande número de atores e áreas terem contribuído para o desenvolvimento e implementação de bibliotecas digitais, além do dinamismo próprio do componente tecnológico envolvido. Como aponta Candela (2007a), biblioteca digital é um espaço sinérgico de um grande número de áreas da Tecnologia da Informação e várias outras disciplinas e campos de pesquisa como Biblioteconomia, Ciência da Informação, Museologia, Arquivologia e Gestão do Conhecimento. Cada definição é influenciada pela percepção e ponto de vista particular de pessoas e organizações de diversas áreas que estiveram envolvidas em empreendimentos voltados para a concepção e implementação de bibliotecas digitais. A diversidade de contribuições que tanto serviu para o enriquecimento da área criou, ao mesmo tempo, uma zona obscura de indefinições (SAYÃO, 2008).

Harter (1997) contrapõe duas visões extremadas sobre a natureza das bibliotecas digitais: uma visão abrangente que toma a biblioteca digital como a *Web* é hoje; e uma visão que considera a biblioteca digital como uma extensão, ou uma metáfora, da biblioteca convencional. Esta segunda visão sobre a natureza das bibliotecas digitais nos parece mais próxima da realidade, e é esta a visão adotada neste trabalho de pesquisa. Ela aponta para uma convergência na direção do enquadramento das bibliotecas digitais aos “cânones biblioteconômicos”, principalmente no que concerne à organização e à representação dos recursos informacionais, assim como às relações orgânicas com suas comunidades-alvo (SAYÃO, 2008). Este movimento de convergência se justifica pelo simples fato: uma biblioteca digital continua sendo uma biblioteca.

O progresso tecnológico mudou a maneira como as bibliotecas fazem o seu trabalho, mas não a razão do seu trabalho. Ainda que os desenvolvimentos tecnológicos mais contundentes – como a conexão de um computador a outro numa cadeia continua pelo mundo a fora – possam alterar o conceito fundamental de biblioteca no século 21, podemos supor que a tecnologia

não vai mudar substancialmente o negócio das bibliotecas que é conectar pessoas com informações. (KUNY; CLEVELAND, 1998)

O que se observa é que todos os valores e funções da biblioteca continuam válidos, o que mudam são os objetos físicos e materiais usados para registrar e transmitir o conhecimento e, naturalmente, o instrumental tecnológico para manipulá-los.

Cunha (2008) observa que o armazenamento digital amplia as possibilidades de pontos de acesso a um determinado documento. Os pontos de acesso tradicionalmente usados nas bibliotecas convencionais são título, autor e alguns cabeçalhos de assunto. Nas bibliotecas digitais, as possibilidades são ampliadas; dezenas de termos de indexação podem ser incluídas e a representação do conteúdo pode ser feita em níveis mais elaborados. Estas características podem agregar um alto grau de flexibilidade e qualidade na busca e recuperação da informação. Com relação à unidade representativa, que na biblioteca convencional era, por exemplo, um livro ou uma tese, e não os seus capítulos, no caso da biblioteca digital, passa a poder ter uma maior granularidade, podendo chegar a parágrafos ou assuntos tratados pelo documento. No caso da proposta do Modelo para Organização Hipertextual de Documentos – MHTX (LIMA, 2004), a unidade representativa são os assuntos tratados, abrindo a possibilidade da navegação em contexto no documento.

Vivemos um estado de transição, no qual grandes estoques de materiais impressos, acumulados durante séculos, convivem com o crescente volume de informações que já nascem na forma digital ou que vêm sendo convertidas para este formato. Neste sentido, alguns autores, entre eles Garcez e Rados (2002, p.44), definem o conceito de biblioteca híbrida, que reflete o estado de transição de uma biblioteca que não é completamente convencional, apresentando também características de uma biblioteca digital:

A biblioteca híbrida é designada para agregar diferentes tecnologias, diferentes fontes, refletindo o estado que hoje não é completamente digital nem completamente impresso, utilizando tecnologias disponíveis para unir, em uma só biblioteca, o melhor dos dois mundos (o impresso e o digital) (GARCEZ; RADOS, 2002, p.44).

Pelo exposto acima, podemos concluir que a maioria dos autores, entre eles Cunha (2008), Sayão (2008) e Zafalon (2008), corrobora o entendimento de que a biblioteca digital possui os mesmos objetivos da biblioteca convencional, e que deveria se basear nos mesmos princípios, teorias e técnicas desenvolvidas pelo campo da Biblioteconomia e Ciência da Informação. No entanto, como foi visto na seção anterior, o desenvolvimento das bibliotecas digitais teve desde suas origens uma forte influência da Ciência da Computação, e mais especificamente da subárea conhecida como Recuperação da Informação. Como

nos sinaliza Hjørland (2000), o que se observa é que as novas tecnologias geralmente vêm ignorando o conhecimento existente. Além disso, a tendência observada é que a Ciência da Informação tem usado a tecnologia de forma passiva, sem contribuir para o seu desenvolvimento. A tecnologia traz vantagens indiscutíveis, principalmente aquelas relacionadas ao armazenamento e acesso, reduzindo custos e derrubando barreiras de tempo e espaço. No entanto, estas vantagens deveriam ser aproveitadas sem relegar os conhecimentos acumulados sobre como organizar acervos e como tratar informações.

2.1.2 Perspectivas para as bibliotecas

Como observa Cunha (2008), a percepção pública das bibliotecas continua muito vinculada ao livro impresso, e a partir do momento em que as pessoas refletem sobre o papel das bibliotecas frente ao avanço tecnológico, é comum a percepção de que o livro impresso irá desaparecer, e com ele a biblioteca. Neste aspecto a biblioteca digital se mostra como o caminho mais natural para evolução e perpetuação da biblioteca enquanto instituição. Segundo o mesmo autor, este é o momento de avançar para além dos aspectos relativos à aquisição, processamento e guarda dos materiais, integrando fontes e materiais digitais nos acervos e serviços. O foco da biblioteca digital é maior no acesso e menor na coleção. Considerando que o material digital pode estar disponível em *web sites* espalhados pela rede, os bibliotecários podem utilizar os seus conhecimentos técnicos no sentido de possibilitar acesso a estes materiais. Neste contexto, os leitores potenciais poderão se beneficiar dessas habilidades técnicas dos bibliotecários, solicitando a assistência destes na busca de informações dispersas na Internet, e na tarefa de fazer conexões entre os recursos informacionais dispersos em diversos locais e mídias. Estes tipos de serviços farão com que as pessoas entendam que a biblioteca abriga todas as mídias informacionais. As bibliotecas digitais podem incorporar este tipo de serviço, aproveitando a tecnologia e o ambiente virtual, para permitir a interação entre bibliotecários e usuários, para que os primeiros possam prestar auxílio técnico nas pesquisas.

Lesk (2005) alerta para a necessidade de incorporação das novas tecnologias pelas bibliotecas convencionais. Se estas bibliotecas ficarem “sentadas esperando” elas fatalmente deixarão escapar a chance de desempenhar o papel de um importante fornecedor de informação digital. Para que as bibliotecas mantenham seu *status* e relevância, elas devem participar do “mundo digital”, como muitas têm feito. Existem muitos

objetivos sociais que são importantes para as bibliotecas como instituições, além da mera capacidade de “empilhar livros ou discos” (LESK, 2005).

O tamanho do acervo mantido por uma biblioteca digital não importa muito. O mais importante é disponibilizar o acesso, independentemente de onde ou em qual acervo se encontra o documento. Trata-se da mudança de paradigma de posse do documento para o de provisão de acesso (CUNHA, 2008). Neste novo contexto, o importante passa a ser a disponibilização do acesso e, com frequência, a confiabilidade da informação disponibilizada. Dada a facilidade de manipulação da informação digital, torna-se especialmente importante saber com segurança quem a produziu, quem a identificou como valiosa, quem a selecionou para disponibilizar e quem garante sua autenticidade.

Com relação ao relacionamento entre a biblioteca convencional e a sua contraparte digital, Cunha (2008) aponta que há indícios de que as bibliotecas continuarão com a custódia dos materiais impressos, com destaque para os livros. Contudo, elas se tornarão também gerenciadoras de linhas de comunicação com outros locais de conhecimento (bibliotecas digitais) e serão responsáveis pelo controle da qualidade das bibliotecas digitais, decidindo quais conteúdos digitais existentes em outras instituições merecem menção pelos selecionadores e hiperorganizadores da biblioteca. Estas ideias se alinham com o conceito de bibliotecas híbridas GARCEZ e RADOS (2002, p.44), apresentado anteriormente. Como afirmou Tennant (2007, p.29),

“[...] sem uma real biblioteca, a digital não será mais que uma pilha de bits. (...) Aqueles que acreditam que uma pilha de bits é suficiente para uma biblioteca robusta nunca foram servidos por um bibliotecário – tanto física como virtualmente. (...) As bibliotecas digitais oferecem muitas coisas, mas sem os serviços providos por uma biblioteca de tijolos e cimento e os bibliotecários que fazem com que elas fiquem abertas, elas não serão mais que uma pilha de bits”.

Lesk (2005) também alerta para o risco do excessivo e exclusivo foco na tecnologia. Neste sentido o autor aponta para experimentos com usuários que mostram que o quanto mais detalhadamente um usuário explica a um bibliotecário sua necessidade de informação, maior a probabilidade do sucesso da busca. Construir uma biblioteca digital não é apenas uma questão de “empilhar” conteúdos digitais; envolve a criação de uma organização de máquinas e pessoas, talvez até mesmo uma cultura, na qual seja possível encontrar informação e usá-la. Não é factível imaginar um mundo no qual a informação é distribuída com a ausência de instituições. O mesmo autor nos aponta diversas vantagens trazidas pelas bibliotecas digitais. Uma biblioteca digital pode ser facilmente pesquisada a partir de múltiplos critérios ou pontos de acesso, incluindo palavras e frases; seus conteúdos podem ser copiados sem o risco de erros. Elas resolvem diversos problemas relacionados à busca, distribuição e preservação da informação. Informação digital ocupa menos espaço e pode

ajudar as bibliotecas a reduzir seus custos. As bibliotecas digitais podem fornecer um nível de serviços nunca antes alcançado – distribuição de informação na mesa do usuário de forma *on-line*, capacidade de busca flexível e informação que não se deteriora fisicamente ao longo do tempo, esteja ela na forma de textos, sons ou imagens. Mas, talvez, a maior vantagem das bibliotecas digitais seja o seu potencial de ajudar a sociedade a tornar a informação mais disponível, melhorando a sua qualidade e aumentando a sua diversidade.

Para Sayão (2008), as bibliotecas digitais caminham rapidamente para se tornar um ponto concentrador de tecnologias e metodologias voltadas para o apoio à pesquisa e a comunicação científica e às novas modalidades de ensino e disseminação de informações para o cidadão comum. Neste sentido, as bibliotecas digitais estariam destinadas a se tornarem uma parte essencial da infraestrutura de informação do século XXI.

O ambiente de serviços de uma biblioteca digital é um espaço de informações em formato digital que suporta diferentes visões e uma gama de usos possíveis. Neste sentido Sayão (2008) apresenta uma breve análise sobre as oportunidades e vantagens vislumbradas para as bibliotecas digitais sob a ótica de diferentes atores. Os profissionais da informação têm uma perspectiva de continuidade evolutiva em relação às bibliotecas digitais. Para eles, as bibliotecas digitais são uma extensão lógica, e com vantagens óbvias, do que as bibliotecas vêm fazendo desde os tempos remotos: adquirindo, organizando e disseminando conhecimento, usando as tecnologias de cada época. O acesso simultâneo a um mesmo documento digital por um número ilimitado de usuários significa o fim da lista de empréstimos. Já os profissionais de Ciência da Computação enxergam as bibliotecas digitais como uma aplicação de computadores em rede com o objetivo de oferecer facilidades informacionais. Tal aplicação envolve também conhecimentos da área de recuperação da informação, banco de dados e multimídia. Os políticos e governantes percebem a biblioteca digital como uma infraestrutura tecnológica que pode ser usada para a superação da desigualdade informacional e de acesso; como um insumo básico para a pesquisa, o ensino superior e a pós-graduação; e como um instrumento para a maior visibilidade de bens e instituições culturais. A percepção da indústria editorial em relação às bibliotecas digitais é ambivalente: por um lado existe a ameaça representada pelas novas formas de autopublicação, a facilidade de cópia, e o movimento crescente em torno do acesso livre. Por outro lado, existe uma visão otimista em que a biblioteca digital se constitui em novo modo de distribuição de conteúdos e um novo mercado a ser conquistado, num contexto de mudança da economia da informação. Para os professores e educadores, que sempre tiveram uma relação de colaboração com a biblioteca convencional, as bibliotecas digitais abrem possibilidades extraordinárias para a educação e o ensino, principalmente em sua modalidade não presencial, mudando paradigmas e estabelecendo novas metodologias

pedagógicas. Para os arquivistas, as bibliotecas digitais acabam com a relação quase antagônica entre a preservação e o acesso, na medida em que a digitalização tanto permite preservar os conteúdos raros, únicos ou frágeis, como ao mesmo tempo proporciona acesso universal às suas representações digitais. Para os pesquisadores, a colaboração é componente fundamental para a pesquisa e o desenvolvimento. Neste sentido eles percebem a biblioteca digital como um espaço que potencializa a geração, o compartilhamento e a disseminação do conhecimento. Finalmente, no campo cultural, o que se observa é que a biblioteca digital é um meio privilegiado de dar visibilidade global a manifestações culturais antes circunscritas às suas comunidades (SAYÃO, 2008).

2.1.3 A pesquisa na área de bibliotecas digitais

A biblioteca digital faz uso de uma rede de computadores e tecnologias relacionadas para prover serviços que estabelecem uma estrutura de intermediação entre recursos informacionais heterogêneos e distribuídos e a comunidade de usuários, que de acordo com Sayão (2008) se mostra um componente bastante amplo, diversificado e mutante. Para cumprir esta tarefa e atender às expectativas dos diversos atores envolvidos, conforme apontado na seção anterior, as pesquisas em torno das bibliotecas digitais precisam superar um conjunto de desafios que se estendem por várias áreas do conhecimento.

Bibliotecas digitais são construídas sob uma infraestrutura tecnológica na qual os artefatos de software são os principais componentes. Neste sentido, se faz necessário o desenvolvimento de uma arquitetura que proporcione uma infraestrutura tecnológica comum, mas que possa ser customizada segundo as necessidades de diferentes atores e aplicações (SAYÃO, 2008). Uma arquitetura de um sistema de software é a estrutura ou estruturas do sistema, que contempla elementos de software, as propriedades visíveis externamente desses elementos e o relacionamento entre eles (ISO/IEC 42010, 2007). Esta infraestrutura tem que incorporar o estado da arte em termos tecnológicos e também permitir a incorporação de novos modelos e técnicas que surgem com a pesquisa e desenvolvimento.

Com relação ao desenvolvimento das coleções, os processos tradicionais desempenhados pelas bibliotecas convencionais precisam ser revisados para acomodar as diferenças determinadas pela natureza digital dos recursos informacionais. Como sugere Sayão (2008), no contexto das bibliotecas digitais, além das políticas e estratégias de seleção e aquisição, devem também ser considerados os problemas inerentes à condição

digital da informação, como a conversão do material impresso para digital, a geração de material unicamente digital, as barreiras tecnológicas que podem impedir o acesso e a usabilidade dos objetos, a sustentabilidade das coleções digitais, as questões do direito autoral, a preservação digital, entre outros. Segundo Brown (2005), um dos maiores desafios que se apresenta na formação das coleções digitais é a integração dos diversos tipos e formatos de objetos digitais com os materiais tradicionais, de modo a oferecer uma visão única e abrangente de todo o acervo. Trata-se do conceito de “coerência digital” (BROWN, 2005), segundo o qual todos os objetos em uma biblioteca digital, seja qual for o formato (sons, imagens, texto, vídeo, etc.) devem ser tratados essencialmente da mesma forma. Como aponta Sayão (2008), esta forma de tratamento uniforme é diferente do que é normalmente praticado pelas bibliotecas convencionais, em que cada mídia recebe um tratamento diferente.

Os metadados representam um conceito familiar para as bibliotecas convencionais, na medida em que sempre foram usados para descrever os documentos nos processos de catalogação e indexação. Para Vellucci (1998), no contexto das bibliotecas digitais os metadados podem ser vistos como dados que descrevem atributos de um documento digital, caracterizam suas relações, possibilitam a sua recuperação, seu uso efetivo e sua existência no ambiente eletrônico. Metadados normalmente consistem em um conjunto de elementos de dados onde cada elemento descreve um atributo do recurso informacional, sua estrutura, sua administração ou uso (VELLUCCI, 1998). Como aponta Cleveland (1998), os metadados são importantes em bibliotecas digitais porque representam a chave para a descoberta de recursos e usos para qualquer documento. Marcondes (2006, p.96) aponta para o fato da literatura brasileira ainda ser bastante escassa sobre o assunto, sendo que as experiências práticas mais importantes foram o desenvolvimento do Padrão Brasileiro de Metadados de Teses e Dissertações Digitais (MTD-BR) no contexto da Biblioteca Brasileira de Teses de Dissertações³ (BDTD), operada pelo IBICT, e a implantação do *Open Archives Protocol for Metadata Harvesting* (OAI-PMH) no contexto do projeto Scielo⁴.

Os metadados representam uma questão crucial no desenvolvimento de bibliotecas digitais. As coleções digitais demandam esquemas de metadados mais elaborados que sejam capazes de descrever os objetos digitais e seus conteúdos em diversos níveis de granularidade, de uma coleção como um todo, a um livro ou até mesmo uma ilustração ou um conceito contido no mesmo. Como aponta Sayão (2008), um dos

³ Ver em: <http://bdttd.ibict.br/bdttd>

⁴ Ver em: <http://www.scielo.org>

maiores desafios em relação aos metadados no contexto das bibliotecas digitais é a diversidade de formatos da informação digital, e a maneira como esta informação deve ser descrita no contexto de diferentes coleções dirigidas a diferentes comunidades. Quando se vislumbra a interoperabilidade entre as bibliotecas digitais, surgem então as importantes questões relacionadas à necessidade do estabelecimento de padrões de metadados, além do mapeamento entre diferentes esquemas de metadados (SHIRI, 2003). Em suma, podemos dizer que a principal finalidade dos metadados é documentar, com elementos descritores, recursos informacionais, de modo a permitir a descoberta e recuperação desses recursos, além da comunicabilidade e interoperabilidade entre sistemas.

Por interoperabilidade entende-se a capacidade de compatibilidade operacional individual ou em conjunto, ou ainda, a habilidade de um sistema se comunicar ou trabalhar com outro sistema. No contexto das bibliotecas digitais, interoperabilidade pode ser definida como “a capacidade de componentes ou serviços de bibliotecas digitais serem funcionalmente e logicamente intercambiáveis em virtude deles terem sido implementados de acordo com um conjunto de interfaces bem definidas e publicamente conhecidas” (PAYETTE *et al.*, 1999, p.2). A interoperabilidade constitui um conceito parceiro de integração entre bibliotecas digitais. Um de seus principais benefícios é a possibilidade de um usuário realizar buscas de recursos informacionais heterogêneos, armazenados em diferentes servidores na rede, utilizando-se de uma interface única sem tomar conhecimento de onde nem de como estes recursos estão armazenados. Em suma, interoperabilidade é a capacidade das bibliotecas digitais trocarem e compartilharem documentos, consultas e serviços. Nos últimos anos, a interoperabilidade tem sido um dos itens mais críticos para quem pensa no desenvolvimento e operação de sistemas de repositórios e de bibliotecas digitais distribuídos e funcionando em rede. No entanto, como bem nos lembram Sayão & Marcondes (2008), o conceito de interoperabilidade está longe de ser uma novidade no domínio das bibliotecas:

Há muito tempo, desde meados do século XX, para fazer frente ao fenômeno social da ‘explosão informacional’, as bibliotecas sempre estabeleceram serviços cooperativos, trocaram informações, criaram um ordenamento universal dessas informações. Toda uma estrutura global foi montada em torno da ideia do compartilhamento e da cooperação entre bibliotecas. (SAYÃO & MARCONDES, 2008, p. 134)

A importância da interoperabilidade tem sido amplamente reconhecida e sua busca vem sendo pesquisada e aplicada nos projetos de bibliotecas e repositórios digitais distribuídos, a exemplo das iniciativas do IBICT (KURAMOTO, 2007), inicialmente em torno da Biblioteca Brasileira de Teses e Dissertações e mais recentemente com os Repositórios

Institucionais (LEITE, 2009), ambos baseados nos conceitos de padrões abertos e de interoperabilidade.

Outra importante linha de pesquisa relacionada à área de bibliotecas digitais, que também é apontada por Sayão (2008), é o desenvolvimento de interfaces e a usabilidade. Usabilidade é um conceito e uma disciplina que busca compreender os aspectos da interação entre pessoas e sistemas computacionais, podendo ser entendida como “a capacidade de um produto ser usado por usuários específicos para atingir objetivos específicos com eficácia, eficiência e satisfação em um contexto específico de uso” (*INTERNATIONAL ORGANIZATION FOR STANDARTIZATION -ISO, 1998*). A usabilidade define a qualidade da interação entre o usuário e o sistema. Sua importância se justifica pelo fato de as bibliotecas digitais se dirigirem para diferentes contextos e audiências, sendo necessário que elas possam ser configuradas de acordo com o perfil de seus usuários. A preocupação com os aspectos relacionados à interface e usabilidade também se justificam pelo fato de as bibliotecas digitais não possuírem o componente humano de interação, representado pelo papel do bibliotecário de referência no contexto das bibliotecas convencionais. Desta forma, torna-se importante desenvolver mecanismos que facilitem e favoreçam o uso destes sistemas, de modo a cumprir seu papel de intermediação entre pessoas e recursos informacionais.

A descoberta de recursos informacionais e sua recuperação, apesar de ser extremamente dependente das etapas anteriores de tratamento, é talvez o aspecto mais visível e importante em uma biblioteca. Como aponta Sayão (2008), os serviços de indexação e busca genéricos, representados pelo seu maior expoente, o Google, oferecem ferramentas básicas que ajudam o usuário a achar a informação que procura. Entretanto, esses serviços não têm o nível de especificidade, de desempenho e, sobretudo, de tratamento biblioteconômico exigido para a maioria dos empreendimentos (SAYÃO, 2008). As bibliotecas digitais não podem prescindir de metodologias de organização do conhecimento e das estruturas semânticas. Sob a perspectiva da interoperabilidade, um dos desafios importantes é o mapeamento e a interoperabilidade entre sistemas de organização de conhecimento distintos (SAYÃO & MARCONDES, 2008).

A questão dos direitos autorais é um dos problemas mais difíceis e desafiadores para a área de bibliotecas digitais. As leis de direitos autorais (*copyright*) foram criadas com a finalidade de se criar um equilíbrio entre os interesses do autor e a necessidade social de facilitar o livre fluxo de informação, salvaguardando os interesses privado e público. O problema se apresenta na medida em que, no ambiente digital, se perde o controle de cópias, na medida em que os objetos são fluidos, facilmente copiados e acessíveis remota e simultaneamente por um número ilimitado de usuários. Como sugere Sayão (2008), as

bibliotecas são, na maioria dos casos, simplesmente custodiantes da informação e não detêm os direitos autorais sobre o material que está sob o seu controle. Desta forma, torna-se necessário o desenvolvimento de procedimentos e mecanismos de gestão de direitos autorais, para gerenciar os direitos de modo a poder disponibilizar informação sem violar as regras. Uma alternativa que se apresenta, e que vem recebendo grande atenção nos últimos anos, é via do acesso aberto à literatura científica (LAGOZE, C.; VAN de SOMPEL, 2000). A Iniciativa dos Arquivos Abertos (*Open Archives Initiative – OAI*) tem como objetivo promover e incentivar o desenvolvimento de mecanismos de autopublicação pelos autores (também conhecido como sistemas de *e-prints*) por meio do desenvolvimento de soluções técnicas e estruturas organizacionais que venham permitir a interoperabilidade entre os arquivos de *e-prints*. Pretende-se que esta interoperabilidade possa estimular a transição dos sistemas de *e-prints* em um novo modelo de comunicação científica baseado no acesso livre.

A preservação é outra questão que se manifesta de forma distinta quando se compara as bibliotecas convencionais com a sua contraparte digital. A informação digital depende de um aparato tecnológico para ser acessada e corretamente interpretada. Este aparato tecnológico se constitui de *hardware*, *software*, mídias e formatos. Como observa Sayão (2008), este aparato tecnológico está em constante evolução e mutação, em ciclos de obsolescência cada vez mais rápidos, determinados principalmente pelo dueto inovação e competição. Neste sentido, são necessárias ações que garantam que os recursos de informação digital produzidos hoje possam ser acessados e corretamente interpretados no futuro. A preservação digital se constitui em um processo permanente, que deve considerar os aspectos e procedimentos técnicos, como migração e emulação de conteúdos. Por outro lado, a preservação também deve ser pensada como um desafio gerencial e organizacional, como apontado por Sayão (2006). Grande parte da informação que é produzida diariamente já nasce no formato digital, e as que foram produzidas em suportes físicos estão sendo convertidas para o formato digital. O problema surge na medida em que esta tendência que parece ser irreversível não vem sendo acompanhada por estratégias para garantir o acesso a longo prazo às informações digitais. Como aponta Sayão (2006, p.114), “a nossa compulsão em produzir informações digitais é infinitamente superior à nossa capacidade de preservar o acesso a elas”.

Na medida em que as bibliotecas digitais passam a ter seu uso mais difundido e passam a atender a um número cada vez maior de usuários, estes se tornam cada vez mais experientes e passam a demandar serviços mais sofisticados e adequados às suas necessidades e habilidades específicas. Smeaton & Callan (2005) sugerem a evolução das bibliotecas digitais no sentido de uma maior pro atividade e capacidade de adaptação aos

usuários individuais. Neste sentido, os autores definem personalização como uma forma de adaptação dos conteúdos e serviços para atender às necessidades de usuários e comunidades específicas. Os autores identificam vários campos de pesquisa que contribuem para este objetivo, entre eles recuperação da informação, interação homem-computador, ambientes colaborativos, modelagem do usuário e hipermídia. A modelagem dos interesses, preferências, conhecimento, e vínculos sociais do usuário requer um gerenciamento mais eficiente do perfil de cada usuário no contexto de uma biblioteca digital. A pesquisa relacionada à personalização em bibliotecas digitais deve ter como foco tanto o aspecto social quanto tecnológico (SMEATON & CALLAN, 2005).

Kruk e outros (2005) defendem que as bibliotecas digitais devem incorporar os desenvolvimentos relacionados aos avanços da Internet, principalmente aqueles relacionados à *web* semântica e redes sociais. O objetivo seria, entre outros, melhorar a usabilidade e a descoberta da informação. Eles argumentam que a Internet evoluiu de uma fonte de informação para um meio de comunicação e interação social. Neste sentido, as bibliotecas digitais devem fazer uso destes avanços, de modo a atender às expectativas da nova geração de usuários da Internet, que usa a rede tanto como uma fonte de informação, quanto uma plataforma de comunicação.

Além dos principais tópicos que foram brevemente mencionados acima, a agenda de pesquisa na área de bibliotecas digitais apresenta várias outras questões também importantes para o pleno desenvolvimento da área. Como observa Sayão (2008), uma parte considerável dessas questões extrapola o plano tecnológico, cobrindo desde modelos econômicos de sustentabilidade, até a administração e gestão de bibliotecas digitais.

As bibliotecas digitais podem ser vistas sob diferentes focos e perspectivas, que abrangem desde a biblioteconomia e ciência da informação até áreas como a recuperação da informação, interação homem computador e usabilidade. Devido a sua natureza interdisciplinar, as bibliotecas digitais devem ser vistas como uma combinação de documentos, tecnologia e trabalho (LEVY; MARSHALL, 1995), e não como simples coleções de recursos de informação, ou tecnologias que viabilizam estas coleções.

No contexto desta pesquisa, o foco será a representação e organização do conhecimento na biblioteca digital, visando o desenvolvimento de mecanismos de recuperação e acesso, que apoiados na organização do acervo, possam trazer benefícios para os usuários, melhorando a descoberta de informação e a transferência do conhecimento nestes sistemas, assim como aumentando a satisfação do usuário da biblioteca digital.

2.1.4 Iniciativas brasileiras relacionadas a bibliotecas digitais

Nesta seção serão brevemente apresentadas as principais iniciativas brasileiras relacionadas à implantação de bibliotecas digitais. Tendo em vista que as principais iniciativas se apoiam em um modelo alternativo de comunicação científica, baseado no movimento que busca o acesso livre ao conhecimento científico, será inicialmente feita uma exposição sobre este modelo, incluindo seus pré-supostos, metodologias e tecnologias de suporte.

A iniciativa dos arquivos abertos (*Open Archives Initiative* - OAI) (LAGOZE, C. & VAN de SOMPEL, 2000) tem como bases o acesso livre, a autopublicação de conteúdos, e o intercâmbio de metadados. Constitui um marco histórico do desenvolvimento da comunicação científica, da publicação eletrônica e das bibliotecas digitais (TOUTAIN, 2006, p.20). Esta iniciativa representa o anseio da comunidade científica em formar um fórum aberto para aprimorar o desempenho do atual modelo de comunicação científica. Pretende-se, assim, acelerar o processo de divulgação dos resultados das pesquisas (CUNHA; MCCARTHY, 2006, p.34). Kuramoto (2006, p.148) sintetiza os principais pontos contemplados pela iniciativa dos arquivos abertos: uso de software livre, autoarquivamento ou autopublicação, criação de repositórios de livre acesso e uso de padrões de preservação de objetos digitais. O uso de software livre, também chamado *open source*, se contrapõem ao modelo proprietário e pago na medida em que oferece o uso gratuito e por vezes a possibilidade de alteração e adaptação. O autoarquivamento ou autopublicação abre a possibilidade de os próprios autores submeterem e arquivarem seus trabalhos nos repositórios, que são bases de dados de trabalhos científicos, com texto integral e metadados, podendo ser temáticos ou institucionais. Brody e Harnad (2004) chamam a atenção para o fato de que o acesso livre a resultados de pesquisa, possibilitado pela iniciativa de arquivos abertos, maximiza o acesso à pesquisa, acelera o seu impacto e, conseqüentemente, aumenta a produtividade, o progresso e as recompensas das pesquisas. O acesso tende a ser a questão crucial para o progresso científico em qualquer área do conhecimento.

Sob o ponto de vista organizacional, a iniciativa dos arquivos abertos estabelece dois papéis distintos a serem desempenhados pelas instituições: o papel de provedor de dados (*data provider*) e o de provedor de serviços (*service provider*). Um provedor de dados é responsável pela gestão de um repositório ao qual são submetidos, diretamente pelos autores, trabalhos técnico-científicos. O provedor de dados se responsabiliza pelo armazenamento e preservação, em longo prazo, dos objetos digitais sob sua guarda, além

de tornar disponíveis, para coleta, os metadados relativos aos trabalhos depositados no repositório. Já o provedor de serviços, também chamado de agregador, se preocupa em coletar os metadados armazenados em diversos provedores de dados, e oferecer uma interface ou ponto de acesso para consulta integrada em uma base de metadados coletados a partir de vários provedores de dados. Desta forma, o provedor de serviços tem como papel integrar vários provedores de dados, além de oferecer o serviço de consulta integrada em todo o acervo. A fim de oferecer suporte técnico a esta forma de organização, foi definido o Dublin Core⁵ como padrão de metadados e foi desenvolvido o protocolo OAI-PMH (*Open Archive Initiative – Protocol of Metadata Harvesting*) como mecanismo a ser usado para interoperar os vários provedores de dados e de serviços. A iniciativa dos arquivos abertos permitiu o surgimento de vários pacotes de software para a implantação de bibliotecas digitais, alinhados aos seus ideais e padrões.

Dentro da mesma linha da iniciativa dos arquivos abertos, o Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT desenvolveu e implantou no Brasil o projeto da Biblioteca Brasileira de Teses e Dissertações (BDTD). A BDTD é uma biblioteca digital de teses e dissertações. Teses e/ou dissertações podem ser definidas como documentos manuscritos, desenvolvidos para preenchimento parcial dos requisitos para a obtenção de um grau acadêmico. Portanto, são primariamente documentos manuscritos que inseridos em uma biblioteca digital terão sua descrição bibliográfica adaptada às normas de descrição para recursos eletrônicos de acesso remoto. Quando um manuscrito se torna um recurso eletrônico, ele passa a necessitar de alguns elementos descritivos além dos previstos para a descrição de manuscritos, que permitam identificar suas novas características enquanto objeto digital (LOURENÇO, 2005).

A implantação da BDTD teve início em 1995, quando o IBICT integrou as bases de teses e dissertações de 17 universidades brasileiras. Em dezembro de 2002, o IBICT disponibilizou o site da BDTD⁶. O IBICT viabilizou ainda a criação de um consórcio de instituições que passaram a publicar e disponibilizar teses e dissertações em formato digital. O projeto parte da premissa de que as instituições provedoras desse tipo de informação, em grande número, podem trabalhar de forma conjunta, proporcionando a multiplicação de pontos de acesso para o usuário, com significativo aproveitamento de recursos materiais e humanos (CUNHA; MCCARTHY, 2006, p.33). Outra consequência positiva foi a incorporação do acervo de teses e dissertações produzidas pelas instituições brasileiras ao

⁵ Ver em: <http://dublincore.org>

⁶ Ver em: <http://bdtd.ibict.br/bdtd>

sistema internacional *Networked Digital Library of Thesis and Dissertations*⁷. Com a finalidade de acelerar a criação de bibliotecas digitais nas instituições brasileiras, o IBICT desenvolveu o Sistema de Publicação Eletrônica de Teses e Dissertações (TEDE). Trata-se de um pacote de programas (*software*) que permitiu a rápida implantação destas bibliotecas dentro das universidades, que passaram a atuar como coletoras e provedoras de dados. O IBICT recebe os dados coletados de cada instituição e faz a sua agregação em nível nacional, fornecendo o serviço de acesso integrado. Outro fator que contribuiu para o sucesso do projeto foi a aprovação, pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), em 2003, de uma recomendação para que o depósito das teses e dissertações fosse obrigatório no âmbito das instituições de ensino superior. Em janeiro de 2013, a BDTD contava com 97 instituições participantes, que mantinham em seus repositórios cerca de 210 mil documentos, entre teses e dissertações. O projeto da BDTD demonstra o valor de iniciativas abrangentes e coordenadas para a implantação de bibliotecas digitais, com foco na interoperabilidade e no consequente acesso uniforme a um amplo conjunto de coleções.

Outra iniciativa na área de bibliotecas digitais, bastante conhecida no Brasil e com abrangência internacional, é o Scielo⁸ (*Scientific Electronic Library On-Line*). Trata-se de um projeto da Bireme (Biblioteca Regional de Medicina), instituição que a partir de 1967 teve sua abrangência geográfica ampliada, passando a funcionar como um centro internacional para o continente e a ter como denominação Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde⁹, mantendo-se a sigla anterior. O projeto Scielo teve início em 1997, contando com o apoio da Fundação de Amparo a Pesquisa de São Paulo (Fapesp). Seu objetivo era desenvolver uma metodologia de digitalização, armazenamento, disseminação e avaliação da literatura científica em formato eletrônico, por meio de uma biblioteca digital de periódicos eletrônicos com texto completo. A coleção que no início contava com periódicos nas áreas da saúde, passou ao longo do tempo a incorporar títulos em outras áreas, mais especificamente nas áreas das ciências sociais e humanidades. A coleção contava, em janeiro de 2013, com 1.008 títulos, sendo 259 títulos brasileiros. Entre as características mais importantes do Scielo estão o acesso aberto e gratuito, a existência de ligações com seus artigos a partir do Web of Science¹⁰, e ligação

⁷ Ver em: <http://www.ndltd.org>

⁸ Ver em: <http://www.scielo.org>

⁹ Ver em: <http://www.bireme.br>

¹⁰ Ver em: <http://www.isiknowledge.com>

do nome do autor do artigo com seu *currículum vitae*, armazenado na Plataforma Lattes¹¹. O Scielo se apresenta como uma alternativa viável para o aumento da visibilidade e do fator de impacto de periódicos de países em desenvolvimento na comunidade internacional (CUNHA; MCCARTHY, 2006, p.36).

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), órgão do Ministério da Educação responsável pela avaliação dos programas de pós-graduação no país, iniciou no ano de 2000 um projeto na área de informação digital com o objetivo de oferecer acesso a diversos sistemas europeus e norte-americanos de periódicos eletrônicos (CUNHA; MCCARTHY, 2006, p.37). Este importante projeto teve como denominação Portal de Periódicos da Capes¹² e permite o acesso de estudantes universitários, professores e pesquisadores de instituições Brasileiras, entre elas universidades, centros de pesquisa e centros federais de ensino técnico. Em 2009, o Portal da Capes foi reestruturado de modo a permitir consulta unificada em todo o acervo, entre outras melhorias. Em janeiro de 2013, o Portal da Capes provia acesso a mais de 31 mil publicações periódicas internacionais e nacionais. O Portal da Capes atua como um fornecedor e agregador de acesso a diversas e importantes bases de dados de periódicos científicos nacionais e internacionais, tendo se tornado uma importante ferramenta para o desenvolvimento da pesquisa no país.

2.1.5 Modelos teóricos para bibliotecas digitais

Modelos são representações abstratas voltadas para a análise e o estudo. Conforme apresentado nas seções anteriores, o universo das bibliotecas digitais se mostra bastante complexo, multifacetado e dinâmico. O desenvolvimento e evolução desta área em termos de abordagens, soluções e sistemas passa pela necessidade de uma base conceitual comum capaz de estabelecer um alicerce para um melhor entendimento, comunicação e estímulo ao desenvolvimento da área. Esta base comum pode ser criada na forma de modelos que atuam como uma espécie de mapas, que guiam os diversos atores envolvidos no domínio das bibliotecas digitais, fazendo com que sigam a mesma rota e compartilhem um entendimento comum ao lidar com as entidades e objetos de pesquisa na área.

Frew e outros (1998) introduziram um modelo genérico para bibliotecas digitais quando apresentaram o projeto da Biblioteca Digital Alexandria. O modelo define três

¹¹ Ver em: <http://lattes.cnpq.br>

¹² Ver em: <http://www.periodicos.capes.gov.br>

componentes para uma biblioteca digital: *neighborhood*, *middleware* e *content*. O primeiro representa o “mundo de fora” ou a vizinhança da biblioteca digital, onde se encontram usuários, agentes e bibliotecários. O segundo componente representa o núcleo do sistema gerenciador da biblioteca digital. O terceiro componente representa os conteúdos da biblioteca digital, incluindo os recursos informacionais (acervo) e o catálogo (metadados).

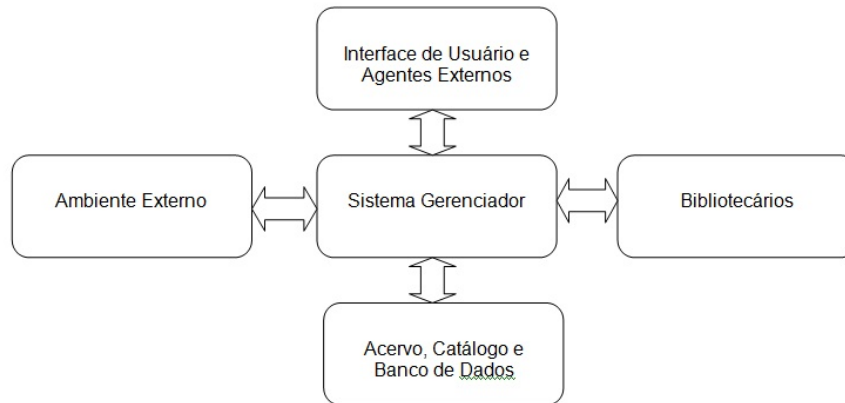


FIGURA 2 - Modelo de arquitetura da Biblioteca Digital Alexandria

Fonte: (FREW *et al.*, 1998) (tradução nossa).

Borbinha e outros (2005) se preocuparam com a representação dos atores e seus papéis no contexto das bibliotecas digitais. Segundo os autores, a abordagem centrada no sistema não considera os canais de comunicação e interação que ligam diferentes classes de atores no contexto de uma biblioteca digital. Neste sentido, eles apresentaram uma abordagem genérica para modelagem dos atores e de seus respectivos papéis. Um ator no contexto de uma biblioteca digital pode ser uma pessoa, organização, ou autômato que desempenha determinado papel na produção, organização, gerenciamento, ou uso da informação digital. Os autores propõem um modelo genérico que consiste de três classes principais de atores que aparecem no cenário das bibliotecas digitais: *Profissionais*, *Usuários* e *Agentes*. Desta forma, a biblioteca digital é vista como um sistema de informação que oferece um conjunto de serviços, gerenciado por *Profissionais*, concebidos para o benefício dos *Usuários* e possivelmente relacionados a fornecedores externos de informação e serviços, referenciados como *Agentes*. No modelo proposto pelos autores, cada classe de ator representa um papel específico. Eles interagem com os serviços da biblioteca digital, mas também uns com os outros. Por exemplo, um *Usuário* é um tipo de ator que interage com a biblioteca digital para consular e acessar informação disponibilizada pela mesma. Exemplos de usuários incluem estudantes ou pesquisadores em uma

biblioteca universitária, cidadãos em uma biblioteca pública, funcionários em uma biblioteca corporativa, etc. No entanto, o papel dos *Usuários* não é necessariamente restrito ao consumo de conteúdos. Os *Usuários* também podem fornecer informações úteis de modo a enriquecer os serviços (por exemplo, fornecendo sugestões ou solicitando novas funcionalidades), ou ainda podem fornecer informações na forma de anotações sobre os documentos recuperados. Em cenários mais sofisticados, os *Usuários* podem interagir uns com os outros (através de serviços específicos da biblioteca) de modo a poderem compartilhar documentos recuperados. Os *Usuários* podem ainda interagir com *Agentes* como, por exemplo, fornecendo comentários, revisões e opiniões aos autores acerca dos documentos da biblioteca. Um cenário para bibliotecas digitais colaborativas, identificando atores e papéis, é apresentado por Candela e outros (2007b). Borbinha e outros (2005) enfatizam a necessidade de uma classificação dos atores e de seus respectivos papéis no contexto de uma biblioteca digital e sugerem o desenvolvimento de uma teoria formal acerca da interação em bibliotecas digitais. A FIGURA 3, a seguir, apresenta o caso de uso genérico de uma biblioteca digital numa perspectiva que enfatiza a interação entre os atores.

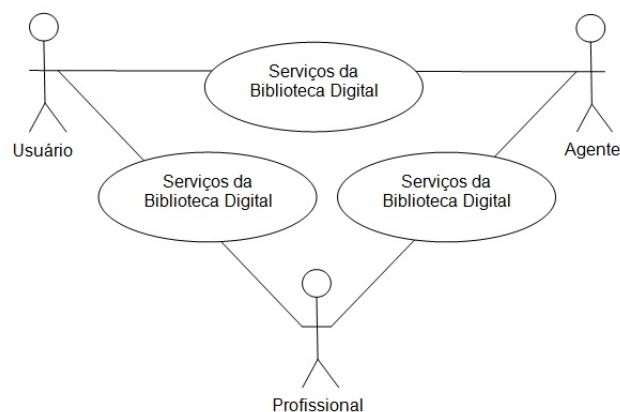


FIGURA 3 - Caso de uso genérico de uma biblioteca digital

Fonte: (BORBINHA *et al.*,2005) (tradução nossa).

Fuhr e outros (2007) apresentaram o Modelo Tríplice de Interação para bibliotecas digitais, que consiste de três componentes principais: *Sistema*, *Conteúdo* e *Usuários*. O primeiro componente, *Sistema*, representa o sistema gerenciador da biblioteca digital, responsável por distribuir *Conteúdo* disponível para os *Usuários*. O modelo também identifica três eixos de interação entre os componentes do modelo: *Performance* relacionada a distribuição de conteúdos pelo sistema, *Utilidade* dos conteúdos para os usuários, e *Usabilidade* do sistema sob a perspectiva dos usuários.

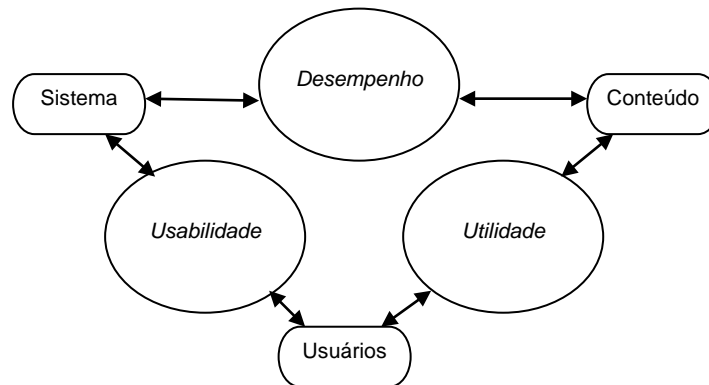


FIGURA 4 - O Modelo Tríplice de Interação

Fonte: (FUHR *et al.*, 2007) (tradução nossa).

A seguir serão apresentados, com maiores detalhes, dois modelos de referência desenvolvidos para o domínio das bibliotecas digitais: o Modelo de Referência DELOS e o Modelo 5S. O Modelo de Referência DELOS identifica os principais componentes no domínio das bibliotecas digitais, introduzindo a distinção entre bibliotecas digitais, sistema de biblioteca digital e sistema gerenciador de biblioteca digital, além de estabelecer as bases para o desenvolvimento da área. O Modelo 5S fornece suporte para a descrição de diferentes conceitos presentes no domínio das bibliotecas digitais, permitindo que se possa melhor compreender tais conceitos. Ambos os modelos são de extrema importância na medida em que estabelecem as bases conceituais para a pesquisa e o desenvolvimento da área.

2.1.5.1 O Modelo de Referência DELOS para bibliotecas digitais

A Rede DELOS de Excelência em Bibliotecas Digitais¹³ tem como objetivo conduzir as pesquisas em tecnologias avançadas para bibliotecas digitais, permitindo o compartilhamento dos resultados destas pesquisas. Um dos produtos deste grupo de pesquisa é o Modelo de Referência DELOS para Sistemas Gerenciadores de Bibliotecas Digitais (CANDELA *et al.*, 2007b). O Modelo DELOS estabelece uma estrutura de três camadas, apresentada na FIGURA 5, na qual (1) *Sistemas Gerenciadores de Bibliotecas Digitais* são instanciados em (2) *Sistemas de Biblioteca Digital* que, por sua vez, permitem a operação de (3) *Bibliotecas Digitais*. De modo a identificar as diferenças e características individuais associadas a cada um dos componentes, seguem as definições apresentadas pelo Modelo:

¹³ Ver em: <http://www.delos.info>

- (1) Uma **Biblioteca Digital** (BD) pode ser vista como uma instituição (podendo ser virtual), que realiza a coleta sistemática, gerencia e preserva *conteúdos digitais*, além de oferecer *funcionalidades* sobre estes conteúdos aos seus *usuários*, funcionalidades estas de *qualidade* mensurável e de acordo com *políticas* bem definidas.
- (2) Um **Sistema de Biblioteca Digital** (SBD) é um software baseado em uma *arquitetura* definida (que pode ser distribuída) e que fornece toda a funcionalidade necessária a uma *Biblioteca Digital* específica. Os *usuários* interagem com uma *Biblioteca Digital* através do *Sistema de Biblioteca Digital* correspondente.
- (3) Um **Sistema Gerenciador de Biblioteca Digital** (SGBD) é um sistema de software genérico que fornece uma infraestrutura apropriada tanto para (i) construir e administrar um Sistema de Biblioteca Digital, incorporando um conjunto de funcionalidades considerado fundamental para *Bibliotecas Digitais* quanto para (ii) integrar softwares adicionais que ofereçam funcionalidades mais refinadas, especializadas ou avançadas.

Enquanto o conceito de *Biblioteca Digital* captura um sistema abstrato formado tanto por componentes físicos como virtuais, os conceitos de *Sistema de Biblioteca Digital* e de *Sistema Gerenciador de Biblioteca Digital* caracterizam sistemas de *software* concretos. Para cada *Biblioteca Digital* existe um determinado *Sistema de Biblioteca Digital* em operação, enquanto que todo *Sistema de Biblioteca Digital* é baseado em algum tipo de *Sistema Gerenciador de Biblioteca Digital*. Uma *Biblioteca Digital* pode ser vista como uma entidade abstrata que possui vida graças aos sistemas de *software*, além das pessoas e organizações envolvidas.

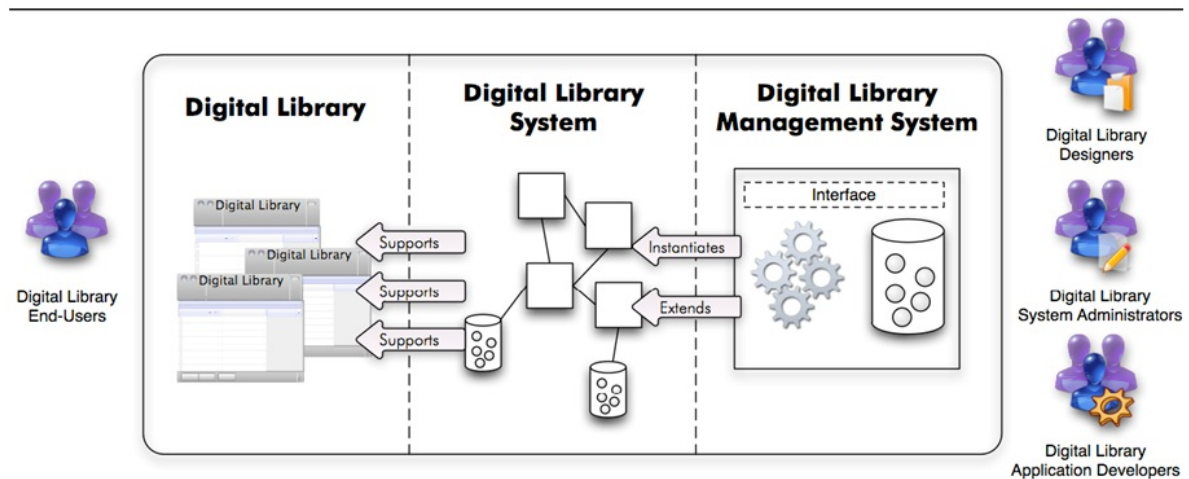


FIGURA 5 - O Modelo de Referência DELOS: componentes e atores

Fonte: (CANDELA *et al.*,2007b)

Candela e outros (2007a) identificam seis conceitos relacionados a bibliotecas digitais, que são apresentados na FIGURA 6. São eles: *conteúdo* (dados e informação); *usuários*; *funcionalidade* (oferecida pela biblioteca digital); *qualidade* (associada a outros conceitos); *política* (regras e condições); *arquitetura* (mapeamento da *funcionalidade* e *conteúdo* da biblioteca digital em componentes de *hardware* e *software*).

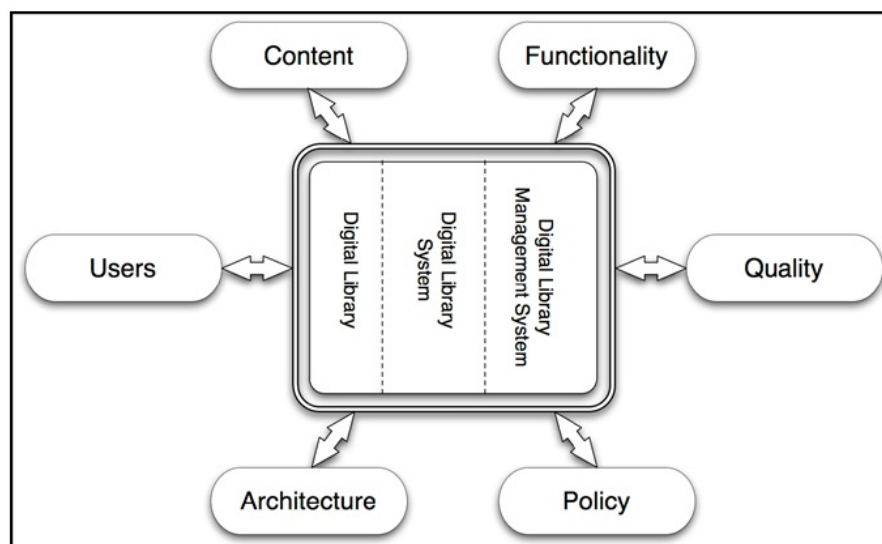


FIGURA 6 - O Modelo de Referência DELOS: principais conceitos

Fonte: (CANDELA *et al.*,2007a)

O Modelo DELOS (CANDELA *et al.*, 2007b) identifica quatro classes de usuários ou atores que estão representadas na FIGURA 5: desenvolvedores de aplicações, administradores do sistema, projetistas de biblioteca digital e usuários finais. Estas classes de usuários se encontram inseridas no universo das bibliotecas digitais. Os autores fornecem definições formais para os conceitos a partir de mapas conceituais. O conceito mais abstrato é *Recurso*, que é estendido para os outros conceitos no modelo:

- *Conteúdo* representa objetos de informação que podem ser distinguidos pelo tipo de dado/informação, tipo da representação da informação (linguagem natural ou formal) e estado da representação digital (nascido digital, digitalizado, ou não digital). A definição de *Conteúdo* segue os princípios do FRBR (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998), que distingue *Obra*, *Expressão*, *Manifestação* e *Ítem*. *Conteúdo* tanto pode ser um objeto primário de informação, metadado ou anotação.
- *Usuário* é um conceito que representa qualquer tipo de ator no domínio das bibliotecas digitais. Estende-se como ator qualquer entidade externa que interaja com a biblioteca digital incluindo pessoas, instituições, agentes de software, etc. A subclasse dos usuários finais exploram as funcionalidades da Biblioteca Digital com o propósito de fornecer, consumir ou gerenciar os seus conteúdos. Neste sentido eles podem ser divididos em *fornecedores de conteúdo*, *consumidores de conteúdo* e *bibliotecários*.
- *Política* é um conceito que engloba as regras de controle de acesso ao sistema.
- *Qualidade* é um conceito abstrato que se aplica aos outros cinco conceitos principais do Modelo DELOS (conteúdo, usuários, funcionalidade, políticas e arquitetura).
- *Arquitetura* é um conceito que fornece propriedades para descrever a estrutura de um dado sistema gerenciador de biblioteca digital, podendo ser entendida como um mapeamento da *funcionalidade* e *conteúdo* da biblioteca digital em componentes de *hardware* e *software*.

O conceito de *Recurso* é referenciado pelo conceito abstrato de *Função* por meio da propriedade *actsOn*, entre outras. Existem vários conceitos funcionais definidos no Modelo DELOS: *access*, *manage resources*, *manage information object*, *collaborate*, *manage digital library*, *manage and configure digital library system*.

O Modelo DELOS também identifica os principais princípios de interoperabilidade relacionados a metadados e protocolos. Neste sentido os autores

identificam um conjunto de conceitos que devem ser mapeados para um dado sistema gerenciador de bibliotecas digitais a fim de permitir interoperabilidade no nível conceitual: *hasMetadata(Resource, Object)*, *hasFormat(Resource, ResourceFormat)*, *associatedWith(Resource, Resource)*, *Transform* como uma especialização de *Process*, *ImportCollection* e *ExportCollection* como especializações de *ManageCollection*, *Compare* como uma especialização de *Analyse*.

O Modelo DELOS constitui-se de um arcabouço abstrato voltado para o entendimento das entidades e de seus relacionamentos no universo das bibliotecas digitais, buscando o desenvolvimento de padrões e especificações consistentes como suporte aos elementos deste universo. O modelo objetiva fornecer uma semântica comum que possa ser usada por diferentes áreas de aplicação para analisar, explicar e organizar bibliotecas digitais, além de servir de suporte a pesquisa e desenvolvimento da área. Os autores reconhecem que devido ao amplo escopo, interdisciplinaridade da área, heterogeneidade dos atores envolvidos e a ausência de iniciativas anteriores em estabelecer as bases da área, o modelo DELOS deve ser considerado como uma iniciativa em constante desenvolvimento, compartilhada pela comunidade envolvida com as bibliotecas digitais.

2.1.5.2 Modelo 5S

O Modelo 5S, proposto por Gonçalves e outros (2004), é formal e consiste em cinco componentes para a descrição do domínio das bibliotecas digitais. Este é citado por Fuhr e outros (2007) como o modelo mais genérico para bibliotecas digitais. De acordo com este modelo, uma biblioteca digital consiste em um repositório (formado por coleções de objetos digitais), catálogos (constituídos por especificações de metadados), serviços e uma sociedade de usuários. A sigla 5S se refere aos cinco componentes do modelo:

- *Streams* são seqüências (ou fluxos) de elementos de um determinado tipo (como por exemplo *bits*, caracteres, imagens) que podem representar conteúdo estático ou dinâmico. *Seqüências estáticas* correspondem à informação representada como elementos simples, como texto numa sequência de caracteres ou como objetos mais complexos, como um livro numa sequência de textos e imagens. *Seqüências dinâmicas* são usadas para modelar qualquer fluxo de informação e, desta forma, são úteis para representar qualquer tipo de comunicação no contexto de uma biblioteca digital. Cada *seqüência* possui um determinado tipo associado, que é usado para definir sua semântica e área de aplicação. As *seqüências* descrevem as propriedades do conteúdo da biblioteca digital.

- *Structures* (ou estruturas) representam a forma como partes de um todo são organizadas. Elas podem ser utilizadas para representar hipertextos e objetos de informação estruturados, além de taxonomias, conexões do sistema e relacionamentos dos usuários. As estruturas definem a organização do conteúdo em uma biblioteca digital.
- *Spaces* (ou espaços) são conjuntos de objetos e operações que operam sobre estes conjuntos sob determinadas restrições. Trata-se de um tipo de construtor bastante poderoso e, como sugerido pelos autores, quando um determinado aspecto da biblioteca digital não puder ser adequadamente descrito pelos demais componentes do modelo, os *espaços* podem ser aplicados. *Espaços* de documentos representam um conceito chave em uma biblioteca digital. No entanto, *espaços* são usados também em outros contextos, como na indexação e visualização. Diferentes tipos de *espaços* são propostos, como os métricos, probabilísticos, vetoriais e topológicos. Os *espaços* especificam a apresentação lógica dos vários componentes de uma biblioteca digital.
- *Scenarios* (cenários ou roteiros) consistem de sequências de ações ou eventos que modificam o estado de uma computação de modo a completar determinado requisito funcional (funcionalidade). Eles descrevem o comportamento externo do sistema a partir do ponto de vista do usuário, servindo de base para a construção de sistemas gerenciadores de bibliotecas digitais. Os *cenários* englobam dois conceitos fundamentais: *state* (estado) e *event* (evento). Os cenários explicam o que acontece com as *sequências*, dentro dos *espaços* e através das *estruturas*. Eles permitem descrever os serviços, atividades, tarefas e operações que, em conjunto, definem as funcionalidades de uma biblioteca digital. Os fluxos de trabalho e de dados são exemplos de *cenários* em uma biblioteca digital. Os *cenários* são usados para definir como os serviços se comportam e mudam o estado do sistema em situações como na busca, navegação, recomendação, etc.
- *Societies* (ou sociedades) são conjuntos de entidades, atividades e relacionamentos entre elas. Entidades podem ser pessoas ou componentes de *hardware* ou *software*. No caso de pessoas, podem ser usuários, autores, editores, desenvolvedores de software ou bibliotecários. No caso de *sociedades* formadas por pessoas, estas possuem papéis, atividades, objetivos e relacionamentos. *Sociedades* representam o conceito de nível mais alto do modelo 5S. De acordo com os autores, uma biblioteca digital deve servir às necessidades de informação de suas *sociedades*. Em uma abordagem alternativa, as *sociedades* também podem desempenhar o papel de fornecedores

de informação, como na proposta de Kruk e McDaniel (2009). O modelo 5S identifica várias questões e políticas que precisam ser tratadas no contexto do conceito das *sociedades*. Estas políticas incluem: uso e reuso da informação, privacidade, propriedade e direitos autorais, gerenciamento do acesso e segurança. As sociedades definem os atores e os serviços utilizados por eles.

Candela e outros (2007b) relacionam os conceitos do modelo 5S com os objetivos de uma biblioteca digital: *Sociedades* definem como uma biblioteca digital ajuda a satisfazer as necessidades de informação de seus usuários; *Cenários* fornecem suporte para a definição e projeto de diferentes tipos de serviços da biblioteca; *Estruturas* permitem a organização da informação de maneiras lógicas e úteis; *Espaços* tratam da apresentação e acesso a informação de formas úteis e efetivas; e *sequencias* lidam com a comunicação e consumo da informação pelos usuários.

Estas abstrações, em conjunto, fornecem uma base formal para definir, relacionar, e unificar conceitos necessários para formalizar e elucidar o domínio das bibliotecas digitais. De acordo com o modelo 5S, uma biblioteca digital consiste em um repositório (formado por coleções de objetos digitais), catálogos (constituídos por especificações de metadados), serviços e uma sociedade de usuários. De acordo com este arcabouço, Golçalves e outros (2004) definem formalmente uma biblioteca digital mínima como uma quádrupla (**R, Cat, Serv, Soc**) onde:

- (1) **R** é um repositório, um serviço que engloba um conjunto de coleções com operações específicas (*obter, armazenar e excluir*) para manipular estas coleções;
- (2) **Cat** é um conjunto de catálogos de metadados, para todas as coleções no repositório;
- (3) **Serv** é um conjunto de serviços que contem pelo menos serviços voltados para a indexação, busca e navegação;
- (4) **Soc** é uma sociedade.

Gonçalves e outros (2004) analisaram três exemplos de aplicações no contexto do Modelo 5S: a *Digital Library Taxonomy*, a *Networked Library of Theses and Dissertations*¹⁴ (NDLTD), e o protocolo OAI-PMH (VAN de SOMPEL *et al.*, 2004).

No caso da *Digital Library Taxonomy*, os autores propõem uma taxonomia para o domínio das bibliotecas digitais, na qual identificam cinco agrupamentos de conceitos que

¹⁴ Ver em: <http://www.ndltd.org>

são relacionados ao Modelo 5S: atores (que interagem com e dentro de uma biblioteca digital), atividades (o que acontece na biblioteca digital), componentes (que formam a biblioteca digital), aspectos socioeconômicos e legais (que cercam a biblioteca digital) e o ambiente (contexto no qual a biblioteca digital se encontra inserida).

No caso da *Networked Library of Theses and Dissertations* (NDLTD), os autores apresentam o uso do Modelo 5S como uma ferramenta analítica que auxilia no entendimento deste empreendimento, na medida em que analisam cada um dos conceitos do modelo no contexto da NDLTD.

A Iniciativa de Arquivos Abertos (*Open Archives Initiative* – OAI) (LAGOZE; VAN de SOMPEL, 2001) não é uma biblioteca digital, mas um projeto multi-institucional que visa a interoperabilidade entre arquivos e bibliotecas digitais a partir da definição de protocolos para o intercâmbio de metadados. A infraestrutura técnica do projeto é definida pelo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) (VAN de SOMPEL et al., 2004), que define mecanismos para a exposição e coleta de metadados. Gonçalves e outros (2004) analisaram esta infraestrutura técnica a partir dos conceitos fornecidos pelo Modelo 5S.

Motivado pelo objetivo maior de construir uma teoria formal para bibliotecas digitais, Golçalves e outros (2004) mostraram que a partir de definições formais, o arcabouço 5S pode ser completamente descrito, permitindo a definição formal de uma biblioteca digital mínima. Os autores mostraram ainda a utilidade do modelo para discutir a terminologia encontrada na literatura, para descrever bibliotecas digitais e para construir uma linguagem declarativa voltada para a especificação de bibliotecas digitais.

2.1.5.3 A importância dos modelos

Modelos formais e teorias têm importância crucial para a especificação e o entendimento claro e sem ambiguidades das características, estrutura, e comportamento de sistemas de informação complexos, como é o caso das bibliotecas digitais. Um modelo formal destaca as características gerais e padrões comuns de um conjunto de sistemas desenvolvidos para tratar problemas similares, explica a sua estrutura e processos, além de incentivar práticas comuns.

A modelagem dos atores e de seus respectivos papéis e interações no contexto de uma biblioteca digital, assim como o desenvolvimento de uma teoria formal acerca deste tema, conforme sugerido por Borbinha e outros (2005), se justifica na medida em que o papel desempenhado pelas bibliotecas digitais mudou, tendo passado de sistemas estáticos de armazenamento e recuperação de documentos para sistemas dinâmicos que facilitam a

comunicação, colaboração, e outras formas de interação entre cientistas, pesquisadores, e público em geral acerca de temas relacionados à informação armazenada em uma biblioteca digital.

O Modelo DELOS captura, organiza e promove o entendimento dos principais conceitos relativos ao domínio das bibliotecas digitais; fornece parâmetros e orientação para a construção de sistemas gerenciadores de bibliotecas digitais e para o desenvolvimento de esquemas de metadados. O Modelo 5S identifica cinco conceitos abstratos que podem ser usados na descrição de taxonomias, sistemas gerenciadores de bibliotecas digitais, e protocolos de comunicação. Ambos os modelos enfatizam a importância da padronização em torno do modelo de objetos e interoperabilidade, assim como identificam o papel dos diversos atores no contexto das bibliotecas digitais.

Candela e outros (2007b), ao analisarem as diferenças e convergências entre o Modelo DELOS e o Modelo 5S apresentaram a FIGURA 7 a seguir.

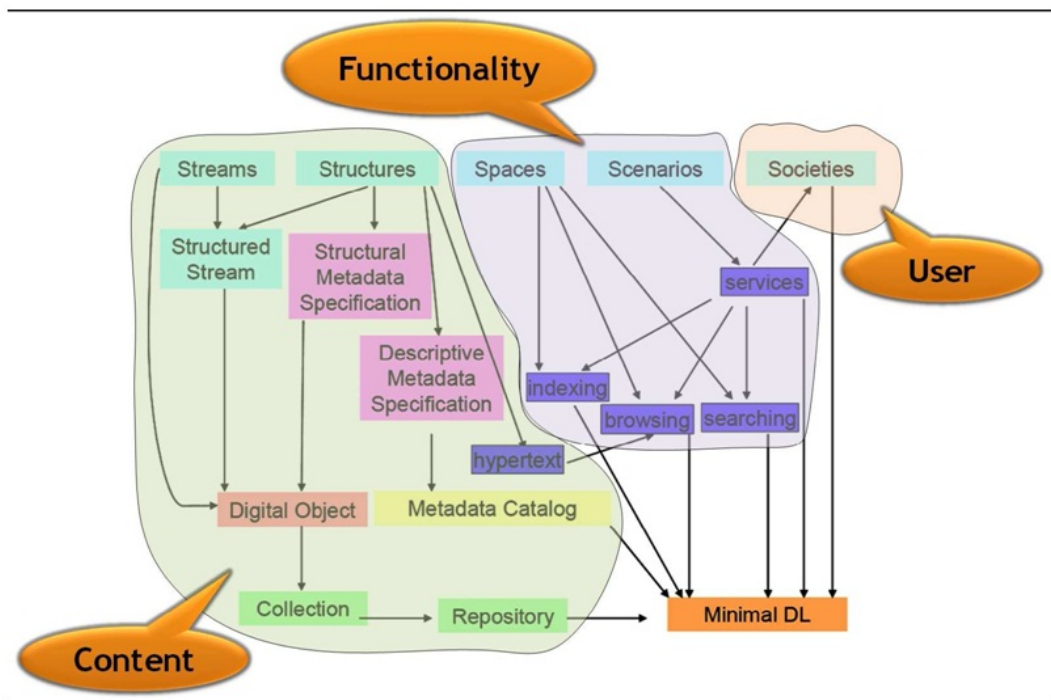


FIGURA 7 - Conceitos do Modelo 5S e sua correspondência com o Modelo DELOS

Fonte: (CANDELA *et al.*, 2007b)

O Modelo 5S cobre basicamente os componentes que no modelo DELOS foram rotulados como *Conteúdo*, *Funcionalidade* e *Usuário*; o conceito de *Qualidade* presente no modelo DELOS foi tratado separada e posteriormente no modelo 5S (GONÇALVES *et al.*, 2007), enquanto o conceito de *Política* fora pouco tratado pelo modelo 5S. Além disso, o

nível de detalhes pode variar bastante em diferentes áreas. Enquanto o Modelo 5S introduz conceitos bem específicos em algumas áreas, em outras ele adota uma abordagem de mais alto nível; estas mesmas considerações valem para o Modelo DELOS.

O Modelo 5S fornece construtores bastante genéricos e desta forma mais difíceis de serem aplicados do que aqueles fornecidos na abordagem mais pragmática adotada pelo Modelo DELOS. Além disso, o Modelo 5S foca na identificação da “biblioteca digital mínima” com o objetivo claro de formalizar seus aspectos, enquanto o modelo DELOS foca em identificar os principais conceitos e relacionamentos que caracterizam todo o universo das bibliotecas digitais, considerando a formalização como um passo futuro. Candela e outros (2007b) chamam a atenção para a convergência observada nos dois modelos em relação ao conceito de *Objeto de Informação*, rotulado como *objeto digital* no modelo 5S. Segundo os autores este fato indica que tal elemento fora mais bem investigado e provavelmente melhor entendido que outros elementos que constituem o universo das bibliotecas digitais.

Os modelos apresentados formam uma base sólida na qual qualquer pesquisa relacionada a bibliotecas digitais deve se apoiar. Eles são especialmente importantes para o projeto uma biblioteca digital que faça uso de mecanismos de representação e organização do conhecimento e metadados semanticamente mais ricos, de acordo com os objetivos desta pesquisa.

2.1.6 Metadados: modelos e padrões

Os metadados representam um conceito familiar para as bibliotecas convencionais, na medida em que sempre foram usados para descrever os documentos no processo de catalogação e indexação. Para Vellucci (1998), no contexto das bibliotecas digitais, os metadados podem ser vistos como dados que descrevem atributos de um documento digital, caracterizam suas relações, possibilitam a sua recuperação, seu uso efetivo e sua existência no ambiente eletrônico. Metadados normalmente consistem de um conjunto de elementos de dados onde cada elemento descreve um atributo do recurso informacional, sua estrutura, sua administração ou uso (VELLUCCI, 1998). Como aponta Cleveland (1998), os metadados são importantes em bibliotecas digitais porque representam a chave para a descoberta de recursos e usos para qualquer documento. Os metadados são parte importante de qualquer biblioteca digital, na medida em que são utilizados para a descrição dos recursos informacionais, servindo tanto para se fazer anotações bibliográficas

nestes recursos, como para expressar relações entre os mesmos. Os metadados semanticamente mais ricos podem permitir a expressão de relacionamentos complexos entre objetos de informação (LAGOZE *et al.*, 2006; KRUK *et al.*, 2005), permitindo também a ligação entre estes objetos e conceitos entendidos pelos usuários. Os serviços oferecidos por uma biblioteca digital, especialmente aqueles relacionados à busca, à descoberta e à recuperação da informação, podem usar estes metadados em mecanismos de navegação, personalização e busca envolvendo semântica.

Modelos são representações abstratas voltadas para a análise e o estudo. No caso no domínio bibliográfico, os modelos são desenvolvidos para facilitar o entendimento acerca das entidades e conceitos envolvidos, visando principalmente formar um entendimento compartilhado do domínio. Neste sentido, os modelos são utilizados para subsidiar a construção de padrões a serem utilizados na descrição bibliográfica.

Padrões podem ser entendidos como ferramentas e normas que permitem que se faça algo da mesma maneira. Servem a um grupo de pessoas e/ou instituições que, consensualmente, aceitam esses padrões. Toda forma de padronização é garantida por códigos, normas, patentes, tabelas, enfim, qualquer documentação que torne os procedimentos em uma determinada área normalizados. Os padrões podem ser definidos pelo mercado; também podem surgir como resultado da iniciativa de grupos de especialistas de uma determinada área; ou como consequência de uma legislação de interesse público. Os padrões também podem surgir como uma expressão de um consenso dentro de uma comunidade. Quando eles nascem por este último motivo tendem a ser mais duradouros, mas este consenso é difícil de ser alcançado (LOURENÇO, 2005, p.59).

Os padrões de metadados podem ser vistos como padrões de descrição bibliográfica modernos, utilizados por bibliotecas digitais e que, em sua maioria, se baseiam nas normas e padrões da representação descritiva tradicional, com o objetivo de padronizar e tornar as bibliotecas digitais mais interoperáveis.

A concepção, projeto e implementação de uma biblioteca digital deve envolver uma análise cuidadosa das necessidades dos atores envolvidos, entre eles os usuários da informação, administradores e bibliotecários, de modo a estabelecer um conjunto de metadados que possa atender a estas necessidades. Neste sentido, modelos que representem o universo bibliográfico, suas entidades, atributos e relacionamentos podem ajudar nesta complexa tarefa. A seguir será apresentado um estudo que gerou importantes contribuições para a descrição bibliográfica, além de alguns padrões de metadados que podem ser utilizados em bibliotecas digitais.

2.1.6.1 FRBR

Os Requisitos Funcionais para Descrição Bibliográfica, sigla em inglês *FRBR* (*Functional Requirements for Bibliographic Records*) é um estudo produzido por um grupo da Federação Internacional das Associações de Bibliotecas, conhecida internacionalmente pela sigla IFLA (*International Federation of Library Associations and Institutions*), especificamente constituído para este propósito (IFLA, 1998). A metodologia usada no desenvolvimento do estudo se baseia na técnica de análise entidade-relacionamento (NAVATHE; ELMASRI, 2003), usada no desenvolvimento de modelos conceituais em aplicações de bancos de dados relacionais. Na construção do modelo produzido pelo estudo, buscou-se inicialmente identificar as entidades, seus atributos e relacionamentos existentes no universo bibliográfico e em seguida relacionar estes elementos com as tarefas genéricas executadas pelos usuários destes registros bibliográficos (em encontrar, identificar, selecionar e obter informação) de modo a recomendar um nível básico de funcionalidade para registros bibliográficos.

O propósito do FRBR é fornecer conceitos para a descrição de uma visão genérica e abrangente do universo bibliográfico (IFLA, 1998). O FRBR descreve a estrutura e os relacionamentos de registros bibliográficos e de autoridade. O modelo foi desenvolvido para ser usado como base na construção de catálogos e sistemas bibliográficos. Neste sentido, o FRBR identifica e define claramente as entidades de interesse no universo bibliográfico, os atributos de cada entidade, assim como os relacionamentos entre as mesmas, tendo dois objetivos principais: (1) fornecer um arcabouço bem definido e estruturado para relacionar os dados disponíveis em registros bibliográficos com as necessidades dos usuários destes registros; e (2) definir e recomendar um nível básico de funcionalidade para registros bibliográficos buscando a padronização e interoperabilidade entre diferentes instituições.

A análise entidade-relacionamento representada pelo modelo também serve como um arcabouço conceitual para o desenvolvimento de modelos e estruturas usadas para armazenar, exibir e transferir dados bibliográficos (metadados) no contexto de sistemas computacionais.

No FRBR, o conceito geral de um dado objeto informacional é definido de forma precisa a partir de quatro entidades: *Item*, *Manifestação*, *Expressão* e *Obra*. Quando se move de *Item* para *Obra*, o entendimento do conceito do objeto informacional se torna mais abstrato e menos atrelado ao objeto físico de informação e mais próximo do trabalho intelectual ou artístico. O modelo organiza as entidades definidas em três grupos. O *Grupo 1* é composto pelas quatro entidades citadas anteriormente.

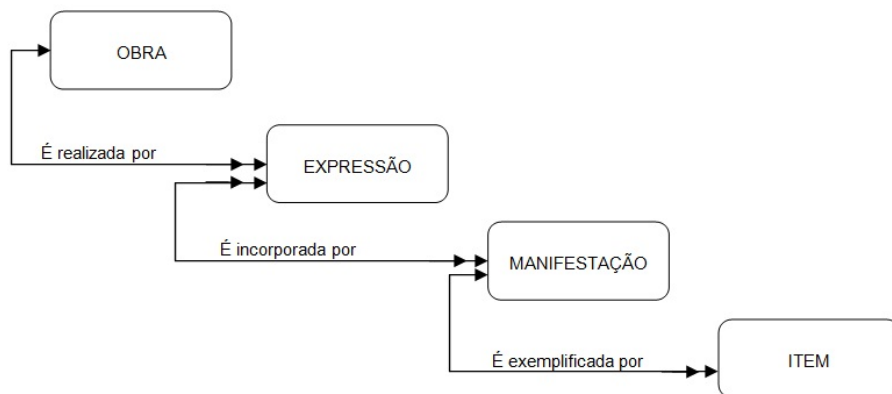


FIGURA 8 - FRBR: entidades e relacionamentos do Grupo 1

Fonte: adaptado de IFLA (1998)

No *Grupo 2* estão as entidades *Pessoa* e *Entidade Coletiva*, que desempenham papéis relacionados às entidades do *Grupo 1*, sendo responsáveis pelo conteúdo intelectual ou artístico, pela produção, disseminação, ou custódia destes conteúdos.

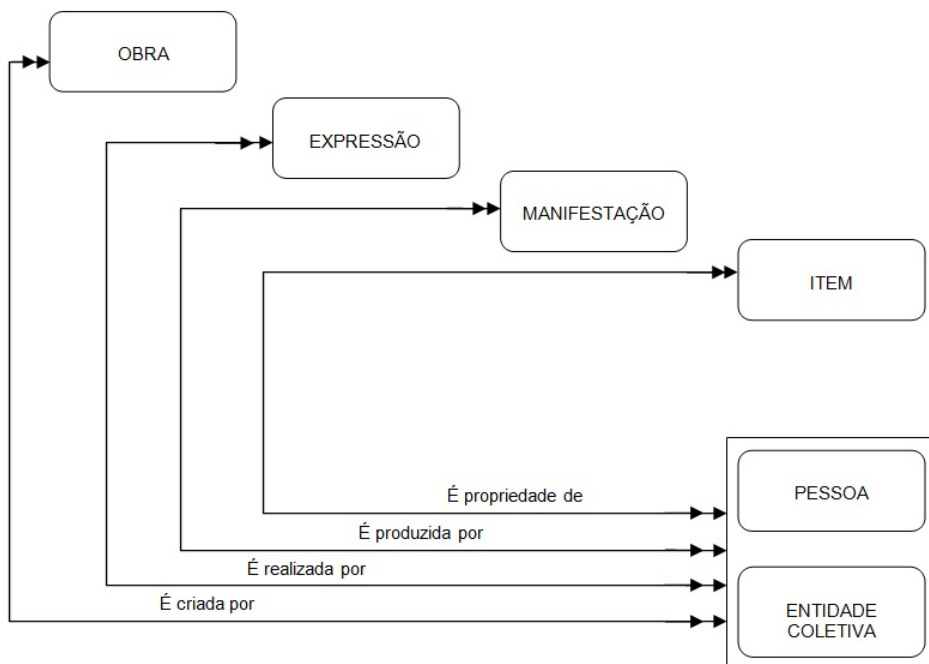


FIGURA 9 - FRBR: entidades e relacionamentos dos grupos 1 e 2

Fonte: adaptado de IFLA (1998)

O *Grupo 3* define um conjunto adicional de entidades que atuam como tema ou assunto de um esforço intelectual ou artístico, quais sejam: *Conceitos*, *Objetos*, *Eventos* e *Lugares*.

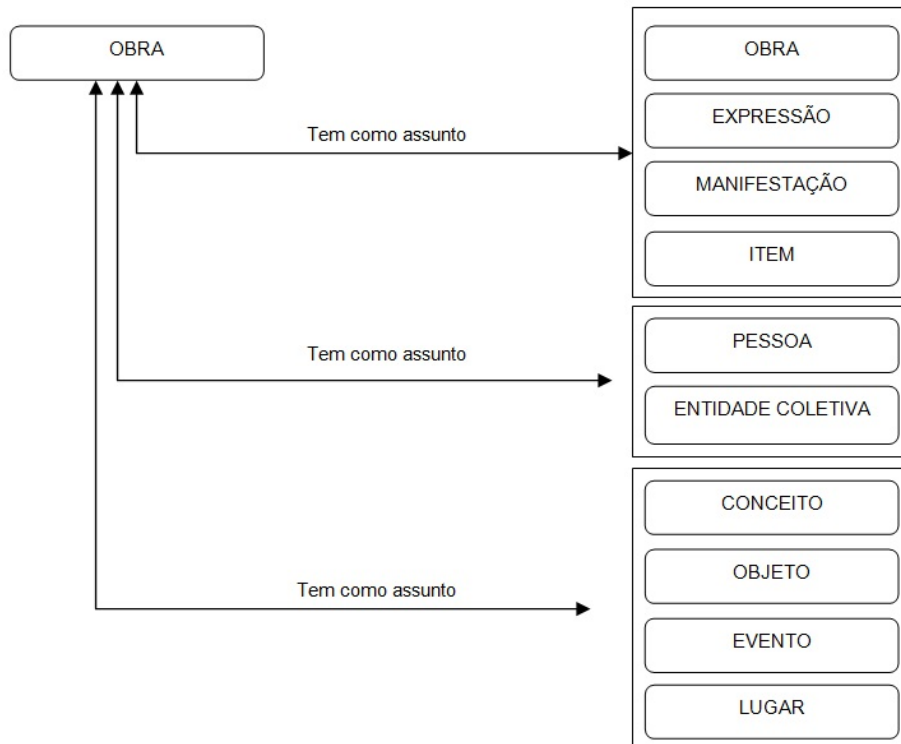


FIGURA 10 - FRBR: entidades e relacionamentos dos Grupos 1, 2 e 3

Fonte: adaptado de IFLA (1998)

O modelo FRBR identifica três tipos de relacionamentos bibliográficos: *inerente* (entre as entidades do *Grupo 1*), *conteúdo*, *todo-parte*, e *parte-parte*. Finalmente o modelo também define algumas tarefas dos usuários em *encontrar*, *identificar*, *selecionar* e *obter* informação, além de relacionar estas tarefas com as entidades, atributos e relacionamentos identificados no universo bibliográfico.

Lourenço (2005), ao analisar o FRBR, ressalta a importância do modelo em trazer novidades para discussão e tornar visíveis categorias e conceitos inerentes às tarefas de descrição bibliográfica, oferecendo uma linguagem através da qual os problemas podem ser discutidos, de modo a melhorar o nível de maturidade e de compreensão acerca das normas de representação descritiva. No entanto, a autora também aponta alguns problemas, entre eles a falta de definição clara dos limites entre as entidades *Expressão* e *Manifestação*, sugerindo um melhor desenvolvimento e aprofundamento de estudos.

A modelagem empreendida pelo FRBR permite reexaminar os princípios fundamentais que estão por trás dos códigos de catalogação e padrões de metadados e assim fixar direções para sua adaptação e desenvolvimento futuro. As propostas de revisão de instrumentos da catalogação de acordo com esses modelos abrem caminho para o efetivo aproveitamento das novas tecnologias. Em especial, elas contribuem para uma mudança de paradigma com relação aos objetivos da catalogação e do uso dos metadados, enfatizando que o principal objetivo da descrição é a recuperação, representada pelas atividades de *encontrar*, *identificar*, *selecionar* e *obter* informação.

O FRBR tem se tornado bastante popular nos projetos relacionados a bibliotecas digitais, inclusive tendo sido desenvolvido um esquema RDF para o FRBR (DAVIS; NEWMAN, 2005), trabalho este que contribui para os esforços de construção de sistemas gerenciadores de bibliotecas digitais que utilizam tecnologias da web semântica.

O FRBR fornece um sólido conhecimento sobre “o que existe” no universo bibliográfico, incluindo as entidades e os relacionamentos. Este conhecimento pode ser utilizado na definição de padrões de metadados específicos que possam atender às necessidades dos diversos atores envolvidos, tanto nas tarefas de encontrar, identificar, selecionar e obter informação, como nas tarefas administrativas, de organização e de preservação de recursos informacionais.

2.1.6.2 Dublin Core

O Dublin Core (DCMI, 2003) é um padrão de metadados criado pela OCLC (*On Line Computer Library*) e pelo NCSA (*National Center for Supercomputer Applications*) para facilitar a descoberta de recursos informacionais na web. O termo “Dublin” se deve ao fato de sua origem ter ocorrido em um *workshop* realizado na cidade de Dublin, Ohio; “Core” devido ao propósito de seu desenvolvimento, em ser um padrão amplo e genérico, para ser utilizado na descrição de vários tipos de recursos informacionais. Entre as principais características do Dublin Core está a sua generalidade de aplicação e uso, podendo ser aplicado para descrever quase todo tipo de recurso informacional.

Embora tenha sido originalmente concebido como um mecanismo que encorajasse os próprios autores de documentos digitais a criarem seus próprios metadados em suas publicações, sua maior utilização tem sido verificada em projetos associados a bibliotecas digitais, instituições culturais e agências do governo. Sua importância no contexto das bibliotecas digitais foi reconhecida em 2003, quando o Dublin Core se tornou uma norma ISO.

Desde 1998, quando os 15 elementos do Dublin Core passaram pelo processo de padronização, o desenvolvimento de alguns conceitos e técnicas no contexto da web semântica, com o objetivo de formalização, adicionou os conceitos de domínio (*domain*) e extensão (*range*) às propriedades representadas pelos 15 elementos do Dublin Core. O Dublin Core incorporou este nível de formalização em uma versão paralela e compatível do padrão. Domínio e extensão especificam, respectivamente, os tipos de recursos que podem ser descritos e a faixa de valores aceitos para determinada propriedade. Desta forma, estes elementos expressam o significado, implícito na linguagem natural, de uma forma explícita que pode ser utilizada no processamento automático de inferências lógicas. Por exemplo, ao se deparar com a propriedade *Creator*, um agente computacional pode inferir que o recurso descrito por esta propriedade é um recurso informacional (domínio) e que o valor atribuído a ela é uma pessoa ou uma instituição (extensão).

O Dublin core possui atualmente 15 elementos descritivos que são:

1. **Title** – nome pelo qual o recurso é formalmente conhecido.
2. **Creator** – entidade principal responsável pelo conteúdo intelectual do recurso. Pode ser uma pessoa ou instituição. Geralmente o nome da entidade é usado para identificá-la.
3. **Subject** – tópico ou assunto relacionado ao recurso descrito. Pode ser uma sequência de palavras-chave ou algum código de classificação. Recomenda-se o uso de algum tipo de vocabulário controlado. Tópicos espaciais ou temporais devem ser descritos com o elemento *Coverage*.
4. **Description** – qualquer tipo de explicação ou descrição do recurso, podendo incluir um extrato, resumo, tabela de conteúdo, representação gráfica, ou qualquer explicação em texto livre.
5. **Publisher** – agente (pessoa ou instituição) responsável por tornar o recurso disponível. Geralmente o nome do agente é usado para identificá-lo.
6. **Contributor** – entidades (pessoas ou instituições) responsáveis por contribuições ao conteúdo intelectual do recurso. Geralmente o nome da entidade é usado para identificá-la.
7. **Date** – um dado ponto ou período no tempo associado com um evento (publicação, registro, disponibilização, etc.) do ciclo de vida do recurso. Geralmente representado por uma data em qualquer nível de granularidade.
8. **Type** – natureza ou gênero do recurso. Para descrever formato de arquivo, meio físico (mídia) ou dimensões (tamanho, duração) relacionadas ao recurso deve-se utilizar o elemento *Format*.

9. **Format** – formato do arquivo, meio físico (mídia), ou dimensões (tamanho, duração) associadas ao recurso.
10. **Identifier** – uma referência unívoca (não ambígua) para o recurso em um dado contexto.
11. **Source** – um recurso relacionado a partir do qual o recurso descrito fora derivado. O recurso descrito pode ser total ou parcialmente derivado a partir de outro recurso. Recomenda-se que o recurso relacionado seja identificado de maneira unívoca e em conformidade com algum esquema de identificação.
12. **Language** – idioma relativo ao conteúdo intelectual do recurso.
13. **Relation** – um recurso relacionado. Recomenda-se que o recurso relacionado seja identificado de maneira unívoca e em conformidade com algum esquema de identificação.
14. **Coverage** – cobertura espacial ou temporal do recurso, aplicabilidade espacial do recurso ou ainda a jurisdição sob a qual o recurso é relevante. Pode ser um período, data ou faixa de datas. Quando representar uma jurisdição, pode ser uma entidade administrativa ou um lugar aonde o recurso se aplica.
15. **Rights** – direitos sobre o recurso. Inclui a declaração sobre os direitos de propriedade associados com o recurso, incluindo os direitos de propriedade intelectual.

Além dos elementos básicos, o Dublin Core possui qualificadores que podem ser utilizados em conjunto com os elementos principais, citados e descritos acima. O Dublin Core foi utilizado como base para diversos outros padrões mais específicos, como o ETD-MS e o MTD-BR, ambos voltados para a descrição de teses e dissertações. Estes padrões serão discutidos a seguir.

2.1.6.3 ETD-MS e MTD-BR

O ETD-MS (*Electronic Thesis and Dissertations Metadata Standard*) (ATKINS *et. al.*, 2008) é um padrão de metadados voltado para a descrição de teses e dissertações, desenvolvido no contexto do projeto da NDLTD¹⁵ (*Networked Digital Library of Theses and Dissertations*) (FOX *et. al.*, 1997). Tendo se baseado no Dublin Core, o padrão EDT-MS aplicou quase todos os seus elementos, com exceção dos elementos *DC.Source* e *DC.Relation*. Em contrapartida, acrescentou um elemento específico para atender o tipo de material a ser descrito: *thesis.degree*. Este elemento foi acompanhado de quatro

¹⁵ Ver em: <http://www.ndltd.org>

qualificadores: *thesis.degree.name*, *thesis.degree.level*, *thesis.degree.discipline* e *thesis.degree.grantor*.

O Padrão Brasileiro de Metadados para Teses e Dissertações, MTD-BR, foi desenvolvido e utilizado no contexto do projeto da BDTD¹⁶ (Biblioteca Digital de Teses e Dissertações) (MARCONDES; SAYÃO, 2003), patrocinado e dirigido pelo Instituto Brasileiro de Informação Científica e Tecnológica, IBICT. O Padrão MTD-BR contém basicamente os elementos apresentados no QUADRO 1 abaixo. O quadro apresenta um comparativo entre os elementos do Dublin Core (DC) e os elementos dos padrões que nele se basearam: EDT-MS e MTD-BR. O quadro apresenta apenas o nome dos elementos principais, sendo excluídos: os qualificadores, atributos, descrição do conteúdo, indicação de repetição, obrigatoriedade e valores padrões adotados para cada atributo.

QUADRO 1 - Comparativo entre os elementos dos padrões Dublin Core, EDT-MS e MTD-BR

MTD-BR	NDLTD-ETD-MS	DC
Controle	dc.title	dc.title
Biblioteca Digital	dc.creator	dc.creator
Biblioteca Depositária	dc.subject	dc.subject
Título	dc.description	dc.description
Arquivo	dc.publisher	dc.publisher
Idioma	dc.contributor	dc.contributor
Grau	dc.date	dc.date
Titulação	dc.type	dc.type
Resumo	dc.format	dc.format
Cobertura	dc.identifier	dc.identifier
Assunto	dc.language	dc.source
LocalDefesa	dc.coverage	dc.language
DataDefesa	dc.rights	dc.relation
Autor	thesis.degree	dc.coverage
Contribuidor		dc.rights
InstituiçãoDefesa		
AgênciaFomento		
Direitos		
Extensão		

Fonte: (LOURENÇO, 2005)

O padrão MTD-BR adotou os mesmos campos do EDT-MS, tendo descartado os mesmos elementos do Dublin Core. Entretanto, como apontado por Lourenço (2005), o padrão brasileiro detalhou a descrição, identificando também: autores, contribuidores, instituições, informações acerca das bibliotecas, dos programas de pós-graduação das universidades e de outras instituições envolvidas. Com isso o MTD-BR oferece mais pontos

¹⁶ Ver em: <http://bdtd.ibict.br>

de acesso, além de informações para administração da biblioteca digital (metadados administrativos).

2.1.6.4 Considerações sobre os modelos e padrões de metadados

A análise dos modelos e padrões de metadados realizada permitiu verificar que o foco tem sido a representação descritiva dos itens bibliográficos. O modelo FRBR apresenta as entidades *Conceito*, *Objeto*, *Evento*, *Lugar* relacionadas como assuntos de uma *Obra*. Além destas, também podem ser assuntos de uma *Obra* instâncias das entidades *Pessoa*, *Entidade Coletiva*, *Item*, *Expressão*, *Manifestação*, e até mesmo outra *Obra*. No entanto, ao identificar os atributos das entidades *Conceito*, *Objeto*, *Evento*, e *Lugar* o FRBR se limita apenas ao atributo *Termo* para descrever estas entidades.

O padrão MTD-BR, assim como os demais, trata a entidade *Assunto*, que faz parte da representação temática de uma tese ou dissertação. No entanto esta entidade fora tratada de forma superficial. Lourenço (2005, p.128) argumenta que o estudo tem se voltado mais para a representação descritiva e, além disso, a representação temática, se trabalhada em profundidade, pode originar outros estudos na área de padrões de metadados e semântica da recuperação em bibliotecas digitais.

A análise dos modelos e padrões de metadados mostra que as abordagens variam muito com relação ao grau de complexidade semântica. Temos desde padrões como o Dublin Core e outros que nele se basearam, como o ETD-MS e o MTD-BR, que oferecem um nível básico de descrição. Enquanto temos modelos baseados em ontologias que abrem a possibilidade de descrições semanticamente mais complexas. A extensão da funcionalidade relacionada à recuperação da informação em uma biblioteca digital está estreitamente relacionada com a completude e complexidade semântica do esquema de metadados usado na descrição dos recursos. Neste ponto é interessante destacar que o suporte a consultas mais expressivas quase sempre requer a utilização de metadados semanticamente mais ricos, como num maior esforço humano na criação das descrições.

2.1.7 Sistemas gerenciadores de biblioteca digital

Conforme definido pelo Modelo DELOS (CANDELA *et al.*,2007b), apresentado na seção 2.1.5.1, um *sistema gerenciador de biblioteca digital* consiste de um pacote de *software* e procedimentos padronizados que podem ser aplicados mediante configurações específicas na forma de um *sistema de biblioteca digital* que, por sua vez, permite o

funcionamento e operação de uma *biblioteca digital*. De acordo com esta definição podemos entender a importância do sistema gerenciador de biblioteca digital na medida em que é nele que estão implementados os serviços e funcionalidades que podem ser aplicados e disponibilizados em uma biblioteca digital. Os sistemas gerenciadores de bibliotecas digitais representam o resultado da pesquisa na área de bibliotecas digitais, sendo objeto de constantes avanços. A seguir são apresentados alguns dos principais sistemas gerenciadores de bibliotecas digitais disponíveis, bem como suas principais características e funcionalidades. A análise das características desses sistemas são importantes para a definição dos requisitos a serem utilizados na implementação de um protótipo que permita a representação e organização do acervo baseada em uma estrutura facetada, bem como a implementação de componentes de busca e acesso baseados nesta mesma estrutura, no contexto de uma biblioteca digital de teses e dissertações. Serão brevemente apresentados os sistemas gerenciadores de biblioteca digital Fedora (Fedora Development Team, 2005), BRICKS (MEGHINI e RISSE, 2005), Greenstone (BAINBRIDGE *et al.*, 2001) e DSpace (DSPACE FOUNDATION, 2009). Este último será apresentado em maiores detalhes devido ao fato de ser amplamente utilizado na implantação de bibliotecas digitais de teses e dissertações no Brasil. Além disso, este sistema será utilizado como plataforma para o desenvolvimento de um protótipo, no contexto da presente pesquisa.

2.1.7.1 Fedora

O projeto Fedora¹⁷ (Fedora Development Team, 2005) oferece uma estrutura para a construção de bibliotecas digitais baseada em uma arquitetura distribuída e orientada a serviços. Desta forma, o sistema Fedora faz a mediação entre aplicações clientes e diversos serviços web distribuídos. De acordo com Payette e Staples (2002), a arquitetura do Fedora, baseada em XML e serviços web, foi definida de modo a contemplar algumas limitações apresentadas por outros sistemas, que:

- Eram excessivamente atrelados a tipos de mídia específicos.
- Não apresentavam suporte a conteúdos estruturados.
- Não ofereciam suporte a fluxos contínuos (*streams*) de multimídia.
- Não ofereciam suporte para gerenciamento de fluxos de trabalho (*workflow*).
- Não eram capazes de tratar relacionamentos complexos.
- Não eram capazes de tratar diversas manifestações de um mesmo trabalho.
- Apresentavam problemas críticos relacionados à interoperabilidade.

¹⁷ Ver em: <http://www.fedora-commons.org>

Cada objeto no Fedora pode ser composto a partir de outros objetos digitais. Além disso, um objeto pode apresentar diversas versões, representadas por arquivos diferentes. Os arquivos (*datastreams*) podem ser gerados dinamicamente através de disseminadores (*disseminators*), aonde cada disseminador tem o seu comportamento e mecanismo formalmente definido.

De acordo com Lagoze (2005), o Fedora oferece suporte tanto para objetos convencionais (livros, dados geoespaciais, ou imagens), quando para objetos compostos, complexos e dinâmicos (vídeo, dados numéricos e áudio). O Fedora utiliza o padrão METS (METS EDITORIAL BOARD, 2007) para representar informação (metadados) sobre os objetos digitais, incluindo arquivos (*datastreams*) e disseminadores (*disseminators*).

O Fedora oferece em seu ambiente gerenciamento integrado de dados e serviços. A interoperabilidade foi possibilitada a partir da implementação de padrões de protocolos como o OAI-PMH (VAN de SOMPEL et al., 2004). O Fedora também disponibiliza formas para descoberta e execução de serviços web. Desta forma as soluções baseadas no Fedora podem ser integradas a outras aplicações e sistemas.

2.1.7.2 BRICKS

O BRICKS¹⁸ é um projeto europeu envolvendo um grande consórcio de fornecedores de conteúdo e instituições de pesquisa. Tal projeto desenvolveu uma infraestrutura aberta, distribuída e completamente descentralizada para acesso ao patrimônio cultural (MEGHINI e RISSE, 2005). Os principais requisitos considerados na arquitetura do BRICKS foram: (1) facilidade de expansão da infraestrutura de *software*, (2) permitir a entrada gradual de novos parceiros no projeto, (3) escalabilidade e disponibilidade e (4) interoperabilidade com várias outras soluções presentes no consórcio.

BRICKS apresenta uma arquitetura distribuída baseada numa rede do tipo P2P (*peer-to-peer*) cuja principal característica é a independência de cada nó da rede, bem como a facilidade para incorporação de novos nós. Cada nó é chamado de *BNode* (ou *brick*, tijolo). Cada nó da rede é capaz de oferecer três tipos de serviços:

- Serviço básico necessário em cada nó da rede, responsável pela comunicação com os demais nós, armazenamento, e registro dos serviços.
- Serviços necessários para prover as funcionalidades básicas para usuários locais do nó, que inclui navegação, busca, controle de usuários, autenticação e autorização.

¹⁸ Ver em: <http://www.brickcommunity.org>

- Serviços opcionais necessários somente para as instituições que fornecem conteúdo para a rede BRICKS, incluindo conteúdo, metadados, gerenciamento de anotações, relatórios e recursos de proteção.

BRICKS oferece a possibilidade de recuperação da informação baseada em conteúdo através de serviços que operam dentro da suposição de que os usuários estão interessados em conteúdos similares com base, por exemplo, em anotações submetidas anteriormente. O sistema utiliza RDF como modelo de dados para manipular anotações, permitir interoperabilidade com serviços de outras bibliotecas digitais, além de permitir personalização nos serviços orientados a comunidades.

2.1.7.3 Greenstone

O Greenstone¹⁹ (BAINBRIDGE *et al.*, 2001) apesar de não ter sido desenvolvido numa arquitetura orientada a serviços (como o Fedora e BRICKS), representa o resultado de importantes pesquisas na área de bibliotecas digitais. Um dos objetivos do projeto foi o desenvolvimento de um sistema gerenciador de biblioteca digital que pudesse ser facilmente configurado em tempo de execução (BAINBRIDGE *et al.*, 2004). Outro objetivo era facilitar a construção de novas coleções. Nos cenários vislumbrados pelo projeto estavam questões de compatibilidade com outros sistemas, complemento de informações no momento de exibição, apresentação de uma biblioteca digital baseada em mapas e a organização hierárquica de documentos em agrupamentos. Com o intuito de facilitar a construção de novas coleções, o Greenstone faz uso dos serviços do GATE (CUNNINGHAM *et al.*, 2002) para processamento de linguagem natural.

De forma semelhante ao Fedora, o Greenstone também utiliza o METS (*METS EDITORIAL BOARD*, 2007) para representar os metadados e a estrutura interna dos recursos. Buchanan e outros (2005) apresentam uma arquitetura distribuída para o Greenstone, exemplificando os processos que podem ser executados em paralelo dentro no fluxo de trabalho (*workflow*) relativo à importação e construção de uma coleção.

Parte da pesquisa em torno do Greenstone se dedicou aos serviços de alerta (BUCHANAN e HINZE, 2005). De acordo com estes autores, o gerenciamento e fornecimento de acesso a informação digital é somente parte da missão de uma biblioteca digital. Os usuários devem ser informados sobre novas publicações. Neste sentido, os usuários terão melhores notificações se o sistema utilizado for extensível e configurável. Eles argumentam que os serviços de alerta devem ser integrados na experiência de busca e navegação do usuário, ou seja, de forma que os usuários possam ter informações referentes

¹⁹ Ver em: <http://www.greenstone.org>

a atualizações recentes sobre os resultados das consultas construídas no passado. Buchanan e Hinze (2005) defendem ainda que os serviços de alerta devem ir além das notificações aos usuários. Os autores discutem outros cenários, como notificações sobre textos recomendados, modificações nos documentos, monitoramento da instalação de novas bibliotecas e modificações na classificação de tópicos. Eles observam que a variedade de soluções para bibliotecas digitais não permite uma fácil integração e uma solução que atenda a todas as necessidades. Neste sentido, os desenvolvedores de sistemas gerenciadores de bibliotecas digitais e os editores precisariam interagir no sentido de estabelecer uma base comum. Os autores descrevem ainda alguns tipos de bibliotecas digitais (arquivos, de referência, pequenas e especializadas) e modelos de distribuição (isolado, distribuído, federado e híbrido). Eles discutem problemas relacionados à identidade dos documentos digitais, controle de versões, exclusão, substituição, além do conceito de trabalhos derivados. A solução híbrida apresentada por Buchanan e Hinze (2005) permite a definição de alertas personalizados dentro de um fluxo de quatro fases: construção, observação, filtragem e notificação. Os autores identificam vários tipos de eventos e sugerem tipos de eventos mais ricos que vão muito além das notificações convencionais sobre novas publicações.

2.1.7.4 DSpace

O DSpace²⁰ é um sistema gerenciador de bibliotecas digitais desenvolvido pelo grupo de bibliotecas do MIT²¹ (*Massachusetts Institute of Technology*) em parceria com a empresa *Hewlett Packard* (HP), amplamente reconhecido e utilizado em todo o mundo. Trata-se de uma plataforma de software de código aberto (*open source*) que permite que qualquer instituição possa capturar e descrever conteúdos digitais a partir de procedimentos configuráveis de submissão, aprovação e incorporação destes conteúdos. Além do módulo de submissão de conteúdo, o sistema oferece outras formas alternativas e programáveis de importação de conteúdo. O DSpace permite que os conteúdos sejam distribuídos na web através de módulos de navegação, busca e recuperação. Finalmente, o sistema possui funcionalidades que permitem que os conteúdos sejam preservados numa perspectiva de longo prazo. A seguir será apresentada uma visão geral do DSpace sob o prisma da funcionalidade oferecida pelo sistema, ou seja, uma visão funcional. Baseia-se no manual do DSpace em sua versão 1.5.2 (DSpace FOUNDATION, 2009), bem como na experiência obtida na instalação e configuração do sistema.

²⁰ Ver: <http://www.dspace.org>

²¹ Ver: <http://libraries.mit.edu>

O modelo de dados, ou seja, a forma como os dados são organizados pelo DSpace procura refletir a estrutura da instituição que dele faz uso. Uma instalação do sistema pode ser configurada ou organizada em termos de *comunidades*, que podem ser divididas em *sub-comunidades*, de modo a espelhar a estrutura típica de instituições educacionais ou centros de pesquisa. Cada *comunidade* pode abrigar uma ou mais *coleções*, que representam grupos de conteúdos relacionados. Uma *coleção* pode aparecer em mais de uma *comunidade*. A FIGURA 11 a seguir apresenta o modelo de dados usado pelo DSpace.

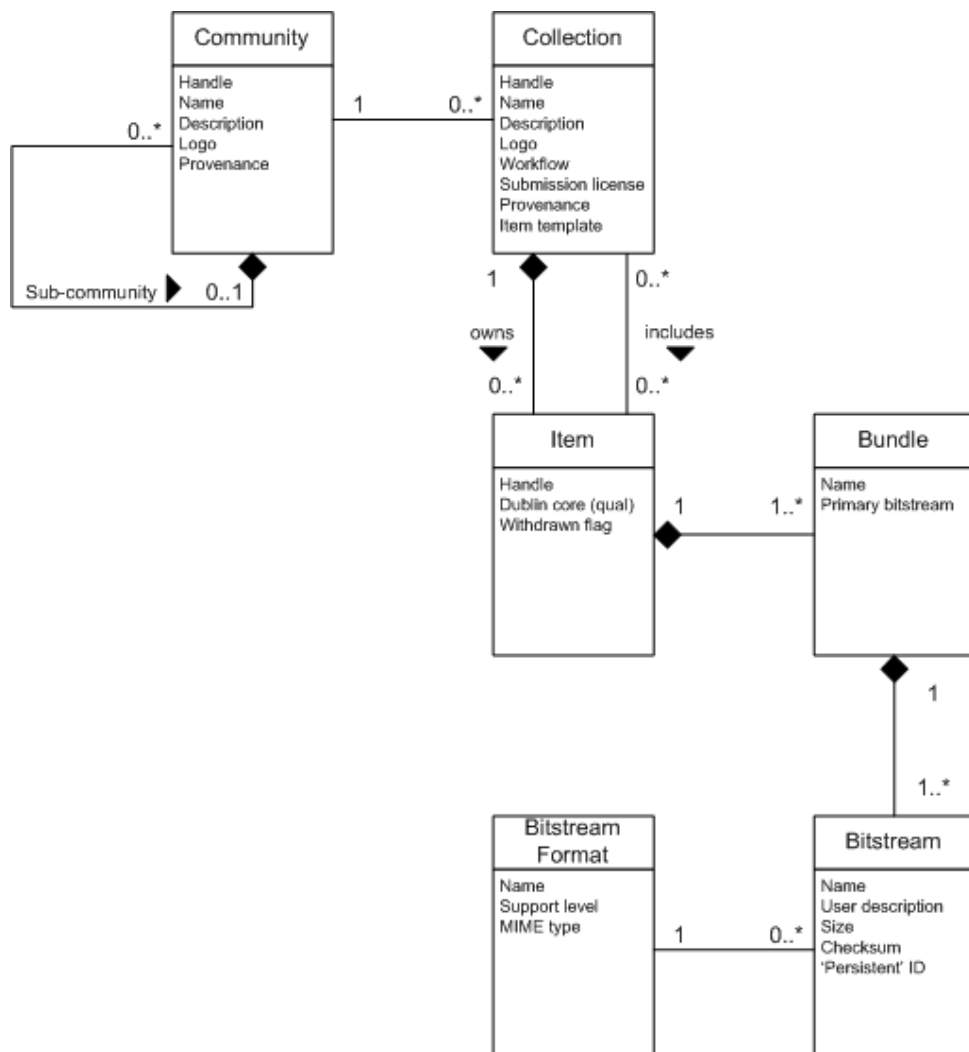


FIGURA 11 - Modelo de dados do DSpace

Fonte: (DSpace FOUNDATION, 2009)

Cada *coleção* é composta de *itens*, que representam os conteúdos armazenados pelo sistema. Um *item* representa uma unidade bibliográfica. Cada *item* é de propriedade de uma única e determinada *coleção*, mas pode aparecer em outras *coleções*. Os *itens* são

compostos de pacotes de arquivos (*bundles of bitstreams*). Estes pacotes podem ser usados para organizar arquivos relacionados, como páginas HTML e imagens associadas que compõem um único documento HTML. Entre os tipos de pacotes mais comuns estão os arquivos originalmente depositados no sistema (ORIGINAL), arquivos contendo o texto completo extraído dos arquivos originais para fins de indexação (TEXT), arquivos contendo licenças concedidas pelo depositário à instituição hospedeira (LICENSE), ou ainda licenças de distribuição concedidas ao usuário final (CC_LICENSE).

A preservação dos conteúdos em longo prazo constitui-se em uma forte preocupação que é contemplada pelo DSpace. Neste sentido, é importante capturar o formato dos arquivos que são submetidos e incorporados ao sistema. A cada arquivo deve estar associado um formato único que descreve implícita ou explicitamente como o seu conteúdo pode ser interpretado e manipulado. Por exemplo, a interpretação para arquivos codificados no padrão JPEG é definido explicitamente pelo padrão ISO/IEC 10918-1. Já a interpretação para arquivos do MS Word 2000 é definido implicitamente pela referência ao aplicativo Word 2000. É importante notar que formatos de arquivos são mais específicos que extensões de arquivos. Por exemplo, a extensão “.doc” pode se referir a variadas versões do MS Word, cada qual com formato próprio.

Cada formato de arquivo deve estar associado a um determinado nível de suporte, que indica a capacidade da instituição hospedeira em preservar conteúdos neste formato para uso futuro. Cada instituição deve determinar o significado exato de cada nível de suporte, após considerar requisitos e custos envolvidos. O QUADRO 2 a seguir mostra os níveis de suporte adotados pelas bibliotecas do MIT.

QUADRO 2 - Exemplos de níveis de suporte a preservação

Suportado (<i>supported</i>)	O formato é reconhecido, e a instituição hospedeira é capaz de garantir que os arquivos neste formato poderão ser manipulados no futuro, usando qualquer combinação de técnicas (como migração, emulação, etc.) que sejam apropriadas ao contexto de uso.
Conhecido (<i>know</i>)	O formato é reconhecido, e a instituição hospedeira se compromete a preservar o arquivo no formato atual (<i>as-is</i>), e permitir que o mesmo seja recuperado no futuro. A instituição hospedeira irá tentar possibilitar que o formato seja elevado ao nível mais alto de suporte: “suportado”.
Não suportado (<i>unsupported</i>)	O formato não é conhecido, mas a instituição hospedeira irá preservar o arquivo no formato atual (<i>as-is</i>), assim como permitir que o mesmo possa ser recuperado.

Fonte: (DSpace FOUNDATION, 2009)

Cada item armazenado no DSpace recebe um registro de metadados Dublin Core qualificado (ver seção 2.1.6.2). Outros metadados podem ser armazenados em um

item através de arquivos anexados. Os metadados Dublin Core são armazenados num formato relacional de banco de dados, visando interoperabilidade e facilidade de busca. Os metadados Dublin Core podem ser fornecidos pelos usuários como parte do processo de submissão de conteúdo ou podem ser derivados de outros metadados como parte de um processo de importação automática.

Os itens podem ser removidos do DSpace de duas maneiras diferentes: 1) eles podem ser desabilitados (*withdrawn*) de modo a permanecer invisíveis no sistema, assim, no momento do acesso será informado ao usuário que o item fora removido; ou 2) eles podem ser completamente removidos (*expunged*), de forma que todos os seus registros sejam completa e definitivamente removidos do sistema. A tabela a seguir ilustra alguns exemplos de objetos presentes no modelo de dados do DSpace.

QUADRO 3 - Exemplos de objetos no modelo de dados do DSpace

Objeto	Exemplos
Comunidade (<i>community</i>)	Escola de Ciência da Informação; Programa de Pós Graduação em Ciência da Informação (PPGCI).
Coleção (<i>collection</i>)	Teses e Dissertações do PPGCI; Revista Perspectiva em Ciência da Informação.
Item	Uma tese de doutorado; um relatório técnico; um artigo de uma revista; uma gravação em vídeo de uma aula; um conjunto de dados acompanhado de descrições.
Pacote (<i>bundle</i>)	Um conjunto de arquivos HTML e de imagens que formam um documento.
Arquivo (<i>bitstream</i>)	Um arquivo HTML; um arquivo PDF de uma dissertação; um arquivo de uma imagem.
Formato de arquivo (<i>bitstream format</i>)	Microsoft Word versão 6.0; formato JPEG para imagem.

Os metadados utilizados pelo DSpace podem ser agrupados em três categorias. A primeira categoria representa os *metadados descritivos* que se prestam a descrever cada item armazenado. O DSpace usa um esquema baseado no padrão Dublin Core qualificado inspirado na *Library Application Profile*²². O Sistema vem pré-configurado com um conjunto de elementos e qualificadores do Dublin Core usado pelas bibliotecas do MIT. No entanto, é possível configurar múltiplos esquemas de metadados, assim como selecionar campos destes esquemas para descrever os itens armazenados. Outros metadados descritivos

²² Ver em: <http://www.dublincore.org/documents/library-application-profile>

podem ser mantidos em arquivos anexados a cada item. Objetos do tipo *Comunidades* e *Coleções* possuem alguns metadados descritivos que são mantidos no banco de dados em formato relacional.

A segunda categoria representa os *metadados administrativos*, que incluem dados sobre preservação, procedência e política de autorização e acesso. A maioria destes metadados é armazenada no esquema relacional do banco de dados do sistema. Metadados de procedência (*provenance*) são armazenados nos registros reservados ao Dublin Core. Outros metadados administrativos, como tamanho e formato de arquivos, são replicados nos registros do Dublin Core de modo a ficarem acessíveis para outros sistemas visando a interoperabilidade.

A terceira e última categoria representa os *metadados estruturais*, que fornecem informação sobre como apresentar um item. A título de exemplo pode-se considerar o caso de uma tese digitalizada como uma sequência de imagens no formato TIFF, cada imagem representando uma página da tese. Os metadados estruturais devem representar o fato de que cada imagem é uma página, bem como a sequência de ordenação das imagens/páginas. Estes metadados são representados no banco de dados do sistema através de pacotes de arquivos que compõem um item. Um pacote pode, opcionalmente, possuir um arquivo principal (*primary bitstream*). A cada arquivo também é atribuído um número sequencial (*sequence ID*) que o identifica e estabelece a ordenação dentro do item.

Usuários (chamados de *e-people* no DSpace) e grupos representam as formas usadas pelo DSpace para identificar seus usuários e atribuir permissões aos mesmos. Apesar de muitas das funcionalidades do DSpace, relacionadas à descoberta e recuperação de documentos poder ser usada de forma anônima, algumas funcionalidades (e mesmo o acesso a alguns documentos) só são permitidos a determinados usuários e grupos. Os usuários finais podem se auto-registrar no sistema fornecendo seus dados para a criação da sua conta de usuário. Esta conta pode ser usada para fins de personalização como, por exemplo, pela seleção de uma lista de coleções para as quais o usuário gostaria de ser notificado via *e-mail* sobre novos itens recém incorporados à coleção. Outra possibilidade é de os usuários serem cadastrados pelo administrador do sistema, com a atribuição de permissões especiais como, por exemplo, submeter novos itens a determinadas coleções, ou mesmo participar do processo (*workflow*) de aprovação de novos itens.

Os grupos são outro tipo de entidade presente no DSpace. Um grupo pode ser uma lista explícita de usuários que recebem determinadas permissões em conjunto. Outra possibilidade é a de um grupo ser formado de acordo com a origem do acesso no contexto de uma rede de computadores. Por exemplo, uma sessão de uso do sistema pode ser atribuída ao grupo "Usuários Locais" se o acesso partir de um computador conectado a rede

local, dando a este usuário acesso a materiais restritos que não estão acessíveis a usuários externos. Outro uso comum dos grupos é como papéis ou perfis para determinadas classes de usuários que possuem um mesmo conjunto de permissões como, por exemplo, participantes do processo de submissão ou aprovação de novos itens em uma determinada coleção.

O DSpace possui um sistema de autorização baseado na associação de determinadas ações a objetos e usuários (ou grupos) que podem executar estas ações. Estas associações são chamadas políticas (*resource policies*). Existem dois grupos especiais: “*Administrators*”, que podem executar qualquer ação no Sistema e “*Anonymous*”, que contem todos os usuários do sistema. Ao atribuir uma política de acesso a um determinado objeto para o grupo *Anonymous* significa que qualquer usuário poderá executar esta ação. O QUADRO 4, a seguir, relaciona, para cada tipo de objeto, as ações que podem ser atribuídas a usuários ou grupos.

QUADRO 4 - Objetos e ações no sistema de autorização do DSpace

Objeto	Ação	Descrição
Comunidade	ADD/REMOVE	Adicionar ou remover coleções ou sub-comunidades.
Coleção	ADD/REMOVE	Adicionar (submeter novos itens) ou remover itens.
	DEFAULT_ITEM_READ	Fornece permissão padrão de leitura (READ) a todos os novos itens submetidos à coleção.
	DEFAULT_BITSTREAM_READ	Fornece permissão padrão de leitura (READ) a todos os novos arquivos submetidos à coleção.
	COLLECTION_ADMIN	Permite a edição de itens dentro da coleção, além da remoção (<i>withdraw</i>) de itens, bem como a vinculação de itens dentro da coleção.
Item	ADD/REMOVE	Adicionar ou remover pacotes (de arquivos).
	READ	Visualizar o item.
	WRITE	Modificar o item.
Pacote (de arquivos)	ADD/REMOVE	Adicionar ou remover arquivos do pacote.
Arquivo	READ	Visualizar o arquivo.
	WRITE	Modificar o arquivo.

Fonte: (DSpace FOUNDATION, 2009)

Além dos grupos especiais *Administrators* e *Anonymous*, o DSpace permite a criação de novos grupos para os quais as políticas de acesso descritas acima poderão ser atribuídas.

O processo de ingestão representa a maneira como novos conteúdos ingressam no DSpace. A FIGURA 12, a seguir, ilustra as maneiras como novos itens podem ser incluídos na base de dados do DSpace.

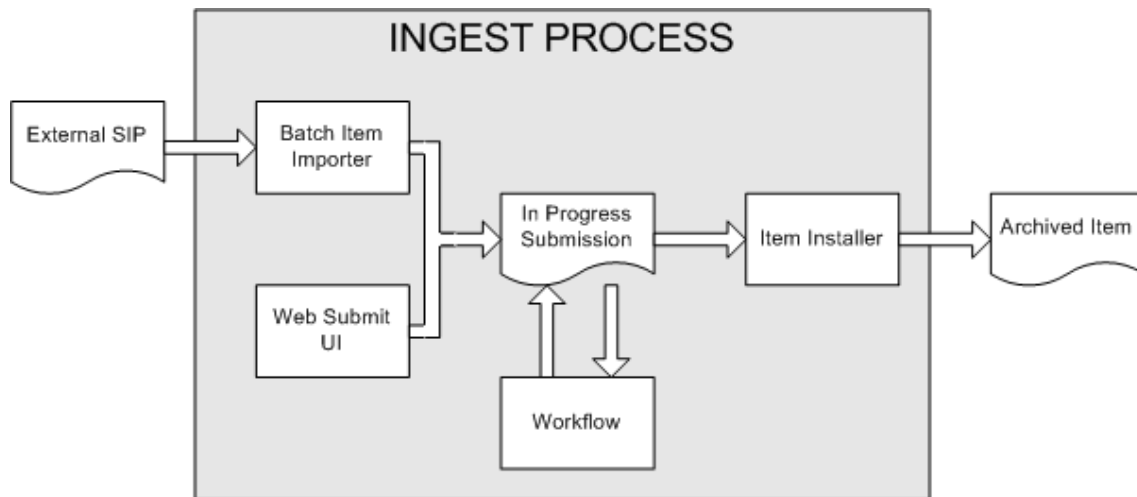


FIGURA 12 - O processo de ingresso de conteúdos no DSpace

Fonte: (DSpace FOUNDATION, 2009)

Existem duas maneiras principais para o ingresso de conteúdo: 1) via processo de importação (*Batch Item Importer*) que transforma um arquivo externo, composto de arquivos de conteúdo e metadados, em um objeto em processo de submissão (*In Progress Submission*) no DSpace; ou 2) via interface web de submissão (*Web Submit UI*), que permite que determinados usuários ou grupos possam submeter novos itens através da interface do sistema.

Dependendo da política da coleção para a qual um novo item esta sendo submetido, um fluxo (*workflow*) de submissão pode ser iniciado. Tal fluxo permite que determinados usuários possam atuar como revisores e aprovem a inclusão do novo item na coleção. Finalmente, na etapa final deste fluxo, o item é instalado e fica disponível para acesso na coleção. No decorrer das etapas que compõem o fluxo de ingestão, vários campos de metadados administrativos são automaticamente gerados, como procedência, datas de submissão e disponibilização, identificadores, política de autorização, índices, etc.

O fluxo de submissão (*workflow*), quando definido para uma determinada coleção, pode incluir até três etapas. A coleção deve possuir um grupo de usuários responsáveis por

executar cada uma das etapas; no caso de não haver um grupo associado a uma dada etapa, a mesma é “pulada”. Se uma coleção não possui nenhum grupo de usuários associado às etapas, então não haverá fluxo de submissão (*workflow*) e os novos itens serão encaminhados diretamente para a instalação e disponibilização na coleção. A FIGURA 13, a seguir, ilustra as três etapas possíveis do fluxo de submissão.

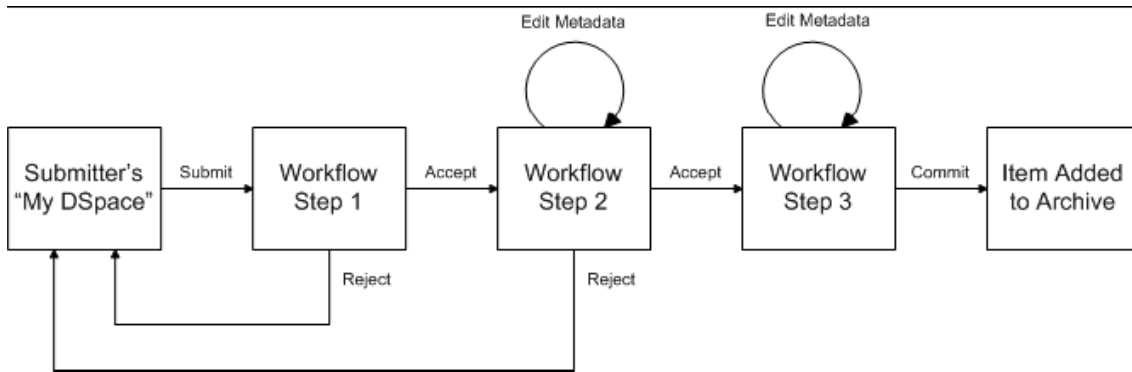


FIGURA 13 - Etapas do fluxo de submissão de conteúdo no DSpace

Fonte: (DSpace FOUNDATION, 2009)

Quando uma etapa é invocada para um novo item recém submetido, a tarefa de executar o respectivo passo do *workflow* é colocada na pasta do grupo de usuários associado. Qualquer usuário membro do grupo poderá então retirar a tarefa para si, e executá-la posteriormente. O usuário poderá executar, de acordo com a etapa, uma das ações apresentadas no QUADRO 5 a seguir.

QUADRO 5 - Ações possíveis para cada etapa do processo de submissão no DSpace

Etapa do workflow	Ações Possíveis
Etapa 1	Pode aceitar o novo item para inclusão, ou rejeitar.
Etapa 2	Pode editar os metadados fornecidos pelo usuário responsável pela submissão do item, mas não pode alterar os arquivos de conteúdo submetidos. Pode aceitar ou rejeitar o novo item para inclusão.
Etapa 3	Pode editar os metadados fornecidos pelo usuário responsável pela submissão do item, mas não pode alterar os arquivos de conteúdo submetidos. Somente poderá aceitar o novo item, não podendo rejeitar a submissão.

Fonte: (DSpace FOUNDATION, 2009)

No caso de o novo item ser rejeitado em qualquer uma das etapas, o motivo (fornecido pelo responsável pela etapa) é enviado via *e-mail* para o usuário que submeteu o

item. O item é retornado para a área restrita deste usuário no DSpace (*My DSpace*), que poderá realizar modificações e submeter novamente o item. O DSpace fornece algumas funcionalidades especiais para tratar os casos em que autores de teses ou dissertações são supervisionados por seus orientadores na preparação da versão digital do seu trabalho (*e-thesis*). Estas mesmas funcionalidades podem ser usadas em situações em que pesquisadores desejam colaborar no processo de submissão de um item.

Um dos problemas encontrados quando se disponibiliza conteúdos digitais na web é que a referência a estes conteúdos (endereços ou *links*) pode mudar, ou simplesmente deixar de funcionar na medida em que as instituições reconfiguram seus servidores e *sites*. Por outro lado, os pesquisadores necessitam de uma referência estável para seus trabalhos. Para resolver este problema, o DSpace permite a criação de identificadores persistentes para cada item, coleção, ou comunidade armazenada no sistema. Para tanto, é necessário um mecanismo, independente de armazenamento e localização, para a criação e manutenção destes identificadores. O DSpace utiliza o *CNRI Handle System*²³ para a criação dos identificadores. O DSpace usa *handles* como uma forma de atribuir identificadores globais e únicos aos seus objetos. Para tanto, cada instalação do DSpace necessita obter da CNRI um prefixo de *handle*, de forma que se possa garantir que ao criar identificadores com tal prefixo eles não irão coincidir com identificadores criados por outras instituições. Estes identificadores são criados para itens, coleções e comunidades armazenadas no DSpace.

O esquema de *handles* fornece uma infraestrutura global para resolução de referências. Um usuário final pode fornecer um *handle* a qualquer servidor registrado no serviço *CNRI Handle System*. De posse do *handle*, tal servidor será capaz de direcionar o usuário para o objeto (item, coleção ou comunidade) identificado pelo mesmo. Para usar o sistema de *handles*, o DSpace precisa fornecer o serviço de resolução (*handle server*) que possa aceitar e resolver solicitações. Esta funcionalidade já se encontra implementada no código fonte do DSpace.

Os *handles* podem ser escritos de duas formas:

(1) `hdl:1721.123/4567`

(2) `http://hdl.handle.net/1721.123/4567`

As duas formas acima representam o mesmo identificador (*handle*). A primeira pode ser usada somente como identificador. Já a segunda forma permite que

²³ Ver em: <http://www.handle.net>

qualquer navegador web possa resolver e acessar o objeto representado pelo *handle*. É importante destacar que o DSpace utiliza a infraestrutura de *handles* somente em nível local. No exemplo acima, a instalação do DSpace recebeu o prefixo “1721.123”. É responsabilidade do DSpace manter a associação entre o *handle* completo (incluindo a terminação “4567”) e o objeto presente na sua base de dados, seja ele um item, coleção ou comunidade.

Ao acessar o endereço “<http://hdl.handle.net/1721.123/4567>” o navegador será redirecionado para o objeto “4567” definido na instalação “17213.123” do DSpace, independente de qual seja o endereço atual usado por esta instalação. Desta forma, mudanças no servidor, ou mesmo no *site* da instituição que abriga o DSpace não afetarão o acesso aos objetos armazenados.

O DSpace fornece aos usuários finais funcionalidades de busca e navegação que permitem a descoberta de conteúdo de variadas formas: 1) via referência externa através de um *handle*; 2) busca por palavra chave nos campos de metadados ou no texto completo; 3) navegação pelos índices de título, autor, data, e assuntos. O escopo das buscas e navegação pode ser todo o conteúdo armazenado, ou pode ser limitado por comunidade ou coleção. O DSpace²⁴ usa o mecanismo de busca Lucene²⁴, que permite que os campos de metadados a serem usados no processo de indexação e navegação possam ser configurados. Além de buscas, o DSpace permite a navegação pelos índices de título, autor, data e assunto. Outros campos de metadados podem ser configurados para busca e navegação.

O DSpace disponibiliza os campos de metadados Dublin Core dos itens que se encontram publicamente disponíveis de modo que outros *sites* possam coletar estes metadados. Além disso, os objetos do tipo coleção também são disponibilizados para coleta. O DSpace faz uso do software de código aberto OAICat²⁵ para prover estas funcionalidades.

Do exposto acima se pode observar que o DSpace é um sistema gerenciador de bibliotecas digitais amplamente configurável e que contempla todas as etapas, desde a ingestão de conteúdo, seja ela realizada por processos programáveis de importação ou via interface do sistema, até mecanismos de descoberta de conteúdo (busca e navegação) e disponibilização de metadados para outros sistemas. Todos os processos abarcados pelo DSpace são amplamente flexíveis e passíveis de serem modificados via parâmetros de configuração. Além disso, a arquitetura aberta do DSpace possibilita a modificação e a adição de novas funcionalidades.

²⁴ Ver em: <http://jakarta.apache.org/lucene>

²⁵ Ver em: <http://www.oclc.org/research/software/oai/cat.shtm>

O DSpace, no entanto, não oferece nenhum tipo de serviço ou funcionalidade que explore a representação e organização semântica do acervo e, conseqüentemente, não há mecanismos de busca e acesso que tirem proveito desta organização do acervo. O grupo de bibliotecas do MIT, responsável pelo projeto e desenvolvimento do DSpace, reconhece a importância das tecnologias semânticas em uma biblioteca digital, e nesse sentido o DSpace pode ser estendido com componentes semânticos desenvolvidos no contexto do projeto SIMILE ²⁶ (*Semantic Interoperability of Metadata and Information in unLike Environments*). Estes componentes também podem ser integrados a outros tipos de serviços, e receberam grande reconhecimento dos usuários, desenvolvedores web, e da comunidade científica no contexto da web semântica (HUYNH *et al.*, 2007).

O grupo responsável pela revisão e evolução do DSpace apresentou um relatório (OCKERBLOOM, 2007) com propostas de mudanças na arquitetura do sistema com o objetivo de atender aos requisitos e aproveitar as oportunidades das tecnologias atuais da Internet. O relatório descreve a necessidade de separação de um núcleo do sistema que seja estável e independente de aplicações específicas. Também é destacada a importância da versão atual do sistema em cenários de uso comuns.

2.2 Organização do conhecimento e a Teoria da Classificação Facetada

Observa-se na literatura certa divergência na utilização de conceitos como informação, conhecimento e organização de informação/conhecimento. Neste sentido parece-nos pertinente apresentar a definição de Anderson (2003) para organização do conhecimento, que exemplifica bem esta dificuldade em relação a tais conceitos:

... a descrição (indexação) e organização (classificação) visando a recuperação de mensagens representando conhecimento, textos através dos quais o conhecimento é codificado e documentos nos quais estes textos são embutidos. O conhecimento propriamente dito reside na mente e cérebro de criaturas vivas. A sua organização para recuperação através da memória de curto e de longo prazo é tópico principal das ciências cognitivas. A ciência da informação e a biblioteconomia lidam com a descrição e organização de artefatos (mensagens, textos, documentos) pelos quais o conhecimento (incluindo sentimentos, emoções, desejos) é representado e compartilhado. Estas fontes de conhecimento são também, muitas vezes, chamadas recursos de informação. Desta forma, “organização do conhecimento” no contexto da ciência da informação e biblioteconomia é o mesmo que “organização de recursos de conhecimento”, muitas vezes chamado também de “organização da informação”. (ANDERSON, 2003, tradução nossa)

²⁶ Ver em: <http://simile.mit.edu>

Contudo, Bräscher e Café (2008) analisaram as ambiguidades no emprego das seguintes expressões no contexto da Ciência da Informação: organização do conhecimento (OC); organização da informação (OI); representação do conhecimento (RC); e representação da informação (RI). Segundo as autoras, “no contexto da OI e da RI, temos como objeto os registros de informação. Estamos, portanto, *no mundo dos objetos físicos*, distinto do mundo da cognição, ou das ideias, cuja unidade elementar é o conceito” (BRÄSCHER; CAFÉ, 2008, p. 5). As autoras apresentam uma proposta de distinção entre os dois processos de organização:

[...] um que se aplica às ocorrências individuais de objetos informacionais – o processo de organização da informação – e outro que se aplica às unidades do pensamento (conceitos) – o processo de organização do conhecimento. A OI compreende, também, a organização de um conjunto de objetos informacionais para arranjá-los sistematicamente em coleções, neste caso, temos a organização da informação em bibliotecas, museus, arquivos, tanto tradicionais quanto eletrônicos. A organização do conhecimento, por sua vez, visa à construção de modelos de mundo que se constituem em abstrações da realidade. (BRÄSCHER; CAFÉ, 2008, p. 6).

O termo *knowledge organization systems* (KOS) é utilizado por Hodge (2000) para englobar dicionários, glossários, arquivos de autoridade, taxonomias, sistemas de classificação, cabeçalhos de assunto, tesouros, redes semânticas e ontologias. Bräscher e Café (2008) ressaltam que estes sistemas de organização de conhecimento são “o coração de toda biblioteca, museu e arquivo”, uma vez que são “mecanismos de organização da informação” (BRÄSCHER; CAFÉ, 2008, p.8).

A representação, seja ela de informação ou conhecimento, é condição necessária para a organização, na medida em que se torna mais fácil manipular as representações do que os objetos representados. A organização da informação ou conhecimento compreende um processo de representação e destina-se prioritariamente à recuperação eficaz por parte dos usuários. Alvarenga (2003) estabelece uma distinção entre representação primária e secundária. Segundo a autora, a representação em nível primário é feita pelos autores no momento da “expressão dos resultados de seus pensamentos, estes derivados de observações metódicas da natureza e dos fatos sociais, utilizando-se das linguagens disponíveis no contexto da produção e comunicação de conhecimentos”. Já a representação secundária, segundo a mesma autora, ocorre no momento em que os “registros de conhecimentos constantes de documentos passam a integrar acervos de arquivos, bibliotecas, serviços ou centros de documentação/informações”. A representação secundária tem como foco a posterior recuperação destes documentos.

O objetivo do processo de organização da informação é possibilitar e facilitar o acesso à informação que, por sua vez, tem a competência e a intenção de produzir conhecimento. No entanto, neste processo de organização da informação, faz-se uso de mecanismos (ou sistemas) de organização do conhecimento. Como afirma Vickery (2010), organizar o conhecimento é reunir o que conhecemos em uma estrutura sistematicamente organizada. Entendemos que a organização do conhecimento é condição necessária para a organização da informação, ou mais especificamente dos recursos informacionais, sob o aspecto do conteúdo dos mesmos. A organização do conhecimento produz representações sistematizadas de conhecimento, que podem ser utilizadas na organização da informação para padronizar as representações dos conteúdos dos recursos informacionais, visando facilitar a recuperação e o uso dos mesmos.

Nesta pesquisa estamos interessados em ambos os processos, organização do conhecimento e da informação. Estamos interessados em analisar e determinar o valor da adoção de um sistema de organização do conhecimento, estruturado na forma de uma classificação facetada, para a organização da informação em uma biblioteca digital de teses e dissertações.

Bibliotecas digitais, assim como as convencionais, fazem uso de diversos mecanismos e estruturas a fim de imprimir certo grau de organização em seu acervo visando facilitar a descoberta e recuperação da informação. A organização do conhecimento pressupõe alguma forma de representação. Esquemas de classificação possuem características que possibilitam a representação de entidades e relacionamentos em estruturas que refletem o conhecimento do domínio sendo classificado. Kwasnik (1999) ao analisar a relação entre as classificações e a representação do conhecimento afirma que na medida em que os conceitos se aglutinam e os relacionamentos entre os mesmos são entendidos, um esquema de classificação pode ser usado como uma rica representação do que é conhecido e, desta forma, ser útil na comunicação e na geração de ciclos sucessivos de exploração, comparação e teorização. A autora conclui que um bom esquema de classificação funciona da mesma forma que uma teoria, conectando conceitos dentro de uma estrutura útil. Uma classificação bibliográfica é um sistema de codificação e organização de materiais bibliográficos de acordo com seus assuntos. Em uma biblioteca convencional, uma classificação consiste geralmente de tabelas de assunto ou estruturas classificatórias utilizadas na atribuição de um número de classe (notação) a cada item do acervo, com base no seu assunto. Existem diversos tipos de estruturas classificatórias. Kwasnick (1999) identifica quatro tipos de estruturas classificatórias: hierarquias, árvores, paradigmas e facetas. Cada tipo de estrutura baseia-se em uma determinada abordagem estrutural e em pressupostos filosóficos visando determinar o conjunto dos conceitos

tratados pelos documentos de uma coleção. Sob o ponto de vista estrutural, nos esquemas de classificação enumerativos, cada assunto é subdividido até que todas as possibilidades (conhecidas) sejam previstas. A cada subdivisão ou classe é associado um número identificador (notação). Numa classificação hierárquica, cada classe é subdividida e as classes resultantes são ordenadas a partir do geral para o específico. Já uma classificação facetada apresenta uma abordagem alternativa, na medida em que divide os assuntos em categorias ortogonais e mutuamente exclusivas a partir do uso da técnica da análise facetada. Segundo Vickery (1960), independente do tipo de estrutura, todas as classificações possuem os mesmos propósitos: (1) normalizar a linguagem dos documentos por um lado e a linguagem das consultas por outro (no contexto de um sistema de recuperação da informação); (2) servir como um mecanismo útil ao indexador na tarefa intelectual de caracterizar o conteúdo temático ou assunto de um documento; (3) atuar como uma ferramenta do usuário na tarefa de analisar e definir suas estratégias de busca ao acervo.

Do ponto de vista filosófico, uma classificação representa uma determinada forma de ver o mundo. Neste sentido, Louie e outros (2003) identificam quatro abordagens para a criação de um sistema de classificação: (1) Racional, que utiliza a lógica; (2) Empírica, que utiliza como base observações no mundo real; (3) Cultural, que se apoia naquilo em todos concordam; e (4) Contextual, que utiliza aquilo que funciona melhor para cada situação. Cada abordagem está ligada a uma escola do pensamento filosófico (HJORLAND; ALBRECHTSEN, 1999) e como sugere a quarta abordagem (Contextual), cada uma possui vantagens e desvantagens. Os esquemas classificatórios facetados são baseados nesta abordagem Contextual (LOUIE *et al.*, 2003).

Nesta seção será apresentada a Teoria da Classificação Faceta e discutida a sua aplicação como mecanismo para organização do conhecimento em uma biblioteca digital. Cabe ainda destacar a importância da classificação facetada no contexto do presente trabalho de pesquisa, visto que é utilizada como base teórica para concepção e desenvolvimento do mesmo.

As classificações facetadas representam uma abordagem alternativa para o processo classificatório através da análise facetada. O processo utilizado na análise facetada tem suas origens em Shiyali Ramamrita Ranganathan (GARFIELD, 1984), um pesquisador indiano que propunha que qualquer entidade complexa poderia ser representada a partir de certo número de perspectivas ou facetas. Para Ranganathan, o problema dos sistemas de classificação tradicionais, como a Classificação Decimal de Dewey, é que as classes ou os termos usados precisam ser elaborados antes dos objetos serem classificados, além disso, estes sistemas não se mostravam adequados para a

representação de assuntos compostos. Com o grande desenvolvimento científico e proliferação da informação observada no século XX, estes sistemas enumerativos ou pré-coordenados não eram adequados. Ao invés de criar compartimentos para depois colocar os objetos dentro dos mesmos, Ranganathan buscava uma forma de partir do objeto e então coletar e arranjar aspectos (ou facetas) que pudessem descrever aquele objeto. Esta abordagem possibilita maior flexibilidade e um alto grau de especificidade no sistema de classificação. Ao longo dos anos as ideias de Ranganathan foram reinterpretadas, desenvolvidas e utilizadas em diversos contextos envolvendo representação, organização e recuperação da informação.

Kwasnick (1999) destaca várias vantagens das estruturas classificatórias facetadas: elas não requerem conhecimento completo sobre as entidades e seus relacionamentos; elas são hospitaleiras, na medida em que podem acomodar facilmente novas entidades; elas são flexíveis; elas são expressivas; elas podem ser criadas para finalidades específicas e de forma livre; e elas permitem variadas perspectivas e abordagens sobre os objetos a serem classificados. A autora também aponta três problemas principais relacionados às estruturas facetadas: a dificuldade em se escolher as facetas corretas; a falta de mecanismos para expressar relacionamentos entre as facetas; e a dificuldade de visualização. A escolha das facetas corretas é fundamental para o sucesso do sistema de classificação e requer conhecimento sobre os itens a serem classificados assim como sobre as necessidades dos usuários, ou seja, as formas pelas quais eles buscam a informação. Isto é verdadeiro para qualquer tipo de estrutura classificatória. A ausência de mecanismos para representar relacionamentos entre facetas é de fato um problema, mas que pode ser amenizado quando os usuários possuem um nível de conhecimento sobre os itens que lhes permita inferir tais relacionamentos por conta própria. Nas abordagens mais recentes para aplicação de estruturas facetadas em ambientes digitais, os relacionamentos entre facetas podem ser facilmente visualizados por evidência empírica, através do uso de estruturas dinâmicas (SACCO E TIZITZICAS, 2009). Com relação à dificuldade de visualização, Kwasnick (1999, 42) destaca que “a tecnologia da informação promete trazer novas formas para possibilitar a visualização multidimensional”, fato que já pode ser comprovado pelas interfaces avançadas disponíveis nos dias atuais.

O entendimento do que vem a ser facetas é bastante facilitado com a apresentação de um caso de uso prático, disponível na Internet²⁷. No caso dos vinhos, cada um possui uma determinada cor, tem origem em um dado local onde é produzido a partir de um tipo específico de variedade (ou combinação de variedades) de uva. Além disso,

²⁷ Ver em: <http://www.wine.com>

também se conhece o ano de produção, a qualidade e denominação do vinho, certificada por alguma autoridade no assunto. O vinho vem ainda engarrafado com certo volume e tem um preço de comercialização. A partir do conceito de facetas, se pode organizar um conjunto de categorias que podem ser combinadas para descrever os vinhos: cor, origem, variedade de uva, ano de produção, denominação, volume, preço, entre outras. Cada categoria representa uma faceta e deve ser povoada com os valores possíveis (termos). O conjunto de categorias (ou facetas) deve ser organizado de forma apropriada. Desta forma, cada garrafa de vinho pode ser classificada a partir da escolha dos termos corretos em cada categoria. Com isso tem-se um esquema de classificação facetado, que nada mais é do que um conjunto mutuamente exclusivo e exaustivo de categorias combináveis, onde cada categoria é obtida a partir do isolamento de uma perspectiva (ou faceta) dos itens a serem classificados, que combinadas descrevem todos os itens em questão e que os usuários podem utilizar, no processo de busca e navegação, de modo a encontrar aquilo que procuram.

É importante deixar claro que o exemplo apresentado acima, apesar de ser elucidativo na medida em que facilita o entendimento dos principais conceitos da classificação facetada, é extremamente limitado, pois apresenta apenas uma faceta relacionada à categoria fundamental *Entidade* (vinho), ou *Personalidade* na visão de Ranganathan. A estrutura resultante se assemelha mais a uma taxonomia do que uma classificação, já que não há nenhum conceito externo à faceta principal e a organização é feita em sequências (*arrays*) dentro de uma única faceta, na medida em que se considera cada atributo da entidade vinho. Uma classificação bibliográfica facetada certamente será bem mais complexa, usando muitas outras facetas, cobrindo uma grande quantidade de conceitos.

Desde os primórdios da classificação e indexação, como aponta Broughton (2006), pode-se observar a recorrência regular de atributos comuns como *lugar*, *tempo* e *forma*. Uma das principais contribuições de Ranganathan foi constatar que não só existiam conceitos recorrentes em praticamente todas as áreas, como também existiam tipos recorrentes de conceitos dentro destas áreas (BROUGHTON, 2006). Alguns representavam atividades ou ações, que Ranganathan chamou de *Energia*; outros conceitos eram relativos a substâncias, ou seja, *Matéria*; os conceitos nucleares que representam os principais objetos de estudo de uma disciplina, representando a sua essência, foram agrupados na categoria *Personalidade* por Ranganathan. Os membros desta categoria são na maioria das vezes entidades como plantas, animais, componentes químicos, corpos celestes, formações geográficas, religiões, objetos manufaturados, etc. Acrescentando as categorias comuns *Espaço* e *Tempo*, chegamos às cinco categorias fundamentais propostas por Ranganathan

e conhecidas pela sigla PMEST (*Personality, Matter, Energy, Space e Time*). Dentro de uma disciplina ou campo de estudo, todos os conceitos ou termos poderiam ser organizados nessas cinco categorias. Ranganathan argumentava que qualquer objeto (ou qualquer conceito ou assunto de um objeto de informação) poderia ser representado através da seleção de elementos destas cinco categorias. Por exemplo, se considerarmos um livro sobre “o projeto de casas de madeira no sul do Brasil no início do século XX”, teríamos as seguintes facetas:

- Personalidade (*Personality*) – casas
- Matéria (*Matter*) – madeira
- Energia (*Energy*) – projeto
- Espaço (*Space*) – sul do Brasil
- Tempo (*Time*) – início do século XX

O exemplo acima é intencionalmente simples. Todas as cinco categorias foram utilizadas, sendo que cada uma foi utilizada uma única vez. Nem todas as categorias precisam ser necessariamente utilizadas e uma mesma categoria pode ser utilizada mais de uma vez para descrever adequadamente um objeto.

O *Classification Research Group* (CRG), grupo de pesquisadores que se estabeleceu na Inglaterra a partir de 1952, com o objetivo de realizar estudos sobre a natureza do processo classificatório e das classificações existentes (SPITERI, 1998), considerou estas categorias limitadas para representar algumas disciplinas e expandiu as categorias fundamentais em 13: *thing, kind, part, property, material, process, operation, agent, patient, product, by-product, space, e time*. Estas categorias poderiam acomodar os conceitos da maioria das disciplinas, apesar de que disciplinas como arte e humanidades requerem categorias adicionais (como *forma, estilo, gênero*) e não há nada que impeça que novas categorias fundamentais possam ser definidas ou inventadas (BROUGHTON, 2006).

A aplicação das categorias fundamentais representa a base da classificação facetada, não importando que sejam as categorias propostas por Ranganathan, as sugeridas pelo CRG, ou ainda categorias criadas especificamente para um determinado domínio ou sistema de classificação. A aplicação de cada categoria como um amplo princípio de divisão dentro de uma disciplina gera um conjunto de conceitos, ou facetas. No exemplo a seguir, adaptado a partir de Broughton (2006), algumas categorias propostas pelo CRG foram aplicadas ao campo da medicina de modo a criar as facetas.

(Thing) Seres humanos

(Kind) Mulher, criança, idoso, etc.

(Part) Cabeça, pernas, músculos, coração, cérebro, pulmão, etc.

(Process) Respiração, digestão, reprodução, doença, etc.

(Operation) Cirurgia, tratamento com medicamento, fisioterapia, etc.

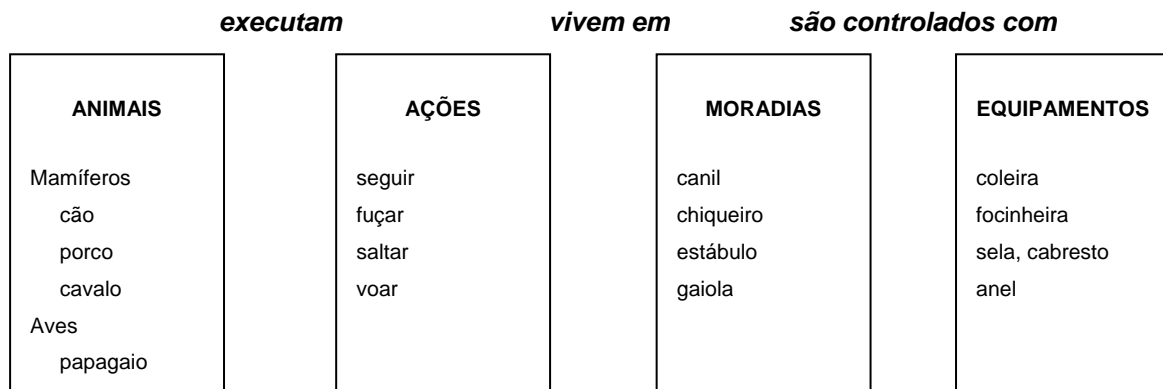
(Agent) Médicos, enfermeiras, equipamento, construção, etc.

Dentro de cada faceta será ainda necessário organizar os conceitos em arranjos ou sub-facetadas com base em seus atributos. A ordem dos termos dentro de cada arranjo deve ser determinada por algum critério prático.

Broughton (2006) alerta para o fato de que a classificação facetada tem como universo de discurso uma disciplina ou campo de estudo específico. Isto significa que antes de se aplicar a análise facetada, deve-se inicialmente selecionar uma disciplina ou campo de estudo. Desta forma, uma classificação facetada geral teria a forma de uma classificação inicial em disciplinas, sendo que, em seguida, cada disciplina teria sua estrutura facetada própria. A faceta inicial não é “disciplina”. Apesar de haver uma divisão inicial em disciplinas, ela é externa a aplicação da análise facetada.

Uma característica pouco explorada nos esquemas de classificação facetados são os relacionamentos entre conceitos. Um esquema facetado tem a capacidade de expressar tais relacionamentos, sejam os relacionamentos entre conceitos dentro de uma mesma faceta (relacionamentos paradigmáticos), bem como os relacionamentos entre conceitos de diferentes facetadas (relacionamentos sintagmáticos). Quando se considera os relacionamentos dentro de uma mesma faceta, todos os conceitos estão numa mesma categoria (são todos ou entidades ou processos ou lugares, etc.). Os relacionamentos entre estes conceitos são basicamente hierárquicos, incluindo conceitos genéricos, específicos e coordenados.


Os esquemas facetados se diferenciam quando se considera seu potencial de combinar conceitos de diferentes facetadas. E é neste ponto que uma das principais características destes esquemas aparece: os relacionamentos entre facetadas e entre termos de diferentes facetadas – os relacionamentos inter-facetadas (BROUGHTON, 2006). O número e a variedade desses relacionamentos são exclusividade dos esquemas facetados e como aponta Broughton (2006), apesar de estes relacionamentos serem raramente objeto de discussão, eles representam um alto grau de sofisticação. Se considerarmos uma estrutura facetada que utilize 8 categorias ou facetadas principais teremos um número total de 56 relacionamentos inter-facetadas em potencial, do tipo “*agente-operação*”, “*processo-produto*”, etc. O QUADRO 6 abaixo ilustra os dois tipos de relacionamentos presentes em um esquema facetado.

QUADRO 6 - Relacionamentos presentes em um esquema facetado

Fonte: Adaptado de Broughton (2006)

A identificação de relacionamentos intra e inter-facetados apresentados por Broughton (2006) se alinha com a distinção entre relacionamentos “paradigmáticos”, que são inerentes à natureza dos conceitos e são válidos em qualquer situação, e relacionamentos “sintáticos” ou “sintagmáticos” que são aqueles que ocorrem quando vários conceitos são reunidos no contexto de um dado documento para a descrição de seu conteúdo. Desta forma os relacionamentos entre conceitos dentro de uma mesma faceta são em geral relacionamentos paradigmáticos, enquanto os relacionamentos entre conceitos de diferentes facetas são normalmente sintagmáticos (BROUGHTON, 2008).

QUADRO 7 - Relacionamentos paradigmáticos e sintagmáticos

Faceta	Entidades	Operações	Agentes
Relacionamentos paradigmáticos entre termos do vocabulário	Cereais	Cultivo	Máquinas agrícolas
	Trigo	Colheita	Colheitadeira
 Relacionamentos sintagmáticos entre termos quando atribuídos a um documento			

Fonte: Adaptado a partir de Broughton (2008)

A ordem de citação determina alguns aspectos importantes em um sistema de classificação facetado na medida em que: fornece regras para a ordem de combinação dos termos no momento da classificação; determina quais aspectos de um assunto são

mantidos próximos ou separados; afeta a estrutura lógica do sistema, a possibilidade de localização de assuntos compostos e com isso a efetividade da recuperação. Ao invés de impor uma determinada ordem, os esquemas facetados geralmente fornecem uma ordem padrão que é recomendada para a maioria dos casos. Esta ordem padrão coincide com a ordem em que as categorias fundamentais são normalmente listadas, tanto no caso do PMEST (*Personality, Matter, Energy, Space e Time*), como no caso das categorias sugeridas pelo CRG (*thing, kind, part, property, material, process, operation, agent, patient, product, by-product, space, e time*). De acordo com Broughton (2006) existem três argumentos teóricos para esta ordenação padrão: (1) progressão do concreto para o abstrato; (2) cada faceta é dependente das precedentes; e (3) fornece o grupamento mais útil (para a recuperação).

Quando se considera a aplicação de esquemas facetados em ambientes digitais a ordem de citação não é importante, dada a flexibilidade e facilidade com que o usuário pode fazer uso da interface do sistema a fim de estabelecer as categorias a serem utilizadas na busca ou navegação. Diversos autores (GODERT, 1991; INGWERSEN; WORMELL, 1992; ELLIS; VASCONCELOS, 2000) reconhecem que os esquemas facetados são adequados para o uso em ambientes digitais, especialmente na web, onde todo o seu potencial pode ser explorado. Ingwersen e Wormell (1992) ao analisar o uso de estruturas facetadas para recuperação da informação em bancos de dados afirmaram que:

... a discussão demonstra a utilidade da categorização facetada, não apenas para documentos textuais, mas também para outras formas de conteúdos. A categorização facetada pode fornecer pontos de acesso multidimensionais e estruturados para o conteúdo dos documentos e desta forma possibilitar o acesso intelectual ao conhecimento gerado e armazenado. (INGWERSEN; WORMELL, 1992, p. 199, tradução nossa).

Segundo Kwasnick (1999, p.39), “A noção de facetadas se assenta na crença de que existe mais de uma maneira de ver o mundo e que mesmo as classificações que são tidas como estáveis, são de fato provisórias e dinâmicas. O desafio é construir classificações que sejam flexíveis e que possam acomodar novos fenômenos”. Denton (2003) acrescenta que depois que o esquema de classificação é construído, o desafio é tornar seu uso fácil e intuitivo. Com o hipertexto e a *Web*, variadas visões podem ser dinamicamente construídas. As facetadas possibilitam um esquema organizacional multidimensional, que com o uso de ferramentas disponíveis em ambientes digitais e na *Web*, pode ser facilmente consultado e navegado em suas diversas dimensões, de modo que todos os benefícios das classificações facetadas podem ser aproveitados nestes ambientes.

Broughton (2006) nos lembra que, no ambiente digital, o arranjo linear dos materiais ou de suas representações não é tão importante. Apesar da ordem de apresentação na tela ser importante, a ordem física dos itens no repositório não é importante. No ambiente digital, o objetivo da organização da informação não é o arranjo físico dos materiais, mas sim a adequada descrição dos mesmos, rotulagem com vistas à recuperação, fornecimento de ferramentas de busca, navegação e recuperação e, finalmente, a apresentação dos resultados. Neste contexto, a autora enumera os benefícios que os esquemas de classificação facetados podem oferecer:

- capacidade de expressar através da síntese a complexidade de assuntos e conteúdos que é típica de documentos digitais;
- uma sintaxe sistemática que assegura o gerenciamento consistente;
- uma estrutura lógica rigorosa que é compatível com a manipulação computacional em qualquer nível;
- uma estrutura que é compatível com uma interface gráfica intuitiva para navegação e formulação de consultas pelo usuário final;
- facilidade para variação do conjunto de critérios utilizados e na ordem de aplicação destes critérios de modo a permitir abordagens múltiplas de consulta e navegação;
- uma estrutura e metodologia que permite a conversão para outros formatos de linguagens de indexação (como cabeçalhos de assunto e tesouros).

As classificações facetadas estão se tornando cada vez mais presentes na *Web*, principalmente em *sites* comerciais (ADKISSON, 2005; LA BARRE, 2006). As facetas e os esquemas que nelas se baseiam representam uma maneira natural, intuitiva e flexível de organizar coisas. Desta forma, os projetistas de *sites* as utilizam na medida em que analisam as diversas formas alternativas nas quais os usuários gostariam de visualizar as informações contidas nos *sites*. Isto sem nada saber sobre Ranganathan, o *Classification Research Group* (CRC), nem sobre as décadas de estudos por traz da Teoria da Classificação Facetada. Por outro lado, os profissionais da informação conhecem a teoria, reconhecem o seu potencial de aplicação, mas não possuem os conhecimentos técnicos necessários para implementar estes sistemas no ambiente digital. Denton (2003) aponta para a necessidade de aproximar estas duas áreas de estudo e, nesse sentido, sugere um processo simplificado para a criação de classificações facetadas e sua implementação na forma de sistemas na *Web*.

Broughton (2006) investigou as formas nas quais as características da classificação facetada vêm sendo total ou parcialmente aproveitadas pelos sistemas e ferramentas na recuperação da informação. A autora concluiu que a classificação facetada vem desempenhando um importante papel em diversos métodos de recuperação da informação. De acordo com a autora, ela se estabeleceu como um método para construção de sistemas de classificação e tesouros, além de ter afetado o desenvolvimento dos sistemas mais conservadores, tradicionalmente utilizados na descrição e organização de documentos. A classificação facetada também vem sendo utilizada na concepção de ferramentas de navegação em *sites* da *Web*, ajudando a organizar todo tipo de objeto e informação sobre estes objetos. O seu uso vem sendo estudado por pesquisadores nas áreas de indexação automática e *web* semântica como uma ferramenta conceitual para o entendimento dos conceitos e relacionamentos presentes em um dado domínio.

A Teoria da Classificação Facetada é importante por fornecer uma metodologia racional para a construção de esquemas de classificação, em contraste com o agrupamento pragmático de classes que caracteriza as classificações desenvolvidas no passado (HJORLAND, 2002). Segundo Hjørland (2007), a classificação facetada é a única forma de organizar os conceitos de um domínio que possui uma base racional e lógica. Neste sentido, a classificação facetada teve sua influência sentida nos principais instrumentos convencionais da classificação bibliográfica, como analisado por Broughton (2006), nas políticas e diretrizes para as versões mais recentes da Classificação Decimal de Dewey, da Classificação Decimal Universal e dos Cabeçalhos de Assunto da Biblioteca do Congresso Americano.

Outra área de influência da classificação facetada pode ser vista nas normas e padrões para construção de tesouros e vocabulários controlados. O padrão NISO para construção de tesouros "*Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies - ANSI/NISO Z39.19-2005*" apresenta uma perspectiva interessante sobre o papel da classificação facetada. Ele reconhece o papel desempenhado pela mesma, afirmando que "vocabulários controlados – especialmente aqueles formados a partir de milhares de termos – podem ser mais fáceis de serem utilizados caso sejam organizados de formas não hierárquicas" (NISO, 2005, tradução nossa). A esta afirmativa segue uma breve descrição das origens da classificação facetada e em seguida um resumo das suas aplicações potenciais:

... a classificação facetada é particularmente útil para: (1) campos de estudo novos e emergentes, nos quais o conhecimento ainda é limitado ou onde os relacionamentos entre os objetos são desconhecidos ou fracamente definidos; (2) áreas interdisciplinares onde existe mais de uma perspectiva sobre como vislumbrar os conteúdos ou onde é necessária a combinação de conceitos; (3) vocabulários onde são necessárias múltiplas hierarquias,

mas que podem ser inadequadas na medida em que não se consegue definir claramente seus limites; ou (4) na classificação de documentos e conteúdos eletrônicos, onde a localização e postura desses materiais não é uma questão importante (NISO, 2005, tradução nossa).

Estas normas trazem ainda uma visão diferenciada sobre o conceito de facetas. Esta visão estende o entendimento do que vem a ser uma faceta de modo a incorporar vários aspectos não semânticos de um documento. A norma NISO afirma que “a análise facetada é às vezes usada para indicar atributos de objetos de informação” (NISO, 2005), e lista em seguida as seguintes facetas potenciais: *tópico*, *autor*, *localização*, *formato*, *idioma* e *local de publicação*. Esta lista tem muita coisa em comum com o projeto FAST (*Faceted Application of Subject Terminology*) (CHAN *et al.*, 2001), na medida em que “*tópico*” (ou assunto) é visto como uma faceta entre várias outras que representam elementos da descrição bibliográfica não relacionados ao assunto. Esta visão de faceta parece ser equivalente ao conceito de atributo de uma entidade (documento) no modelo Entidade-Relacionamento (NAVATHE; ELMASRI, 2003) ou no FRBR (IFLA, 1998). Neste sentido, estas “facetas” podem ser facilmente mapeadas nos campos do MARC ou nos elementos do padrão Dublin Core. Broughton (2006) assinala que este tipo de aplicação da classificação facetada parece ser exclusivo dos pesquisadores americanos.

Pelo exposto acima, parece haver duas visões distintas da classificação facetada:

- A. Um conjunto de “entidades” é classificado de acordo com suas propriedades. Por exemplo, um conjunto de detergentes é classificado por “*marca*”, “*forma*”, “*aroma*”, “*agente*”, “*efeito no agente*” e “*propriedades especiais*” (DENTON, 2003). Neste caso, cada uma destas propriedades representa uma “faceta” através da qual um conjunto de instâncias (detergente) pode ser vislumbrado.
- B. Um conjunto de “conceitos” é agrupado de acordo com o seu tipo mais primitivo. Por exemplo, o conceito “*madeira*” é colocado na faceta “*materiais*”. O conceito “*inseto*”, por sua vez, é colocado na faceta “*organismos*”. Neste caso, uma faceta é essencialmente uma classe primitiva e cada membro do agrupamento representado pela faceta é uma instância ou subclasse desta classe primitiva.

As duas visões parecem ser consequência da diferença entre “classificar entidades” e “classificar conhecimento”. Quando levada para a classificação de documentos, a confusão fica ainda mais complexa, visto que um documento tanto pode ser visto como uma entidade, possuindo propriedades (ou atributos) como *forma*, *idioma*, *data de publicação*, etc., como pode ser visto como uma forma de representação de informação e

conhecimento a ser compartilhado. Neste último caso, este conhecimento pode ser classificado de modo a facilitar o acesso ao documento e seu conteúdo.

Os pesquisadores da tradição biblioteconômica, em geral, reservam a expressão “classificação facetada” para classificação do conhecimento apenas (visão B), ou seja, conhecimento embutido nos documentos. Neste sentido parece estranho falar de facetadas em detergentes, vinhos, etc. Estes pesquisadores reconhecem a existência da visão A, mas quando se referem a uma classificação como “facetada” eles sugerem que além da divisão em propriedades ou atributos básicos de uma entidade deve ser utilizado algum conjunto de categorias fundamentais num nível superior (visão B). Na medida em que os tesouros passaram a ser projetados com base na classificação facetada, a atenção foi voltada exclusivamente para a estrutura superior de acordo com a visão B. Isto se deve a própria natureza dos tesouros: linguagens de indexação que não necessitam de uma estrutura classificatória apurada e que podem trabalhar com uma estrutura organizacional mais flexível, apesar de se basear no princípio lógico da exclusividade mútua para a divisão em categorias fundamentais.

Em suma, se o intuito é a classificação de objetos/entidades (vinhos, detergentes, carros, softwares etc.) então a estrutura classificatória resultante será baseada na visão A. Por outro lado, se a intenção é classificar o conhecimento (conceitos), então existirão facetadas fundamentais num nível mais alto de abstração de modo a permitir a organização dos conceitos de acordo com a visão B.

Uma vantagem de uma visão ampliada (visão A + B) da classificação facetada pode estar na possibilidade de desenvolvimento de um modelo único de busca e navegação, baseado na Teoria da Classificação Facetada, que incorpore tantos os aspectos semânticos (temáticos) como os aspectos descritivos dos documentos. Neste sentido, a Teoria da Classificação Facetada atuaria como alicerce comum, na medida em que serviria tanto a representação temática dos documentos em um determinado campo de estudo, como também poderia incorporar os elementos da representação descritiva, usualmente representados por meio de padrões de metadados. Esta possibilidade foi explorada nesta pesquisa.

La Barre (2006) observou o crescente interesse pela análise e classificação facetada entre profissionais como arquitetos da informação, atuando no projeto de *sites web*, e especialistas em gestão do conhecimento, atuando na construção de mecanismos para facilitar a gestão e o acesso ao conhecimento organizacional. Parte destes profissionais utiliza explicitamente a terminologia sugerindo o embasamento nos princípios inicialmente propostos e desenvolvidos por Ranganathan, outros nem tanto. Apesar de o uso da terminologia não significar necessariamente evidência de que as estruturas organizacionais

desenvolvidas por estes profissionais estejam alinhadas com os princípios teóricos da classificação facetada, seu uso é indicativo da adoção destes princípios e pode indicar áreas nas quais as estruturas de organização e acesso podem ser aperfeiçoadas pela aplicação consistente da teoria. Esta situação também pode revelar novas áreas da prática profissional que podem servir para aperfeiçoar a teoria, numa abordagem simbiótica entre a teoria e a prática (LA BARRE, 2006). A pesquisa da autora buscou fazer uso da Teoria da Classificação Facetada como ferramenta para analisar as estruturas classificatórias utilizadas em aplicações e *sites* na Internet, bem como avaliar até que ponto estas aplicações estão alinhadas aos princípios da classificação facetada. La Barre (2006) concluiu que a Teoria da Classificação Facetada vem sendo amplamente aplicada na arquitetura da informação de grande parte dos *sites*. Mais especificamente, ela vem sendo usada como mecanismo de organização do conteúdo e navegação. A autora aponta para a necessidade de estudos de usabilidade que possam avaliar a *performance* destas interfaces, bem como apontar soluções que venham a reduzir a sua complexidade. Outro ponto destacado pela autora e de interesse desta pesquisa é relativo à possibilidade de exploração, sistematização e formalização dos relacionamentos entre facetas.

Os esquemas de classificação são muitas vezes considerados pouco efetivos e inadequados para a organização de recursos digitais (KOCH, 1998). As razões para isso estão na falta de estrutura lógica observada nos esquemas de classificação oriundos da tradição biblioteconômica. Estes esquemas são construídos numa estrutura hierárquica, com ênfase na sucessiva divisão em classes mais específicas. Neste tipo de estrutura, os únicos relacionamentos identificados são os hierárquicos, sendo que não existe suporte para relacionamentos sintáticos BROUGHTON (2002). Como consequência, o esquema de classificação é geralmente extenso, não havendo mecanismos para a combinação de classes ou para a representação de assuntos compostos e complexos. Estes problemas se acentuam quando se considera domínios interdisciplinares ou conteúdos complexos, como é o caso da área científica.

As classificações construídas de acordo com o modelo analítico facetado representam uma alternativa, na medida em que operam sob princípios distintos, partindo dos conceitos constituintes de um assunto, em contrapartida a uma visão totalizada e rígida do conhecimento. Estas classificações podem ser usadas em ferramentas para controle de vocabulário, descrição temática de documentos e na recuperação. Uma classificação facetada traz diversas vantagens quando aplicada na forma de um mecanismo de organização para o acesso a recursos digitais. O formato lógico e previsível da classificação facetada se torna bastante intuitivo e útil ao usuário, podendo ser usada na construção da interface de navegação e exploração do acervo, como alternativa à estrutura hierárquica de

diretórios, típica das classificações tradicionais. As classificações facetadas podem fornecer pontos de acesso multidimensionais e altamente estruturados aos documentos.

No caso de uma biblioteca digital de teses e dissertações, a representação dos assuntos dos documentos, assim como a organização do acervo visando à navegação e exibição, não podem ser satisfatoriamente conseguidas com os esquemas classificatórios tradicionais. Nessas bibliotecas, o conteúdo intelectual dos documentos é bastante complexo e é necessário um nível mais profundo de indexação, tanto em termos de especificidade como de exaustividade. Os tesouros representam uma solução viável para a representação dos assuntos nestes casos. No entanto, uma classificação facetada fornece uma estrutura mais rigorosa para a organização dos documentos, que pode ser utilizada pelo usuário nos processos de navegação e exploração do acervo. Além disso, uma classificação facetada pode ser usada como base para construção de mecanismos que possibilitam buscas estruturadas ao acervo. Uma classificação facetada pode representar uma alternativa tanto para a representação de assuntos, como para a organização do acervo na medida em que pode permitir a exploração pelo usuário através da combinação de perspectivas multidimensionais, que podem ainda envolver uma combinação de aspectos semânticos e descritivos dos documentos.

Neste ponto é interessante destacar as considerações de Broughton (2008) quando analisa a possibilidade de controle de vocabulário através de uma classificação facetada. Segundo a autora, uma classificação rotula cada classe com um código notacional, e é este código que representa o conceito e facilita a busca e o acesso ao acervo. O código atua como um identificador único para o conceito representado pela classe, incluindo todos os sinônimos que não são adequadamente controlados. Desta forma, é discutível a afirmação de que um esquema de classificação possa atuar como um vocabulário controlado. Apesar de a terminologia ser um importante componente da classificação, o controle existente é geralmente feito através da notação. O controle de vocabulário existente em um tesouro, por exemplo, é somente parcialmente implementado pela classificação. Os sinônimos são controlados apenas na medida em que são identificados e inseridos na mesma classe, mas não há nada que indique os termos preferíveis. Esta dificuldade ocorre devido ao fato de uma classificação não dar importância ao *termo*, mas apenas ao *conceito* representado por uma classe e seu respectivo código (BROUGHTON, 2008). Estas considerações nos levam a questionar qual seria a melhor estrutura a ser usada para a organização do conhecimento em uma biblioteca digital de teses e dissertações. Acreditamos que uma abordagem mista seria a mais adequada, na medida em que possa permitir a organização do acervo visando a navegação e busca estruturada por meio de uma estrutura classificatória facetada mas que, na medida em que abre mão da notação, possa

oferecer um nível adequado de controle de vocabulário. Outra possibilidade é agregar à estrutura de navegação um mecanismo de busca textual, de forma que o usuário tenha facilidade em localizar na estrutura os conceitos a partir de sinônimos ou termos relacionados.

Compartilhamos do entendimento de Lancaster (2004) de que a indexação é uma forma de classificação. Na medida em que se atribui um termo, que representa um conceito específico, a um documento estamos colocando este documento na classe representada pelo conceito. No entanto, a simples atribuição de termos de um vocabulário controlado não oferece uma estrutura adequada para organização do acervo que possa ser utilizada na navegação e em buscas estruturadas. Uma possível solução seria fazer uso de uma estrutura facetada, que possa definir regras para a atribuição de termos que representem conceitos representativos dos variados aspectos temáticos do documento (facetadas), assim como para a combinação de conceitos na representação de assuntos compostos. Desta forma estaríamos elevando a indexação a um novo patamar, na medida em que a estrutura resultante poderia ser utilizada na navegação e formulação de buscas estruturadas através da seleção e combinação de conceitos.

A título de exemplo, consideremos o caso de uso de uma biblioteca digital de teses e dissertações. O usuário deseja consultar “teses” defendidas a partir de “1990” cujo objeto de pesquisa seja a “biblioteca digital”. Além disso, o usuário tem interesse somente nas pesquisas cujo objetivo geral tenha sido melhorar a “recuperação da informação” e tenha adotado uma abordagem “experimental” como metodologia. A busca do usuário envolve tanto aspectos descritivos dos documentos (documentos do tipo “tese”, publicados a partir de “1990”), quanto aspectos temáticos (melhorar a “recuperação da informação” na “biblioteca digital”, usando metodologia “experimental”). Supondo que o acervo tenha sido classificado utilizando uma abordagem facetada, o usuário poderia construir sua estratégia de busca a partir da seleção dos termos “teses” na faceta TIPO DO DOCUMENTO, e “>1990” na faceta DATA DE PUBLICAÇÃO. Em seguida, o usuário poderia selecionar o termo “biblioteca digital” na faceta OBJETO DE PESQUISA, “recuperação da informação” na faceta OBJETIVO e, finalmente, “experimental” na faceta METODOLOGIA.

Com uma abordagem teórica sólida, a classificação facetada tem o potencial de produzir estruturas organizacionais versáteis, facilitando o acesso aos catálogos *on-line* e aos acervos das bibliotecas digitais. Estas estruturas são adequadas para o acesso e exploração do acervo, tanto para usuários novatos como para os experientes. Novas formas inovadoras de busca e navegação, possivelmente inspiradas na aplicação da Teoria da Classificação Facetada, podem ser incorporadas nas bibliotecas digitais e nos catálogos *on-line* das bibliotecas convencionais, como alternativa que possa superar a hegemonia das

buscas textuais (*full-text*) a La Google. Na próxima seção serão apresentados os princípios e metodologia utilizados na análise facetada, que são importantes na medida em que definem o processo pelo qual são construídas as classificações facetadas.

2.2.1 Princípios e metodologia para a construção de classificações facetadas

Louise Spiteri (1998) analisou os complexos conjuntos de regras desenvolvidos inicialmente por Ranganathan e trabalhados posteriormente pelo *Classification Research Group* (CRG) sobre como construir sistemas de classificação facetados. O propósito da autora foi resolver o problema relacionado à complexidade inerente ao texto de Ranganathan no qual ele apresenta 46 cânones, 13 postulados e 22 princípios. Além disso, outro fator que dificultava o entendimento da Teoria da Análise Faceta é que o CRG desenvolveu seus próprios princípios para a análise facetada, que apesar de serem baseados em Ranganathan, em alguns pontos eles contrariam algumas suposições feitas por ele, relacionadas, por exemplo, a escolha das categorias fundamentais e a ordem destas categorias. Os trabalhos desenvolvidos pelo CRG estão espalhados em diversos textos publicados isoladamente ou em conjunto por seus membros. Tendo como base os trabalhos de Ranganathan e do CRG, Spiteri (1998) formulou um conjunto simplificado de regras com o objetivo inicial de servir de ferramenta de ensino para estudantes de biblioteconomia e ciência da informação. No entanto, tal modelo também pode ser utilizado por projetistas de sistemas de classificação facetados e de tesouros, visto que, de outra forma estes profissionais teriam de consultar uma grande variedade de fontes para obter os princípios da análise facetada, necessários para o desenvolvimento de seus trabalhos. Os princípios propostos por Spiteri são bastante completos e claros e serão brevemente explicados a seguir.

Spiteri (1998) seguiu as ideias de Ranganathan no que se refere à divisão do processo de desenvolvimento do sistema de classificação em três partes: “o Plano das Ideias, que envolve o processo de analisar uma área de conhecimento em suas partes componentes; o Plano Verbal, que envolve o processo de escolher uma terminologia apropriada para expressar estas partes componentes; e o Plano Notacional, que envolve o processo de expressar estas partes componentes por meio de um mecanismo notacional”. Partindo do Plano das Ideias para o Verbal e em seguida para o Notacional nos leva da ideia (conceito) para a palavra (termo) e finalmente para o número (notação), ou seja, do conceito geral sobre o qual uma entidade diz respeito para a expressão deste conceito em

um vocabulário controlado e finalmente para uma notação. É importante observar que o Plano Notacional é menos importante que os demais quando se considera o ambiente digital e a *Web*.

O QUADRO 8, a seguir, apresenta os princípios de Spiteri (1998) para a criação de classificações facetadas.

QUADRO 8 - Princípios para a criação de classificações facetadas

Plano das Ideias: Princípios para a escolha das Facetas
<p>a) Diferenciação: ao dividir uma entidade em suas partes componentes, é importante usar características de divisão que possam claramente distinguir entre estas partes componentes. Por exemplo, dividir “humanos” por “sexo”.</p> <p>b) Relevância: ao escolher as facetas a serem usadas para dividir as entidades, é importante ter certeza de que estas facetas refletem o propósito, a natureza, e o escopo do sistema de classificação. Por exemplo, em um sistema de classificação para a área de “Educação” o critério “Grau Escolar” se mostra relevante.</p> <p>c) Verificação: é importante escolher facetas que sejam claras e verificáveis, e nunca subjetivas ou que não possam ser comprovadas. Por exemplo, a faceta “raça” para a entidade “cão” se mostra adequada sob este aspecto, na medida em que pode ser verificada e certificada por veterinários e criadores de cães.</p> <p>d) Permanência: as facetas devem representar qualidades permanentes das entidades sendo divididas. Por exemplo, um “vinho” que foi produzido no “Chile” jamais terá esta característica alterada.</p> <p>e) Homogeneidade: as facetas devem ser homogêneas, no sentido que devem representar uma única característica de divisão.</p> <p>f) Exclusividade Mútua: as facetas devem ser mutuamente exclusivas na medida em os termos em uma faceta devem derivar de seu universo imediato uma e somente uma característica. Por exemplo, ao dividir a entidade “cão”, os termos que aparecem na faceta “tamanho” não devem aparecer sob a faceta “cor”, e vice versa. Esta característica faz com que cada termo tenha uma localização única e bem definida no sistema de classificação.</p> <p>g) Categorias Fundamentais: argumenta que não existem categorias fundamentais válidas para todas as áreas; as categorias devem ser derivadas com base na natureza da área sendo classificada e no propósito do sistema de classificação.</p>
Plano das Ideias: Princípios para a Ordem de Citação de Facetas e Termos
<p>Estes princípios coordenam a organização das facetas e dos termos dentro das respectivas facetas.</p>
<p>a) Sucessão Relevante: a ordem de citação das facetas (e dos termos) deve ser relevante de</p>

<p>acordo com a natureza, propósito, e escopo do sistema de classificação.</p> <p>b) Sucessão Consistente: a partir do momento em que uma ordenação tenha sido estabelecida para um sistema de classificação, ela não deve ser modificada, a menos que ocorra uma mudança na natureza, propósito ou escopo do sistema.</p>
Plano Verbal
<p>a) Contexto: o significado de um termo individual é dado pelo seu contexto com base na sua posição no sistema de classificação.</p> <p>b) Uso Corrente: a terminologia usada no sistema de classificação deve refletir o uso corrente na área. Isto significa que o sistema de classificação requer revisões regulares.</p>
Plano Notacional
<p>a) Sinônimos: cada assunto pode ser representado somente por um único número de classe.</p> <p>b) Homônimos: cada número de classe pode representar somente um único assunto.</p> <p>c) Hospitalidade: a notação deve permitir a adição de novos assuntos, facetas, e termos em qualquer ponto do sistema de classificação.</p> <p>d) Ordem de fichamento: um sistema de classificação deve refletir a ordem de classificação dos assuntos. Este tipo de notação deve refletir a ordem de citação na qual o sistema de classificação se baseia.</p>

Fonte: Adaptado a partir de Spiteri (1998)

O modelo simplificado de Spiteri, apesar de ser derivado tanto das ideias de Ranganathan, quanto das ideias desenvolvidas pelo CRG, apresenta traços mais fortes do CRG na medida em que abre mão das categorias fundamentais fixas propostas por Ranganathan. A autora argumenta que tal decisão se baseou na análise de diversos sistemas de classificação e tesouros facetados, que revelou a popularidade da abordagem proposta pelo CRG, devido em parte a flexibilidade que esta abordagem apresenta. Além disso, Ranganathan não definiu claramente como as categorias fundamentais (PMEST) poderiam ser aplicadas a todas as áreas de conhecimento. A abordagem do CRG, por outro lado, permite aos projetistas construir sistemas de classificação e tesouros facetados mais adaptados às necessidades dos usuários e as especificidades de uma determinada área do conhecimento.

O principal resultado dos estudos do CRG foi a construção da *Bliss Bibliographic Classification*, um esquema de classificação que incorpora os princípios da Teoria da Classificação Facetada desenvolvidos pelo grupo. Broughton (2008) apresenta de forma resumida uma metodologia para construção de uma classificação facetada de acordo com o

modelo utilizado na segunda edição da *Bliss Bibliographic Classification* (BC2). A metodologia proposta pela autora será brevemente apresentada a seguir:

- Os conceitos de um dado domínio são identificados e organizados em grupos ou facetas, que correspondem às categorias fundamentais propostas pelo CRG (*thing, kind, part, property, material, process, operation, agent, patient, product, by-product, space, e time*), ou identificadas no domínio específico (como, por exemplo, *forma* e *gênero* no domínio das artes).
- As facetas formadas são então subdivididas em sequencias (*arrays*), ou seja, grupos de conceitos com características comuns. Por exemplo, uma faceta “*pessoas*” pode ser organizada em sequencias: por idade (criança, adolescente, adulto); por sexo (masculino, feminino); por estado civil (solteiro, casado, divorciado); por ocupação (gerencial, administrativo, operacional).
- Os relacionamentos hierárquicos são identificados dentro de cada faceta e são apresentados a partir de classes e subclasses representadas através da indentação no esquema.
- A ordem de combinação entre termos de diferentes facetas é estabelecida. Na maioria dos casos esta ordem será a ordem padrão de citação, baseada na sequencia das categorias fundamentais conforme listadas acima.
- Todos os termos usados para descrever um determinado conceito (sinônimos ou quase sinônimos) são apresentados juntos na identificação do conceito.

O processo descrito acima, aplicado no desenvolvimento da BC2, gerou um sistema classificatório altamente estruturado e lógico. Regras para a combinação de assuntos também foram definidas e explicitadas, permitindo que assuntos complexos pudessem ser representados e precisamente localizados no sistema. Apesar de uma classificação facetada teoricamente consistir apenas de assuntos elementares (como “*inflamação*” ou “*articulação*”), na BC2 ocorre certo grau de pré-coordenação ou combinação de assuntos visando apresentar assuntos compostos (como “*artrite*”), que seriam provavelmente utilizados pelos indexadores e pelos usuários nas buscas. Estes conceitos são adequadamente localizados no esquema e aparecem no índice alfabético.

Na próxima seção será discutida uma possível abordagem visando a utilização de uma classificação facetada para descrever o conteúdo temático de documentos e a sua representação por meio do padrão de metadados Dublin Core.

2.2.2 A representação temática usando o padrão de metadados Dublin Core

As bibliotecas aplicam varias técnicas como a indexação, sumarização e catalogação para criar representações dos documentos. Quando se considera o ambiente digital, estas representações são mais conhecidas atualmente como metadados. Como o sentido dos documentos não está, em geral, prontamente acessível a partir das palavras e frases que constituem o seu conteúdo, estas representações podem oferecer acesso semântico aos documentos e, assim, ajudar os usuários na localização e acesso.

A análise dos modelos e padrões de metadados realizada na seção 2.1.6 permitiu verificar que o foco tem sido a representação descritiva dos itens bibliográficos. A representação temática também deve ser incorporada nestes padrões, visando principalmente à padronização e interoperabilidade entre bibliotecas digitais. Entre os padrões de metadados, o Dublin Core foi o que mais se destacou dada a sua ampla adoção nos mais variados contextos, bem como ter servido de base para a definição de outros padrões bastante semelhantes, como ETD-MS e MTD-BR.

Quando se considera a representação temática, dentre os quinze elementos do Dublin Core, o elemento SUBJECT se destina a representar o assunto de um documento. Este elemento foi mantido pelos padrões ETD-MS e MTD-BR e é descrito como:

...tópico que representa o conteúdo do recurso. Geralmente é expresso através de palavras chave, frases ou códigos de classificação que descrevem o assunto tratado por um recurso. Recomenda-se que seus valores sejam selecionados a partir de um vocabulário controlado ou um esquema de classificação. (adaptado de DCMI, 2003, tradução nossa).

Além do SUBJECT, alguns outros elementos do Dublin Core também se destinam a representar informações relacionadas ao conteúdo de um documento: TITLE, DESCRIPTION, TYPE, LANGUAGE e COVERAGE. Chan e outros (2001) sugerem que juntos estes elementos pode ser vistos como uma abordagem facetada pós-coordenada para a representação do conteúdo. Em outras palavras, forma (TYPE), idioma (LANGUAGE), espaço e tempo (COVERAGE) aparecem separados do assunto propriamente dito (SUBJECT). Estes elementos podem então ser reagrupados nas seguintes categorias:

- Informações sobre o assunto
 - SUBJECT
 - TITLE
 - DESCRIPTION
- Informações sobre a forma

- TYPE
- Informações sobre o idioma
 - LANGUAGE
- Informações espaciais ou temporais
 - COVERAGE

O padrão Dublin Core foi concebido para ser extremamente flexível; nenhum dos seus elementos é obrigatório e cada elemento pode ser aplicado mais de uma vez para descrever adequadamente um recurso. Além disso, o Dublin Core pode ser estendido a partir do uso de qualificadores.

Tendo em vista uma representação temática mais apurada através da adoção de uma estrutura facetada, o elemento SUBJECT pode ser estendido com a aplicação de qualificadores de modo a contemplar diferentes aspectos (facetadas) do assunto. Se considerarmos as categorias fundamentais de Ranganathan (PMEST) teríamos os seguintes qualificadores:

SUBJECT.Personality

SUBJECT.Matter

SUBJECT.Energy

SUBJECT.Space

SUBJECT.Time

Tendo em vista que os aspectos espaciais e temporais do conteúdo já estão contemplados pelo elemento COVERAGE, estes aspectos poderiam então ser retirados do elemento SUBJECT e serem inseridos como qualificadores no elemento COVERAGE:

COVERAGE.Space

COVERAGE.Time

É importante ressaltar que o PMEST foi escolhido somente para ilustrar a aplicação de categorias fundamentais a partir de qualificadores no Dublin Core. A escolha de um conjunto de categorias fundamentais, seja o PMEST ou aquelas propostas pelo CRG, pode ser interessante na medida em que possibilita interoperabilidade semântica entre as diversas áreas de conhecimento. Por outro lado, a escolha de categorias específicas a uma determinada área de conhecimento pode resultar num esquema facetado mais intuitivo para o usuário e, assim, tornar a recuperação mais efetiva em uma determinada aplicação. Esta questão precisa ser analisada caso a caso, tendo em vista as vantagens de desvantagens de cada abordagem.

Tendo em vista o uso de uma classificação facetada na representação e organização do conhecimento no contexto de uma biblioteca digital de teses e dissertações, existe a necessidade de construir tal estrutura, além de sua formalização visando o processamento automático característico do ambiente digital. As próximas seções irão tratar especificamente destas duas questões.

2.3 Construção de uma taxonomia facetada

Taxonomias são estruturas hierárquicas de categorias, desenvolvidas de forma sistemática, utilizadas como ferramenta de organização intelectual. São geralmente empregadas em portais institucionais e bibliotecas digitais como um mecanismo de consulta e navegação, ao lado de ferramentas de busca. Além dessas aplicações, a taxonomia é um dos componentes em Ontologias (CAMPOS; GOMES, 2007). A organização das informações, através de uma taxonomia, permite alocar, recuperar e comunicar informações dentro de um sistema de maneira lógica, pela navegação.

As taxonomias são geralmente desenvolvidas utilizando a abordagem enumerativa, através da segmentação do universo, na qual as categorias são dispostas em uma estrutura hierárquica e os níveis das subcategorias são definidos. A estrutura hierárquica é de fácil compreensão e uso, porém não é flexível, na medida em que cada objeto deve ser situado em uma única categoria na hierarquia.

Já as taxonomias facetadas são projetadas de acordo com o modelo analítico-sintético descrito na seção 2.2. Campos e Gomes (2007) também indicam princípios de categorização, baseados na classificação facetada, a serem aplicados no desenvolvimento de taxonomias policotômicas, ou seja, onde um elemento é associado a tantas classes e subclasses quantas necessárias, dentro de um domínio especializado ou uma tarefa. Tzitzikas e outros (2002) definem formalmente uma taxonomia facetada como um conjunto finito de facetas. Cada faceta consiste de uma terminologia, ou seja, um conjunto finito de termos estruturados por relação de subsunção (hierárquica). Cada faceta é projetada separadamente e modela um aspecto distinto do domínio. Com o uso de uma taxonomia facetada, cada objeto pode ser classificado a partir da seleção de um ou mais termos a partir de diferentes facetas.

Maculan (2011) tratou em sua pesquisa de mestrado²⁸ da construção de uma taxonomia facetada para a área de Organização e Uso da Informação (OUI). O objetivo da autora foi a criação de uma taxonomia facetada para navegação, a ser utilizada como interface de busca, a partir da representação semântica do conteúdo de teses e dissertações, visando facilitar a posterior recuperação desses documentos.

O universo de pesquisa da autora consistiu do acervo de teses e dissertações defendidas no Programa de Pós-Graduação em Ciência da Informação, da Escola de Ciência da Informação, da Universidade Federal de Minas Gerais (PPGCI/ECI/UFMG), referentes à linha de pesquisa Organização e Uso da Informação (OUI). O corpus utilizado foi composto por 62 documentos, entre teses e dissertações, defendidas no PPGCI/ECI/UFMG no período compreendido entre os anos de 1998 a 2009. O objeto empírico utilizado para a construção da taxonomia facetada foi formado pelos termos oriundos das palavras-chave indicadas pelos autores dos documentos, além dos descritores utilizados para a indexação dos assuntos dos documentos no contexto do Sistema Pergamum, assim como os termos selecionados a partir da análise conceitual dos documentos, realizada pela autora.

O resultado obtido por Maculan (2011) foi uma taxonomia facetada que pode ser utilizada para facilitar a recuperação dos documentos no contexto da Biblioteca Digital de Teses e Dissertações da Escola de Ciência a Informação da UFMG (BDTD/ECI/UFMG). Posteriormente, a autora revisou e ampliou a taxonomia construída, de modo a abarcar as demais linhas de pesquisa do PPGCI/ECI/UFMG. As categorias principais (facetadas) utilizadas pela autora na construção da taxonomia foram:

1. **Tema** - Qual o assunto de que trata o documento?
2. **Objeto Empírico** - Qual o objeto empírico do estudo em questão? Qual objeto foi utilizado e/ou analisado na pesquisa?
3. **Ambientação** - O tema, objeto e/ou ação são considerados no contexto de um lugar específico ou ambiente?
4. **Tipo de Pesquisa** - Tendo em vista a natureza (básica, aplicada), a abordagem (quantitativa, qualitativa), os objetivos (exploratória, descritiva, explicativa) ou os procedimentos (bibliográfica, experimental, documental, estudo de caso, pesquisa-ação, levantamento, ex-post-facto, participante), quais as classificações podem tipificar a pesquisa realizada?

²⁸ O autor participa do mesmo grupo de pesquisa da referida autora.

5. **Coleta de Dados** - Quais instrumentos específicos (questionários, entrevistas, registros áudios visuais, coleta de documentos, etc.) foram utilizados para realizar a ação?
6. **Métodos** - Quais modos específicos foram utilizados para realizar a ação (por exemplo, técnicas ou métodos para tratamento dos dados, que podem ser do tipo: modelagem estatística, análise estrutural, codificação, análise de conteúdo, indexação, análise semiótica, retórica ou de discurso)?
7. **Fundamento Teórico** - Houve alguma corrente ou abordagem teórica específica (teorias, hipóteses, pressupostos, etc.) utilizada em função da natureza do objeto a ser pesquisado e dos objetivos pretendidos, que foram descritos na pesquisa?
8. **Fundamento Histórico-Contextual** - Quais temas foram tratados e revisados, a partir de pesquisa bibliográfica, para contextualizar o tema pesquisado de forma profunda e consistente?
9. **Resultados** - Quais pontos a pesquisa alcançou, levando em consideração os objetivos propostos? Houve formulação ou reformulação de teoria, criação de um método ou de um produto?

A taxonomia desenvolvida por Maculan (2011), incluindo a sua revisão posterior, foi utilizada no contexto da presente pesquisa, servindo como instrumento de organização do conhecimento, a ser utilizado pelos componentes de busca e acesso de um protótipo computacional.

2.4 Formalização de uma taxonomia facetada

Visando a manipulação computacional do esquema classificatório no contexto de uma biblioteca digital, faz-se necessário representar formalmente tal esquema. Uma classificação facetada apresenta várias vantagens sobre as classificações estritamente hierárquicas conforme amplamente discutido na seção 2.2.

Buscando formalizar matematicamente o esquema classificatório facetado, nos apoiamos em Tzitzikas e outros (2002), que trabalham o conceito de taxonomias facetadas. Tzitzikas e outros (2002) definem formalmente uma *taxonomia facetada* como um conjunto finito de *facetas*. Cada *faceta* consiste de uma *terminologia*, ou seja, um conjunto finito de *termos* estruturados por relação de *subsunção* (hierárquica). Cada faceta é projetada separadamente e modela um aspecto distinto do domínio. A FIGURA 14 ilustra uma

taxonomia faceta voltada para a organização de informações turísticas, possuindo as facetas: *Local*, *Tipo de Acomodação* e *Atividade*. Cabe destacar que a taxonomia apresentada não segue os princípios básicos de divisão de classes, adotando uma abordagem mais livre, proposta para ser utilizada especificamente para recuperação de informações relacionadas a pacotes turísticos.

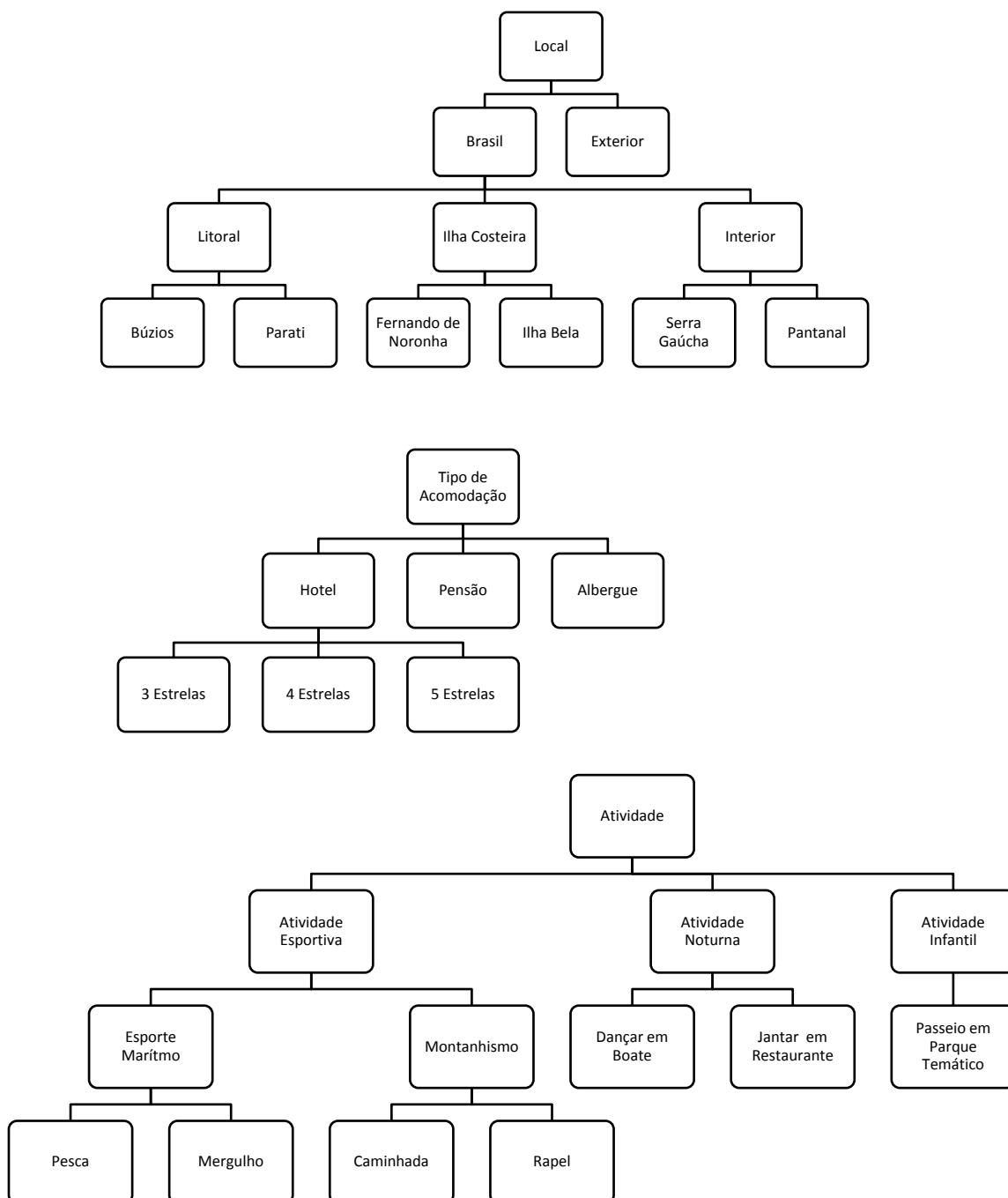


FIGURA 14 - Taxonomia faceta para organização de informações turísticas

Fonte: adaptado de Tzitzikas e outros (2002)

Seguem as definições propostas pelos autores.

DEFINIÇÃO 1: Uma *taxonomia* é um par (T, \leq) onde T é uma *terminologia*, ou seja, um conjunto finito de *termos*, e \leq é uma relação de *subsunção* sobre T , ou seja, uma relação reflexiva e transitiva sobre T . Na figura acima, $Esporte \leq Atividade$, ou seja, esporte é um tipo de atividade.

DEFINIÇÃO 2: Uma *taxonomia facetada* é um conjunto finito de *taxonomias* $F = \{F_1, \dots, F_k\}$ no qual cada $F_i = (T_i, \leq_i)$ é chamado de *faceta*.

Utiliza-se TT para denotar a união das terminologias de todas as facetas de F , ou seja, $TT = \bigcup_{i=1}^k T_i$, e \leq para denotar a união de todas as relações de subsunção \leq_i , ou seja, $\leq = \bigcup_{i=1}^k \leq_i$. O par (TT, \leq) é uma *taxonomia* de acordo com a Definição 1.

A *descrição* de um objeto no processo de indexação ou classificação usando uma *taxonomia facetada* é dada pela associação do objeto com uma combinação de *termos* das diferentes *facetadas*.

DEFINIÇÃO 3: Uma *descrição* d sobre F é um *termo* $t \in TT$ ou uma sequência de *termos* separados por ".", ou seja, sequências derivadas da seguinte gramática $d ::= d.t \mid t$

Considerando-se o exemplo de *taxonomia facetada* apresentado na FIGURA 14, as seguintes descrições são válidas de acordo com a *Definição 3*:

- Parati.Albergue
- Parati.Albergue.Mergulho
- Serra Gaúcha.5 Estrelas.Restaurante
- Serra Gaúcha.Restaurante

Na próxima seção será apresentado um modelo genérico visando a implementação de mecanismos de busca envolvendo *taxonomias facetadas*, conforme proposto por Sacco e Tzitzicas (2009). Os autores trabalham o conceito de *taxonomias dinâmicas*, que representam uma abordagem mais flexível, que pressupõe apenas a classificação multidimensional dos objetos, ou seja, a possibilidade de um objeto poder ser classificado por mais de um conceito ou classe da *taxonomia*, em qualquer nível da estrutura taxonômica.

2.4.1 Modelo de implementação – Taxonomias Dinâmicas

Taxonomias dinâmicas (SACCO; TZITZICAS, 2009), também chamadas atualmente de sistemas de busca facetada, representam um modelo genérico de conhecimento, baseado numa classificação multidimensional de objetos ²⁹, que é usado na exploração/navegação de bases de dados de uma forma guiada, porém não restritiva, por meio de uma interface visual. O modelo é voltado principalmente para o acesso centrado no usuário.

O esquema conceitual de uma taxonomia dinâmica é composto de uma taxonomia simples projetada por especialistas do domínio: uma hierarquia de conceitos partindo de conceitos genéricos para conceitos mais específicos, apresentando apenas relacionamentos hierárquicos. Um conceito A é um sub-conceito do conceito B ($A \leq B$) se o conjunto de instâncias classificadas sob o conceito A é um subconjunto do conjunto formado pelas instâncias classificadas sob o conceito B : $Conj(A) \subseteq Conj(B)$. Tais relacionamentos hierárquicos representam relacionamentos taxonômicos do tipo É-UM. Neste caso, $A \leq B$ significa que ou $A \cong B$ (A equivale a B) ou que A é um descendente de B na taxonomia, de forma que os relacionamentos de subsunção definem uma ordem parcial entre conceitos. Taxonomias formadas por grafos acíclicos direcionados representando herança múltipla também são suportadas pelo modelo.

Os objetos que constituem a extensão (instâncias) são genéricos, de modo que qualquer tipo ou formato de objeto pode ser gerenciado pelo modelo de maneira uniforme. Estes objetos podem ser livremente classificados sob n ($n \geq 1$) conceitos em qualquer nível de abstração, ou seja, em qualquer nível na estrutura conceitual da taxonomia. Esta classificação multidimensional representa uma generalização do esquema de classificação monodimensional usado nas taxonomias convencionais. Num primeiro caso, objetos podem ser classificados tematicamente sob diferentes conceitos: por exemplo, um livro pode ser classificado sob “recuperação da informação”, “banco de dados”, “bibliotecas digitais”, etc. Num segundo caso, os objetos a serem classificados geralmente possuem diferentes características, perspectivas, ou facetas (tempo, localização, etc.), cada qual podendo ser descrita por uma taxonomia independente. Taxonomias com uma classificação

²⁹ O termo *objeto* é usado para denotar um item de informação abstrato que é atômico e cujo conteúdo, formato e meio físico de armazenamento são transparentes ao modelo. Objetos são geralmente chamados na literatura de documentos, itens ou recursos.

multidimensional são chamadas de taxonomias multidimensionais ou facetadas (SACCO; TZITZICAS, 2009).

Um conceito C pode ser visto como um rótulo abstrato (*label*) que representa todos os objetos classificados sob o respectivo conceito. Conceitos não devem ser confundidos com rótulos textuais (termos ou palavras). Apesar de os conceitos serem geralmente apresentados externamente como rótulos textuais, estes rótulos não necessitam ter qualquer conexão com os termos contidos nos objetos representados pelo conceito. Além disso, apesar de rótulos representando conceitos serem usados para carregar o significado do conceito para os usuários, este significado não é explorado pelo modelo. Desta forma, para propósitos de modelagem, cada conceito pode ser representado por um identificador numérico único.

Os conceitos são definidos a partir de suas extensões, ao invés das propriedades específicas que possuam. São definidos dois tipos diferentes de extensões para um dado conceito C . A extensão rasa de C (denotada por $Ext_Rasa(C)$) é definida como o conjunto de objetos diretamente classificados sob C . A extensão profunda de C (denotada por $Ext_Prof(C)$) inclui todas as extensões rasas dos conceitos que compõem a sub-árvore com raiz em C :

$$Ext_Prof(C) = \{d \mid d \in Ext_Rasa(C') \wedge (C' = C \vee C' \text{ é um descendente de } C)\}$$

Ou de forma equivalente:

$$Ext_Prof(C) = \{d \mid d \in Ext_Rasa(C') \wedge (C' \leq C)\}$$

Pode-se facilmente observar que no caso de conceitos terminais (representados pelas folhas da árvore taxonômica), as extensões rasas e profundas se equivalem. A FIGURA 15, a seguir, apresenta um exemplo de cada tipo de extensão. A *intensão* (ou estrutura conceitual) é apresentada acima da linha, enquanto a *extensão* (instâncias) é apresentada abaixo da linha. Os círculos representam conceitos, enquanto os objetos são representados por retângulos. Arcos sólidos representam relacionamentos hierárquicos entre conceitos, enquanto arcos pontilhados representam a classificação de objetos.

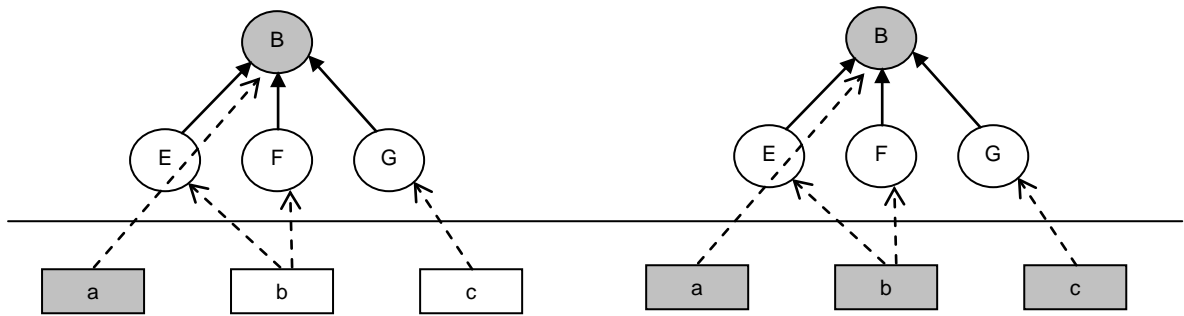


FIGURA 15 - Extensão rasa (a esquerda) e profunda (a direita) do conceito B.

Fonte: adaptado de (SACCO; TZITZICAS, 2009).

Cabe destacar que os conceitos de intensão e extensão são apresentados pelos autores no escopo restrito do modelo proposto para taxonomias dinâmicas, sendo que podem diferir do que é entendido e utilizado pela lógica e pela terminologia que se apropria da mesma.

Sacco e Tzitzicas (2009) afirmam que a semântica natural de um conceito C é dada pela sua extensão profunda, visto que um dado nível de abstração engloba todas as suas especializações. Quando nos referimos a animais, incluímos cães, gatos, aves, etc., e todos os objetos classificados sob estes conceitos. Desta forma a semântica de um conceito C será dada pelo conjunto dos objetos descritos pelo conceito C ($Objetos(C)$), que corresponde a extensão profunda do conceito ($Ext_Prof(C)$). Nesta pesquisa adotaremos esta visão.

Os relacionamentos hierárquicos presentes na taxonomia requerem que uma restrição de inclusão seja mantida. Se o conceito C' é um descendente de C na taxonomia, estão $Objetos(C') \subseteq Objetos(C)$. Se ambas as extensões, rasa e profunda, forem explicitamente armazenadas, temos uma forma de herança retroativa. Um objeto na extensão rasa de um conceito C é também classificado na extensão profunda de cada antecessor de C . A extensão profunda do conceito raiz da taxonomia inclui todo o universo U de objetos classificados pela taxonomia.

A extensão rasa somente é necessária se os objetos puderem ser classificados em qualquer nível da árvore conceitual. No caso mais simples, em que os objetos são classificados somente sob conceitos terminais (folhas), a extensão rasa não é necessária,

visto que ela será sempre vazia para conceitos não terminais e equivalente à extensão profunda para conceitos terminais.

Uma consequência imediata da interpretação de conceitos como conjuntos de objetos é que as operações lógicas em conceitos podem ser executadas por operações correspondentes, aplicadas aos conjuntos formados a partir de suas extensões. Desta forma, a base de dados pode ser manipulada e novos conceitos podem ser derivados a partir da combinação de outros conceitos através de operações lógicas (*and, or, not*).

2.4.1.1 Regra de Inferência Extensional

Nas taxonomias convencionais, os conceitos estão relacionados apenas por relacionamentos hierárquicos. Já numa taxonomia dinâmica, outros relacionamentos (não hierárquicos) podem ser inferidos através da extensão (instâncias), a partir da seguinte regra básica de inferência extensional:

Dois conceitos A e B são relacionados (denotado por $A \leftrightarrow B$) se existe ao menos um objeto d na extensão que é classificado simultaneamente sob o conceito A, ou um de seus descendentes, e sob o conceito B ou um de seus descendentes.

Por exemplo, um relacionamento entre “terrorismo” e “Nova York” pode ser inferido se existir um objeto classificado sob ambos os conceitos. Como “Nova York” é um conceito descendente de “EUA”, também poderá ser inferido um relacionamento entre “terrorismo” e “EUA”. A regra de inferência extensional pode ser vista como um mecanismo para inferir relacionamentos com base em evidência empírica. Seguem definições equivalentes para a regra de inferência extensional:

$A \leftrightarrow B$ se, somente se, $\exists o \in U: o \in \text{Objetos}(A) \wedge o \in \text{Objetos}(B)$

$A \leftrightarrow B$ se, somente se, $\text{Objetos}(A) \cap \text{Objetos}(B) \neq \emptyset$

Os relacionamentos hierárquicos na taxonomia são, por definição, um caso especial da regra de inferência extensional.

A regra básica de inferência extensional pode ser estendida de modo a cobrir os relacionamentos entre um dado conceito C e um conceito expresso por um subconjunto arbitrário S de objetos do universo: C está relacionado à S se, somente se, existir ao menos um objeto o em S que também pertença a $\text{Objetos}(C)$ ou, de forma equivalente, se $\text{Objetos}(C) \cap S \neq \emptyset$.

Desta forma, a regra de inferência extensional pode inferir relacionamentos não apenas entre conceitos básicos, mas também para qualquer combinação lógica de conceitos. O conjunto S poderia inclusive ser produzido a partir de outros métodos de recuperação da informação, como busca textual por exemplo. Desta forma, a taxonomia dinâmica pode inferir relacionamentos entre um conceito e um conjunto de objetos produzido por outros métodos de recuperação da informação. A regra de inferência extensional se reduz à regra básica quando o conjunto S corresponde à extensão profunda de um conceito C' .

Dada uma taxonomia, o conjunto de conceitos relacionados a um conjunto S de objetos é chamado de *conjunto relacionado de conceitos* ($CR(S)$) e é definido por: $CR(S) = \{C \mid \text{Objetos}(C) \cap S \neq \emptyset\}$. Devido à regra de inclusão para relacionamentos hierárquicos, se $C \in CR(S)$ e C' é um antecessor de C , então $C' \in CR(S)$. Por outro lado, se $C \notin CR(S)$ e C' é um descendente de C , então $C' \notin CR(S)$.

Durante a interação com a taxonomia, nos processos de busca e navegação, é extremamente útil saber, para cada conceito C relacionado à S , quantos objetos existem na interseção $\text{Objetos}(C) \cap S$. Chamamos este número de contagem relacionada, $cr(C/S)$: $cr(C/S) = |\text{Objetos}(C) \cap S|$. Por definição, $cr(C/U) = |\text{Objetos}(C)|$. O conjunto relacionado de conceitos $CR(S)$ pode ser reformulado em termos da contagem relacionada: $CR(S) = \{C \mid cr(C/S) > 0\}$.

2.4.1.2 Taxonomias Reduzidas e Exploração

Dado um conjunto de objetos S , a regra de inferência extensional pode ser usada para produzir um sumário conceitual de S de acordo com a taxonomia original, retirando da taxonomia todos os conceitos C que não são relacionados à S , ou seja, todo $C \notin CR(S)$. A taxonomia obtida a partir deste processo é chamada taxonomia reduzida, $TR(S)$.

A ideia principal para permitir a exploração centrada no usuário é usar a taxonomia para:

- (a) definir um *foco de interesse* a partir da combinação lógica de conceitos ou a partir de uma consulta externa, e
- (b) sumarizar os conceitos relacionados ao *foco de interesse* através da taxonomia reduzida, na qual os conceitos não relacionados são eliminados.

Desta forma, a taxonomia original pode se adaptar e sumarizar qualquer subconjunto de objetos do universo. Daí decorre o fato de ser chamada de taxonomia dinâmica. As taxonomias tradicionais (estáticas) somente são capazes de sumarizar o universo inteiro.

O foco de interesse inicial do usuário F_0 corresponde a todo o universo U , ou seja, todos os objetos na base de dados. O usuário tem inicialmente a seu dispor uma representação em forma de árvore (hierárquica) da taxonomia inicial, abarcando toda a base de dados. Nesta representação, cada conceito C possui uma contagem relacionada $cr(C/F_i)$. A contagem relacionada é uma função do foco de interesse atual F_i . Como inicialmente $F_0 = U$, $cr(C/F_0)$ é equivalente a $|Objetos(C)|$, isto é, a cardinalidade da extensão profunda do conceito C ou, em outras palavras, o número total de objetos classificados sob o conceito C ou um de seus descendentes.

No caso mais simples, o usuário seleciona um conceito C na taxonomia, definindo um novo foco de interesse. Esta operação é chamada de *zoom* (SACCO; TZITZICAS, 2009). A operação de *zoom* altera o estado atual de duas maneiras. Primeiro, o novo foco de interesse F_i corresponde a $F_{i-1} \cap Objetos(C)$. Os objetos que não pertencem ao novo foco são descartados. Segundo, a representação hierárquica da taxonomia é modificada para sumarizar o novo foco de interesse. Apenas os conceitos relacionados à F_i são mantidos e a contagem relacionada a cada conceito C' é atualizada para refletir o número de objetos no foco F_i que são classificados sob o conceito C' , ou seja, $cr(C'/F_i)$. A taxonomia reduzida é derivada a partir da taxonomia inicial a partir da retirada (“poda”) de todos os conceitos não relacionados à F_i , representando assim um sumário conceitual do conjunto de objetos que compõem o foco de interesse atual F_i , da mesma forma que a taxonomia original representava um sumário conceitual de todo o universo.

Por motivo de simplicidade, foi assumido que os conceitos retirados não são exibidos para o usuário na taxonomia reduzida. No entanto, o simples fato de um conceito ter sido retirado pode representar informação importante para o usuário. Neste caso, uma alternativa é apresentar os conceitos que foram retirados num formato diferenciado dos demais, indicando que eles possuem uma contagem relacionada igual a zero e não são passíveis de serem selecionados nas operações seguintes.

O processo de recuperação da informação pode ser visto como um processo interativo de redução ou filtragem da base de dados: o usuário seleciona um foco de interesse, que restringe a base de dados a partir do descarte de todos os objetos que não pertencem ao respectivo foco. Apenas os conceitos usados para classificar os objetos

presentes no foco de interesse e seus antecessores na hierarquia são mantidos na taxonomia reduzida. Estes conceitos, que resumam o foco de interesse atual, são somente aqueles que podem ser utilizados para refinamentos posteriores. A partir da perspectiva de interação homem-máquina, o usuário é efetivamente guiado para atingir seu objetivo através de uma lista clara e consistente de alternativas possíveis. Este tipo de interação é geralmente chamado de navegação guiada (SACCO; TZITZICAS, 2009).

As figuras a seguir exemplificam o mecanismo de interação ora descrito. A FIGURA 16 representa uma taxonomia dinâmica. A *intensão* (estrutura conceitual) aparece acima da linha, enquanto a *extensão* (objetos ou instâncias) abaixo da mesma. Os arcos sólidos denotam relacionamentos hierárquicos entre os conceitos, enquanto os arcos pontilhados representam a classificação dos objetos.

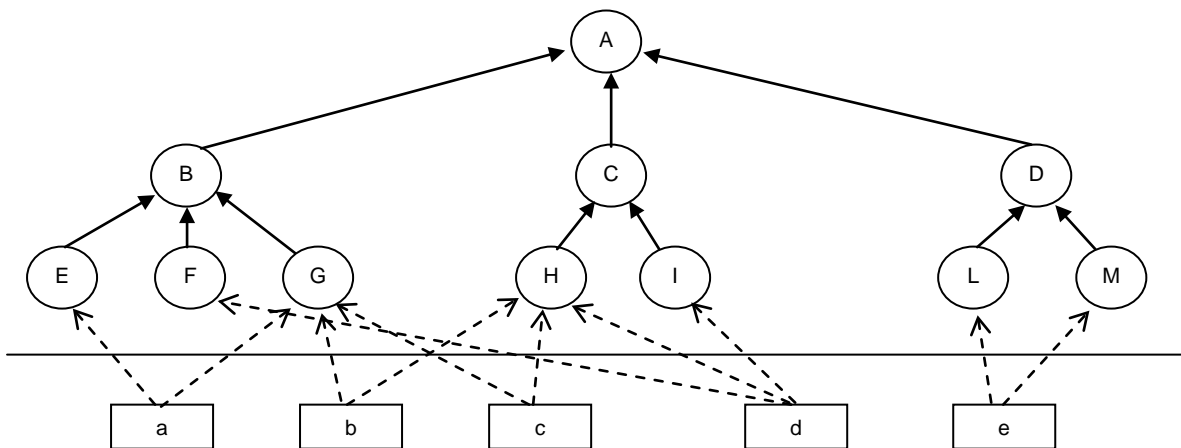


FIGURA 16 - Taxonomia dinâmicas

Fonte: adaptado de (SACCO; TZITZICAS, 2009).

Inicialmente, o usuário seleciona o conceito C, caracterizando uma operação de *zoom* a partir do respectivo conceito. Primeiro, o conjunto de objetos de interesse (extensão do foco de interesse) é computado. Neste caso, o conjunto de interesse corresponde a extensão profunda do conceito C, $Ext_Prof(C) = \{b, c, d\}$ (FIGURA 17). Este conjunto é computado a partir dos arcos pontilhados que chegam ao conceito C ou a um de seus descendentes $\{H, I\}$. Todos os objetos que não pertencem ao conjunto de interesse podem ser removidos da extensão, que é reduzida.

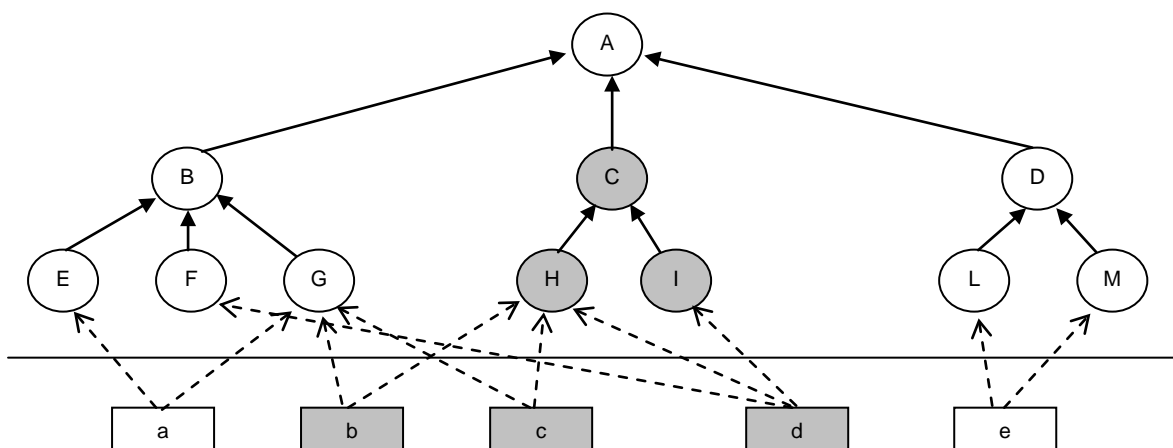


FIGURA 17 - Cálculo da extensão profunda do conceito C.

Fonte: adaptado de (SACCO; TZITZICAS, 2009).

Em seguida, a taxonomia reduzida para o conjunto de interesse é computada. Em primeiro lugar, todos os conceitos relacionados ao foco de interesse ($CR(C)$) é computado. De acordo com a regra de inferência extensional, eles correspondem aos conceitos sob os quais pelo menos um objeto do conjunto de interesse é classificado. Na FIGURA 18, este conjunto é computado a partir dos arcos originados nos objetos que compõem o conjunto de interesse. O conjunto de conceitos relacionados ao foco de interesse é $CR(C) = \{F, G, H, I, B, C, A\}$. Lembrando que, se um conceito é relacionado a um conjunto de objetos S, então todos os seus antecessores também o serão.

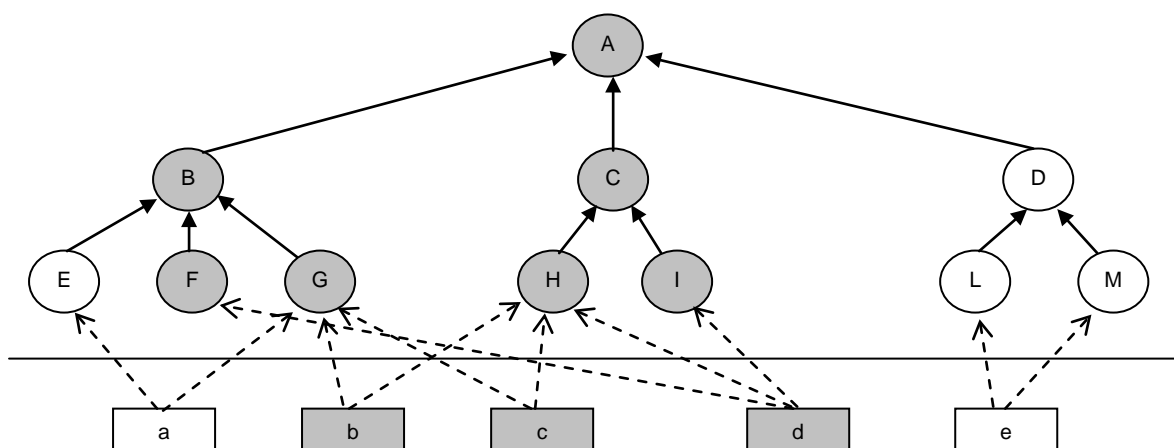


FIGURA 18 - Conjunto relacionado de conceitos para C ($CR(C)$).

Fonte: adaptado de (SACCO; TZITZICAS, 2009).

Finalmente, na FIGURA 19, todos os conceitos não pertencentes a $CR(C)$ são removidos da taxonomia, produzindo uma taxonomia reduzida que descreve somente os objetos pertencentes ao foco de interesse atual, definido a partir da seleção do conceito C.

Uma operação subsequente de *zoom* somente poderá ser executada a partir da seleção de conceitos presentes na taxonomia reduzida, ou seja, conceitos pertencentes a $CR(F_1)$. Isto garante, por construção, que nunca haverá resultados vazios e que, em qualquer estágio de interação, os conceitos não relacionados (irrelevantes) serão descartados. A seleção de um conceito C_2 a partir da taxonomia reduzida irá determinar um novo foco de interesse $F_2 = F_1 \cap Ext_Prof(C_2)$. Por exemplo, a seleção do conceito G no contexto do foco F_1 (FIGURA 19) irá determinar $F_2 = \{b, c\}$ e $CR(F_2) = \{G, H, B, C, A\}$.

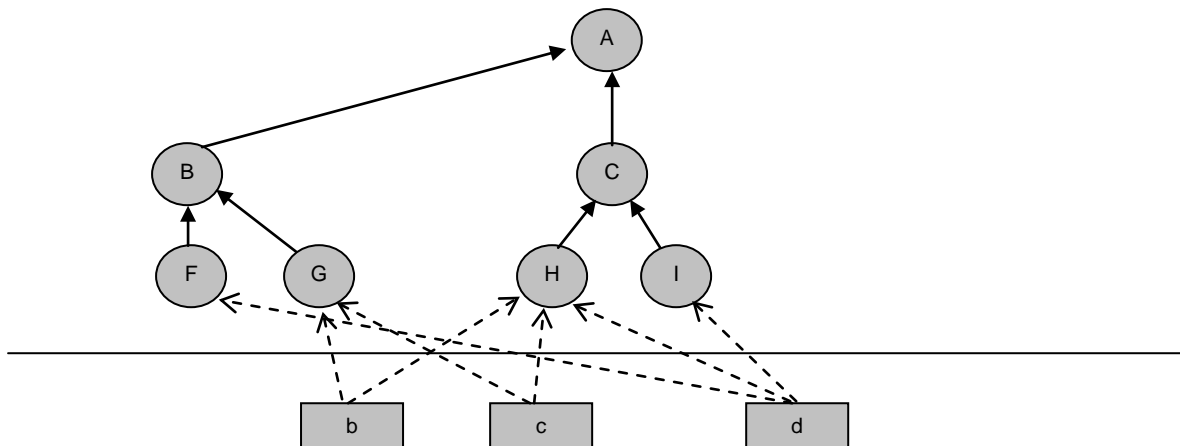


FIGURA 19 - Taxonomia reduzida para o foco definido a partir de C.

Fonte: adaptado de (SACCO; TZITZICAS, 2009).

A derivação de relacionamentos entre conceitos a partir da regra de inferência extensional traz importantes implicações para a modelagem conceitual da taxonomia. Em primeiro lugar, um ponto fundamental é que os relacionamentos entre conceitos não precisam ser antecipados, podendo ser inferidos a partir da classificação atual. Isto simplifica o projeto e manutenção da taxonomia. Nas abordagens tradicionais, somente os relacionamentos explicitamente descritos no esquema conceitual podem ser utilizados pelos usuários para navegação e recuperação. Desta forma, todos os relacionamentos possíveis precisam ser antecipados e descritos: tarefa difícil, senão impossível. Nas taxonomias dinâmicas, cujo modelo de operação fora descrito acima, nenhum tipo de relacionamento, além dos relacionamentos hierárquicos (paradigmáticos), precisa ser definido. Os relacionamentos conceituais (sintagmáticos) são automaticamente derivados a partir da

classificação. Por esta razão, as taxonomias dinâmicas se adaptam facilmente a novos relacionamentos e são capazes de permitir a descoberta de relacionamentos novos e não previstos.

Em segundo lugar, as taxonomias dinâmicas são capazes de sintetizar assuntos compostos, que não precisam ser representados explicitamente. Esta propriedade permite lidar com o problema do crescimento combinatorial (SACCO, 2000) presente nas taxonomias tradicionais, nas quais os assuntos compostos são explicitamente representados. Sacco (2000) propõem diretrizes para a construção de taxonomias dinâmicas compactas e de fácil entendimento para os usuários. A maioria destas diretrizes é similar a aquelas propostas para a construção de classificações facetadas (ver seção 2.2.1): a taxonomia deve ser organizada como um conjunto de sub-taxonomias independentes e ortogonais, cada uma representando uma perspectiva ou faceta, a ser usada para a descrição dos objetos. Outras diretrizes propostas por Sacco (2000) são voltadas para a usabilidade na interação homem-máquina, recomendando um *fan-out* (número de descendentes diretos de um dado conceito) de no máximo 10, e um número de níveis (profundidade da taxonomia) de no máximo 4. Um *fan-out* maior resulta numa lista muito longa de conceitos que precisa ser percorrida pelo usuário para encontrar um dado conceito; já uma profundidade maior que 4 níveis torna difícil percorrer a hierarquia. Estes limites para *fan-out* e profundidade sugerem taxonomias com um número de conceitos terminais (folhas) entre 1.000 e 10.000, que são adequadas para a manipulação pelo usuário no contexto de uma taxonomia facetada implementada numa interface computacional.

Num exemplo que segue as diretrizes para a construção facetada da taxonomia, e considerando o assunto composto “pinturas francesas do século XV”, teríamos: uma sub-taxonomia de locais (da qual o conceito “França” é descendente), uma sub-taxonomia temporal (da qual “Século XV” é descendente) e uma sub-taxonomia de artes (da qual “Pintura” é descendente). Os objetos a serem classificados sob o assunto composto seriam classificados pela taxonomia da seguinte forma: *Local>França*, *Tempo>Século XV* e *Arte>Pintura*. A regra de inferência extensional poderia então estabelecer o relacionamento entre estes conceitos fazendo com que o assunto composto pudesse ser recuperado por operações de *zoom* executadas sob qualquer combinação das facetas.

O projeto facetado da taxonomia dinâmica torna simples para o usuário focar num determinado conceito como, por exemplo “Século XV”, e visualizar imediatamente, a partir da aplicação da regra de inferência extensional, todos os conceitos relacionados como literatura, pintura, política, etc. Numa abordagem que faça uso do assunto composto, estas

correlações não seriam possíveis devido ao fato de estarem ocultas sob o rótulo do assunto composto.

Outra consequência do projeto facetado para a taxonomia é que ele possibilita a quebra dos relacionamentos em conceitos primitivos e independentes (SACCO, 2000). Estes conceitos primitivos tendem a ser mais estáveis no tempo, fazendo com que o esquema da taxonomia necessite de pouca manutenção. São os relacionamentos entre estes conceitos primitivos que tendem a variar ao longo do tempo. As taxonomias dinâmicas podem facilmente acomodar esta variação já que são capazes de descobrir automaticamente estes relacionamentos a partir da regra de inferência extensional, ou seja, por evidência empírica.

As taxonomias dinâmicas projetadas de forma facetada requerem um esquema conceitual simples e são capazes de se adaptar a relacionamentos novos e não previstos, permitindo aos usuários descobrir estes relacionamentos nas operações de navegação e recuperação. O controle do usuário e o mecanismo adaptativo claro e consistente proporcionado pelas taxonomias dinâmicas encorajam o usuário a experimentar e explorar, proporcionando a sensação de ter considerado todas as alternativas para chegar ao resultado. Sacco e Tizizicas (2009) afirmam que a maior facilidade de interação é resultado do uso de uma única representação para consulta (definição do foco de interesse) e sumarização (apresentação da taxonomia reduzida). Desta forma, o usuário lida com uma representação conceitual única da base de dados. Esta propriedade é chamada pelos autores de exploração autoadaptável (SACCO; TZITZICAS, 2009).

2.4.1.4 Taxonomias Dinâmicas e Busca Facetada

Neste capítulo, temos usado os termos taxonomias dinâmicas, sistema de busca facetada e taxonomias facetadas como sinônimos. No entanto, o uso dos princípios da classificação facetada para a construção da taxonomia dinâmica é apenas uma diretriz de projeto. O modelo para taxonomias dinâmicas requer somente uma taxonomia multidimensional (que permita a classificação de um objeto por mais de um conceito), que não precisa ser necessariamente facetada. Os conceitos de inferência extensional e de taxonomias reduzidas, que são fundamentais para caracterizar uma taxonomia dinâmica, não estão implícitos (nem são necessários) em uma classificação facetada. Além disso, como afirmam Sacco e Tizizicas (2009), os mecanismos de busca facetada são apresentados na literatura através de exemplos, não tendo sido formalmente definidos, o que causa certa confusão.

Conforme discutido anteriormente, o projeto baseado na classificação facetada contribui para a construção de taxonomias dinâmicas compactas e de fácil entendimento e uso pelos usuários. Outra vantagem observada diz respeito à estabilidade e como consequência a pouca necessidade de manutenção da taxonomia, além de proporcionar uma estrutura mais enxuta, intuitiva e com melhor capacidade de convergência. Conclui-se desta forma que o projeto da taxonomia dinâmica baseado nos princípios da classificação facetada é desejável, mas não obrigatório.

Neste trabalho de pesquisa optamos por utilizar um modelo formal para taxonomias dinâmicas, conforme proposto por Sacco e Tiziticas (2009). Este modelo foi utilizado para a implementação do mecanismo de busca em um protótipo computacional. O projeto e construção da taxonomia, realizado no contexto do trabalho de pesquisa de Maculan (2011) e utilizado na presente pesquisa, se baseia nos princípios da classificação facetada.

2.5 Avaliação de bibliotecas digitais

Em alinhamento com o objetivo de avaliar as vantagens da adoção de uma abordagem facetada para a representação e organização do acervo de uma biblioteca digital de teses e dissertações, bem como dos novos mecanismos de recuperação da informação implementados no protótipo computacional, esta seção trata do tema avaliação de bibliotecas digitais. São apresentados e discutidos os principais modelos de avaliação propostos na literatura.

Entende-se por avaliação o processo sistemático de determinar o mérito, valor, e importância de algo. Uma avaliação possui objetivo, metodologia e mecanismos ou técnicas. O objetivo da avaliação sugere alguns critérios de avaliação (como por exemplo, desempenho), que às vezes podem ser subdivididos em parâmetros de avaliação (como por exemplo, tempo de resposta, disponibilidade). Algumas vezes estes parâmetros são quantitativos, ou seja, uma unidade de medida numérica pode ser definida e associada ao parâmetro. Outras vezes são qualitativos e não quantificáveis numericamente. No caso dos parâmetros quantitativos, os valores para suas medidas podem ser comparados em diferentes avaliações. Uma boa avaliação deve ser transferível (pode ser aplicada a outros contextos) e verificável (pode ser confirmada por outras pessoas).

Fuhr e outros (2007), ao analisar a avaliação de bibliotecas digitais, sugerem três tipos principais de avaliação:

- *Comparativa*. Uma atividade de avaliação que busca denotar o grau com que um sistema ou componente possui uma propriedade ou qualidade maior ou menor que outro sistema ou componente semelhante.
- *Formativa*. Uma atividade de avaliação que é conduzida em paralelo com as fases de desenvolvimento de um sistema de biblioteca digital. Sendo parte do processo de desenvolvimento, a avaliação formativa busca minimizar as imperfeições do sistema antes que o mesmo seja terminado.
- *“Somativa” (summative)*. Uma atividade de avaliação conduzida depois de terminado o processo de desenvolvimento de modo a medir o desempenho do sistema em condições de uso real.

Segundo Saracevic (2000), a avaliação é raramente objeto de discussão da pesquisa e prática relacionada a bibliotecas digitais. O autor destaca que, apesar do grande desenvolvimento da área, ocorrido a partir da década de 1990, a avaliação se destaca pela sua ausência ou presença mínima nos diversos projetos relacionados à área. O autor enumera algumas possíveis explicações para este fato:

- Talvez ainda seja cedo demais no estágio de evolução das bibliotecas digitais para que se possa definir e realizar avaliações formais.
- Avaliações informais podem ser suficientes neste estágio de desenvolvimento.
- Não existe interesse na avaliação.
- A pressão pela evolução não permite tempo para avaliação.
- Talvez a avaliação de bibliotecas digitais seja complexa demais, de modo que mesmo quando desejada, não possa ser implementada devido ao estágio ainda imaturo do seu desenvolvimento conceitual.

O autor acredita que este último ponto seja o principal limitador das atividades de avaliação e discute várias questões conceituais e teóricas relacionadas à avaliação de bibliotecas digitais, além de propor alguns conceitos e abordagens para avaliação destes sistemas.

Entre as poucas iniciativas de avaliação formal, Saracevic (2000) destaca o projeto da Biblioteca Digital Alexandria (FREW *et al.*, 1998). A abordagem utilizada incluiu uma série de estudos de usuários envolvendo diferentes comunidades e se concentrando na funcionalidade e usabilidade proporcionada por diferentes estratégias de projeto. Os estudos

de usabilidade se tornaram uma das abordagens mais populares para a avaliação de bibliotecas digitais, mas, como destaca Saracevic (2000), usabilidade é apenas um dos critérios possíveis e necessários. Segundo o autor, os estudos de usuários, apesar de úteis para o entendimento de como as pessoas usam os sistemas, não constituem avaliação por si só, mesmo que tenham implicações e forneçam importantes critérios a serem utilizados na avaliação de bibliotecas digitais.

As principais questões que permeiam qualquer tipo de avaliação são: Porque avaliar? O que avaliar? Como avaliar? Para quem avaliar? A resposta para a primeira questão (porque avaliar) serve de base para a escolha da abordagem a ser utilizada na avaliação. Saracevic (2000) se concentra numa abordagem centrada no sistema, o que se justifica por ser a abordagem mais indicada e utilizada para a avaliação de sistemas de informação, incluindo bibliotecas digitais. No entanto, o autor reconhece as vantagens e desvantagens desta abordagem. O pressuposto básico das abordagens centradas no sistema é o de que a avaliação parte de algum aspecto relacionado ao desempenho (*performance*) e então define objetivos e adota opções mais específicas.

Um sistema pode ser considerado um conjunto de elementos que interagem entre si para atingir determinados objetivos. Estes elementos ou componentes interagem para executar determinadas funções ou processos, de modo a atingir os objetivos. Qualquer sistema, incluindo uma biblioteca digital, existe dentro de um determinado ambiente e interage com o mesmo. Saracevic (2000) observa que geralmente é difícil e arbitraria a definição dos limites entre o sistema e o ambiente. Esta dificuldade impacta diretamente a avaliação de uma biblioteca digital na medida em que coloca questões como onde começa e termina a avaliação? O que deve ser incluído e o que deve ser excluído da avaliação? Segundo Saracevic (2000), a resposta para estas questões determinam os elementos constituintes de uma biblioteca digital. Neste contexto, a avaliação representa uma estimativa do desempenho ou funcionalidade do sistema, ou parte dele, com relação a um conjunto de objetivos. O desempenho pode ser avaliado como:

- Eficácia: quão bem o sistema (ou parte dele) atinge o objetivo para o qual foi projetado?
- Eficiência: com que custo (financeiro, tempo, esforço) ele opera?
- Uma combinação dos anteriores: relação entre efetividade e custo.

A avaliação deve determinar os critérios de desempenho que serão utilizados. Segundo Saracevic (2000) uma avaliação deve atender a certos requisitos, envolvendo escolhas e decisões relacionadas a:

- Elementos para avaliação. O que avaliar? O que é uma biblioteca digital? Quais os elementos (componentes e processos) que serão avaliados?
- Contexto da avaliação. Seleção de um objetivo, estrutura, ponto de vista, ou níveis de avaliação. Qual o nível para a avaliação? O que é crítico para um dado nível selecionado? Finalmente, quais os objetivos a serem selecionados para um dado nível?
- Quais critérios de desempenho estão relacionados aos objetivos selecionados. Quais os parâmetros de desempenho? Qual a dimensão ou característica a ser avaliada?
- Quais as medidas a serem utilizadas para os critérios de desempenho. Que medida específica será utilizada para um dado critério?
- Metodologia da avaliação. Quais os instrumentos de medição serão utilizados? Quais as amostras? Quais os procedimentos para a coleta de dados e para a análise dos mesmos?

Segundo Saracevic (2000), uma clara resposta para cada uma destas questões é requisito para a avaliação de uma biblioteca digital. No entanto, o autor reconhece que estas respostas ainda não estão claras, afirmando que não existe consenso relacionado aos critérios, medidas e metodologias para a avaliação de bibliotecas digitais, e até mesmo para o que deve ser avaliado e em que contexto. Buscando clarificar os conceitos e fornecer possíveis respostas para as três primeiras questões, o autor discute aspectos conceituais e teóricos relacionados à avaliação de bibliotecas digitais.

Para definir o que deve ser avaliado é necessária uma visão clara do que consiste uma biblioteca digital. A partir de várias definições encontradas na literatura, Saracevic (2000) buscou determinar os elementos passíveis de serem avaliados e apresentou a seguinte lista:

- Coleção digital, recursos;
- Seleção, coleta, propriedade;
- Distribuição, conexões, links;
- Organização, estrutura, armazenamento;
- Interpretação, representação, metadados;
- Gerenciamento;
- Preservação, persistência;
- Acesso;
- Redes físicas, distribuição;
- Interfaces, interação;

- Busca, recuperação;
- Serviços, disseminação, entrega;
- Disponibilidade;
- Assistência, orientação;
- Uso, usuários, comunidades;
- Segurança, privacidade, políticas, aspectos legais, licenças;
- Gerenciamento, operação, equipe;
- Custo, economia;
- Integração, cooperação com outras fontes, bibliotecas ou serviços.

A partir da análise da lista acima fica claro que não existe “avaliação de biblioteca digital” por completo, ou seja, nenhuma avaliação pode cobrir a maioria dos elementos ou constructos presentes em uma biblioteca digital. A avaliação deve selecionar e explicitar os elementos que serão avaliados, reconhecendo e justificando a ênfase que será dada aos elementos incluídos na avaliação. A qualidade de um sistema complexo nunca será melhor do que a qualidade de seu componente mais fraco (FUHR *et al.*, 2007). Desta forma quando mais ampla a avaliação, ou seja, quando maior o número de elementos ou componentes avaliados, melhores os resultados. Fuhr e coletas (2007), ao analisar o estado da arte, reconhecem que as bibliotecas digitais representam um domínio complexo e interdisciplinar, sendo que quando se considera uma definição mais ampla, os seguintes aspectos de uma biblioteca digital devem ser considerados:

- uma infraestrutura tecnológica (redes, softwares e serviços);
- uma coleção estruturada de objetos digitais (coleções); e
- uma organização com pessoas desempenhando diferentes papéis e atuando sob diversos padrões de interação.

Os autores concluem que a avaliação ampla de bibliotecas digitais é uma tarefa bastante difícil devido à multiplicidade e complexidade dos seus elementos constituintes.

Com relação ao contexto de avaliação, ou seja, a seleção do nível de objetivos a serem tratados pela avaliação, Saracevic (2000) divide os objetivos do sistema (biblioteca digital) em sete níveis. Estes níveis não são mutuamente exclusivos, como observa o autor. Os primeiros três níveis são mais centrados no usuário, enquanto os três últimos são centrados no sistema, sendo que existe o nível de *interface* entre os dois grupos. O autor aponta questões de desempenho referentes a cada nível.

- *Nível social.* Quão bem a biblioteca digital oferece suporte às necessidades, demandas e práticas de uma comunidade? É de difícil avaliação devido à diversidade de objetivos de uma sociedade ou comunidade.
- *Nível institucional.* Quão bem a biblioteca digital oferece suporte a missão e objetivos institucionais? Quão bem ela se integra a outros recursos institucionais? Está fortemente ligado os objetivos institucionais da organização. Também difícil de ser avaliado.
- *Individual.* Quão bem a biblioteca digital (ou um serviço dela) oferece suporte à necessidade de informação, tarefas e atividades de pessoas como usuários individuais ou grupos de usuários com características comuns? A maioria das avaliações é realizada neste nível, principalmente por ser mais direto e simples de ser avaliado, no entanto, diferenças nas percepções individuais podem gerar dificuldades de generalização.
- *Interface.* Quão bem uma dada interface do sistema fornece suporte ao acesso, busca, navegação, exploração do acervo e interação com a biblioteca digital? Estas questões podem tanto ser aplicadas ao usuário, como serem respondidas usando-se dados coletados diretamente a partir do sistema.
- *Engenharia.* Quão bem atuam o *hardware*, redes, e outros componentes da infraestrutura computacional? Estas questões produzem medidas mais replicáveis e generalizáveis.
- *Processamento.* Quão bem atuam os procedimentos, técnicas, algoritmos, operações, etc.? Também geram medidas bastante sistemáticas, no entanto são passíveis de variações devido a diferenças de configuração, capacidade, etc.
- *Conteúdo.* Quão bem a coleção ou os recursos de informação são selecionados, representados, organizados, estruturados, e gerenciados? Apesar de também ser sistemático, gera outras questões relacionadas como para quem e com que propósito?

Saracevic (2000) observa ainda que as bibliotecas digitais geralmente são avaliadas em apenas um nível. Segundo o autor, este isolamento entre os níveis representa um grande desafio para a área.

Devem ser determinados critérios para avaliação em cada nível. Nas avaliações analisadas por Saracevic (2000), um determinado nível era sempre implícita ou explicitamente selecionado e, com ele, uma série de critérios eram adotados. Segundo o autor, o nível mais utilizado nas avaliações era o nível *Individual*, sendo que, o critério mais

utilizado era a *Usabilidade*. O autor, ao apresentar o ponto de vista de outros autores que defendem as abordagens centradas no usuário, relegando as demais, defende as abordagens mistas: "... a questão não é se devemos adotar abordagens centradas no sistema ou nas pessoas (usuários). A questão é menos ainda relacionada à competição entre estas abordagens. A questão é como fazer com que ambas possam trabalhar juntas" SARACEVIC (2000).

Após analisar os critérios de avaliação utilizados em bibliotecas e serviços de informação, sistemas de recuperação da informação e interfaces homem-máquina, Saracevic (2000) enumerou uma lista de critérios que podem ser adaptados para a avaliação de bibliotecas digitais:

- Critérios biblioteconômicos tradicionais.
 - Coleção: propósito, assunto, escopo, autoridade, cobertura, atualidade, audiência, custo, formato, tratamento, preservação, persistência;
 - Informação: exatidão, conveniência, interconexões, representação, exclusividade, comparabilidade, apresentação, disponibilidade, propriedade;
 - Uso: acessibilidade, disponibilidade, "encontrabilidade", usabilidade;
 - Padronização: relacionada a vários elementos e processos.
- Critérios da área de recuperação da informação.
 - Relevância (conduz a medidas de precisão e recuperação);
 - Satisfação, sucesso;
 - Indexação, busca, características da saída.
- Critérios da área de interfaces e interação homem-máquina.
 - Usabilidade, funcionalidade, esforço;
 - Conveniência das tarefas, falhas;
 - Conectividade, restauração;
 - Características de projeto;
 - Navegação, exploração;
 - Serviços, ajuda.

Fuhr e outros (2007) realizaram um amplo estudo sobre as atividades relacionadas à avaliação de bibliotecas digitais. O estudo resultou em uma descrição do estado da arte nesta área. Os autores também propõem um novo arcabouço conceitual para a avaliação de bibliotecas digitais, que pode ser usado para registrar, descrever e analisar as pesquisas nessa área. O arcabouço inclui uma metodologia para classificação dos procedimentos de avaliação. O objetivo foi fornecer um conjunto de diretrizes flexíveis e adaptáveis para a avaliação de bibliotecas digitais. Os autores reconhecem que os métodos e métricas para

avaliação de bibliotecas digitais podem variar de acordo com a perspectiva adotada. Neste sentido uma biblioteca digital pode ser vista como uma instituição, um sistema de informação, como um recurso tecnológico, como uma coleção de recursos de informação, ou como serviços.

Uma biblioteca digital representa um tipo particular de sistema de informação composto de um conjunto de elementos: uma ou mais coleções de recursos informacionais, um sistema computacional que oferece diversos serviços relacionados às coleções, pessoas, e o ambiente (uso) para o qual o sistema foi desenvolvido. Fuhr e outros (2007), assim como Saracevic (2000) defendem a necessidade do estabelecimento de acordos com relação aos conceitos centrais, pressupostos, parâmetros e critérios para a avaliação de bibliotecas digitais.

Em uma biblioteca digital os casos de uso mais comuns e importantes são aqueles relacionados à interação entre os usuários e o conteúdo, ou seja, o acesso. Os esquemas de avaliação oriundos da área de recuperação da informação se baseiam no conceito de *relevância*, que é central para o entendimento da interação entre usuários, uso, necessidade de informação e conteúdos. Segundo Fuhr e outros (2007), relevância – qualidade momentânea de um texto que faz com que o mesmo tenha valor para a seleção e leitura – varia em função da tarefa, características do texto, preferências e conhecimento prévio do usuário, situação, restrições de tempo e vários outros fatores. A relevância conforme tratada pelas avaliações na área de recuperação da informação não leva em consideração a satisfação do usuário, qualidade da informação, relações entre documentos, confiabilidade da informação ou o contexto do usuário. A partir do conceito de relevância, o desempenho do sistema de recuperação da informação é geralmente avaliado utilizando-se as métricas relacionadas à *precisão* e *recuperação*. Fuhr e outros (2007) argumentam que os esquemas de avaliação devem ser projetados especificamente para a área de bibliotecas digitais. Neste sentido, a sessão de uso da biblioteca digital deve ser tomada como unidade de avaliação, e não as consultas isoladamente. Isto se justifica devido ao fato de os usuários interagirem com a informação, representada pelos conteúdos que eles acessam. Eles recuperam conteúdos por uma determinada razão; eles avaliam continuamente a informação que acessam; eles formulam, reformulam e alteram seus objetivos dinamicamente na medida em que acessam os conteúdos; a informação que eles acessam altera suas perspectivas. Os usuários julgam os documentos a partir de uma leitura detalhada, ou apenas pela visualização rápida em comparação com outros documentos disponíveis. Uma sessão de acesso à informação representa uma unidade real e perceptível, mas os sistemas geralmente não consideram a interação em níveis de abstração acima de ciclos de consulta-resposta (FUHR *et al.*, 2007). Os autores sugerem a desconstrução e o

aprimoramento do conceito de relevância de modo que possa considerar aspectos relacionados ao contexto.

Fuhr e outros (2001) partem das definições iniciais do modelo conceitual para bibliotecas digitais desenvolvido pela Rede DELOS (ver seção 2.1.5.1), para definir critérios e métricas a serem utilizadas na avaliação de bibliotecas digitais. O modelo DELOS identifica três componentes não ortogonais: *Usuários*, *Dados/Coleções* (conteúdo), e *Sistema/Tecnologia*. A FIGURA 20, a seguir, ilustra tais componentes no domínio das bibliotecas digitais. A definição do conjunto de usuários determina a extensão e o conteúdo das coleções (seta cheia interligando *Usuários* a *Dados/Coleções*). A natureza das coleções determina as possíveis tecnologias a serem empregadas (seta cheia interligando *Dados/Coleções* a *Tecnologia*). A atratividade das coleções para os usuários e a facilidade com que eles podem utilizar a tecnologia determina a extensão do uso da biblioteca digital (setas finas mostram a interação homem-máquina, enquanto as setas pontilhadas mostram a contribuição coletiva do usuário, das coleções e da tecnologia para o uso observado da biblioteca digital).

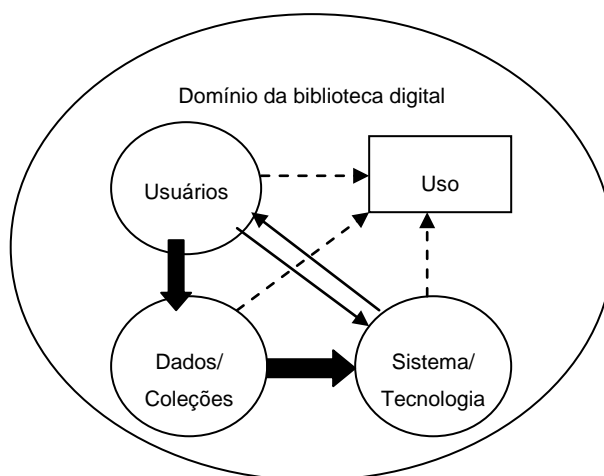


FIGURA 20 - Componentes de uma biblioteca digital

Fonte: adaptado de Fuhr e outros (2001)

Com base na análise das interações entre os componentes de uma biblioteca digital, ou seja, nas relações entre *Usuário-Conteúdo*, *Conteúdo-Sistema* e *Usuário-Sistema*, Fuhr e outros (2007) definiram um arcabouço conceitual composto de três dimensões para a avaliação de bibliotecas digitais: o Modelo Tríplice de Interação (FIGURA 21). Neste modelo o componente *Usuário* representa ambos os componentes *Usuários* e *Uso* da FIGURA 20.

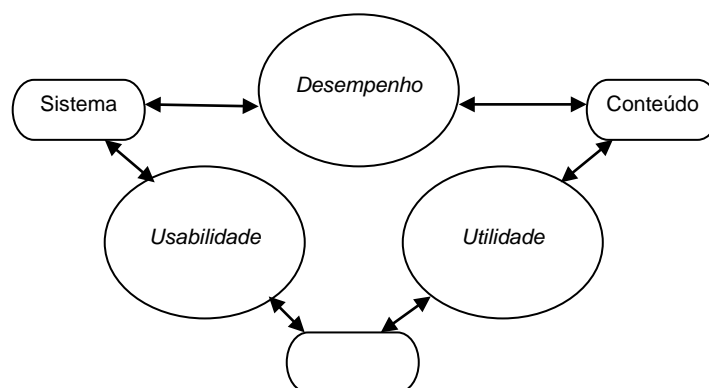


FIGURA 21 - Modelo Tríplice de Interação

Fonte: adaptado e Fuhr e outros (2001)

O par *Conteúdo-Sistema* está relacionado a atributos de desempenho ou *performance* (precisão, recuperação, tempo de resposta, etc.), o par *Usuário-Sistema* está relacionado a aspectos de usabilidade (efetividade, satisfação, etc.), e o par *Usuário-Conteúdo* está relacionado a aspectos de utilidade. Estes pares representam as relações – interações entre componentes de uma biblioteca digital – e definem uma estrutura composta de três dimensões para avaliação de bibliotecas digitais chamada Modelo Tríplice de Interação (FUHR *et al.*, 2007).

O *Usuário* representa um componente complexo e em constante evolução, segundo os autores. Independente da metodologia selecionada, em uma avaliação orientada aos usuários, o objetivo é obter um entendimento de suas necessidades para o projeto tecnológico e das coleções. Marchionini (2000) destaca duas questões que devem ser consideradas: a avaliação deve ser longitudinal para que se possa capturar um conjunto de dados rico e confiável para análise, e deve ser multifacetada, ou seja, deve utilizar uma combinação de métodos. Cada avaliação deve definir um conjunto de medidas e métodos de coleta de dados.

O *Conteúdo* é de extrema importância, já que representa a principal razão da interação dos usuários com a biblioteca digital, na medida em que eles (usuários) procuram satisfazer suas necessidades de informação. A utilidade percebida no conteúdo é o principal critério de seleção do usuário. Mas como já discutido anteriormente, durante uma sessão de busca, a percepção da utilidade do conteúdo para uma necessidade de informação pode se alterar. Segundo Fuhr e outros (2007) a avaliação do conteúdo deve ser realizada sob regras e padrões claros, de modo a garantir o acesso do usuário a informações de qualidade e apropriadas às suas necessidades.

O *Sistema* é o componente mais visível do processo de interação, e está sob o controle de seus desenvolvedores. O objetivo do sistema, no contexto de uma biblioteca digital, deve ser possibilitar o acesso de pessoas (usuários) ao conhecimento (representado pelo acervo) a qualquer hora e a partir de qualquer local, de uma forma amigável e flexível, superando barreiras de distância, língua e cultura, e usando vários dispositivos de redes interconectados (CANDELA *et al.*, 2007b). A avaliação do componente tecnológico (sistema) de uma biblioteca digital tem maior sentido quando executada em relação aos demais componentes: usuários e conteúdos (FUHR *et al.*, 2007).

Conforme afirmado anteriormente, segundo Fuhr e outros (2007), o processo de avaliação deve se basear nas relações entre os componentes da biblioteca digital:

- *Usabilidade*, que define a qualidade da interação entre *Usuário* e *Sistema*. Ela ajuda o usuário a manipular o sistema com eficácia, de uma forma eficiente e agradável, de modo a explorar todas as funcionalidades disponíveis. Um sistema com boa usabilidade é fácil de aprender, flexível, e se adapta às preferências e habilidades do usuário.
- *Utilidade*, que relaciona os componentes *Usuário* e *Conteúdo*. A utilidade do conteúdo e sua relevância para as necessidades e tarefas do usuário são as razões principais que determinam a seleção e uso de uma biblioteca digital.
- *Desempenho* (ou *performance*), relaciona os componentes *Conteúdo* e *Sistema*. O desempenho do sistema está fortemente relacionado aos formatos, estruturas e representação dos conteúdos (incluindo os metadados).

No modelo conceitual apresentado, *Usabilidade* representa uma área de pesquisa que investiga a interação entre o usuário e o sistema. Ela engloba uma série de princípios que buscam garantir que o sistema possa ajudar o usuário a realizar suas tarefas de uma forma eficiente, efetiva e satisfatória. De acordo com a definição mais utilizada, *Usabilidade* representa a “medida na qual um produto pode ser utilizado por usuários específicos para atingir objetivos específicos com eficácia, eficiência e satisfação, em um contexto de uso específico” (INTERNATIONAL ORGANIZATION FOR STANDARTIZATION -ISO, 1998). Um sistema com boa usabilidade deve ser atraente para o usuário, mas, acima de tudo, deve ser de fácil aprendizado, seguro e confiável, tolerante a falhas, consistente, adaptável às preferências terminológicas e operacionais dos usuários, fornecer suporte, ser fácil de memorizar e visualmente claro. Blandford e outros (2004) alertam para o risco de se produzir sistemas de bibliotecas digitais com recursos novos e avançados, mas que acabem sendo rejeitados pelos usuários devido à dificuldade de entendimento de sua operação, aprendizado e uso. A avaliação de usabilidade (ou a identificação de questões que a afetam)

pode ser entendida como a medida dos atributos de usabilidade de um sistema ou dispositivo com relação a determinados usuários, executando determinadas tarefas num dado contexto (HILBERT; REDMILES, 2000). Fuhr e outros (2007) agrupam as técnicas de avaliação de usabilidade de bibliotecas digitais em cinco categorias, além de indicarem trabalhos que adotaram técnicas em cada categoria:

- *Técnicas automatizadas* que utilizam algum tipo de software específico para produzir dados quantitativos. Esta categoria inclui análise de *logs*, que examina a atividade dos usuários a partir dos registros de uso coletados automaticamente pelo sistema durante um período de tempo.
- *Técnicas baseadas em pesquisa empírica* que requerem a participação do usuário, observação e registro de suas ações.
- *Técnicas comparativas* que comparam criticamente o sistema com determinados padrões, diretrizes, ou com outro sistema.
- *Entrevistas, grupos focais, surveys e questionários* são os principais métodos para a coleta de dados qualitativos, como preferências dos usuários, opiniões explanatórias, e taxas de satisfação. Segundo os autores existe grande variação na forma de aplicação destes métodos (*on-line*, pessoalmente) e também com relação ao grau de estrutura (estruturado, semi-estruturado, entrevistas não estruturadas).
- *Técnicas de análise* que requerem a participação de especialistas em usabilidade, que examinam o sistema sozinhos ou em cooperação com especialistas do domínio.

Alguns autores consideram que a avaliação de usabilidade possui limitações. Um dos principais motivos apontados é fato de o procedimento de teste ser uma situação artificialmente criada. Uma possível forma de amenizar este problema é a aplicação de múltiplos métodos combinados. Esta combinação de métodos é chamada de “paradigma convergente” que, além de melhorar a usabilidade dos sistemas, serve para a comparação dos métodos de avaliação de usabilidade (FUHR *et al.*, 2007).

Utilidade pode ser vista como o grau com que as necessidades de informação dos usuários são atendidas pelos conteúdos acessados na biblioteca digital. No Modelo Tríplice de Interação, a utilidade representa a relação entre *Usuário* e *Conteúdo*. A estimativa da utilidade depende tanto de aspectos relacionados aos usuários e suas necessidades de informação, quanto da natureza dos conteúdos presentes no acervo. O monitoramento dos usuários pode fornecer informações sobre como eles buscam e utilizam a informação no contexto de uma biblioteca digital. Fuhr e outros (2007) identificam duas abordagens principais utilizadas na avaliação da utilidade: *Estudos de Usuários* e *Comportamento*

Informacional. Estudos de Usuários é uma forma amplamente adotada para coletar e avaliar aspectos relacionados às preferências de uma comunidade de usuários. Várias técnicas de pesquisa podem ser utilizadas para a coleta de dados de modo a produzir representações estatísticas sobre as preferências dos usuários relacionadas aos conteúdos, além das formas e modos de acesso e uso destes conteúdos. A área de *Comportamento Informacional* investiga o processo de busca, recuperação e uso da informação em um nível profundo de detalhes. Este processo é dividido em estágios e analisados em relação às ações, estados cognitivos e emoções. Fuhr e outros (2007) observam que é necessária uma visão clara sobre o conteúdo e suas propriedades, como parte da avaliação de uma biblioteca digital. Além das informações relacionadas aos usuários e seus perfis, é necessário saber quais os tipos de conteúdos disponíveis e como o acesso e o uso da biblioteca digital é afetada pela natureza destes conteúdos.

O *Desempenho* ou *performance* de uma biblioteca digital relaciona os componentes *Sistema* e *Conteúdo* dentro do Modelo Tríplice de Interação proposto por Fuhr e outros (2007). O desempenho está relacionado a uma abordagem centrada no sistema (componente tecnológico) dentro do arcabouço genérico para avaliação de bibliotecas digitais apresentado por Saracevic (2004). Esta abordagem é uma das mais utilizadas e envolve o estudo de aspectos relacionados ao desempenho, bem como estimativas da eficiência de uma característica, projeto específico ou componente tecnológico. Esta avaliação de sistemas de informação é geralmente realizada em ambientes controlados, usando diferentes critérios e medidas, com o objetivo fornecer parâmetros de comparação e melhorar o desempenho. De acordo com Fuhr e outros (2007), a avaliação do desempenho de uma biblioteca digital é uma questão complexa que deve cobrir vários aspectos como: arquitetura, capacidade de armazenamento, acesso, gerenciamento de coleções, etc. Os autores observam que, apesar das bibliotecas digitais serem sistemas modulares e distribuídos onde vários componentes interagem usando uma infraestrutura de redes, as avaliações raramente exploram a natureza distribuída destes sistemas. As atividades de avaliação geralmente apresentam aos participantes um ou mais ambientes de testes, uma ou mais coleções, uma série de tarefas a serem executadas, além de um método pelo qual o desempenho do sistema, ou de um componente do mesmo, pode ser avaliado com relação às tarefas e coleções disponíveis.

Com relação às questões de ordem prática a serem consideradas na implementação de uma avaliação, Fuhr e outros (2007) enumeram: custo da avaliação (recursos financeiros, tempo, pessoal, etc.); efetividade (potencial de a avaliação revelar problemas e oferecer resultados confiáveis); infraestrutura (armazenamento, softwares específicos para análise); local (ambientes controlados, condições contextuais, avaliação remota); momento da

avaliação (estágio do desenvolvimento da biblioteca digital); ritmo (frequência da avaliação, objetivos e circunstâncias); avaliação colaborativa (ambientes de teste necessários). Com relação ao local de avaliação, pode-se adotar um ambiente de laboratório altamente controlado ou realizar uma avaliação empírica em ambiente real de uso. No segundo caso, onde o nível de controle é reduzido, pode ser necessária a aplicação de mais de um método de coleta de dados. Neste caso, pode ser interessante integrar métodos estatísticos (quantitativos) com narrativas (qualitativas) (FUHR *et al.*, 2007). Os autores destacam ainda algumas dificuldades relacionadas à coleta de dados qualitativos remotamente, entre elas a dificuldade de recrutamento e estímulo dos usuários para participação. Outra questão importante diz respeito ao fato de que as bibliotecas digitais fornecem modos de acesso remoto e assíncrono à informação, necessitando desta forma de métodos que avaliem efetivamente as bibliotecas digitais em casos de uso de acesso remoto, onde os usuários não possuem nenhuma conexão física com os responsáveis pela administração da biblioteca. A seleção do melhor método de avaliação depende das questões citadas anteriormente, assim como da sua combinação com o grau de estrutura dos dados a serem coletados. Entre os métodos que envolvem dados qualitativos, Fuhr e outros (2007) destacam os protocolos *think-aloud*, observação, entrevistas, questionários e simulação, enquanto que para dados quantitativos os autores sugerem métodos como análise de *logs* e análise de erros e tempos, como candidatos a serem empregados na coleta de dados.

Pelo exposto até o momento fica claro que a avaliação de bibliotecas digitais é uma tarefa complexa, que envolve a consideração de variados aspectos ou dimensões para a definição de uma metodologia de avaliação adequada para uma determinada circunstância.

Kovács e Micsik (2004) propuseram um mecanismo analítico para a descrição e classificação de procedimentos de avaliação, chamado Modelo de Avaliação Computacional. O modelo se baseia na classificação facetada (ver seção 2.2) e permite a criação de descrições de avaliações que podem ser comparativamente listadas de modo a verificar sua profundidade e cobertura. De acordo com este modelo, o processo de avaliação é definido a partir da seleção de aspectos em um espaço multidimensional (facetado). O modelo define os seguintes aspectos principais: *avaliação*, *usuário*, *organização*, *conteúdo* e *sistema*. Cada aspecto é subdividido em facetas. Desta forma o modelo é capaz de registrar sistematicamente as principais características relacionadas ao processo de avaliação (aspectos de avaliação), componentes da biblioteca digital (aspectos relacionados aos usuários, conteúdos e sistema) e as condições contextuais (aspectos organizacionais). A seguir estão listados os aspectos e principais facetas do modelo proposto por Kovács e Micsik (2004).

- Aspectos de avaliação
 - Coleta de dados: especialistas, **logs**, entrevistas, *surveys*.
 - Tipo de avaliação: formativa, “**somativa**”, comparativa.
- Aspectos relacionados aos usuários
 - Tipo de usuários: estudantes, professores, internos, **gerais**, profissionais, pesquisadores.
 - Escopo: nacional, **mundial**, regional, comunidade.
- Aspectos organizacionais
 - Setor: governo, **educação**, setor privado, setor público.
- Aspectos relacionados ao conteúdo
 - Forma do conteúdo: 3D, imagens, **texto**, áudio, vídeo.
 - Conteúdo: somente metadados, **texto completo**, resumos.
- Aspectos relacionados ao sistema
 - Tipo de Aplicação: dispositivos móveis, desktop, **interface web**, realidade virtual, interface multimodal.
 - Arquitetura: **cliente-servidor**, *peer-to-peer*, serviços web.

Os termos destacados em negrito representam um exemplo de uma possível configuração para um dado procedimento de avaliação. Neste exemplo, a natureza da avaliação é “somativa” e o método utilizado é a análise de *logs*. A avaliação utiliza os *logs* resultantes do uso do sistema pelo público em geral, sem restrição geográfica (escopo mundial). A comunidade mundial de usuários utiliza a biblioteca digital com propósitos educacionais (aspecto organizacional). A avaliação considera a interação dos usuários com conteúdos na forma de objetos textuais em formato completo (aspectos relacionados ao conteúdo). Os usuários interagem com a biblioteca digital a partir de sua interface *web*, enquanto que a arquitetura do sistema é do tipo cliente-servidor (aspectos relacionados ao sistema). O exemplo ilustra a capacidade do modelo em representar (e classificar) um procedimento de avaliação, tanto no nível de processo quanto no nível dos componentes da biblioteca digital considerados.

Fuhr e outros (2007) após constatar a complexidade em se propor um arcabouço conceitual para a avaliação de bibliotecas digitais devido à complexidade e multiplicidade de modelos, formalismos e contribuições de diversas áreas relacionadas, decidiram se apoiar no trabalho de Saracevic (2004). Este autor projetou um arcabouço conceitual composto de quatro dimensões ou componentes (*constructos*, *contexto*, *critérios* e *metodologia*) para a descrição das atividades de avaliação. O arcabouço conceitual proposto por Saracevic

(2004) serve tanto como um esquema de classificação para os estudos de avaliação existentes, como um modelo para os novos estudos nessa área. Fuhr e outros (2007) observam que um conjunto de diretrizes podem ser extraídas de modo a tornar as avaliações menos complexas e controversas, assim como permitir que possam ser replicadas e comparadas com outras avaliações similares de modo que, no futuro, os experimentos de avaliação e seus resultados não se tornem exercícios meramente individualistas e possam servir a toda a comunidade de pesquisa, trazendo progressos e avanços onde necessário. O modelo proposto por Saracevic (2004) tem se mostrado útil na medida em que fornece quatro categorias claras e bem formuladas, ao mesmo tempo em que preserva a expressividade necessária. Estas quatro categorias (ou dimensões) serão apresentadas a seguir.

Constructos representam o que será avaliado e podem ser utilizados em diferentes níveis de abstração. Podem ser usados inclusive em cenários onde a equipe de avaliação define sua própria interpretação para o que seja uma biblioteca digital. Podem ser utilizados para definir os aspectos e/ou componentes da biblioteca digital que serão alvo do procedimento de avaliação, de modo a gerar a resposta para a pergunta: O QUE avaliar?

Contexto é possivelmente a dimensão mais rica na medida em que representa tudo aquilo que se qualifica como motivações e contexto para uma dada avaliação, cobrindo cenários, atores, objetivos, abordagens e perspectivas adotadas no estudo. Responde a pergunta: POR QUE avaliar?

Crítérios representam o núcleo de um estudo de avaliação, cobrindo parâmetros, fatores e medidas usadas para estimar a qualidade do que está sendo avaliado e todo aspecto de uma biblioteca digital sendo avaliada. Os critérios podem ser vistos como padrões para se julgar algo, podendo ser usados para definir medidas (por exemplo, tempo é um critério, minuto é uma medida e cronômetro é o instrumento de medição; relevância é um critério, precisão e cobertura são medidas e o julgamento de relevância realizado por pessoas é o instrumento de medição). Os critérios adotados respondem parcialmente a pergunta: COMO avaliar?

A *metodologia*, em conjunto com os *crítérios*, respondem a questão de COMO avaliar. Fornece uma forma de descrever os procedimentos e protocolos adotados tanto para coleta como para a análise dos dados.

Nenhuma das quatro dimensões é independente das demais, mas tomadas em conjunto elas fornecem um arcabouço autocontido no qual as várias questões relacionadas a avaliação de bibliotecas digitais podem ser exploradas e descritas sem o risco de sobreposições desnecessárias (FUHR et al., 2007).

Os três modelos apresentados, Modelo DELOS (FUHR *et al.*, 2001), Modelo Tríplice de Interação (FUHR *et al.* 2007) e o Modelo de Avaliação Computacional (KOVÁCS; MICSIK, 2004) foram analisados por Fuhr e outros (2007), tomando como base a estrutura geral proposta por Saracevic (2004). O QUADRO 9, a seguir apresenta o quadro comparativo resultante.

QUADRO 9 - Comparação entre os modelos de avaliação

	Modelo DELOS	Modelo Tríplice de Interação	Modelo de Avaliação Computacional
Objetivo geral	Classificar e apresentar um esquema de avaliação. Identifica deterministicamente as principais entidades do domínio das bibliotecas digitais.	Identificar as principais entidades, assim como suas relações de modo a definir eixos (perspectivas) de avaliação e respectivos critérios, adotando um ponto de vista baseado na interação dos usuários com a biblioteca digital.	Formular perfis (descrições) de avaliação e encorajar a comparação entre eles. Serve como uma estrutura referencial para comparação entre experimentos de avaliação.
Porquê avaliar? <i>Contexto.</i>	Define quatro componentes de modo a descrever a operação de uma biblioteca digital (usuários, uso, dados/coleções e sistema), Trata de objetivos de avaliação centrados no sistema e na melhoria de seu desempenho.	Considera a biblioteca digital como a interação entre três entidades (usuários, conteúdo e sistema). Considera que estas três entidades definem três eixos de avaliação (utilidade, usabilidade, desempenho). O foco da avaliação é a melhoria da interação do usuário.	Um perfil de avaliação é considerado como a aplicação de um determinado procedimento de avaliação a uma instância contextualizada de uma biblioteca digital. Uma instância contextualizada da biblioteca digital é quádrupla composta de valores referentes aos aspectos: usuário, conteúdo, organização e sistema.
O que avaliar? <i>Constructo.</i>	Sugere o desenvolvimento de um ambiente de testes que possa agregar os requisitos para operação e avaliação da biblioteca digital.	Considera a avaliação como um processo composto e analisado em três dimensões (utilidade, usabilidade, desempenho).	Analisa a operação de uma biblioteca digital a partir de quatro componentes (usuário, conteúdo, organização e sistema). Combina diferentes métodos de avaliação com estes

			quatro componentes de modo a definir perfis de avaliação.
Como avaliar? <i>Critério.</i>	Trata-se de um modelo genérico que encoraja os pesquisadores a definir critérios específicos a serem aplicados a cada experimento de avaliação.	Para cada eixo ou dimensão de avaliação é fornecido um conjunto de critérios e medidas.	Fornece um mecanismo para a geração de perfis de avaliação.

Fonte: Adaptado de Fuhr e outros (2007)

Duas questões principais surgem a partir da comparação entre os modelos:

- Orientação: O Modelo DELOS é centrado no sistema, enquanto o Modelo Tríplice de Interação é centrado no usuário e o Modelo de Avaliação Computacional é neutro;
- Enquanto os dois primeiros modelos propõem a avaliação de entidades, tanto numa visão ampla quanto restrita da operação de uma biblioteca digital, o Modelo de Avaliação Computacional fornece uma visão da configuração do procedimento de avaliação.

O arcabouço conceitual para avaliação de bibliotecas baseado na análise dos três modelos apresentados, e analisados sob a estrutura geral proposta por Saracevic (2004), pode ser utilizado como base para definição do experimento de avaliação a ser conduzido. Com isto espera-se que o experimento de avaliação possa trazer resultados efetivos que possam ser comparados com estudos similares de modo a maximizar sua utilidade. No estágio inicial de preparação e configuração do experimento de avaliação, o arcabouço conceitual apresentado pode ser utilizado para agrupar e analisar estudos similares de modo que se possa obter um melhor entendimento sobre o foco que vem sendo utilizado pela comunidade de pesquisa, bem como os pontos que vêm sendo negligenciados. Num estágio posterior, quando for necessário decidir quais os aspectos deverão ser tratados pela avaliação, o arcabouço conceitual pode ser usado como base de um processo de reflexão e definição, de modo a responder as principais questões da avaliação:

- PORQUE avaliar; ou seja, determinar os objetivos da avaliação. O objetivo principal da avaliação no contexto da presente pesquisa é verificar os benefícios da adoção de um mecanismo de organização do conhecimento baseado na classificação facetada, e dos novos mecanismos de recuperação da informação, baseados em

taxonomias dinâmicas. Mais especificamente, pretende-se avaliar as melhorias a serem obtidas na interação dos usuários com a biblioteca digital nos casos de uso relacionados ao acesso e recuperação da informação.

- O QUE avaliar; ou seja, determinar os constructos (componentes da biblioteca digital e seus relacionamentos). Os componentes de acesso e recuperação da informação da BDTD deverão ser avaliados em comparação ao protótipo desenvolvido. Mais especificamente, a interação dos usuários com estes componentes deverá ser avaliada no contexto de um estudo comparativo.
- COMO avaliar; ou seja, determinar os critérios, métricas, amostras e o método de avaliação. O experimento deverá envolver a participação de uma amostra de usuários em tarefas de busca e recuperação da informação, envolvendo ambos os sistemas tratados.

Na medida em que o experimento tenha sido projetado e os procedimentos operacionais definidos, as quatro dimensões propostas por Saracevic (2004) podem ser usadas na análise dos dados coletados. Desta forma, haverá um direcionamento para a avaliação, além de poder gerar resultados que possam ser comparados com outros estudos que adoraram o mesmo arcabouço conceitual. Pelo exposto, o arcabouço conceitual apresentado deverá ser utilizado tanto no projeto, implementação e análise dos resultados do experimento de avaliação.

2.5.1 Avaliação de usabilidade em bibliotecas digitais

Usabilidade é um critério amplamente utilizado na avaliação de bibliotecas digitais, apesar de não existir uma definição uniforme do que é coberto por este critério no contexto das bibliotecas digitais (SARACEVIC, 2004). Trata-se de um critério bastante genérico, que inclui vários critérios mais específicos. De acordo com a definição mais utilizada, usabilidade representa a “medida na qual um produto pode ser utilizado por usuários específicos para atingir objetivos específicos com eficácia, eficiência e satisfação em um contexto de uso específico” (ISO, 1998). Este é o “guarda chuva” sob o qual a usabilidade é tratada na avaliação de bibliotecas digitais.

Existem várias formas de avaliar um sistema sob o ponto de vista da interação homem-máquina. Como citado acima, o padrão ISO 9241-11 (ISO, 1998) identifica três aspectos da usabilidade: efetividade, eficiência e satisfação. Shneiderman e Plaisant (2004)

identificam cinco medidas de usabilidade: tempo necessário para o aprendizado, velocidade de interação, taxa de erros do usuário (precisão), tempo de retenção e satisfação subjetiva. Estas medidas são interdependentes como observado por Chin e outros (1988). Para várias tarefas, a velocidade e precisão são duas medidas de desempenho relacionadas que afetam a atitude do usuário com relação ao sistema. O tempo necessário para o aprendizado e a retenção do conhecimento adquirido ao longo do tempo também afetam a utilidade do sistema para o usuário. A aceitação do sistema pelo usuário, ou seja, a satisfação subjetiva é também uma medida crítica para o sucesso de um sistema. Chin e outros (1988) observam que mesmo que um sistema seja bem avaliado nas medidas relacionadas ao desempenho, ele pode não ser utilizado pelos usuários devido à insatisfação dos mesmos com o sistema e sua interface. Os autores apresentaram um questionário para avaliar a satisfação do usuário com a interface de um sistema. Eles identificaram várias dimensões para se avaliar a satisfação do usuário: reação geral ao software, visualização, terminologia e informações do sistema, aprendizado e desempenho. Para cada dimensão, Chin e outros (1988) utilizaram um conjunto de medidas apresentadas no questionário em escalas de dez pontos, com limites representados por pontos antagônicos do tipo *terrível – maravilhoso* e *difícil - fácil*.

Hornbaek e Law (2007), com o objetivo de melhorar o entendimento sobre como avaliar a usabilidade, apresentaram uma meta-análise de várias medidas de usabilidade. A pesquisa envolveu a análise da correlação entre medidas de usabilidade obtidas a partir dos dados primários de 73 estudos publicados entre os anos de 2003 e 2005. Os autores analisaram a relação entre efetividade (alta precisão e baixa taxa de erros), eficiência (tempo reduzido para completar tarefas) e satisfação. As correlações encontradas pelos autores foram baixas, em geral, o que sugere que todas as medidas devem ser utilizadas. Entre as conclusões dos autores estão: questionários padronizados para medir a satisfação são aparentemente mais confiáveis do que questionários desenvolvidos especificamente para cada caso; e a medida da percepção dos usuários sobre o fenômeno geralmente não se relaciona com medições objetivas do mesmo fenômeno. Os questionários a serem preparados para a avaliação a ser realizada no contexto desta pesquisa deverão ser basear nos estudos acima citados de Shneiderman e Plaisant (2004), Hornbaek e Law (2007) e Chin e outros (1988).

Freyne e outros (2007) realizaram uma avaliação dos serviços de busca e navegação, baseados em técnicas sociais, aplicadas à biblioteca digital da ACM (*Association for Computing Machinery*). Durante a avaliação, 10 estudantes de doutorado foram solicitados a encontrar (dentro do tempo máximo de 20 minutos) artigos sobre o tópico “web social” que fossem adequados ao uso por estudantes de graduação, numa

disciplina sobre este mesmo tópico. Os autores concluíram que as dicas visuais relacionadas à popularidade da navegação, relevância da busca ou popularidade das anotações, desempenhavam um importante papel durante o processo de descoberta da informação. Kruk e outros (2008) também realizam um estudo de avaliação buscando comprovar a utilidade de tecnologias da web semântica aplicadas no contexto de uma biblioteca digital. Os autores realizaram um estudo comparativo que envolveu um grupo de usuários que utilizaram o sistema gerenciador de biblioteca digital JeromeDL (KRUK et al., 2005), que incorpora várias tecnologias semânticas e sociais, e outro grupo de usuários (grupo de controle) que utilizou o sistema DSpace (ver seção 2.1.7.4), que fornece apenas os mecanismos convencionais encontrados na maioria das sistemas gerenciadores de bibliotecas digitais. Os objetivos e a configuração da avaliação a ser conduzida nesta pesquisa são similares aos utilizados nos trabalhos apresentados por Freyne e outros (2007) e Kruk e outros (2008).

Bohmerwald (2005) levanta a questão de que autores da Ciência da Computação e da Ciência da Informação têm se dedicado aos estudos sobre a interação dos usuários com os sistemas de informação e, em especial, com as bibliotecas digitais, usando abordagens distintas. A autora propõe que os estudos das duas áreas sejam usados conjuntamente, complementando-se, com o intuito de analisar de forma abrangente um sistema de informação. Nesse sentido ela aponta para a possibilidade de os estudos de usuários (mas especificamente do comportamento de busca), da Ciência da Informação, serem combinados com os estudos de usabilidade da Ciência da Computação. No caso dos estudos de usuários, estes são o foco. Usuários são estudados para ver porque eles usam o sistema e como interagem com o mesmo. Usuários são geralmente observados fazendo suas próprias tarefas no sistema. Já nos estudos de usabilidade, o sistema é o foco. Ele é estudado através do comportamento dos usuários. São dadas tarefas para os usuários executarem no sistema. Eles são observados para se identificar como pensam e usam o sistema para completar as tarefas. O estudo do comportamento de busca por informação engloba questões importantes que não estão inseridas nos testes de usabilidade, como a motivação, o contexto e a própria individualidade do usuário, com a análise das atividades realizadas livremente no sistema, ou seja, sem tarefas predeterminadas (BOHMERWALD, 2005).

No caso da presente pesquisa, o objetivo é avaliar comparativamente aspectos técnicos relacionados ao sistema de biblioteca digital, mais especificamente deseja-se avaliar os benefícios para o usuário relacionados a um novo mecanismo de recuperação da informação, baseado em taxonomias dinâmicas e que se apoia numa estrutura classificatória facetada para a organização do acervo.

Pereira (2011) apresentou, em sua pesquisa de mestrado, um estudo de caso sobre a avaliação de usabilidade aplicada sobre a Biblioteca Digital Brasileira de Teses e Dissertações. Em seu trabalho a autora buscou: evidenciar a contribuição dos princípios de usabilidade para melhoria das interfaces de bibliotecas digitais; mapear os problemas de usabilidades encontrados na interface da referida biblioteca; propor soluções para os problemas encontrados. A interface da biblioteca foi avaliada por meio de um método analítico (avaliação heurística), combinado com um método empírico (teste com usuários). A avaliação realizada pela autora considerou apenas os serviços e funcionalidades já disponíveis na referida biblioteca. No caso da presente pesquisa, o objetivo da avaliação é revelar possíveis benefícios advindos da implementação de novos mecanismos de busca e exploração do acervo, implementados e incorporados em um protótipo para um sistema de biblioteca digital.

Lima e outros (2011) pesquisaram as abordagens usadas nas avaliações de bibliotecas digitais, a partir da bibliografia de Cunha (2009) sobre a temática. A identificação das abordagens tomou como base aquelas levantadas por Saracevic (2005), em pesquisa sobre avaliação de bibliotecas digitais. Os autores buscaram identificar as abordagens mais utilizadas nas avaliações. Para isso, adotaram as seguintes etapas: identificar na bibliografia os artigos que tratam de avaliação de bibliotecas digitais; identificar os elementos característicos determinantes das abordagens; apresentar a abordagem mais adotada para avaliação de biblioteca digital a partir dos modelos encontrados na literatura. Os resultados da pesquisa apontaram a abordagem centrada no homem como a mais usada nos estudos de avaliação de bibliotecas digitais. O uso desta abordagem se justifica na medida em que tem a capacidade de apontar questões importantes para atender às necessidades do usuário, lembrando que ele é a razão de ser da biblioteca digital.

Os resultados da pesquisa de Lima e outros (2011) corroboram os resultados obtidos por Saracevic (2000), mencionados anteriormente, que aponta o nível *Individual* como o mais utilizado nas avaliações de bibliotecas digitais, sendo que o critério mais utilizado era a *Usabilidade*. A abordagem centrada na usabilidade pode ser considerada uma ponte entre as abordagens centradas em sistemas e no homem (FUHR *et al.* 2007). O uso dessa abordagem na avaliação de bibliotecas digitais cresceu nos últimos anos, uma vez que objetiva detectar problemas e dificuldades de interação entre o usuário e o sistema.

2.6 Trabalhos relacionados e considerações

Nesta seção serão brevemente apresentados alguns trabalhos anteriores que estão fortemente relacionados com a presente pesquisa, assim como algumas considerações sobre a fundamentação teórica e metodológica apresentada nas seções anteriores.

Pontes (2002) utilizou em sua pesquisa de mestrado o tesouro AGROVOC para facilitar o acesso dos usuários ao acervo de uma biblioteca digital temática (PONTES *et al.* 2001). O AGROVOC é um tesouro multilíngua para o domínio agrícola, desenvolvido e mantido pela FAO³⁰ (*Food and Agriculture Organization of the United Nations*) por meio do WAICENT³¹ (*World Agricultural Information Centre*). O tesouro foi utilizado para navegação sobre os itens do acervo, mediante um processo de indexação automática. Também foi utilizado para a expansão das buscas textuais, mediante sugestão de termos relacionados pela estrutura do tesouro. Os resultados obtidos indicaram melhorias significativas na recuperação, além de melhorias na precisão quando se considera os primeiros itens retornados pelas buscas.

Lima (2004), em sua pesquisa de doutorado, criou um modelo para facilitar a recuperação da informação contida em documentos do tipo teses e dissertações a partir de elementos semânticos, estruturados com base na teoria da classificação facetada e representados na forma de um mapa conceitual. Tal estrutura foi construída individualmente para cada documento (tese ou dissertação), permitindo a navegação em contexto dentro do documento. Nessa pesquisa, a autora propôs uma nova forma de recuperação de informação, em contexto, com a construção de um modelo hipertextual para a organização de documentos acadêmicos do tipo teses e dissertações, visando a sua implantação tecnológica no contexto de uma biblioteca digital. A pesquisa da autora teve continuidade com a formação do grupo de pesquisa MHTX³² e as experiências têm sido feitas no recorte da linha de pesquisa Organização e Uso da Informação, do Programa de Pós Graduação em Ciência da Informação da Universidade Federal de Minas Gerais.

O trabalho de Pontes (2002), de forma semelhante a esta pesquisa, também tinha como objetivo o uso de uma estrutura semântica de representação e organização do conhecimento (um tesouro no caso) visando melhorar a recuperação da informação no contexto de uma biblioteca digital. No entanto, a abordagem utilizada se valeu de processos

³⁰ Ver em: <http://www.fao.org>

³¹ Ver em: <http://www.fao.org/waicent>

³² Ver em: <http://www.gercinalima.com/mhtx>

automáticos para fazer a associação dos documentos com os termos do tesauro (descritores). Apesar de serem úteis para a representação temática, possibilitando a especificidade e exaustividade adequadas, os tesauros não possibilitam a organização dos documentos em uma estrutura que pudesse ser facilmente utilizada pelo usuário nos processos de navegação e exploração do acervo. Uma classificação facetada, conforme proposta da presente pesquisa, pode representar uma alternativa tanto para a representação de assuntos, como para a organização do acervo na medida em que possa permitir a exploração pelo usuário através da combinação de perspectivas multidimensionais, que podem ainda envolver uma combinação de aspectos semânticos e descritivos dos documentos, conforme foi discutido na seção 2.2.

Lima (2004) também buscou a construção de estruturas semânticas para a representação do conteúdo dos documentos, visando facilitar a recuperação no contexto de uma biblioteca digital. No entanto, o foco da autora foi a representação do conteúdo de cada documento isoladamente, visando a navegação em contexto. Pesquisas posteriores buscaram realizar a compatibilização das estruturas individuais dos documentos visando possibilitar a busca e navegação integrada em todo o acervo (SILVA, 2008). No entanto, os resultados obtidos ainda não foram satisfatórios.

A abordagem utilizada na presente pesquisa propõe o uso de uma estrutura semântica construída na forma de uma taxonomia facetada e, desta forma, possibilitar a busca e navegação em todo o acervo. Nada impede que a abordagem proposta por Lima (2004) seja posteriormente integrada, de modo a também permitir a busca e navegação interna ao documento, em contexto.

Nas seções anteriores foram apresentadas as origens históricas das bibliotecas digitais, suas bases teóricas, perspectivas de evolução, além das principais linhas de pesquisa na área. Também foram apresentadas as principais iniciativas brasileiras nesta área. Em seguida foram apresentados os principais modelos teóricos de bibliotecas digitais, enfatizando-se a importância destes modelos para a pesquisa e desenvolvimento da área. Foram brevemente apresentados os principais modelos e padrões de metadados utilizados para a representação dos acervos, além dos principais sistemas gerenciadores de bibliotecas digitais disponíveis. A Teoria da Classificação Faceta foi apresentada, bem como discutida a sua aplicação como mecanismo para organização do conhecimento em uma biblioteca digital. Apresentamos também um modelo formal para taxonomias dinâmicas, utilizado para a implementação do mecanismo de busca em um protótipo computacional. Em seguida, foram apresentados e discutidos os principais modelos de avaliação de bibliotecas digitais propostos na literatura, com vistas à elaboração de um experimento de avaliação do protótipo computacional desenvolvido. Por fim, foram brevemente

apresentados alguns trabalhos anteriores estreitamente relacionados com a presente pesquisa.

Passamos agora a destacar alguns pontos importantes, apresentados nesse capítulo, bem como algumas considerações acerca de suas implicações para a presente pesquisa.

A maioria das bibliotecas digitais se limita a fornecer valores para os atributos dos documentos se valendo de elementos de metadados como título, autor, palavras chave, etc. Além disso, estes sistemas se apoiam fortemente em técnicas de busca textual nos campos de metadados e ou no texto completo (*full text*). Podemos dizer que os metadados utilizados são na verdade uma representação digital da velha e conhecida ficha catalográfica. As técnicas desenvolvidas na área de recuperação da informação, especialmente aquelas relacionadas à busca textual foram, sem dúvida, importantes e amplamente adotadas pelas bibliotecas digitais. Contudo estas técnicas apresentam os problemas relacionados à ambiguidade da linguagem natural, além de requerer que o usuário explicita a sua necessidade de informação na forma de uma expressão de consulta, não oferecendo mecanismos adequados para a exploração do acervo.

Algumas ideias que precederam e deram origem às bibliotecas digitais, como o Memex (BUSH, 1945), já enfatizavam a necessidade de interconexões semânticas entre os recursos informacionais. O sucesso atual de mecanismos de buscas como o Google se deve em grande parte a exploração das interconexões entre os recursos web, na medida em que os algoritmos fazem uso dos *links* entre as páginas web como parâmetro usado para ordenação dos resultados de uma busca (*ranking*). Projetos como o SALT (GROZA et al., 2007) propõem o uso de metadados semanticamente mais ricos que possam permitir representar as argumentações em publicações científicas. No entanto, o que se observa é que a maioria das publicações disponíveis em meio digital possui as mesmas propriedades das publicações em papel. Além disso, a maioria das bibliotecas digitais tem negligenciado a representação temática e a organização do acervo com base no assunto dos documentos. Predominam as abordagens baseadas nos metadados e na busca textual aplicada aos elementos destes metadados, assim como ao texto completo. Um dos argumentos é que as estruturas classificatórias tradicionalmente utilizadas pelas bibliotecas convencionais para organização de seus acervos impressos não seriam adequadas ao ambiente digital (KOCH, 1998). Esta questão foi analisada e discutida anteriormente, quando analisamos a aplicação de estruturas classificatórias facetadas no ambiente digital.

Outro ponto que merece destaque é o fato, já mencionado na seção 2.1, de que as origens teóricas e práticas das bibliotecas digitais tiveram forte influência da computação (recuperação da informação), tendo sido pouco influenciada pela biblioteconomia. Este fato tem grandes implicações nos sistemas atualmente usados para a implementação de

bibliotecas digitais. Em geral estes sistemas não incorporam as técnicas desenvolvidas e aplicadas pelos bibliotecários. Esta constatação tem grande importância na justificativa e direcionamento deste trabalho de pesquisa.

A classificação facetada, conforme apresentado na seção 2.2, representa uma abordagem mais adequada para a representação e organização do acervo em ambientes digitais, quando comparada aos outros tipos de estruturas classificatórias utilizadas pelas bibliotecas convencionais. Estruturas facetadas já vêm sendo amplamente utilizadas para organização da informação na *web*, principalmente em *sites* comerciais (ADKISSON, 2005; LA BARRE, 2006). No caso das bibliotecas digitais, a representação temática utilizando uma estrutura facetada pode ser incorporada aos padrões de metadados existentes, conforme sugerido na seção 2.2.2, para o caso do Dublin Core. Neste caso, tanto a representação temática quanto a representação descritiva podem ser combinadas nos metadados de modo a proporcionar um modelo unificado de busca e navegação, conforme sugerido na seção 2.2, que incorpore tanto os aspectos semânticos (temáticos) como os aspectos descritivos dos documentos.

Uma classificação facetada pode ser formalmente representada de modo que tanto os procedimentos de indexação e classificação, assim como os mecanismos de busca e navegação podem se valer do auxílio computacional, conforme mostrado na seção 2.4. Esta formalização é condição necessária para a implementação de um protótipo computacional que possa prover mecanismos de busca, recuperação e exploração do acervo da biblioteca digital.

Finalmente, com base em modelos teóricos e metodologias sugeridas por diversos autores para avaliação de bibliotecas digitais (SARACEVIC, 2000, 2004; KOVÁCS; MICSIK, 2004; FUHR *et al.*, 2007), é possível construir um experimento que possa determinar comparativamente as vantagens da adoção de uma abordagem facetada para a representação e organização do acervo de uma biblioteca digital de teses e dissertações, bem como dos novos mecanismos de recuperação da informação implementados no protótipo computacional.

No próximo capítulo são apresentados os procedimentos e métodos adotados para o desenvolvimento do trabalho de pesquisa, visando atingir os objetivos traçados no Capítulo 1, a partir da fundamentação teórica e metodológica apresentada no presente capítulo.

3 METODOLOGIA

Neste capítulo são apresentados os métodos e procedimentos adotados para o desenvolvimento do trabalho de pesquisa. A importância do método se justifica na medida em que as conclusões e achados de uma pesquisa terão pouca validade científica se a forma como se chegou a eles carecer de cientificidade. O pesquisador precisa usar métodos cientificamente aceitos na sua busca pela compreensão do mundo. Com base nos objetivos desta pesquisa, apresentados na seção 1.3, será apresentado, a seguir, o seu delineamento metodológico sob três enfoques: quanto aos objetivos; quanto aos procedimentos técnicos; e quanto à forma de abordagem do problema.

Em relação aos seus objetivos, esta pesquisa se enquadra como descritiva que, de acordo com Gil (1991), visa descrever determinado fenômeno ou estabelecer relações entre variáveis. Descobrir e classificar a relação entre variáveis e a relação de causalidade entre fenômenos é o que caracteriza a pesquisa descritiva. Neste sentido, a presente pesquisa busca determinar o impacto da utilização de uma estrutura classificatória facetada, assim como da implementação de mecanismos de recuperação da informação baseados nesta estrutura, na melhoria da recuperação da informação e na satisfação do usuário da biblioteca digital. Também é de caráter aplicado que, segundo Lakatos e Marconi (1991), caracteriza-se por seu interesse prático, isto é, que os resultados sejam aplicados imediatamente na solução de problemas que ocorrem na realidade.

Segundo Gil (1991), o delineamento de uma pesquisa se refere ao “planejamento da pesquisa em sua dimensão mais ampla, envolvendo tanto a diagramação quanto a previsão de coleta de dados, análise e interpretação”. Kerlinger (1980, p.94) explica que “a palavra delineamento focaliza a maneira pela qual um problema de pesquisa é concebido e colocado em uma estrutura que se torna um guia para a experimentação, coleta de dados e análise”. Nesse sentido, o delineamento é intrínseco à pesquisa científica, norteando o pesquisador na busca de uma resposta para determinado problema. O delineamento expressa em linhas gerais o desenvolvimento da pesquisa, com ênfase nos procedimentos técnicos de coleta e análise de dados. Com relação aos procedimentos técnicos a serem empregados, ou seja, seu delineamento geral, a presente pesquisa pode ser enquadrada na modalidade experimental. Kerlinger (1980, p.125) diz que “um experimento é um estudo no qual uma ou mais variáveis independentes são manipuladas e no qual a influência de todas ou quase todas as variáveis relevantes possíveis não pertinentes ao problema da investigação é reduzida ao mínimo”. No caso da presente pesquisa, deseja-se determinar a influência ou impacto da organização do acervo que utiliza uma estrutura classificatória

facetada. Para tanto foi constituído um experimento objetivando manipular a variável independente e analisar os efeitos nas variáveis dependentes relacionadas à eficácia da recuperação da informação e satisfação do usuário da biblioteca.

Todas as etapas da pesquisa foram desenvolvidas no contexto da Biblioteca Digital de Teses e Dissertações da Escola de Ciência da Informação da UFMG (BDTD/ECI/UFMG). Desta forma, deste ponto em diante, a sigla BDTD será utilizada como referência a esta biblioteca específica.

A organização do acervo pode ser vista como uma variável independente que é analisada em duas situações: 1) baseada nos metadados e em indexação automática (que é a situação atual na BDTD); e 2) baseada nos metadados e numa estrutura classificatória facetada (que foi implementada no protótipo computacional). Deseja-se estudar as variáveis dependentes relacionadas à eficácia da recuperação da informação e satisfação dos usuários.

Quanto à forma de abordagem, a presente pesquisa se baseia na abordagem quantitativa e também qualitativa. Foram coletados dados numéricos a partir do uso da biblioteca digital por uma amostra de seus usuários. Estes dados buscam capturar o ponto de vista dos usuários e representam variáveis relacionadas à eficácia da recuperação da informação e satisfação do usuário. No entanto, a abordagem qualitativa também é empregada, na medida em que são aplicados questionários contendo questões não estruturadas, bem como são analisadas as interações dos usuários com o sistema de biblioteca digital, a partir da gravação das atividades durante a execução de tarefas.

3.1 Etapas metodológicas

Em alinhamento com os objetivos apresentados na seção 1.3, o QUADRO 10, a seguir, descreve as principais etapas desta pesquisa.

QUADRO 10 - Principais etapas da pesquisa

Nº	ETAPA
1	Pesquisa bibliográfica e análise de ferramentas de <i>software</i> e esquemas de metadados existentes para a construção de bibliotecas digitais.
2	Desenvolvimento do protótipo computacional , envolvendo as seguintes sub-etapas: 2.1. Formalização matemática e implementação computacional de um esquema classificatório baseado numa taxonomia facetada. 2.2. Desenvolvimento de componentes de software para permitir a classificação das teses e dissertações do acervo da BDTD com base no esquema classificatório implementado. 2.3. Desenvolvimento de componentes de software para prover mecanismos de busca e acesso no acervo da BDTD, com base no esquema classificatório facetado.
3	Realização de uma avaliação comparativa entre os mecanismos de busca e acesso atualmente disponíveis na BDTD e os mecanismos de busca e acesso implementados no protótipo computacional.

A **Etapa 1** envolveu pesquisa bibliográfica visando a fundamentação da pesquisa, tendo sido apresentada no Capítulo 2. Os principais temas pesquisados foram bibliotecas digitais, incluindo aspectos históricos (SAYÃO, 2008; CUNHA, 2008), modelos teóricos (FREW *et al.*, 1998; GONÇALVES *et al.* 2004; BORBINHA *et al.*, 2005; CANDELA *et al.*, 2007a, 2007b; FUHR *et al.*, 2007), além de metadados e software de gerenciamento. Também foi tratada a aplicação de classificações facetadas em ambientes digitais (DENTON, 2003; ADKISSON, 2005; LA BARRE, 2006; BROUGHTON, 2006, 2008), além de taxonomias dinâmicas (SACCO, 2000; TZITZIKAS *et al.*, 2002; SACCO e TZITZIKAS, 2009) e avaliação de bibliotecas digitais (MARCHIONINI, 2000; SARACEVIC, 2000, 2004; FUHR *et al.*, 2001, 2007; KOVÁCS e MICSIK, 2004; SHNEIDERMAN e PLAISANT, 2004; HORNBAEK e LAW, 2007; FREYNE *et al.*, 2007; KRUK *et al.*, 2008). Ainda na Etapa 1, foi realizado o levantamento e análise das principais ferramentas de *software* e esquemas de metadados utilizados para a construção de bibliotecas digitais. Trata-se de um estudo exploratório que se encontra descrito nas seções 2.1.6 e 2.1.7. Foram analisados os sistemas gerenciadores de biblioteca digital FEDORA (Fedora Development Team, 2005), BRICKS (MEGHINI; RISSE, 2005), Greenstone (BAINBRIDGE *et al.*, 2001, 2004) e DSpace (DSPACE FOUNDATION, 2009). Este último foi analisado em um nível mais profundo por ser amplamente utilizado na implantação de bibliotecas digitais de teses e dissertações no contexto de projetos do IBICT (ver seção 2.1.4). Com relação aos metadados, foi analisado o modelo FRBR (IFLA, 1998), assim como os padrões de metadados Dublin Core (DCMI, 2003), ETD-MS (ATKINS *et al.*, 2008) e MTD-BR (LOURENÇO, 2005).

A **Etapa 2** tratou do desenvolvimento de um protótipo computacional, envolvendo três sub-etapas. A **sub-etapa 2.1** trata da formalização matemática e da implementação computacional de um esquema classificatório baseado numa taxonomia facetada. Esta taxonomia foi inicialmente desenvolvida no contexto da pesquisa de mestrado de Maculan (2011), tendo sido posteriormente revisada e ampliada pela mesma autora. O desenvolvimento da taxonomia facetada encontra-se descrito na seção 2.3. A formalização matemática da taxonomia, visando a sua implementação computacional, se baseou no trabalho de Tzitzikas e outros (2002), apresentado e descrito na seção 2.4. A formalização e implementação foi genérica, ou seja, independente da taxonomia e podendo ser utilizada em conjunto com outras taxonomias facetadas desenvolvidas para outras áreas de conhecimento, desde que possuindo a mesma estrutura geral.

A **sub-etapa 2.2** trata do desenvolvimento de componentes de software para permitir a utilização do esquema classificatório facetado, implementado na sub-etapa 2.1, para a classificação das teses e dissertações do acervo da BDTD. Os metadados da BDTD foram utilizados em seu formato atual, baseado no padrão de metadados Dublin Core (ver seção 2.1.6.2). Para poder acomodar a estrutura classificatória facetada, permitindo a classificação dos documentos com base nessa estrutura, estes metadados foram estendidos conforme apresentado na seção 2.2.2. O conjunto de documentos classificados foi composto de 195 documentos, sendo 54 teses e 141 dissertações, defendidas no PPGCI/ECI/UFMG no período compreendido entre os anos de 1995 a 2011. Parte dos documentos (62), pertencentes à linha de pesquisa Organização e Uso da Informação, foi classificada, usando a taxonomia facetada, no contexto da pesquisa de Maculan (2011), descrita na seção 2.3. Outra parte foi posteriormente classificada pela mesma autora, a partir da revisão e ampliação da taxonomia facetada, conforme mencionado anteriormente.

A **sub-etapa 2.3** trata do desenvolvimento de componentes de software para implementar mecanismos de busca, navegação e exploração do acervo. Estes mecanismos se apoiaram e tiraram proveito da estrutura classificatória facetada formalizada na sub-etapa 2.1 e aplicada ao acervo da BDTD no contexto da sub-etapa 2.2, mediante o processo de classificação. Estes mecanismos tratam especificamente do acesso, visando facilitar as tarefas dos usuários relacionadas à recuperação de documentos para atender suas necessidades de informação em tarefas específicas. Os mecanismos foram implementados utilizando o modelo de taxonomias dinâmicas (SACCO; TZITZIKAS, 2009), descrito na seção 2.4.1.

O protótipo computacional desenvolvido utilizou a estrutura de metadados já disponível na BDTD, estendida na sub-etapa 2.2, conforme apresentado na seção 2.2.2. Os mecanismos de busca, navegação e exploração do acervo foram implementados de

forma acoplada ao DSpace, em sua versão 1.8.2, utilizando a linguagem de programação Java. Algumas novas relações (tabelas) foram criadas para prover o armazenamento de metadados adicionais, necessários aos novos mecanismos de busca implementados.

A **Etapa 3** trata de uma avaliação comparativa, realizada a partir de dados coletados no contexto de um experimento. Foi realizado um estudo comparativo entre os mecanismos de busca e acesso atualmente disponíveis na BDTD e os mecanismos de busca e acesso implementados no protótipo computacional desenvolvido na Etapa 2. O objetivo foi verificar as possíveis melhorias relacionadas à recuperação da informação e satisfação do usuário, em alinhamento com os objetivos desta pesquisa. A avaliação se pautou nos principais modelos para avaliação de bibliotecas digitais (SARACEVIC, 2000, 2004; KOVÁCS; MICSIK, 2004; FUHR et al., 2007), apresentados na seção 2.5.

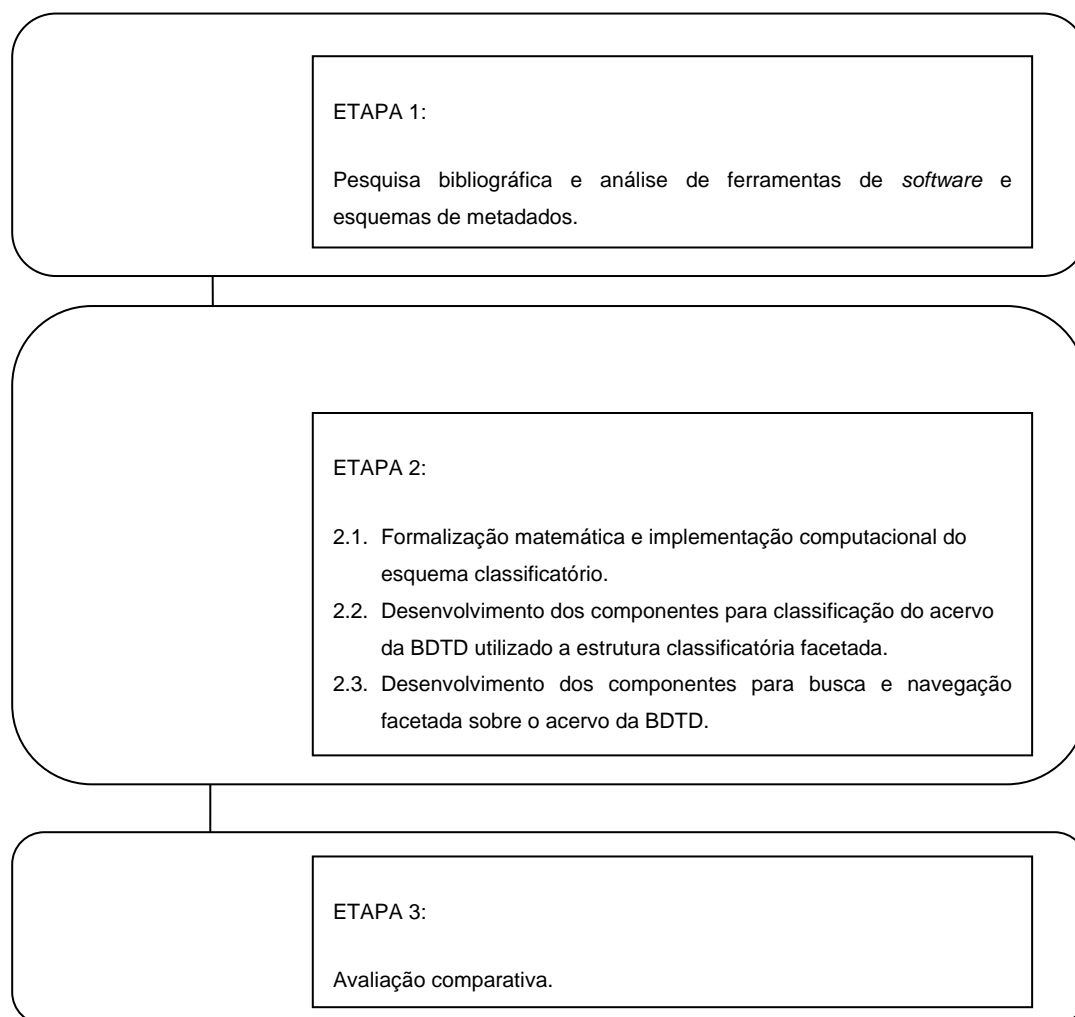


FIGURA 22 - Etapas metodológicas.

Conforme indicado na FIGURA 22, a Etapa 1 consistiu de um estudo exploratório que, além da pesquisa bibliográfica, buscou identificar e analisar as principais ferramentas utilizadas na construção do componente tecnológico da biblioteca digital, incluindo sistemas gerenciadores de biblioteca digital e metadados. A Etapa 2 consistiu da implementação de componentes computacionais (*software*) que foram desenvolvidos para possibilitar a organização do acervo da BDTD com base em uma estrutura classificatória facetada, além da criação de mecanismos de busca e acesso que se apoiam nessa mesma estrutura. Estas etapas foram fortemente influenciadas pela Teoria da Classificação Facetada e pelo modelo proposto para Taxonomias Dinâmicas, conforme apresentado nas seções 2.2 e 2.4, respectivamente. Finalmente, a Etapa 3 consistiu da realização de um experimento, envolvendo testes de uso do protótipo computacional desenvolvido na Etapa 2. Adotou-se uma abordagem comparativa, buscando avaliar as melhorias proporcionadas pelos novos mecanismos de busca e recuperação da informação, implementados no protótipo.

As principais etapas da pesquisa serão apresentadas em maiores detalhes nas próximas seções.

3.2 Desenvolvimento do protótipo computacional

Esta seção pretende apresentar e discutir a estratégia de desenvolvimento dos novos mecanismos de busca e acesso, implementados na forma de um protótipo computacional. Na seção 2.4.1 foi apresentado o modelo genérico para Taxonomias Dinâmicas, conforme proposto por Sacco e Tzitzicas (2009), que foi utilizado como base para a implementação desses mecanismos. Uma das características principais do modelo utilizado são a sua simplicidade e o fato de ser centrado no usuário, que contrasta com linhas de pesquisa que propõem soluções altamente complexas e centradas no sistema. Estas características tornam o modelo facilmente entendido pelos usuários, sem a necessidade de mediação por qualquer tipo de agente.

3.2.1 Arquitetura e estratégia de implementação

Entre os requisitos considerados para a implementação está a *performance*, relacionada à necessidade de resposta em tempo real do sistema para as operações

executadas a partir da interação do usuário. Outro requisito considerado é que o acesso a partir da taxonomia pudesse ser integrado como um componente de acesso no contexto do sistema de biblioteca digital. Um tempo excessivo de espera iria impactar negativamente a experiência do usuário na exploração do acervo. A principal operação é a de *zoom* (ver seção 2.4.1), ou seja, a definição do foco de interesse do usuário com base na seleção de um ou mais conceitos, a subsequente computação dos conceitos relacionados ao foco de interesse e, finalmente, a redução da taxonomia.

Outro requisito considerado neste trabalho diz respeito à necessidade de se explorar a capacidade das taxonomias dinâmicas poderem se apresentar em mais de um idioma. Neste sentido, cada conceito da taxonomia foi representado por um identificador numérico único, abrindo a possibilidade para a existência de rótulos textuais em diferentes idiomas.

3.2.1.1 Arquitetura lógica

Entre os principais objetivos considerados para o desenvolvimento da arquitetura apresentada a seguir estão a flexibilidade e *performance*. Nesse sentido foram considerados três pontos básicos:

1. Taxonomias dinâmicas são voltadas para o acesso somente, e não para a classificação; desta forma, assume-se que a classificação dos objetos com base na taxonomia é realizada por uma entidade externa (um especialista humano no caso deste trabalho de pesquisa).
2. O acesso é baseado em descrições conceituais dos objetos (metadados), que são independentes do conteúdo dos objetos (dados).
3. A definição de um conceito *C* como um conjunto de objetos classificados sob o mesmo (ver seção 2.4.1) indica que o “nome” do conceito é apenas um rótulo textual, podendo inclusive existir diversos rótulos para o mesmo conceito em diferentes idiomas ou até mesmo como sinônimos.

A arquitetura apresentada prevê uma separação clara de diversos aspectos da taxonomia dinâmica. Em primeiro lugar, a identificação de um objeto está separada do conteúdo do mesmo, de modo que a taxonomia dinâmica possa ser usada para gerenciar qualquer tipo ou formato de objetos heterogêneos. Cada objeto é identificado no sistema por um identificador único que chamamos *ID-O* (IDentificador do Objeto). Assume-se que o conteúdo de cada objeto seja armazenado em um repositório, que é responsável por oferecer mecanismos para o armazenamento, extração e exibição do objeto, com base no *ID-O*. Desta forma separamos os metadados do conteúdo dos objetos, sendo que a comunicação entre os dois níveis se dá pelo identificador do objeto (*ID-O*).

Em segundo lugar, adota-se uma separação da camada de metadados na *intensão* (representada pela estrutura conceitual da taxonomia) e na *extensão* (objetos classificados pelos conceitos da taxonomia). A comunicação entre a *intensão* e a *extensão* é dada pela identificação de cada conceito com um identificador numérico único que chamamos de *ID-C* (IDentificador do Conceito). Desta forma, a *intensão* compreende os relacionamentos entre os conceitos na estrutura taxonômica, enquanto que a *extensão* compreende a classificação dos objetos, sendo cada objeto representado pelo seu *ID-O* e classificado sob conceitos que são representados pelos seus respectivos *ID-C*.

O uso de um identificador numérico para cada conceito (*ID-C*) permite ainda separar os rótulos textuais do conceito do seu conteúdo (que é representado pelo conjunto de objetos que são classificados sob o conceito). Os rótulos textuais de cada conceito podem ainda ser armazenados separadamente em dicionários, um para cada idioma. A FIGURA 23 representa a arquitetura ora descrita.

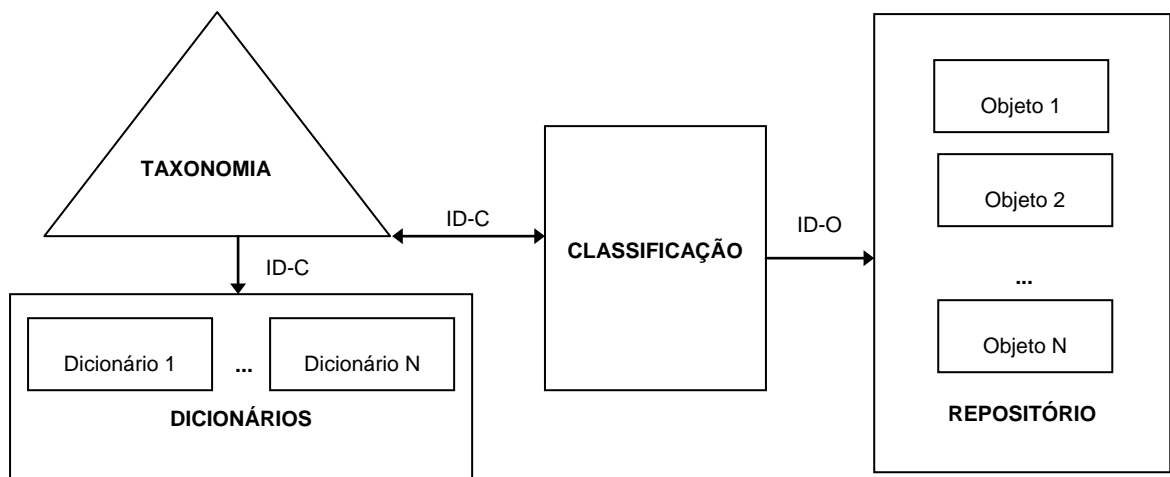


FIGURA 23 - Arquitetura lógica.

Fonte: elaborado a partir de (SACCO; TZITZICAS, 2009)

As estruturas necessárias para implementar a *intensão* são basicamente:

1. estrutura Pai-para-Filho, *PF*, que relaciona para cada conceito *ID-C* à sequência de seus descendentes, ordenada pela ordem de exibição. Esta estrutura pode ser usada para exibir a taxonomia partindo da raiz até as folhas (*top-down*);
2. estrutura Filho-para-Pai, *FP*, que relaciona para cada conceito *ID-C* o conjunto de seus antecessores, ou seu único antecessor caso herança múltipla não seja

utilizada. Esta estrutura permite a navegação partindo de um conceito até a raiz da taxonomia (*botton-up*);

3. estrutura *DESCENDENTES* que relaciona para cada conceito o conjunto formado por todos os seus descendentes; e
4. estrutura *ANTECESSORES* que relaciona para cada conceito o conjunto formado por todos os seus antecessores.

A estrutura *PF* será sempre necessária, sendo que a estrutura *DESCENDENTES* pode ser dinamicamente construída a partir dela, caso necessário. A navegação partindo das folhas até a raiz (*botton-up*) é necessária para implementar herança retroativa, conforme será apresentado posteriormente. Todas as estruturas armazenadas devem ser atualizadas no caso da *intensão* (estrutura taxonômica) ser alterada.

Com relação à *extensão*, a principal operação é o *zoom* (ver seção 2.4.1), que pode ser expressa com base nas seguintes operações primitivas:

1. *definição do foco*, isto é, cálculo do conjunto dos objetos que satisfazem uma condição booleana a partir de conceitos selecionados;
2. *redução da taxonomia* por meio da retirada de todos os conceitos não relacionados ao foco atual e, opcionalmente, o cálculo do número de objetos no foco que pertencem a extensão profunda de cada conceito *C* remanescente na taxonomia reduzida.

Estas operações serão apresentadas, em detalhes, na sequência.

Definição do Foco

O cálculo do foco de interesse, com base numa condição booleana aplicada nos conceitos selecionados pelo usuário, utiliza a extensão profunda destes conceitos (ver seção 2.4.1). Desta forma, torna-se um requisito de performance que a extensão profunda dos conceitos seja armazenada. Caso somente a extensão rasa seja explicitamente armazenada, a extensão profunda de um dado conceito *C* precisaria ser calculada dinamicamente a partir da união de todas as extensões rasas dos conceitos descendentes de *C*. Esta operação é muito custosa, principalmente no caso de taxonomias complexas e

de grandes bases de dados. Desta forma, uma requisito importante para a *performance* é manter a extensão profunda de cada conceito explicitamente armazenada.

A extensão profunda de cada conceito pode ser construída de maneira eficiente no momento da classificação através do mecanismo de herança retroativa: se o objeto d é classificado sob o conceito C , então d é inserido na extensão rasa de C , na extensão profunda de C , e também na extensão profunda de todos os conceitos antecessores de C . Este método constrói a extensão profunda de maneira *bottom-up* (dos filhos para o pai) uma única vez, no momento da classificação, e dispensa a necessidade de realizar a mesma operação de maneira *top-down* por ocasião da definição do foco (na recuperação). A extensão profunda do conceito raiz da taxonomia equivale, por construção, a todo o universo dos objetos classificados pela mesma, não sendo necessário armazená-la explicitamente. A forma mais fácil de implementar o mecanismo de herança retroativa descrito é utilizar a estrutura *ANTECESSORES* presente na *intensão*.

O cálculo do foco de interesse pode se dar de duas formas:

1. Usando o operador *OR* (disjunção), de modo que sejam selecionados, para compor o foco, os objetos que pertençam à extensão profunda de pelo menos um dos conceitos selecionados pelo usuário na taxonomia;
2. Usando o operador *AND* (conjunção), de modo que sejam selecionados, para compor o foco, somente os objetos que pertençam à extensão profunda de todos os conceitos selecionados pelo usuário na taxonomia.

A extensão profunda é armazenada na estrutura (relação) *EXTENSAO_PROFUNDA(id-c, id-o)*, onde *id-c* corresponde ao identificador do conceito, e *id-o* é o identificador do objeto que faz parte da extensão profunda do respectivo conceito. O comando em SQL³³ a seguir calcula o foco de interesse com base no operador *OR*:

```
SELECT DISTINCT id-o
FROM EXTENSAO_PROFUNDA
WHERE id-c IN (c1, c2, ..., cn)
AND id-o IN Fi-1
```

³³ Linguagem SQL estendida para suportar operações em conjuntos – pseudo SQL.

Onde c_i representa os conceitos selecionados na taxonomia, e F_{i-1} representa o foco anterior. No caso da primeira iteração da busca, o foco inicial F_0 corresponde a todo o universo de objetos classificados pela taxonomia.

No caso da opção ser pelo operador *AND*, o comando SQL (pseudo SQL) a seguir pode ser usado para o cálculo do foco de interesse:

```

SELECT DISTINCT id-o
FROM EXTENSAO_PROFUNDA E1
WHERE NOT EXISTS (
    SELECT *
    FROM EXTENSAO_PROFUNDA E2
    WHERE id-c IN (c1, c2, ..., cn)
        AND NOT EXISTS (
            SELECT *
            FROM EXTENSAO_PROFUNDA E3
            WHERE E3.id-o = E1.id-o
                AND E3.id-c = E2.id-c
        )
    )
AND E1.id-o IN Fi-1

```

Onde c_i representam os conceitos selecionados na taxonomia, e F_{i-1} representa o foco anterior, lembrando que na primeira iteração da busca, o foco inicial F_0 corresponde a todo o universo de objetos classificados pela taxonomia.

A extensão rasa somente é necessária no caso de os objetos poderem ser classificados sob conceitos não terminais. Para os conceitos terminais (folhas), as extensões rasa e profunda se equivalem. Além disso, se os objetos forem classificados somente sob conceitos terminais, a extensão rasa dos conceitos não terminais será obviamente vazia.

Como mencionado anteriormente, as taxonomias dinâmicas podem ser facilmente integradas com outras estratégias de recuperação da informação, como banco de dados e busca textual. Esta integração se dá de duas maneiras:

1. uso da taxonomia para sumarizar conceitualmente o resultado da busca;
2. uso da taxonomia para definir um contexto conceitual sob o qual a busca deverá se restringir.

Para que a integração seja possível, o mecanismo externo de busca deve operar com os mesmos identificadores de objetos (*id-o*) usados pela taxonomia dinâmica. Desta forma, considerando o caso 2 acima, é possível calcular o foco de interesse (conjunto de objetos) pela interseção do foco F , definido pelos conceitos selecionados na taxonomia, com o conjunto de resultados R obtidos a partir do mecanismo externo de busca. Para o caso 1, o foco de interesse é dado pelo conjunto R formado pelo resultado da busca realizada pelo mecanismo externo.

Redução da Taxonomia

A taxonomia reduzida $TR(S)$ (ver seção 2.4.1) pode ser calculada a partir do foco de interesse S , através da retirada de todos os conceitos que não estão relacionados aos objetos que compõem o foco de interesse S , ou seja, conceitos que não pertencem a $CR(S)$. O conjunto de conceitos relacionados a S , $CR(S)$, pode ser representado pelo conjunto de identificadores de conceitos (*id-c*) que estão relacionados a S . Esta é a chamada estratégia sem contagem. Numa estratégia que incorpore a contagem, cada elemento em $CR(S)$ deve conter, além do *id-c*, a cardinalidade de $Ext_Prof(C) \cap S$. Esta cardinalidade representa o número de objetos em S que são classificados sob o conceito C . Apesar de a estratégia com contagem ser computacionalmente mais custosa, a contagem representa uma informação de extrema importância para a interação do usuário. Ela indica se uma operação adicional de *zoom* pode ser útil, ou se a base de dados já se encontra reduzida o bastante de modo a permitir a inspeção individual dos resultados.

O conjunto $CR(S)$ pode ser computado de duas maneiras:

1. método guiado pelo foco;
2. método guiado pela taxonomia.

O método guiado pelo foco utiliza o conjunto de objetos que compõem o foco de interesse S e recupera todos os conceitos usados para classificar estes objetos. Para calcular a taxonomia reduzida $TR(S)$ é necessária uma estrutura responsável por armazenar, para cada objeto d na base de dados, os conceitos usados para classificar d : $CLASSIFICAÇÃO(id-o, id-c)$, onde *id-o* representa o identificador do objeto d , e *id-c*

representa o identificador do conceito usado para classificar d . O conjunto de conceitos relacionados $CR(S)$ pode ser, então, computado na estratégia sem contagem através do seguinte comando SQL:

```
SELECT DISTINCT T2.id-c
FROM FOCO T1, CLASSIFICAÇÃO T2
WHERE T1.id-o = T2.id-o
```

Onde $FOCO(id-o)$ representa a estrutura que armazena o foco de interesse S .

No caso da estratégia que incorpora a contagem, é necessário o seguinte comando SQL mais complexo:

```
SELECT T2.id-c, COUNT(*)
FROM FOCO T1, CLASSIFICAÇÃO T2
WHERE T1.id-o = T2.id-o
GROUP BY T2.id-c
```

Vale mencionar que o sistema Flamenco³⁴, de acordo com Yee e outros (2003), utiliza esta mesma estratégia apresentada, baseada em banco de dados relacional. No entanto, o sistema Flamenco não armazena a extensão profunda dos conceitos, tendo que calcular a mesma dinamicamente a partir da extensão rasa dos conceitos descendentes.

O método guiado pela taxonomia (SACCO; TZITZICAS, 2009) para o cálculo da taxonomia reduzida se baseia numa formulação alternativa da regra de inferência extensional:

Dado um conceito C e um segundo conceito expresso por um subconjunto arbitrário S de objetos do universo, C é relacionado à S , $C \leftrightarrow S$ se, somente se $Objetos(C) \cap S \neq \emptyset$

Vale lembrar que $Objetos(C)$ (objetos que são classificados sob o conceito C) equivale a $Ext_Prof(C)$ (extensão profunda do conceito C). Esta formulação mostra que o

³⁴ Ver em: <http://flamenco.berkeley.edu>

conjunto de conceitos relacionados $CR(S)$ pode ser calculado a partir de cada conceito da taxonomia, pela interseção entre a extensão profunda de C e o foco de interesse S . Caso esta interseção não seja vazia, então C pertence ao conjunto de conceitos relacionados $CR(S)$. O método guiado pela taxonomia requer apenas o uso da estrutura $EXTENSAO_PROFUNDA(id-c, id-o)$. No caso da estratégia sem contagem a operação de interseção pode parar no momento em que se encontre o primeiro objeto em comum, enquanto que na estratégia com contagem é necessário determinar a cardinalidade da interseção.

Lembrando que pela regra da restrição de inclusão, característica das relações hierárquicas presentes na taxonomia, se $C \notin CR(S)$ e C' é um descendente de C , então $C' \notin CR(S)$. Desta forma, o cálculo de $CR(S)$ pode ser otimizado através da adoção de uma abordagem *top-down*, ou seja, partindo da raiz para os conceitos terminais. No caso da extensão profunda de um conceito C ter uma interseção vazia com o foco atual S , todos os seus descendentes também terão interseção vazia, não sendo necessário realizar o processamento. A estrutura $DESCENDENTES$, presente na *intensão*, pode ser usada para determinar rapidamente todos os descendentes de C .

Na presente pesquisa, optou-se por utilizar o método guiado pelo foco, utilizando-se a relação $CLASSIFICAÇÃO(id-o, id-c)$, dada a facilidade de sua implementação em banco de dados relacional.

Estratégias de Apresentação

Na seção anterior foram descritas as operações utilizadas para a implementação do mecanismo de recuperação da informação utilizando taxonomias dinâmicas, principalmente no que se refere a aplicação da regra de inferência extensional de forma eficiente. Nesta seção são discutidos os aspectos referentes a apresentação da taxonomia dinâmica, ou seja, como a taxonomia reduzida é gerada e apresentada ao usuário durante o processo de interação com o sistema. Neste trabalho de pesquisa optou-se pelo uso de uma arquitetura *web* para a implementação da taxonomia. Desta forma, assume-se que o mecanismo responsável pelo processamento da taxonomia é executado no servidor *web*, e a interação entre o navegador *web* (cliente) e o servidor *web* se dá por conexões sem a manutenção de estado (*stateless*), que é o padrão neste tipo de arquitetura computacional.

Quando o usuário define ou altera o foco atual (pela seleção de conceitos), a taxonomia precisa ser recalculada (ou reduzida) de modo a retirar os conceitos não relevantes. Existem duas estratégias que podem ser adotadas:

1. *estratégia de recarga total*, na qual toda a árvore taxonômica é recalculada e reenviada ao cliente (navegador *web*). Qualquer interação adicional que não envolva novas operações de *zoom* pode ser executada no próprio cliente, sem a necessidade de novo acesso ao servidor *web* para processamento.
2. *estratégia de recarga sob demanda*, na qual somente o primeiro nível da taxonomia é recalculado, ou seja, os descendentes diretos da raiz. Os demais níveis serão recalculados sob demanda de acordo com a navegação do usuário. Esta estratégia requer um número maior de envio de pequenos pacotes de informação para processamento no servidor *web*, na medida que a expansão de um dado nível da taxonomia requer o cálculo do conjunto de descendentes diretos daquele nível.

Apesar de ambas as estratégias poderem ser empregadas em conjunto, tanto com o método de redução da taxonomia guiado pelo foco, quanto pelo guiado pela taxonomia, a estratégia de recarga sob demanda beneficia somente o método guiado pela taxonomia, visto que o método guiado pelo foco sempre recalcula toda a árvore taxonômica reduzida.

A tecnologia relacional para banco de dados representa um padrão de fato em aplicações comerciais. Dada a sua ampla difusão e uso, inclusive na implementação de sistemas de repositórios e bibliotecas digitais, optamos pelo seu uso na implementação da taxonomia dinâmica e nos mecanismos de interação do usuário com a mesma. Outro fator motivador para esta opção é o fato do Sistema DSpace, usado na implementação da BDTD/ECI/UFMG, também ser baseado nesta mesma tecnologia, o que facilitou a integração da taxonomia com os metadados já disponíveis nessa biblioteca.

Nesta pesquisa optamos pelo uso do método guiado pelo foco, em conjunto com a estratégia de recarga total. Esta opção se justifica pela maior facilidade de implementação das estruturas de armazenamento necessárias em formato relacional de banco de dados e, também, por requerer um número menor de conexões de comunicação entre cliente e servidor *web*.

Conceitos Virtuais

Em geral, as taxonomias dinâmicas assumem que toda a árvore taxonômica deva ser projetada e desenvolvida antes do seu uso, assim como ocorre com o esquema conceitual de um banco de dados. A taxonomia poderá ser modificada posteriormente, mas os conceitos deverão ser definidos para que possam ser incorporados e utilizados na nova versão da taxonomia. Existem diversas situações práticas nas quais este requisito pode se mostrar difícil de ser atendido, principalmente nos casos em que um dado ramo ou faceta da taxonomia ser voltado para a descrição de um atributo que possua um grande número de valores diferentes. Entre os exemplos podemos citar atributos como preços, datas, autores, etc. Apesar de ser possível acomodar estes valores como conceitos na abordagem apresentada, eles certamente irão impactar no tamanho da árvore taxonômica e irão requerer o projeto antecipado de todos os valores possíveis.

Como uma solução para este problema, Sacco e Tzitzicas (2009) propõem, como alternativa, o que eles chamam de *conceito virtual simples*. Na proposta desses autores, no caso de um conceito virtual, não é necessário armazenar nem os seus descendentes (conjunto de valores), nem a sua extensão (conjunto de objetos que possuem um dado valor). Tanto os descendentes quando a extensão é calculada por operações definidas a partir de um mecanismo externo.

Um conceito virtual simples *CV* (*data de publicação*, por exemplo) pode ser caracterizado pelas quatro operações seguintes:

1. Dado o conceito virtual *CV*, encontre todos os seus descendentes. No exemplo, isto significa encontrar todos os valores distintos, existentes na base de dados, para o atributo *data de publicação*;
2. Dado o conceito virtual *CV*, encontre a sua extensão profunda, ou seja, encontre todos os objetos para os quais o conceito *data de publicação* é definido;
3. Dado um descendente direto *f* (filho) do conceito virtual *CV*, encontre a sua extensão profunda, ou seja, encontre todos os objetos que possuam o valor específico *f* para *data de publicação*; e
4. Dado um objeto *o*, encontre todos os descendentes do conceito virtual *CV* sob os quais *o* é classificado, ou seja, encontre a *data de publicação* de um dado objeto.

Os conceitos virtuais oferecem uma forma simples para “virtualizar” partes específicas da taxonomia e materializar estas partes, quando necessário, a partir de um mecanismo externo que forneça as operações descritas acima.

Os conceitos virtuais simples descrevem uma ramo raso (ou faceta) da taxonomia e seus valores atuais. Sacco e Tzitzicas (2009) também definem os *conceitos virtuais derivados*, que permitem que os valores atuais de um conceito virtual simples possam ser agrupados de maneiras que sejam úteis aos usuários. Um *conceito virtual derivado* $D(CV)$ é derivado a partir de um conceito virtual CV pela especificação de restrições adicionais sobre o mesmo. Por exemplo, no caso da faceta (ou conceito virtual) *data de publicação*, seus valores não foram definidos explicitamente na taxonomia, mas obtidos a partir de um mecanismo externo usando as operações definidas acima. Com um conceito virtual derivado é possível definir uma hierarquia através da incorporação de restrições adicionais. Por exemplo, é possível especializar *data de publicação* em “*década de 1970*”, “*década de 1980*”, etc. através de operações adicionais sobre o conceito virtual. Desta forma, temos as restrições “ $1970 \leq \text{data de publicação} < 1980$ ” para a “*década de 1970*”, “ $1980 \leq \text{data de publicação} < 1990$ ” para a “*década de 1980*”, etc.

Os conceitos virtuais derivados podem ser obtidos a partir de outros conceitos virtuais derivados, de modo a definir uma hierarquia. Neste caso, as restrições adicionais são compostas pelo operador *AND*. Também é importante destacar que os conceitos virtuais derivados são dinâmicos por natureza, de modo que é interessante que possam ser dinamicamente especificados pelo usuário, a partir de valores fornecidos pelo mesmo, de modo a definir os grupamentos.

Com relação a implementação do protótipo na presente pesquisa, as operações que definem um conceito virtual puderam ser implementadas a partir da seguinte estrutura (relação) para cada conceito virtual CV : $CV(\text{Valor}, \text{id-o})$. Desta forma as quatro operações puderam ser implementadas com os seguintes comando SQL:

1. `SELECT DISTINCT Valor FROM CV;`
2. `SELECT DISTINCT id-o FROM CV;`
3. `SELECT DISTINCT id-o FROM CV WHERE Valor = f ;`
4. `SELECT DISTINCT Valor FROM CV WHERE id-o = o ;`

No caso da abordagem com contagem, ela pôde ser facilmente acomodada. A estrutura CV não precisa ser explicitamente armazenada, podendo ser calculada a partir de outras relações do banco de dados. Caso seja explicitamente armazenada, é possível abrigar mais de um conceito virtual na mesma relação. No momento do cálculo da taxonomia reduzida para um dado foco de interesse F , as operações foram aplicadas na junção das relações CV e F , através do campo *id-o* (identificador dos objetos).

No caso dos conceitos virtuais derivados, eles puderam ser tratados, usando as mesmas operações, pela simples adição de uma cláusula WHERE em cada comando SQL. Por exemplo, considere o conceito virtual *data de publicação*. Ao invés de definir seus descendentes diretos como sendo os valores das datas, estes valores podem ser agrupados em faixas correspondentes á décadas por exemplo. Desta forma, as quatro operações necessárias para definir o conceito virtual derivado “*década de 1990*” serão as mesmas descritas acima, acrescida da cláusula “WHERE Valor \geq 1990 and Valor $<$ 2000”.

No caso da presente pesquisa, para a implementação dos mecanismos de busca no contexto da implementação do protótipo, os atributos “orientador”, “tipo do documento”, “data de publicação” e “linha de pesquisa” foram tratados como conceitos virtuais, sendo que “data de publicação” fora tratado como um conceito virtual derivado.

3.2.2 O protótipo desenvolvido

Nesta seção são brevemente apresentados alguns aspectos técnicos relacionados à implementação do protótipo computacional. A descrição detalhada do protótipo será apresentada no Capítulo 4 (Resultados).

O protótipo computacional, que foi chamado TDF-Biblio³⁵, foi desenvolvido na forma de uma biblioteca digital de teses e dissertações. Utilizou-se a plataforma DSpace, em sua versão 1.8.2, para a implementação. Desta forma, os novos mecanismos de busca e recuperação da informação foram desenvolvidos de forma acoplada ao DSpace. Adotou-se a linguagem de programação Java³⁶ e o banco de dados PostgreSQL³⁷.

³⁵ O protótipo pode ser acessado no endereço: <http://www.tdf-biblio.ufv.br>

³⁶ Ver em: <http://www.java.com>

³⁷ Ver em: <http://www.postgresql.org>

O acervo utilizado foi importado a partir da BDTD/ECI/UFMG, utilizando as ferramentas de importação/exportação de conteúdos disponibilizadas pelo DSpace. A importação foi realizada em novembro de 2011, quando a biblioteca contava com 195 documentos, entre teses e dissertações.

O mecanismo de recuperação da informação foi chamado de Busca Dinâmica-Facetada. O mecanismo utiliza um modelo baseado na classificação multidimensional (facetada) dos documentos do acervo e permite a exploração e a navegação no acervo de uma forma guiada, porém não restritiva, por meio de uma interface visual fácil e intuitiva. O mecanismo de busca foi inicialmente implementado em duas modalidades: “Pelos Metadados” e “Pela Taxonomia Facetada”. A modalidade “Pelos Metadados” utiliza os atributos descritivos, representados nos elementos de metadados Dublin Core: Orientador (advisor), Tipo do Documento (type) e Ano de Publicação (date.issued). Além desses atributos, já presentes nos metadados importados, foi também acrescentado o atributo referente a Linha de Pesquisa de cada documento, representado pelo elemento “subject.linha-de-pesquisa”. Além dos atributos descritivos, foram também representados os aspectos temáticos dos documentos, modelados pela taxonomia facetada e representados através de novos elementos de metadados que foram criados através da qualificação do elemento “subject”, quais sejam: “subject.tema”, “subject.objeto”, “subject.ambientacao”, “subject.tipo-de-pesquisa”, “subject.coleta-de-dados”, “subject.metodos”, “subject.fundamento-teorico”, “subject.fundamento-historico-contextual” e “subject.resultados”.

Já na modalidade de busca “Pela Taxonomia Facetada”, foram utilizados apenas os aspectos temáticos das teses e dissertações que compõem o acervo. Nessa modalidade, a Taxonomia Facetada é exibida para o usuário, permitindo que ele possa navegar pela estrutura da Taxonomia, selecionando os conceitos a serem utilizados na busca, com a possibilidade de utilizar os operadores AND ou OR para combinação dos conceitos selecionados. A FIGURA 24 e a FIGURA 25 apresentam, respectivamente, as telas iniciais de ambas as modalidades de busca inicialmente implementadas no protótipo.

TDF-Biblio
ORGANIZAÇÃO DO CONHECIMENTO EM BIBLIOTECAS DIGITAIS:
Uma abordagem baseada na classificação facetada e na taxonomia dinâmica.

TDF-Biblio → Página Inicial → Busca pelos Metadados

Busca Dinâmica-Facetada

- ▶ Pelos Metadados
- ▶ Pela Taxonomia Facetada
- ▶ Ajuda

Busca pelos Metadados

- Orientador
 - ▶ Ricardo Rodrigues Barbosa (17)
 - ▶ **Maria Aparecida Moura (15)**
 - ▶ **Lidia Abarenho (14)**
 - ▶ Beatriz Valadares Cendon (11)
 - ▶ Alcinor Soares dos Reis (10)
 - ▶ Ana Maria Rezende Caira (9)
 - ▶ Eduardo Jose Wense Dias (9)
 - ▶ Maria Guilomar da Cunha Frota (9)
 - ▶ Marlene de Oliveira (9)
 - ▶ Marcello Peixoto Bar (8)
 - ▶ (mais ...)
- Tipo de Documento
 - ▶ Dissertação de Mestrado (141)
 - ▶ Tese de Doutorado (54)
- Data de Publicação
 - ▶ 2010 - 2011 (24)
 - ▶ 2000 - 2009 (167)
 - ▶ 1995 - 1999 (4)
- Tema
 - ▶ Aspectos teóricos e gerais da ciência da informação (4)
 - ▶ Gerência de serviços e unidades de informação (4)
 - ▶ Estudos de usuário, demanda e uso da informação e de unidades de informação (10)
 - ▶ Tecnologias da informação (2)
 - ▶ Processamento, recuperação e disseminação da informação (20)
 - ▶ Assuntos correlatos e outros (1)

Busca pelos Metadados

Busca pelos Metadados
Abrangência da busca:

Filtros
Atualizar Filtros:
Use filtros para refinar a pesquisa.
Título

Ordenação
Resultados por página: Classificar itens por: em ordem:

Busca pelos Metadados

Exibindo itens 1-10 de 199 1 2 3 4 ... 20 Próxima Página

MSI **Análise conceitual sobre as relações semânticas em ciência da informação: contribuições para o desenvolvimento de ontologias**
Julia Aparecida Gonçalves Campos; Renato Rocha Souza (UFMG, 2009-05-27)
Para o desenvolvimento de ontologias, verifica-se a falta de métodos sistematizados para a modelagem conceitual resultado do processo intelectual de análise conceitual que, associado a outras técnicas, envolve a identificação, ...

MSI **Avaliação de usabilidade em bibliotecas digitais: um estudo de caso**
Fernanda Pereira; Gercina Angela Borem de Oliveira Lima (UFMG, 2011-04-27)
As bibliotecas digitais têm por objetivo disponibilizar informações para seus usuários com maior rapidez e facilidade de busca e acesso. Nessas bibliotecas, a interação dos usuários com o acervo realiza-se por meio de suas ...

FIGURA 24 - Modalidade de busca pelos metadados.

Fonte: elaborado pelo autor.

TDF-Biblio
ORGANIZAÇÃO DO CONHECIMENTO EM BIBLIOTECAS DIGITAIS:
Uma abordagem baseada na classificação facetada e na taxonomia dinâmica.

TDF-Biblio → Página Inicial → Busca pela Taxonomia Facetada

Busca Dinâmica-Facetada

- ▶ Pelos Metadados
- ▶ Pela Taxonomia Facetada
- ▶ Ajuda

Taxonomia

Expandir - Reduzir

- TEMA (41)
- OBJETO (41)
 - Gerência de serviços e unidades de informação (10)
 - Comunicação, divulgação e produção editorial (11)
 - Avaliação de periódicos (2)
 - Documentação científica (6)
 - Literatura (1)
 - Produção audiovisual (1)
 - Scripts de atendimento (1)

AND OR

Histórico
OBJETO

Busca pela Taxonomia Facetada

Resultados da busca

Sua requisição produziu 41 resultado(s).

Exibindo item 1-10 de 41. Primeira Anterior Próxima Última

MSI **A análise de assunto na literatura ficcional infantil: categorias para ler o que você tem**
Margareth Epitima Moreira; Eduardo Jose Wense Dias (UFMG, 10/04/2006)
Considerando a necessidade de investigação, na biblioteca escolar, da análise de assunto na literatura ficcional infantil, buscou-se identificar os procedimentos de indexação, relativos à leitura técnica, identificação e seleção de conceitos de três títulos literários. Utilizou-se o método de Protocolo [...]

MSI **A indexação temática de recursos fundamentada por estrutura profunda e abordagem objeto-relacionamento**
Maria Aparecida Lourenço Santana; Madalena Martins Lopes Naves (UFMG, 21/02/2005)
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 máquina [...]

MSI **A interação dos usuários da UFMG com o catálogo on-line do Sistema Pergamum**
Carla Cristina Vieira de Oliveira; Eduardo Jose Wense Dias; Beatriz Valadares Cendon (UFMG, 30/04/2008)
Os catálogos online, também conhecidos como Online Public Access Catalogue, mudaram a rotina dos usuários de bibliotecas, possibilitando a utilização de diversos recursos por vários usuários ao mesmo tempo, sem limite de espaço. Conhecer a interação dos usuários de bibliotecas universitárias com o ca [...]

MSI **A utilização de mapas de tópicos na compatibilização de conteúdos hipertextuais semanticamente estruturados**
Guilherme Baiao Salgado Silva; Gercina Angela Borem de Oliveira Lima (UFMG, 26/08/2008)
Considerando as características dos sistemas de hipertexto e os problemas da desorientação e da sobrecarga cognitiva dos usuários desses sistemas, este projeto aborda como objeto de estudos os documentos hipertextuais que passaram pelo processo de estruturação conceitual, utilizando estruturas facet [...]

FIGURA 25 - Modalidade de busca pela taxonomia facetada.

Fonte: elaborado pelo autor.

Posteriormente, a modalidade de busca “Pela Taxonomia” foi retirada do protótipo, sendo que somente a modalidade “Pelos Metadados” fora mantida, tendo a sua denominação sido alterada para “Busca Dinâmica Facetada”. Esta decisão se deveu a dois motivos principais: 1) maior facilidade de entendimento e uso por parte dos usuários, visto que a exibição de toda a taxonomia poderia sobrecarregar o usuário e dificultar a sua interação; 2) funcionalidade mais ampla proporcionada pela modalidade que foi mantida, visto que tantos os aspectos descritivos quanto os temáticos dos documentos são contemplados de uma forma uniforme pela mesma.

3.3 Avaliação

Nesta seção, são apresentados o ambiente e a configuração da avaliação conduzida no contexto desta pesquisa (Etapa 3). O objetivo da avaliação é determinar a influência ou impacto, do ponto de vista dos usuários, da organização do acervo que utiliza uma estrutura classificatória facetada, bem como dos novos mecanismos de recuperação da informação, baseados no conceito de taxonomias dinâmicas e implementados em um protótipo de sistema de biblioteca digital de teses e dissertações. As decisões relacionadas à configuração do experimento de avaliação são baseadas em Saracevic (2000, 2004), Kovács e Micsik (2004) e Fuhr e outros (2001, 2007), conforme apresentado na seção 2.5.

Com relação ao nível da avaliação, de acordo com os seis níveis propostos por Saracevic (2000), este estudo adota o nível *Individual*, ou seja, quão bem os componentes de acesso e recuperação da informação oferecem suporte à necessidade de informação, tarefas e atividades de pessoas como usuários individuais ou grupos de usuários com características comuns. Esta escolha se justifica por ser este o nível mais direto e fácil de ser avaliado e que, ao mesmo tempo, tem a capacidade de revelar o benefício a ser obtido para os usuários na medida em que se adota um mecanismo alternativo de organização do acervo baseado na classificação facetada.

Quando se considera o Modelo de Avaliação Computacional (KOVÁCS; MICSIK, 2004) de modo a determinar o perfil da avaliação a ser empregada nesta pesquisa, chegamos à configuração apresentada no QUADRO 11 abaixo.

QUADRO 11 - Configuração da avaliação baseada no Modelo de Avaliação Computacional

Aspecto	Dimensões/Facetadas	Escolhas	Justificativa
Avaliação	Coleta de dados	Questionários com questões estruturadas e semiestruturadas; observação a partir de filmagem da interação do usuário.	Perguntas semiestruturadas, assim como a observação da interação do usuário podem oferecer dados qualitativos valiosos; questões estruturadas podem oferecer maior capacidade de generalização.
	Tipo da avaliação	Comparativa	Em alinhamento com o objetivo de avaliar o benefício a ser obtido com a adoção de um mecanismo alternativo de organização do acervo baseado na classificação facetada.
Usuários	Tipo de usuários	Estudantes, pesquisadores internos à instituição.	Estudantes de pós-graduação do PPGCI/ECI/UFMG são os principais usuários da BDTD.
	Escopo	Comunidade	Comunidade formada pelos estudantes de pós-graduação do PPGCI/ECI/UFMG.
Organização	Setor	Acadêmico, pesquisa	Ambiente acadêmico universitário em nível de pós-graduação.
Conteúdo	Forma	Texto	Formato textual das teses e dissertações que constituem o acervo da BDTD.
	Tipo de conteúdo	Metadados e texto completo	Texto completo das teses e dissertações disponíveis, assim como seus respectivos metadados.
Sistema	Tipo de aplicação	Interface web	O acesso a BDTD é realizado atualmente através de sua interface web. O protótipo desenvolvido para o estudo comparativo adota o mesmo tipo de interface.
	Arquitetura	Cliente-servidor	Tanto a BDTD quanto o protótipo desenvolvido são baseados na arquitetura cliente-servidor.

Fonte: construído pelo autor a partir do modelo proposto por Kovács e Micsik (2004).

O perfil ou configuração da avaliação foi, conforme sintetizado pelo QUADRO 11, de natureza comparativa, empregando questionários com questões semiestruturadas e estruturadas para a coleta de dados. Também foi realizada a filmagem da interação dos usuários com o sistema. A avaliação considera uma amostra da comunidade formada pelos usuários da BDTD enquanto estudantes de pós-graduação do Programa de Pós Graduação em Ciência da Informação (PPGCI) da Escola de Ciência da Informação (ECI) da UFMG. O conteúdo do acervo compreende objetos textuais (teses e dissertações defendidas no PPGCI/ECI/UFMG, na forma de texto completo com seus respectivos metadados). O componente tecnológico corresponde à interface *web* da BDTD e utiliza a arquitetura cliente-servidor. Cabe ainda destacar que, no contexto do estudo comparativo empreendido, todos os componentes da biblioteca digital (conteúdo/acervo, usuários e sistema) são exatamente os mesmos, com exceção dos componentes responsáveis pelo acesso e recuperação da informação, que são o alvo do estudo comparativo.

Quando se considera o Modelo Tríplice de Interação (FUHR *et al.* 2007), o experimento conduzido tem seu foco na interação entre o componente *Sistema* e o componente *Usuário* da biblioteca digital, sob o eixo da usabilidade. O objetivo foi avaliar até que ponto a adoção de um mecanismo alternativo baseado na classificação facetada para organização do acervo afeta o desempenho da interação *Usuário-Sistema* na medida em que o usuário utiliza a biblioteca para satisfazer suas necessidades de informação.

3.3.1 Definição operacional: perguntas a serem respondidas pela avaliação

Em alinhamento com os objetivos da pesquisa, apresentados na seção 1.3, foram definidas duas perguntas que deverão ser respondidas a partir da análise dos dados produzidos pelo experimento conduzido. O ambiente e procedimentos do experimento foram projetados para fornecer respostas a estas perguntas:

Pergunta 1: A organização do acervo com base numa estrutura facetada e os novos mecanismos de busca e acesso implementados no protótipo TDF-Bíblia facilitam a recuperação da informação e melhoraram seus resultados, considerando o ponto de vista dos usuários da BDTD/ECI/UFMG, em tarefas de busca e recuperação da informação, visando atender às suas necessidades de informação? O objetivo é determinar se o desempenho dos usuários nas tarefas de busca e recuperação da informação é melhorado,

no caso de uso dos novos mecanismos de busca e acesso, implementados no protótipo desenvolvido.

Pergunta 2: A organização do acervo com base numa estrutura facetada e os mecanismos de busca e acesso implementados no protótipo TDF-Bíblia melhoram a satisfação dos usuários da BDTD/ECI/UFMG? O objetivo é determinar se os usuários ficam mais satisfeitos ao utilizar um sistema de biblioteca digital com os mecanismos de busca e acesso implementados no protótipo desenvolvido.

Outro requisito considerado é que a avaliação fosse simples, de modo a ser executada pelos usuários da biblioteca digital. Com base nestes requisitos, foram definidas as seguintes atividades a serem desenvolvidas no contexto do experimento:

1. *Treinamento dos usuários* – necessário para que os usuários adquiram familiaridade com a biblioteca digital; teve como objetivo possibilitar com que os usuários conheçam o ambiente do sistema de biblioteca digital sendo utilizado, e que sejam capazes de usá-lo nas tarefas subsequentes.
2. *Tarefas do tipo pergunta-resposta* – estas tarefas requerem que os usuários realizem operações de busca e navegação na biblioteca de modo a responder determinadas questões que lhes são oferecidas; estas tarefas têm o objetivo de ajudar a responder à primeira pergunta da avaliação.
3. *Questionário de satisfação* - Após a realização de cada tarefa, os usuários foram solicitados a preencher um questionário que forneceu subsídios para responder à segunda pergunta (satisfação do usuário).

Os dados foram coletados a partir do uso dos novos mecanismos de busca e acesso, implementados no protótipo desenvolvido na Etapa 2 desta pesquisa. Estes dados foram comparados com os dados coletados a partir do uso dos mecanismos de busca e acesso convencionais, disponíveis atualmente na BDTD/ECI/UFMG. Cada usuário realizou as mesmas tarefas, utilizando ambos os sistemas. A seguir será apresentado o projeto da implementação do experimento.

3.3.2 Experimento

O objetivo é determinar a influência ou impacto da organização do acervo que utiliza uma estrutura classificatória facetada, bem como dos novos mecanismos de recuperação da informação, baseados no conceito de taxonomias dinâmicas e implementados em um protótipo de sistema de biblioteca digital de teses e dissertações. Para tanto foi constituído um experimento com base comparativa, contando com a participação de usuários e envolvendo testes de uso de ambos os sistemas.

O sistema de biblioteca digital utilizado pelos usuários pode ser visto como uma variável independente, que será analisada em duas situações distintas: 1) sistema atualmente utilizado na Biblioteca Digital de Teses e Dissertações da Escola de Ciência da Informação da UFMG - BDTD/ECI/UFMG³⁸, que tem a organização de seu acervo baseada nos metadados e que oferece mecanismos de recuperação da informação convencionais, permitindo a navegação e a busca textual no acervo; e 2) protótipo de sistema de biblioteca digital TDF-Bíblia³⁹, implementado na Etapa 2 desta pesquisa, que utiliza, além dos metadados convencionais, uma estrutura classificatória facetada, além de taxonomias dinâmicas para compor novos mecanismos de recuperação da informação. Deseja-se estudar as variáveis dependentes, relacionadas à eficácia da recuperação da informação e satisfação dos usuários da biblioteca.

Foi selecionado um grupo de usuários para a realização de um teste de uso envolvendo ambos os sistemas: 1) sistema atual da BDTD/ECI/UFMG, e 2) protótipo TDF-Bíblia.

3.3.2.1 População e Amostra

População:

1. População Alvo: Usuários da BDTD/ECI/UFMG – pode incluir pessoas externas ao Programa de Pós Graduação em Ciência da Informação da UFMG – PPGCI/UFMG.
2. População Acessível: Estudantes de pós-graduação do PPGCI/UFMG – mais fácil para a definição da amostra, sendo que são eles os principais usuários da BDTD/ECI/UFMG.

Amostra:

³⁸ Ver em: <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/477>

³⁹ Ver em: <http://www.tdf-biblio.ufv.br>

- Estudantes de pós-graduação do PPGCI/UFMG selecionados para participar do experimento.

Nos estudos comparativos, categoria na qual se enquadra o experimento proposto, o principal objetivo é comparar certas características da amostra da população em estudo, sob diferentes condições artificialmente impostas. Neste contexto, a principal preocupação no plano de amostragem foi obter amostras comparáveis, ou seja, que se diferenciem somente com respeito ao fator de comparação, representado pelo sistema de biblioteca digital utilizado.

Partimos de uma hipótese sobre a população em estudo e queremos verificar a sua validade, com base em uma amostra. A hipótese é de que haverá variação de desempenho relacionado à eficácia da recuperação da informação e satisfação do usuário, durante o uso de ambos os sistemas, no contexto do experimento a ser aplicado, e que esta variação será positiva em favor do protótipo TDF-Bíblia.

3.3.2.2 Preparação do experimento

Nesta seção, serão apresentados os procedimentos de preparação para a realização do experimento, envolvendo o recrutamento dos participantes, agendamento dos testes e preparação do ambiente de testes.

3.3.2.2.1 Levantamento Inicial

Foi realizado um levantamento inicial com o objetivo de obter informações demográficas, hábitos, necessidades e formas de acesso à informação para pesquisas, entre os estudantes do PPGCI/UFMG. Para tanto foi enviado um *e-mail* para todos os estudantes matriculados, contendo *link* para um questionário *on-line*. O e-mail enviado aos estudantes do PPGCI/UFMG pode ser consultado no Apêndice A. O Questionário de Levantamento Inicial (Q1), além de uma síntese dos dados coletados a partir do mesmo, encontram-se no Apêndice B. O questionário também teve como objetivo identificar os estudantes dispostos a participar do experimento. Mais especificamente, o levantamento inicial teve como objetivo obter informações sobre:

- Estimativa do número de estudantes potenciais para o experimento, bem como seus dados demográficos e áreas de formação;
- Principais fontes de informação utilizadas;
- Estimativa do uso da BDTD/ECI/UFMG pelos estudantes;

- Necessidades de informação dos estudantes relacionadas com o acervo de teses e dissertações da BDTD/ECI/UFMG.
- Disponibilidade do estudante para participar do experimento.

A partir de um conjunto de 47 registros de respostas do Questionário de Avaliação Inicial – Q1, foram identificados um total de 35 respondentes que ofereceram informações potencialmente úteis para a elaboração do experimento. Estes 35 respondentes foram classificados conforme o QUADRO 12 a seguir.

QUADRO 12 - Categorização dos respondentes quanto ao grau de contribuição relacionada ao preenchimento do Questionário Q1

CATEGORIA	OBSERVAÇÕES	Nº DE RESPONDENTES
1	Respondentes que se identificaram e se dispuseram a participar do experimento de pesquisa, além de terem fornecido informações sobre suas necessidades informacionais, relacionadas ao acervo da BDTD/ECI/UFMG, que foram utilizadas na formulação de questões a serem oferecidas aos participantes do experimento. Estes respondentes foram posteriormente contactados via e-mail para saber da disponibilidade para a realização do teste na semana de 17 a 21 de dezembro de 2012.	17
2	Respondentes que, apesar de não terem se identificado e ou não demonstrado interesse em participar do experimento, ofereceram informações sobre suas necessidades informacionais, relacionadas ao acervo da BDTD/ECI/UFMG, que também foram utilizadas na formulação de questões a serem oferecidas aos participantes do experimento.	4
3	Respondentes que não se identificaram e ou não demonstraram interesse em participar do experimento. No entanto, formularam questões específicas de seus projetos de pesquisa, que fornecem uma visão geral sobre as necessidades informacionais dos estudantes do PPGCI no que se refere ao acervo de teses e dissertações da BDTD/ECI/UFMG.	14
	TOTAL:	35

Os 17 respondentes que se identificaram e demonstraram interesse em participar do experimento (categoria 1) foram contactados por *e-mail*, a fim de explicar os objetivos e procedimentos previstos para os testes, bem como verificar a sua disponibilidade para realização do teste no período previsto. O modelo do *e-mail* individual e personalizado encontra-se no Apêndice C. Por fim, foram recrutados um total de 7 participantes (usuários) para a realização dos testes no contexto do experimento proposto.

3.3.2.2.2 *Preparação do ambiente de testes*

Durante a preparação dos testes, as tarefas de busca e recuperação da informação foram formuladas, levando-se em consideração os temas, interesses e respectivos projetos de pesquisa de cada respondente, além de suas necessidades informacionais, que

poderiam ser atendidas com base em consultas ao acervo de teses e dissertações da BDTD/ECI/UFMG. Trata-se de uma abordagem mais centrada nas reais necessidades informacionais do usuário, o que pode contribuir para a sua participação e engajamento nas atividades no contexto do experimento.

Para a elaboração das tarefas, pensou-se inicialmente em preparar um conjunto de tarefas específico para cada participante, com base no seu projeto de pesquisa e nas necessidades de informação informadas pelo mesmo no Questionário Q1. No entanto, esta abordagem, apesar de apresentar a vantagem ser focada nas reais necessidades informacionais do usuário individual, não iria possibilitar a comparação dos resultados dentro do conjunto de participantes do experimento. Por esse motivo, optou-se por elaborar um conjunto de tarefas único a ser apresentado a todos os participantes do experimento. Estas tarefas foram elaboradas tendo como base os dados coletados a partir das respostas do Questionário Q1, mas especificamente nas características gerais identificadas nas respostas das questões relacionadas aos projetos de pesquisa e necessidades de informação declaradas pelos respondentes.

Cada respondente do Questionário de Levantamento Inicial – Q1 – elaborou uma questão ou pergunta representativa de suas necessidades de informação, que poderia ser respondida a partir de consultas ao acervo de teses e dissertações da BDTD/ECI/UFMG. Eles foram orientados a descrever a pergunta com suas próprias palavras de modo que pudesse ser passada a um bibliotecário de referência, responsável pelo atendimento na biblioteca. Os respondentes também foram orientados a fornecer algumas informações sobre seu projeto de pesquisa, como título, descrição sucinta, metodologia e pesquisas relacionadas. Estas informações foram utilizadas para entender melhor as necessidades de informação do respondente.

As questões elaboradas pelos respondentes puderem ser consideradas representativas das necessidades informacionais dos estudantes do PPGCI/UFMG quando se dirigem a BDTD/ECI/UFMG. A abordagem utilizada buscou obter questões que fossem relevantes para os participantes do experimento, de modo a suscitar o envolvimento e o engajamento dos mesmos na busca pelas respostas.

Com base nas respostas da questão número 9 (necessidades de informação) do Questionário de Levantamento Inicial (Q1), apresenta-se, a seguir, o QUADRO 13 com as principais categorias de respostas identificadas.

QUADRO 13 - Categorização das necessidades de informação relatadas pelos respondentes do Questionário Q1

CATEGORIA	EXEMPLOS REPRESENTATIVOS DA CATEGORIA	Nº DE OCORRÊNCIAS
1) Temas de pesquisa específicos	"Preciso encontrar trabalhos desenvolvidos na ECI sobre a temática da acessibilidade Web para pessoas com deficiência."	20
2) Perguntas específicas	"Quais indicadores de inovação são utilizados pelas organizações brasileiras?"	7
3) Trabalhos que utilizaram uma dada metodologia	"Você teria alguma dissertação ou tese que realizam uma análise de cluster na área da Ciência da Informação?"	4
4) Correlações entre temas de pesquisa, orientadores e data de publicação	<p>"Quais orientadores trabalham com as contribuições da linguística para a Ciência da informação?"</p> <p>"Que publicações recentes acerca do tema 'direito de propriedade intelectual' estão disponíveis no acervo?"</p> <p>"Teses ou dissertações publicadas nos últimos 5 anos que fazem referência ao uso de expressões multipalavras/n-gramas em sistemas de recuperação da informação."</p>	5
5) Cobertura do acervo	<p>"No âmbito da biblioteca digital de teses e dissertações da UFMG gostaria de saber, exatamente, qual a cobertura com relação ao PPGCI: a partir de qual ano as teses e dissertações estão disponibilizadas?"</p> <p>"Gostaria de saber qual o período de cobertura da BDTD/PPGCI e se todas as teses e dissertações disponíveis no formato impresso na biblioteca (referente a esse período de cobertura) estão, também, na BDTD."</p>	3
6) Questões exploratórias: trabalhos em uma dada linha de pesquisa; como um dado tema é tratado; etc.	<p>"Por que as linhas de pesquisas do PPGCI não são utilizadas para categorizar melhor as buscas a serem realizadas pelos diversos estudantes da Escola?"</p> <p>"Que trabalhos tratam da atuação do profissional bibliotecário? A pergunta é de caráter exploratório, ou seja, desejo obter todos os trabalhos que tratam sobre a atuação do profissional bibliotecário para constituir um panorama e identificar sob quais abordagens ou critérios o tema é recorrentemente tratado."</p>	3

Tendo como base as necessidades de informação identificadas e categorizadas no quadro acima, adotaram-se os seguintes critérios adicionais para a elaboração das questões a serem apresentadas aos participantes no contexto do experimento proposto:

- Cada questão deveria ser formulada como uma pergunta representativa de uma necessidade de informação, a ser apresentada ao participante.
- Para executar a tarefa e responder a respectiva questão, o participante deveria realizar buscas e navegação, utilizando o sistema de biblioteca digital.
- Cada questão deveria representar uma necessidade de informação válida e útil para o participante, de modo a suscitar o seu engajamento na busca pela resposta.
- Cada questão deveria ser representativa de pelo menos uma das categorias de necessidades de informação identificadas e apresentadas anteriormente (ver QUADRO 13).

Documentos do tipo teses e dissertações são extensos, o que inviabiliza a sua leitura pelo participante durante a realização dos testes. Por este motivo, as respostas para as questões elaboradas devem prescindir da leitura dos documentos. Deveriam ser questões que pudessem ser respondidas a partir da localização e exame rápido dos documentos, pela leitura dos títulos dos mesmos ou, em alguns casos, pela leitura do resumo.

A partir da realização de um pré-teste, chegou-se a um número de 5 questões, como adequado para realização dos testes num tempo total aproximado de 40 minutos. Lembrando que cada tarefa de busca e recuperação da informação deveria ser executada duas vezes: usando o sistema atual da BDTD e usando o protótipo TDF-Bíblia. Com a finalidade de avaliar o comportamento em um contexto de maior liberdade, foi incluída uma questão de busca livre (quinta questão).

Com base nos critérios apresentados acima, o QUADRO 14 apresenta as questões formuladas e posteriormente apresentadas aos participantes no contexto do experimento.

QUADRO 14 - Questões apresentadas aos participantes

Nº	QUESTÃO
1	Quais os documentos (testes ou dissertações) orientados por seu orientador(a), que tenham sido publicados nos últimos 5 anos.
2	Qual foi o orientador (a) que mais orientou trabalhos dentro da sua “Linha de Pesquisa”?
3	Quais as “Teses” defendidas na linha de pesquisa “Gestão da Informação e do Conhecimento”, nos últimos 5 anos?
4	Quais os documentos (teses ou dissertações), que relatam a utilização do “Questionário” para “Coleta de Dados”?
8	Elabore livremente uma tarefa de recuperação da informação e utilize a biblioteca digital em busca de documentos relevantes.

Conforme anteriormente apresentado na definição operacional do experimento, após a execução das tarefas de busca e recuperação da informação, utilizando cada sistema de biblioteca digital, os usuários foram solicitados a responder a um Questionário de Satisfação – Q2. Para a elaboração do Questionário de Satisfação, com base nos trabalhos de Chin e outros (1988) e Kruk (2008), foram utilizados três conjuntos de métricas de satisfação:

- *Métricas relacionadas à satisfação com as tarefas realizadas.* Um conjunto de três métricas foi utilizado para avaliar a opinião do usuário relacionada à tarefa sendo executada: (1) difícil de entender X fácil de entender; (2) difícil de executar X fácil de executar; e (3) não intuitiva X intuitiva.
- *Métricas relacionadas à satisfação com o uso do sistema.* Um conjunto de seis métricas foi utilizado para medir a satisfação geral do usuário com os componentes de busca e acesso à informação oferecidos pelo sistema de biblioteca digital: (1) fácil de usar X difícil de usar; (2) complexo, intimidador X simples, claramente organizado; (3) difícil de aprender X direto, descomplicado; (4) tedioso, maçante X interessante; (5) não atraente, pouco interessante X atraente, interessante; e (6) inútil X útil, conveniente.

O questionário de satisfação pode ser consultado no Apêndice D.

Tendo sido apresentados os principais componentes do ambiente de testes do experimento, o QUADRO 15, a seguir, resume e apresenta a estimativa de tempo previsto para a realização de cada atividade do experimento.

QUADRO 15 - Atividades do experimento e previsão de tempo

ATIVIDADE	DESCRIÇÃO	TEMPO
Treinamento	Tutorial explicativo do uso de ambos os sistemas envolvidos.	10 minutos
Tarefas de busca e recuperação da informação usando o Sistema 1	Tarefas de busca e recuperação da informação, fazendo uso do Sistema atual da BDTD/ECI/UFMG.	10 minutos
Questionário de satisfação de uso do Sistema 1	Questionário de satisfação de uso do Sistema atual da BDTD/ECI/UFMG.	5 minutos
Tarefas de busca e recuperação da informação usando o Sistema 2	Tarefas de busca e recuperação da informação, fazendo uso do Protótipo TDF-Bíblia.	10 minutos
Questionário de satisfação de uso do Sistema 2	Questionário de satisfação de uso do uso do Protótipo TDF-Bíblia.	5 minutos
	TEMPO TOTAL:	40 minutos

3.3.2.3 Aplicação dos testes e coleta de dados

Os participantes foram convidados a participar do experimento num dia e horário previamente acertado com cada participante. Os testes foram aplicados com sete participantes (sujeitos), individualmente, entre os dias 19 e 21 de dezembro de 2012. Os procedimentos foram detalhadamente explicados a cada participante, antes do início dos trabalhos. Para o controle da execução das atividades e gravação (em áudio e vídeo) da interação do usuário, foi utilizado o software Morae⁴⁰. O referido software possibilitou a gravação, em primeiro plano, das atividades na tela, bem como a filmagem do participante num quadro específico (formato *Picture in Picture* – PIP). As atividades foram gravadas visando possibilitar, posteriormente, a coleta detalhada dos dados.

Na data e horário agendado para a aplicação, antes do início dos testes de uso dos sistemas, foi realizado o treinamento dos participantes. O treinamento teve como objetivo o aprendizado do uso do sistema de biblioteca digital, para possibilitar que o participante

⁴⁰ Ver em: <http://www.techsmith.com/morae.html>

conheça as funcionalidades e adquira fluência no seu uso. Foi preparado um tutorial de uso do Protótipo TDF-Bíblia e um outro do Sistema atual da BDTD/ECI/UFMG, ambos apresentados aos participantes.

Na sequência, os participantes foram orientados a utilizar ambos os sistemas (sistema atual da BDTD/ECI/UFMG e o protótipo TDF-Bíblia, respectivamente) a fim de localizar os documentos relevantes, buscando responder às questões apresentadas (tarefas). Tanto o sistema atual da BDTD/ECI/UFMG, como o protótipo desenvolvido, forneceu acesso ao mesmo acervo, composto de metadados e do texto integral dos documentos. O acervo é constituído por cerca de 200 documentos, entre teses e dissertações, defendidas no PPGCI no período de 1995 a 2011. No caso do protótipo TDF-Bíblia, estes documentos foram classificados de acordo com uma estrutura classificatória facetada desenvolvida por Maculan (2011). O experimento foi realizado em um ambiente controlado e de forma individualizada com cada participante. O computador utilizado ofereceu acesso ao protótipo TDF-Bíblia e também ao sistema atual da BDTD/ECI/UFMG.

Após a realização das tarefas de recuperação da informação, em cada sistema, os participantes foram orientados a preencher o Questionário de Satisfação – Q2. Na próxima seção serão apresentados os procedimentos utilizados para a codificação dos dados coletados, visando à sua análise posterior, com vista a responder às perguntas definidas anteriormente.

3.3.3 Codificação dos dados coletados

Nesta seção, são apresentados os procedimentos utilizados para a codificação e formatação dos dados coletados durante a realização dos testes de uso, no contexto do experimento proposto.

3.3.3.1 Tarefas de busca e recuperação da informação

As métricas tradicionalmente utilizadas para avaliar a recuperação da informação são revocação e precisão, ambas baseadas no critério de relevância dos documentos para uma dada necessidade de informação. No entanto, algumas questões de ordem prática e outras de ordem conceitual inviabilizaram o uso destas métricas no contexto da presente pesquisa. As questões de ordem prática se referem às características do acervo em questão: 1) número reduzido de documentos (cerca de 200) e 2) documentos extensos que

inviabilizam a sua leitura pelo participante no decorrer do experimento. Estas questões dificultaram a elaboração de tarefas de recuperação da informação que pudessem avaliar adequadamente a recuperação da informação utilizando as métricas tradicionais.

Além disso, existem também questões de ordem conceitual, já mencionadas anteriormente, que põem em cheque a adequação do critério de relevância, no qual se baseiam as métricas tradicionais (revocação e precisão). Relevância pode ser entendida como uma medida da adequação entre um item do acervo e a “necessidade de informação” do usuário. No entanto argumenta-se que o critério de relevância se mostra bastante subjetivo (ELLIS, 1996). Uma “necessidade de informação” pode ser vista como a carência de conhecimento que impede com que a pessoa dê prosseguimento a uma tarefa qualquer e, como tal, por definição, é de difícil diagnóstico. Além disso, o retorno obtido pode mudar a sua noção sobre sua própria necessidade de informação anteriormente declarada. Uma “necessidade de informação” se revela uma entidade bastante dinâmica, cujo dinamismo e mudança se apresentam durante o próprio processo de interação com o sistema de recuperação da informação. Além disso, as métricas tradicionais são geralmente aplicadas aos mecanismos de busca. No entanto, no contexto das bibliotecas digitais, a qualidade dos mecanismos de recuperação da informação deve ser avaliada em um contexto mais amplo (FUHR *et al.*, 2007), como em uma sessão de uso da biblioteca digital, que compreende e mescla atividades de busca e navegação.

Pelo exposto e tendo como base as características específicas do acervo e das questões formuladas e apresentadas aos participantes durante as tarefas de busca e recuperação da informação, optou-se por adotar uma forma alternativa para avaliar e eficácia da recuperação da informação no contexto do experimento proposto.

Para cada resposta fornecida pelo participante durante as tarefas de busca e recuperação da informação, foi atribuída uma pontuação variando de 0 a 3 pontos.

Pontuação $(p, q) \in [0...3]$, onde *p*: participante, *t*: questão.

Caso a resposta para a questão apresentada esteja correta e completa, será conferida a pontuação 3, caso contrário a pontuação será menor ou igual a 2. A pontuação 0 (zero) será atribuída caso não seja oferecida nenhuma resposta pelo participante para a questão, mais especificamente:

- *Pontuação 3*: Participante recuperou e apresentou os documentos corretos.

- *Pontuação 2:* Participante recuperou e apresentou uma relação de documentos que necessita inspeção individual para verificar a adequação.
- *Pontuação 1:* Participante apenas apresentou uma estratégia de busca, não tendo executado até o final.
- *Pontuação 0:* Participante sequer apresentou estratégia de busca.

Também foram utilizadas as medições dos tempos despendidos para execução das tarefas pelos participantes. O tempo decorrido durante as atividades foram tomados para cada questão apresentada e respondida pelo participante.

Tempo (p,q), onde p: participante, q: questão.

3.3.3.2 Questionário de Satisfação

No que se refere ao Questionário de Satisfação (Q2), os participantes puderam expressar sua opinião em cada métrica apresentada, fazendo uma marca numa escala de 7 pontos.

Satisfação (p, m) \in [1...7] onde p: participante, m: métrica de satisfação.

3.3.4 Análise e interpretação dos dados

Nesta seção, são brevemente apresentados os procedimentos a serem utilizados para análise dos dados coletados. Os testes de uso propostos no contexto do experimento devem permitir avaliar a eficácia da recuperação da informação, no que se refere aos documentos recuperados pelos participantes para as questões apresentadas. Além disso, a satisfação do usuário também deverá ser avaliada a partir do Questionário de Satisfação. Tanto no caso da pontuação conferida à resposta de cada questão e os respectivos tempos despendidos, assim como nas métricas de satisfação, deverá ser utilizada a média normalizada para cada um dos sistemas de biblioteca digital utilizado.

Nos estudos comparativos, categoria na qual se enquadra o experimento proposto, o principal objetivo é comparar certas características da amostra da população em estudo, sob diferentes condições artificialmente impostas. Nesse sentido, deverá ser adotado um nível de significância estatística adequado, a fim de verificar se os dados coletados fornecem

evidência suficiente sobre possíveis diferenças de desempenho, entre os dois sistemas avaliados, relacionadas às tarefas de recuperação da informação e satisfação do usuário.

As medidas de qualidade relacionadas à recuperação da informação são geralmente aplicadas aos mecanismos de busca. No entanto, no contexto das bibliotecas digitais, a qualidade dos mecanismos de recuperação da informação deve ser avaliada em um contexto mais amplo (FUHR *et al.*, 2007), como em uma sessão de uso da biblioteca digital. A qualidade e eficiência da recuperação, quando se considera as respostas fornecidas pelos participantes para as questões apresentadas, estão fortemente relacionadas com a qualidade das ferramentas que os mesmos utilizam para obter tais respostas; neste caso, os componentes de busca e acesso do sistema de biblioteca digital. Portanto, a qualidade dos componentes de busca e acesso fornecidos por cada sistema de biblioteca digital será indiretamente avaliada com base na qualidade das respostas fornecidas pelos participantes, assim como no tempo despendido pelos mesmos para obter tais respostas.

De modo a poder responder às perguntas definidas anteriormente (ver seção 3.3.1), e com base nos resultados da avaliação conduzida na forma do experimento proposto, será necessário comparar:

- A pontuação conferida às respostas de cada questão apresentada aos participantes durante as tarefas de busca e recuperação da informação, assim como os tempos despendidos pelos mesmos para obtenção das respostas, durante o uso de ambos os sistemas.
- Medidas relacionadas à satisfação do usuário, apuradas após o uso de cada sistema.

Os objetivos desta pesquisa foram analisados de modo a identificar duas questões ou perguntas para as quais o experimento de avaliação deverá fornecer respostas. Os procedimentos e a configuração do experimento foram projetados para fornecer tais respostas. Além disso, foram discutidas as formas como deverão ser codificados, processados e analisados os resultados do experimento.

4 RESULTADOS

Neste capítulo são descritos os resultados alcançados por nossa pesquisa. Inicialmente será apresentado o protótipo para um sistema de biblioteca digital de teses e dissertações desenvolvido na Etapa 2 da pesquisa. O protótipo foi chamado TDF-Bíblia e será apresentado com ênfase nas funcionalidades proporcionadas pelos novos mecanismos para recuperação da informação implementados. Em seguida, são apresentados os resultados da avaliação conduzida na Etapa 3 da pesquisa, na forma de um experimento que envolveu testes de uso com a participação de usuários. No que se refere à avaliação realizada, são brevemente apresentados e discutidos os dados obtidos a partir de um levantamento inicial que buscou colher informações demográficas e também relacionadas às necessidades de informação dos estudantes do PPGCI/ECI/UFMG, no que se refere ao acervo de teses e dissertações da BDTD/ECI/UFMG. Em seguida, são apresentados os resultados relacionados às tarefas de recuperação da informação executadas pelos usuários no contexto do experimento. Também são apresentados os resultados relacionados à satisfação do usuário, obtidos a partir do Questionário de Satisfação – Q2 (ver Apêndice D), aplicado após as tarefas de busca e recuperação da informação. Finalmente, os resultados da pesquisa são discutidos frente aos objetivos inicialmente traçados, bem como são apresentadas as considerações finais e propostas de trabalhos futuros.

4.1 O protótipo TDF-Bíblia

O protótipo computacional, que foi chamado TDF-Bíblia⁴¹, foi desenvolvido no formato de uma biblioteca digital de teses e dissertações. O mecanismo de recuperação da informação implementado foi chamado de Busca Dinâmica Facetada. Este utiliza um modelo baseado na classificação multidimensional (facetada) dos documentos do acervo e permite a exploração e a navegação no acervo de uma forma guiada, porém não restritiva, por meio de uma interface visual fácil e intuitiva. A FIGURA 26 apresenta a tela inicial do protótipo.

⁴¹ O protótipo pode ser acessado no endereço: <http://www.tdf-biblio.ufv.br>

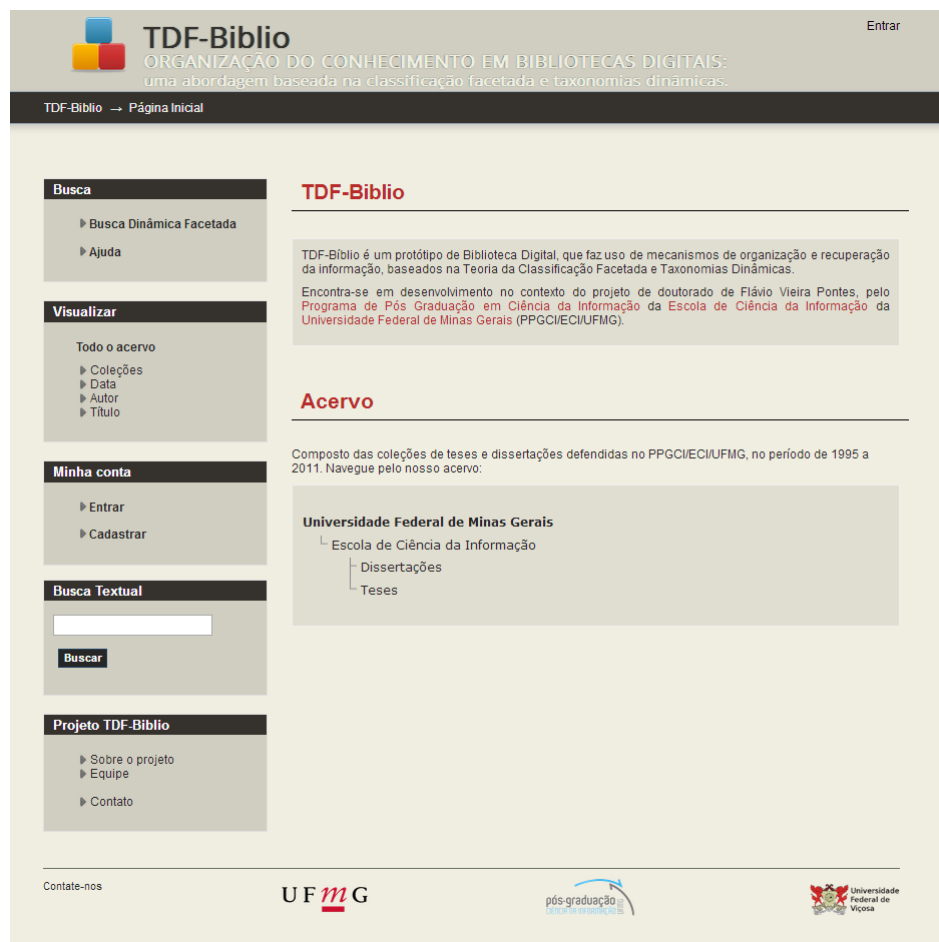



FIGURA 26 - Tela inicial do protótipo TDF-Bíbio.

Fonte: elaborado pelo autor.

A interface da Busca Dinâmica Facetada apresenta atributos descritivos, representados pelos elementos de metadados: Orientador, Tipo do Documento, Ano de Publicação e Linha de Pesquisa. Além desses atributos descritivos, são também utilizados aspectos relacionados ao conteúdo dos documentos, modelados pela taxonomia facetada e representados através de novos elementos de metadados que foram criados através da qualificação do elemento “subject”, quais sejam: Tema, Objeto de Pesquisa, Ambientação, Tipo de Pesquisa, Coleta de Dados, Métodos, Fundamento Teórico, Fundamento Histórico-Conceitual e Resultados. As categorias são apresentadas ao usuário seguidas de um número, que aparece entre parênteses, representando o total de documentos classificados na respectiva categoria. O usuário pode elaborar a sua busca, progressivamente, ao mesmo tempo em que pode observar o efeito que sua escolha por uma categoria determina quais outras categorias estarão disponíveis para ele, na sequência. O processo de recuperação da informação pode ser visto como um processo interativo de redução ou filtragem do

acervo: o usuário seleciona um foco de interesse, que restringe o acervo a partir do descarte de todos os documentos que não pertencem ao foco em questão. Apenas as categorias usadas para classificar os documentos presentes no foco de interesse, além de seus antecessores na hierarquia, são mantidas na taxonomia reduzida. Essas categorias, que resumem o foco de interesse atual, são somente aquelas que podem ser utilizadas para refinamentos posteriores. A partir da perspectiva de interação homem-máquina, o usuário é efetivamente guiado para atingir seu objetivo por meio de uma lista clara e consistente de alternativas possíveis.

Na medida em que o usuário seleciona o valor em uma dada categoria (um Orientador, por exemplo), os resultados são prontamente exibidos, e a estrutura é dinamicamente reconstruída para representar somente os documentos que tiveram a participação do respectivo orientador. Desta forma, o usuário pode dar sequência em sua busca, através da seleção de valores em outras categorias (Tipo de Documento, Ano de Publicação, Linha de Pesquisa, Tema, etc.). Tomando como exemplo a FIGURA 27, a seleção da professora “Lídia Alvarenga” produzirá como resultado 14 documentos em que o nome da professora aparece na qualidade de orientadora. Nesse momento, a estrutura será reconstruída para representar o resultado parcial, permitindo ao usuário visualizar a distribuição dos documentos nas demais categorias, como pode ser observado na FIGURA 28. Esse resultado pode ainda ser filtrado na sequência através das categorias Tipo do Documento (tese ou dissertação), Data de Publicação, Linha de Pesquisa, ou ainda algum aspecto referente ao conteúdo do documento, como Tema, Objeto de Pesquisa, Fundamento Teórico, Metodologia, entre outros.



TDF-Biblio
 ORGANIZAÇÃO DO CONHECIMENTO EM BIBLIOTECAS DIGITAIS:
 uma abordagem baseada na classificação facetada e taxonomias dinâmicas.

Entrar

TDF-Biblio → Página Inicial → Busca Dinâmica Facetada

Busca

- ▶ Busca Dinâmica Facetada
- ▶ Ajuda

Busca Dinâmica Facetada

☑ **Orientador**

- ▶ Ricardo Rodrigues Barbosa (17)
- ▶ Maria Aparecida Moura (15)
- ▶ Lidia Alvarenga (14)
- ▶ Beatriz Valadares Lendon (11)
- ▶ Alcenir Soares dos Reis (10)
- ▶ Ana Maria Rezende Cabral (9)
- ▶ Eduardo Jose Wense Dias (9)
- ▶ Maria Guiomar da Cunha Frota (9)
- ▶ Marlene de Oliveira (9)
- ▶ Marcello Peixoto Bax (8)
- ▶ (mais ...)

☑ **Tipo de Documento**

- ▶ Dissertação de Mestrado (141)
- ▶ Tese de Doutorado (54)

☑ **Data de Publicação**

- ▶ 2010 - 2011 (24)
- ▶ 2000 - 2009 (167)
- ▶ 1995 - 1999 (4)

☑ **Linha de Pesquisa**

- ▶ Gestão da Informação e do Conhecimento (44)
- ▶ Informação Gerencial e Tecnológica (23)
- ▶ Informação, Cultura e Sociedade (68)
- ▶ Mestrado em Biblioteconomia (1)
- ▶ Organização e Uso da Informação (46)
- ▶ Tratamento da Informação e Bibliometria (8)
- ▶ (1)

☑ **Tema**

Busca Dinâmica Facetada

Busca Dinâmica Facetada

Abrangência da busca: Todo o acervo ▼

Buscar

Filtros

Atualizar Filtros:
Use filtros para refinar a pesquisa.

Título ▼ **Add**


Ordenação

Resultados por página 10 ▼ Classificar Itens por Relevância ▼ em ordem descendente ▼

Aplicar

Busca Dinâmica Facetada

Exibindo itens 1-10 de 199 1 2 3 4 . . . 20 Próxima Página

 **Análise conceitual sobre as relações semânticas em ciência da informação: contribuições para o desenvolvimento de ontologias**
 Julia Aparecida Gonçalves Campos; Renato Rocha Souza (UFMG, 2009-05-27)
 Para o desenvolvimento de ontologias, verifica-se a falta de métodos sistematizados para a modelagem conceitual resultado do processo intelectual de análise conceitual que, associado a outras técnicas, envolve a identificação, ...


 **Avaliação de usabilidade em bibliotecas digitais: um estudo de caso**
 Fernanda Pereira; Gercina Angela Borem de Oliveira Lima (UFMG, 2011-04-27)
 As bibliotecas digitais têm por objetivo disponibilizar informações para seus usuários com maior rapidez e facilidade de busca e acesso. Nessas bibliotecas, a interação dos usuários com o acervo realiza-se por meio de suas ...

FIGURA 27 - Exemplo da Busca Dinâmica-Facetada.

Fonte: elaborado pelo autor.

TDF-Biblio ORGANIZAÇÃO DO CONHECIMENTO EM BIBLIOTECAS DIGITAIS:
uma abordagem baseada na classificação facetada e taxonomias dinâmicas.

Entrar

TDF-Biblio → Página Inicial → Busca Dinâmica Facetada

Busca

- ▶ Busca Dinâmica Facetada
- ▶ Ajuda

Busca Dinâmica Facetada

Orientador

- ▶ Lidia Alvarenga (14)

Tipo de Documento

- ▶ Dissertação de Mestrado (8)
- ▶ Tese de Doutorado (6)

Data de Publicação

- ▶ 2010 (4)
- ▶ 2009 (1)
- ▶ 2008 (1)
- ▶ 2007 (1)
- ▶ 2006 (2)
- ▶ 2005 (3)
- ▶ 2004 (1)
- ▶ 2003 (1)

Linha de Pesquisa

- ▶ Organização e Uso da Informação (11)
- ▶ Tratamento da Informação e Bibliometria (3)

Tema

- ▶ Catalogação (processo) (1)
- ▶ Comunicação, divulgação e produção de informação (ciência, redes e afins) (1)
- ▶ Fontes de informação (tipologia, características e uso) (1)
- ▶ Gestão de conteúdo organizacional (1)
- ▶ Indexação automática e manual (processo, critérios, método e outros) (3)
- ▶ Instrumentos de processamento (organização, recuperação e assemelhados) (3)
- ▶ Padrões, linguagens e protocolos (1)
- ▶ Preservação e restauração de informação (impresso e digital) (1)
- ▶ Processamento, recuperação e disseminação da informação (2)

Busca Dinâmica Facetada

Abbrangência da busca: **Buscar**

Filtros

Filtros selecionados:

- Orientador: Lidia Alvarenga

Atualizar Filtros

Atualizar Filtros:

Use filtros para refinar a pesquisa.

Título **Add**

Ordenação

Resultados por página: Classificar Itens por: em ordem:

Aplicar

Exibindo itens 1-10 de 14 **RESULTADOS** [1](#) [2](#) Próxima Página

Educação Física no PPGMH/UFRGS: uma visão a partir da análise de citações e perfil dos pesquisadores

Ivone Job; Lidia Alvarenga (UFMG, 2006-06-21)

Este trabalho analisa o perfil dos professores e 1.171 citações de teses do Programa de Pós-graduação em Ciência do Movimento Humano da Escola de Educação Física da Universidade Federal do Rio Grande do Sul defendidas entre ...

Comparação entre termos de indexação e palavras dos títulos dos artigos do periódico "Cadernos de Saúde Pública (2000/2005)"

Elizabeth Andrade Duarte; Lidia Alvarenga (UFMG, 2007-05-30)

A presente pesquisa tem como problema verificar se os termos de indexação atribuídos aos artigos publicados em periódico científico na área da saúde pública estariam contidos nos títulos destes mesmos artigos. Sendo assim, ...

FIGURA 28 - Resultado parcial da Busca Dinâmica Facetada.

Fonte: elaborado pelo autor.

Outra funcionalidade, que também foi implementada no protótipo, diz respeito à integração da Busca Dinâmica Facetada com a busca textual, conforme foi apontado no Capítulo 2, seção 2.4.1. Essa integração acontece de duas formas. No primeiro caso, o usuário pode iniciar fornecendo uma expressão para o mecanismo de busca textual, sendo que os resultados retornados pela busca textual serão dinamicamente descritos pela taxonomia. Dessa maneira, o usuário poderá continuar a sua busca utilizando o mecanismo de Busca Dinâmica Facetada, partindo dos resultados parciais fornecidos pela busca textual. No segundo caso, o usuário pode iniciar utilizando o mecanismo de Busca Dinâmica Facetada, sendo que, a qualquer momento, ele poderá fazer uso da busca textual, tendo como escopo os resultados parciais obtidos até o momento.

A FIGURA 29 ilustra um exemplo da integração descrita acima. Neste caso, o usuário iniciou a sua busca fornecendo o termo “ontologia” para o mecanismo de busca textual. Conforme ilustrado pela FIGURA 29, foi obtido um conjunto de 12 documentos como resultado. Conforme pode ser observado, esse resultado parcial aparece distribuído nas diversas categorias da taxonomia, à esquerda. Desta forma, o usuário poderia dar sequência à busca a partir do mecanismo de Busca Dinâmica Facetada, selecionando opções sob as categorias Orientador, Tipo de Documento, Data de Publicação, Linha de Pesquisa, Tema, etc.

TDF-Biblio ORGANIZAÇÃO DO CONHECIMENTO EM BIBLIOTECAS DIGITAIS: uma abordagem baseada na classificação facetada e taxonomias dinâmicas. Entrar

TDF-Biblio → Página Inicial → Busca Dinâmica Facetada

Busca

- Busca Dinâmica Facetada
- Ajuda

Busca Dinâmica Facetada

- Orientador
 - Lidia Alvarenga (3)
 - Marcello Peixoto Bax (3)
 - Mauricio Barcellos Almeida (2)
 - Renato Rocha Souza (2)
 - Marlene de Oliveira (1)
 - Ricardo Rodrigues Barbosa (1)
- Tipo de Documento
 - Dissertação de Mestrado (8)
 - Tese de Doutorado (4)
- Data de Publicação
 - 2010 - 2011 (2)
 - 2000 - 2009 (10)
- Linha de Pesquisa
 - Gestão da Informação e do Conhecimento (4)
 - Informação Gerencial e Tecnológica (3)
 - Organização e Uso da Informação (4)
 - Tratamento da Informação e Bibliometria (1)
- Tema
 - Avaliação (métodos, modelos, processos e afins) (1)
 - Gestão de conteúdo organizacional (2)
 - Indexação automática e manual (processo, critérios, método e outros) (1)
 - Instrumentos de processamento (organização, recuperação e assemelhados) (4)
 - Processamento, recuperação e disseminação da informação (2)
 - Sistema de informação (criação e uso) (1)
 - Tecnologias da informação e comunicação (evolução, implicações e efeitos de uso) (1)
- Objeto

Busca Dinâmica Facetada

Busca Dinâmica Facetada
Abrangência da busca:

Filtros
Atualizar Filtros:
Use filtros para refinar a pesquisa.

Ordenação
Resultados por página: Classificar Itens por: em ordem:

Busca Dinâmica Facetada

Exibindo itens 1-10 de 12 [1 2 Próxima Página](#)

Ontologias no suporte a portais semânticos
Marcelo Alvim Jorge; Marcello Peixoto Bax (UFMG, 2005-07-15)
Portais semânticos aparecem como uma evolução natural dos tradicionais portais web e podem ser entendidos como portais de informações que utilizam os padrões de representação de informação propostos pelo W3C para a chamada ...

Conceitualização na construção de ontologias: relações semânticas no âmbito do Blood Project
Livia Marangoni Duffles Teixeira; Mauricio Barcellos Almeida (UFMG, 2009-12-11)
Nos últimos anos a pesquisa em ontologias tem recebido destaque pelas possibilidades que oferece na organização da informação para uso por pessoas e em sistemas. No desenvolvimento de ontologias, uma de suas etapas é a ...

Uma proposta metodológica para construção de ontologias: uma perspectiva interdisciplinar entre as Ciências da Informação e da Computação
Daniela Lucas da Silva; Renato Rocha Souza (UFMG, 2008-06-27)
A pesquisa apresenta princípios metodológicos para desenvolvimento de ontologias formais, construídos através da análise da literatura sobre metodologias para construção de ontologias e de normas internacionais para ...

FIGURA 29 - Integração da Busca Dinâmica Facetada com a busca textual.

Fonte: elaborado pelo autor.

Conforme pôde ser verificado pelos testes do protótipo desenvolvido e também a partir dos exemplos apresentados acima, a utilização de uma classificação facetada, assim como a sua implementação utilizando taxonomias dinâmicas para compor o mecanismo de recuperação da informação, permitem melhorar a experiência do usuário na busca e exploração do acervo da biblioteca digital. Essa melhoria pode ser observada, principalmente, em três cenários de uso:

1. **Seleção pragmática de documentos:** neste caso, o usuário deseja acessar um conjunto de documentos com base em um conjunto de critérios pré-estabelecidos. Por exemplo, deseja-se encontrar todos os documentos de tipo “Tese de Doutorado” publicados a partir do ano de “2000”, que tenham tido a participação de um determinado “Orientador”, em uma determinada “Linha de Pesquisa”.
2. **Exploração do acervo:** neste caso, o usuário deseja explorar o acervo de forma livre, buscando aumentar seu nível de conhecimento acerca dos documentos e seus relacionamentos. Por exemplo, no caso de uma biblioteca de teses e dissertações, um estudante, no início de seu trabalho de pesquisa, deseja saber quais as “Metodologias” utilizadas em pesquisas sobre um determinado “Tema” ou, ainda, quais os principais “Fundamentos Teóricos” adotados nessas pesquisas.
3. **Análise do acervo:** neste caso, o usuário deseja descobrir e analisar “leis internas” ao acervo, de modo a perceber e compreender questões relacionadas ao mesmo. Por exemplo, no caso de uma biblioteca de teses e dissertações, o coordenador do Programa de Pós Graduação pode estar interessado em analisar a produção de uma determinada “Linha de Pesquisa”, quais os “Orientadores” mais produtivos, quais os “Temas” mais pesquisados, como a produção encontra-se distribuída de acordo com os documentos publicados nos últimos anos, etc.

Esses cenários demonstram a importância da capacidade exploratória oferecida pela abordagem adotada nesta pesquisa, que não pode ser encontrada nos mecanismos de recuperação da informação tradicionalmente oferecidos pelas bibliotecas digitais.

4.2 Avaliação

Nesta seção são apresentados os resultados da avaliação conduzida na forma de um experimento (Etapa 3 da pesquisa). O objetivo da avaliação foi determinar a influência

ou impacto, do ponto de vista dos usuários, da organização do acervo que utiliza uma estrutura classificatória facetada, bem como dos novos mecanismos de recuperação da informação, baseados no conceito de taxonomias dinâmicas e implementados no protótipo de biblioteca digital de teses e dissertações TDF-Bíblia, apresentado acima.

Inicialmente, são brevemente apresentados e discutidos os dados obtidos a partir de um levantamento inicial que buscou colher informações demográficas, hábitos de acesso à informação, além de informações relacionadas às necessidades informacionais dos estudantes do PPGCI/ECI/UFMG, no que se refere ao acervo de teses e dissertações da BDTD/ECI/UFMG. Devido ao fato de se tratar de informação acessória, utilizada para o projeto do experimento e não diretamente relacionada aos objetivos desta pesquisa, estes resultados serão apresentados de forma resumida. No entanto, os dados completos, incluindo gráficos, podem ser consultados no Apêndice B. Lembrando ainda que se trata de informações colhidas junto aos respondentes do Questionário de Levantamento Inicial – Q1, enviado a todos os estudantes matriculados no PPGCI/UFMG, tendo sido obtidos 47 registros de resposta, perfazendo uma taxa de retorno de aproximadamente 40%.

4.2.1 Levantamento Inicial

Com relação às informações demográficas, 57% dos respondentes eram do sexo feminino e 43% do sexo masculino. Com relação à faixa etária, a maior parte dos respondentes se enquadra na faixa de 31 a 35 anos (34%), seguido de 46 anos ou mais (21%). 57% dos respondentes eram doutorandos e 43% mestrandos. Com relação à área de formação (graduação), a maior parte dos respondentes (57%) era graduada em Biblioteconomia, seguido de Informática / Ciência da Computação (15%) e Administração (11%).

Com relação aos hábitos de acesso a informação, 34% dos respondentes disseram visitar as instalações físicas da biblioteca da ECI/UFMG, em média, pelos menos uma vez por semana; 28% disseram visitar a referida biblioteca somente uma vez por mês; os demais respondentes informaram frequências menores. No que se refere à biblioteca digital de teses e dissertações, 30% informaram visitar a biblioteca pelo menos uma vez por mês, sendo que 47% informou visitar a referida biblioteca com menos frequência. Já com relação ao tempo diário de acesso a Internet, 68% dos respondentes disseram utilizar em média mais de 4 horas diárias. Quando perguntados sobre as principais fontes de

informação utilizadas para estudos/pesquisa, as respostas se distribuíram de forma relativamente uniforme: biblioteca de ECI/UFMG (17%), Portal de Periódicos da Capes (18%), outros periódicos eletrônicos (20%), biblioteca digital de teses e dissertações (20%), Google/Google Acadêmico (23%), outros (2%).

Quando questionados sobre suas necessidades de informação, relacionada ao acervo de teses e dissertações, grande parte dos respondentes (cerca de 50%) formulou questões envolvendo temas específicos de pesquisa, relacionados a seus respectivos projetos. No entanto, um número expressivo de respondentes (cerca de 36%) formulou perguntas de caráter exploratório, envolvendo metodologias utilizadas, correlações entre temas de pesquisa e orientadores, linhas de pesquisa, cobertura do acervo, etc. Estas questões de caráter exploratório se beneficiam em maior grau da abordagem utilizada na presente pesquisa. Foi apresentada na seção 3.3.2.2 uma categorização das perguntas formuladas pelos respondentes, na medida em que estas informações foram utilizadas na elaboração das questões posteriormente apresentadas aos participantes no contexto do experimento.

4.2.2 Tarefas de busca e recuperação da informação

O planejamento da pesquisa, conforme descrito no Capítulo 3 (Metodologia), foi realizado de modo a gerar dados pareados, ou seja, cada participante do experimento está associado a pares de medidas: uma associada ao uso do sistema atual da BDTD e outra associada ao uso do protótipo TDF-Bíblia. Desta forma, procuramos garantir que o único fator que afeta sistematicamente os dados (relacionados à eficácia da recuperação e satisfação dos usuários) é o fator em estudo (sistema de biblioteca digital utilizado). Desta forma procurou-se ter maior controle sobre a variabilidade aleatória.

Conforme apresentado na seção 3.3.3, para cada resposta fornecida pelo participante durante as tarefas de busca e recuperação da informação, foi atribuída uma pontuação variando de 0 a 3 pontos. Apresentamos na FIGURA 30 os gráficos com as frequências das pontuações obtidas para cada questão apresentada aos participantes durante o uso de cada sistema de biblioteca digital, bem como a frequência geral das pontuações, considerando o conjunto total de questões. O eixo horizontal apresenta as pontuações atribuídas e o eixo vertical apresenta o número de participantes que receberam a pontuação na respectiva questão.

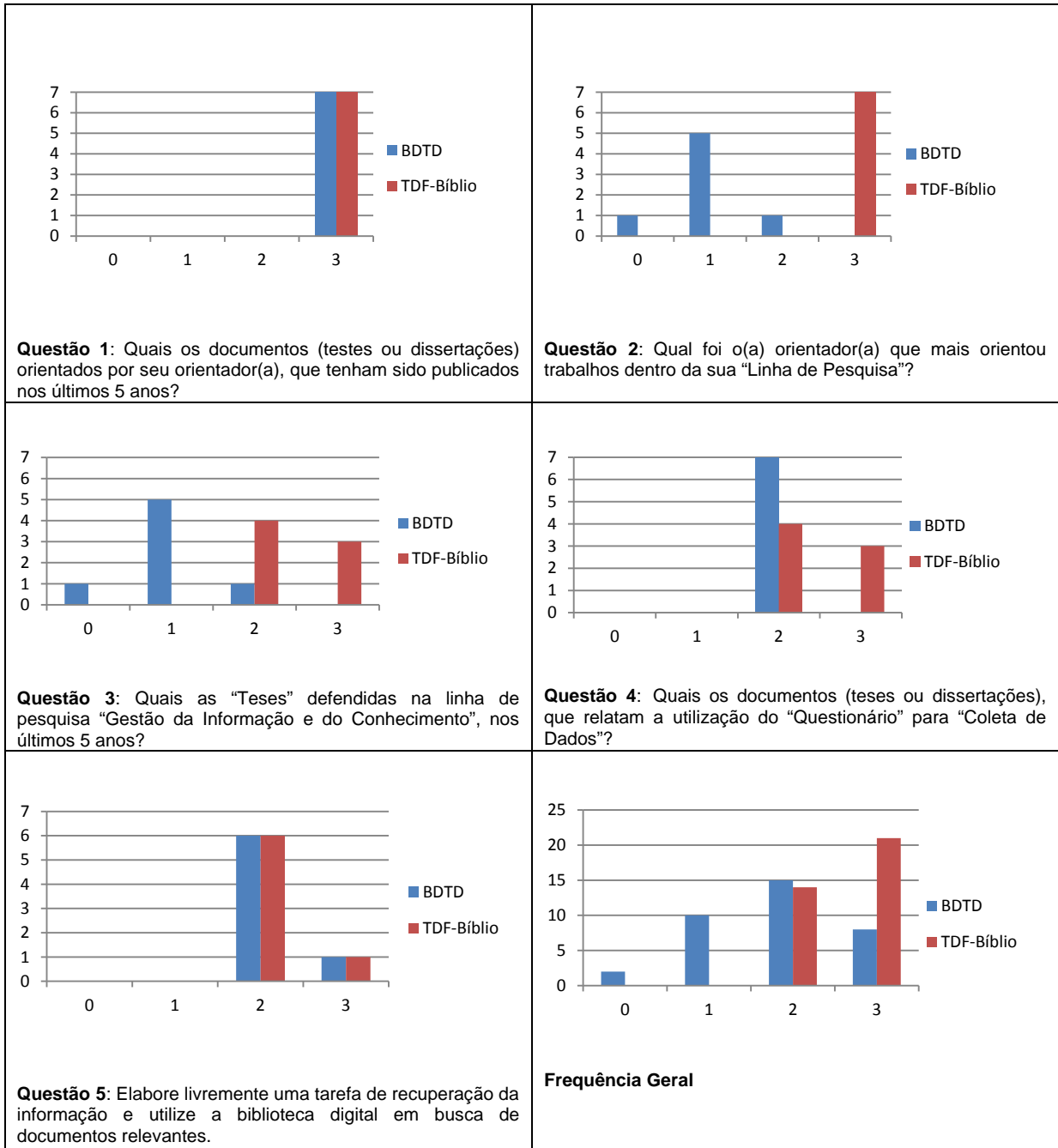


FIGURA 30 - Gráficos com as frequências das pontuações obtidas nas tarefas de busca e recuperação da informação.

No caso da Questão 1, não houve qualquer variação do desempenho, sendo que todos os participantes obtiveram a pontuação máxima na questão, ao utilizarem ambos os sistemas. No caso da BDTD, os participantes utilizaram o índice de orientadores para chegar aos resultados. Já no caso do uso do protótipo TDF-Bíblia, os participantes utilizaram a categoria "Orientador", tendo chegado aos mesmos resultados.

Já na Questão 2, todos os sete participantes obtiveram pontuação máxima no caso do uso do protótipo TDF-Bíblia. Eles utilizaram as categorias “Orientador” e “Linha de Pesquisa”, em conjunto. No caso da BDTD, a maioria dos participantes (cinco) obteve pontuação 1, ou seja, apresentaram uma estratégia para a busca, mas não executaram até o final devido ao fato da estratégia elaborada ser muito trabalhosa de ser executada no referido sistema de biblioteca digital. Devido ao fato de o sistema da BDTD não utilizar a informação da linha de pesquisa para a categorização dos documentos, a estratégia traçada pela maioria dos participantes foi a de consultar os trabalhos orientados por cada orientador que conhecidamente atuam na respectiva linha de pesquisa, pressupondo-se que cada orientador tenha sempre trabalhado na mesma linha de pesquisa.

No caso da Questão 3, três dos participantes obtiveram nota máxima ao usarem o protótipo TDF-Bíblia, sendo que quatro obtiveram nota 2, ou seja, apresentaram como resposta uma relação de documentos que necessita inspeção individual para confirmar a relevância. A maioria dos participantes utilizou corretamente as categorias “Tipo do Documento” e “Linha de Pesquisa”, tendo obtido a relação dos documentos distribuída nos anos de publicação. No entanto, alguns participantes apresentaram dificuldade em ordenar os resultados pela data de publicação de modo a obter o resultado final em ordem decrescente. Já no caso da BDTD, a maioria dos participantes (cinco) obteve nota 1, ou seja, apresentaram somente a estratégia de busca, não tendo executado a mesma até o final. A dificuldade dos participantes deve ao fato de o sistema da BDTD não permitir relacionar tipo de documento com as linhas de pesquisa. A estratégia traçada pela maioria foi realizar a busca a partir dos orientadores que conhecidamente atuam na linha de pesquisa e, a partir daí, consultar cada documento individualmente.

No caso da Questão 4, três participantes obtiveram nota máxima ao usarem o protótipo TDF-Bíblia, sendo que quatro obtiveram nota 2, ou seja, apresentaram como resposta uma relação de documentos que necessita inspeção individual para confirmar a relevância. A maioria dos participantes (quatro) apresentou dificuldade em localizar o termo “Questionário” dentro da categoria “Coleta de Dados” da taxonomia. A dificuldade foi relacionada ao fato do referido termo se encontrar sob o conceito mais amplo “Coleta de documentos (impressos e digitais)” na hierarquia taxonômica. Alguns participantes tiveram dúvidas entre utilizar as categorias “Coleta de Dados” ou “Métodos”. Já no caso da BDTD, a totalidade dos participantes obteve nota 2, ou seja, apresentaram uma relação de documentos que necessita inspeção individual para confirmar a relevância. Estes participantes utilizaram a busca textual a partir da expressão “Questionário”.

Finalmente, no caso da Questão 5, o desempenho dos participantes foi exatamente o mesmo, quando se compara o uso de ambos os sistemas. A maioria dos participantes (seis) apresentou uma relação de documentos que necessita inspeção individual para confirmar a relevância. Tanto no caso do uso da BDTD como do protótipo TDF-Bíblia, a maioria dos participantes optou pelo uso da busca textual.

A observação e análise exploratória dos gráficos de frequência apresentados na FIGURA 30 e brevemente descritos acima permite concluir que, no caso das questões 2, 3 e 4, o desempenho médio dos participantes foi superior quando utilizaram o protótipo TDF-Bíblia. A análise estatística dos dados relacionados às tarefas de busca e recuperação da informação foi realizada utilizando o teste T para dados pareados. O teste T é apropriado para comparar dois conjuntos de dados quantitativos, em termos de seus valores médios, sendo bastante utilizado em pesquisa social (BARBETTA, 2011). Para tanto, a pergunta a ser respondida a partir da análise dos dados produzidos pelo experimento (ver seção 3.3.1) pode ser formatada em termos da hipótese nula (H_0) e da correspondente hipótese alternativa (H_1), conforme segue:

Pergunta 1: A organização do acervo com base numa estrutura facetada e os novos mecanismos de busca e acesso implementados no protótipo TDF-Bíblia facilitam a recuperação da informação e melhoraram seus resultados, considerando o ponto de vista dos usuários da BDTD/ECI/UFMG, em tarefas de busca e recuperação da informação, visando atender as suas necessidades de informação?

- **H_0 :** os resultados médios relacionados à eficácia da recuperação da informação não se alteram quando se compara ambos os sistemas.
- **H_1 :** o protótipo TDF-Bíblia melhora os resultados médios relacionados à eficácia da recuperação da informação.

A fim de determinar a probabilidade de significância (p) dos resultados obtidos, adotamos o nível de significância de 5% ($\alpha = 0,05$), que é um valor adequado no caso de pesquisas sociais (BARBETTA, 2011). Lembrando que, ao supor H_0 verdadeira, a probabilidade de significância, ou valor p obtido, é a probabilidade de a estatística do teste acusar um resultado tão ou mais distante do esperado por H_0 , como resultado da amostra observada. Desta forma, para poder negar a hipótese nula (H_0) e aceitar a hipótese alternativa (H_1) é necessário obter uma probabilidade de significância $p < 0,05$. Em outras palavras, o valor p pode ser interpretado como o risco de se tomar a decisão errada após a

observação da amostra, caso se rejeite H_0 . Nesta pesquisa, fixamos esse risco máximo em 5%.

A TABELA 1 abaixo apresenta a compilação dos resultados obtidos nas tarefas de busca e recuperação da informação. São apresentadas as pontuações médias obtidas em cada uma das 5 questões (Q1 a Q5) apresentadas aos participantes durante os testes, bem como a média geral obtida para cada sistema de biblioteca digital utilizado. Além das médias, são apresentados os desvios padrão D para cada questão. Também são apresentadas as probabilidades de significância (p) e as variações Δ obtidas em termos percentuais para cada questão a partir da razão entre as diferenças das pontuações médias obtidas para cada sistema e a pontuação obtida pela BDTD. A variação Δ denota o ganho percentual obtido pelo protótipo TDF-Bíblia em relação à BDTD.

TABELA 1 - Desempenho nas tarefas de busca e recuperação da informação.

Sistema	Q1	Q2	Q3	Q4	Q5	Média	D _{Q1}	D _{Q2}	D _{Q3}	D _{Q4}	D _{Q5}
BDTD	3,00	1,00	1,00	2,00	2,14	1,83	0,00	0,58	0,58	0,00	0,38
TDF-Bíblia	3,00	3,00	2,43	2,43	2,14	2,60	0,00	0,00	0,53	0,53	0,38
p	1	0,000048	0,000200	0,039070	1	0,000003					
Δ	0,00%	200,00%	142,86%	21,43%	0,00%	42,19%					

A análise estatística dos dados utilizando o teste T, apresentada na TABELA 1 acima, mostra que o desempenho dos participantes foi melhorado (no caso do uso do protótipo TDF-Bíblia) quando se analisa as questões 2, 3 e 4. Já para as questões 1 e 5 não houve variação. Na média, o desempenho dos participantes foi melhorado em 42,19%, com o valor de p indicando uma probabilidade de erro de 0,0003% apenas, menor que o nível de significância adotado de 5%. Desta forma, podemos concluir que, no contexto do experimento realizado e considerando a amostra analisada, o protótipo TDF-Bíblia melhorou os resultados médios relacionados à eficácia da recuperação da informação.

Os dados relacionados aos tempos despendidos pelos participantes nas tarefas de busca e recuperação da informação, apesar de terem sido coletados e tabulados, não foram analisados. Esta decisão se justifica pelo fato de, em grande parte dos casos, os participantes não terem executado a busca até o final, tendo se limitado a relatar a estratégia elaborada. Na maioria dos casos isso aconteceu quando a estratégia elaborada pelo participante demandaria uma maior quantidade de tempo para ser executada,

inviabilizando a sua execução no tempo estipulado para os testes. Desta forma, a análise baseada nos tempos ficou prejudicada, tendo sido descartada.

4.2.3 Satisfação do usuário

Após a execução das tarefas de busca e recuperação da informação, utilizando cada sistema de biblioteca digital, os usuários foram solicitados a responder a um Questionário de Satisfação – Q2 (ver Apêndice D). Foram utilizados dois conjuntos de métricas de satisfação. O primeiro busca medir a satisfação relacionada à execução das tarefas no sistema de biblioteca digital. Apesar de essas métricas serem relacionadas com as tarefas executadas, as respostas dos usuários, colhidas logo após a execução das tarefas, refletem a satisfação do usuário no que se refere ao uso do sistema de biblioteca digital durante a tarefa. O segundo conjunto de métricas é mais específico, buscando avaliar a satisfação do participante com os mecanismos de busca e acesso à informação oferecidos pelo sistema de biblioteca digital. Os participantes foram orientados a expressar o seu grau de satisfação, em cada métrica, numa escala gradativa de 1 a 7 pontos. O Questionário de Satisfação (Q2) pode ser consultado no Apêndice D. A FIGURA 31 abaixo apresenta as frequências das respostas obtidas em cada uma das métricas do primeiro conjunto (relacionada à satisfação com as tarefas executadas), assim como a compilação destas frequências. O eixo horizontal apresenta os valores possíveis para o grau de satisfação e o eixo vertical apresenta o número de participantes que informaram cada valor para a respectiva métrica.

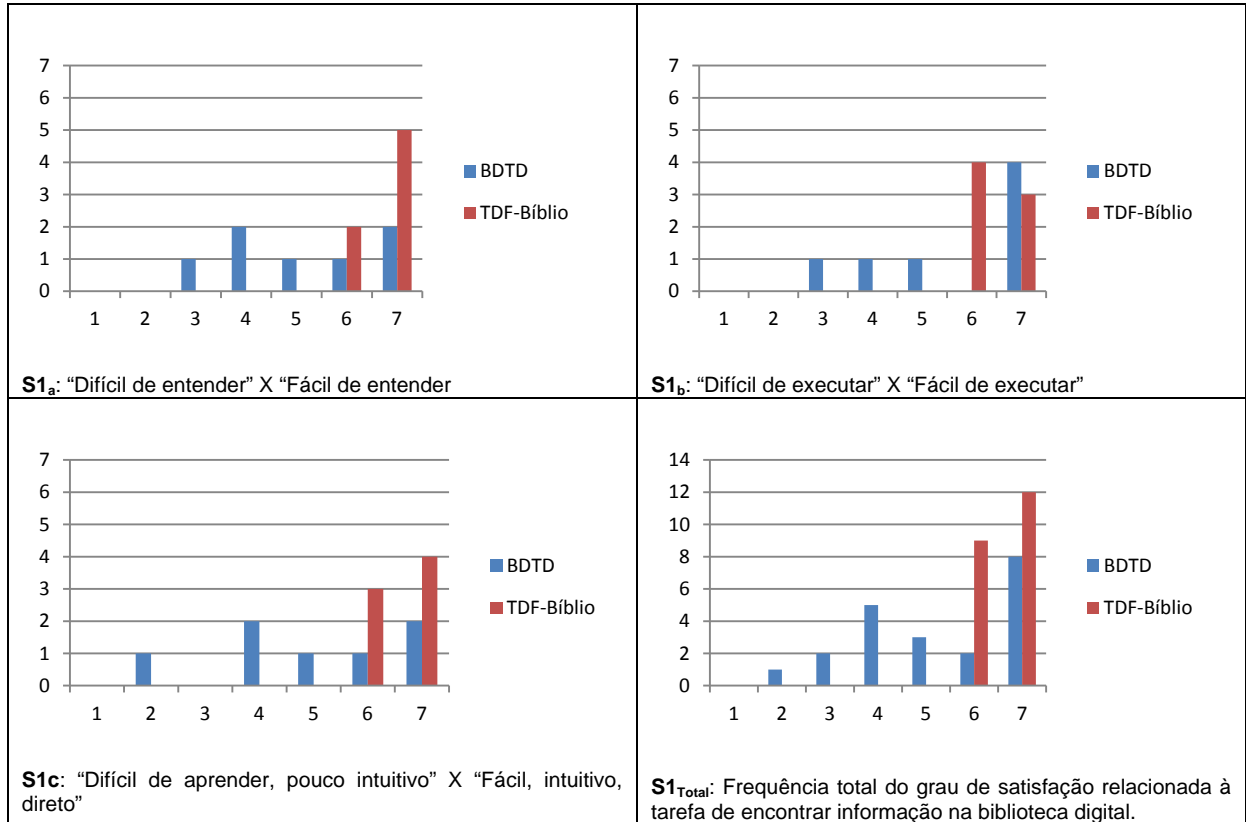


FIGURA 31 - Gráficos com as frequências do grau de satisfação com a tarefa de encontrar informação na biblioteca digital.

A FIGURA 32 abaixo apresenta as frequências do grau de satisfação relacionada a cada métrica do segundo conjunto (relacionada com os mecanismos de busca e acesso à informação oferecidos pelo sistema de biblioteca digital), assim como a compilação destas frequências.

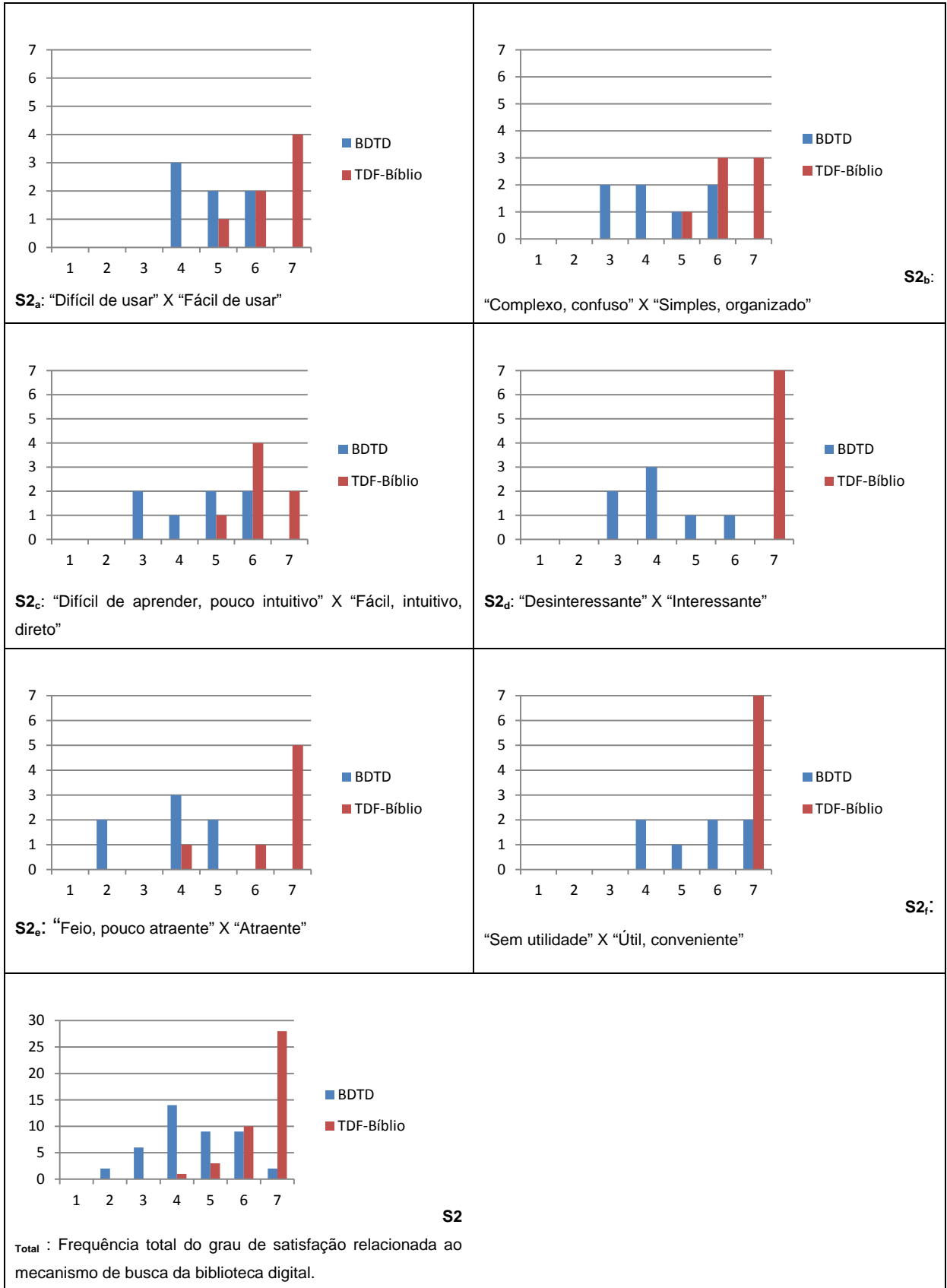


FIGURA 32 - Gráficos com as freqüências do grau de satisfação com o mecanismo de busca da biblioteca digital.

A observação e análise exploratória dos gráficos de frequência apresentados na FIGURA 31 e 32 mostra que a satisfação dos participantes foi maior, no caso do uso do protótipo TDF-Bíblia, para todas as métricas utilizadas. Cabe ainda destacar as métricas S_{2d} (“Desinteressante” X “Interessante”) e S_{2f} (“Sem utilidade” X “Útil, conveniente”), ambas relacionadas à satisfação com o mecanismo de busca e acesso oferecido pelo sistema de biblioteca digital. No caso dessas duas métricas, a totalidade dos participantes expressou grau máximo de satisfação nas mesmas, no caso do protótipo, indicando a utilidade e conveniência do Mecanismo de Busca Dinâmica Facetada oferecida pelo protótipo TDF-Bíblia.

Assim como foi feito para os dados coletados nas tarefas de busca e recuperação da informação, a análise dos dados relacionados à satisfação dos usuários foi realizada utilizando o teste T para dados pareados, com nível de significância $\alpha = 0,05$. Para tanto, a pergunta a ser respondida a partir da análise dos dados produzidos pelo experimento (ver seção 3.3.1) pode ser formatada em termos da hipótese nula (H_0) e da correspondente hipótese alternativa (H_1), conforme segue:

Pergunta 2: A organização do acervo com base numa estrutura facetada e os mecanismos de busca e acesso implementados no protótipo TDF-Bíblia melhoram a satisfação dos usuários da BDTD/ECI/UFMG?

- H_0 : a satisfação média do usuário é a mesma para ambos os sistemas.
- H_1 : o protótipo TDF-Bíblia melhora a satisfação média dos usuários.

A TABELA 2, abaixo, apresenta a compilação dos resultados relacionados à satisfação dos usuários. São apresentados os graus médios de satisfação (numa escala de 1 a 7 pontos) para cada uma das métricas utilizadas, bem como o grau médio de satisfação geral obtido para cada sistema de biblioteca digital utilizado. Além das médias, são apresentados os desvios padrão D para cada métrica. Também são apresentadas as probabilidades de significância (p) e as variações Δ obtidas em termos percentuais para cada métrica a partir da razão entre as diferenças das médias obtidas para cada sistema e a média obtida pela BDTD. A variação Δ denota o ganho percentual obtido pelo protótipo TDF-Bíblia em relação à BDTD.

TABELA 2 - Satisfação dos usuários.

Métrica	S1.a	S1.b	S1.c	S2.a	S2.b	S2.c	S2.d	S2.e	S2.f	Média
BDTD	5,14	5,71	5,00	4,86	4,43	4,57	4,14	3,71	5,57	4,79
TDF-Bíblia	6,71	6,43	6,57	6,43	6,29	6,14	7,00	6,43	7,00	6,56
p	0,0085	0,1416	0,0259	0,0085	0,0076	0,0166	0,0002	0,0036	0,0125	0,0052
Δ	30,56%	12,50%	31,43%	32,35%	41,94%	34,38%	68,97%	73,08%	25,64%	36,75%
Desvios	D1 _a	D1 _b	D1 _c	D2 _a	D2 _b	D2 _c	D2 _d	D2 _e	D2 _f	
BDTD	1,57	1,70	1,83	0,90	1,27	1,27	1,07	1,25	1,27	
TDF-Bíblia	0,49	0,53	0,53	0,79	0,76	0,69	0,00	1,13	0,00	

A análise estatística dos dados utilizando o teste T, apresentada na TABELA 2, acima, mostra que o grau médio de satisfação dos participantes foi melhorado (no caso do uso do protótipo TDF-Bíblia) para todas as métricas utilizadas. Com exceção da métrica S1_b, que apresentou probabilidade de significância $p = 0.1416$, maior que o nível de significância estabelecido $\alpha = 0.05$, em todas as demais métricas a melhoria na satisfação pôde ser comprovada pelo teste estatístico. Na média, o grau de satisfação dos participantes foi melhorado em 36,75%, com o valor de p indicando uma probabilidade de erro de 0,52% apenas, menor que o nível de significância adotado de 5%. Desta forma, podemos concluir que, no contexto do experimento realizado e considerando a amostra analisada, o protótipo TDF-Bíblia melhorou o grau médio de satisfação dos participantes.

Na parte final do Questionário de Satisfação (Q2), os participantes foram solicitados a responder a três questões discursivas:

- **S3:** Eu penso que teria uma melhor experiência caso... (complete a sentença).
- **S4:** Você gostaria de continuar a utilizar o sistema de biblioteca digital avaliado? (justifique).
- **S5:** Comentários gerais sobre a avaliação.

O QUADRO 16, abaixo, apresenta as respostas obtidas.

QUADRO 16 - Respostas obtidas para as questões abertas do Questionário de Satisfação (Q2)

	BDTD	TDF-Bíbio
Participante 1	<p>S3: “A interface fosse melhor constituída e se houvesse mais recursos de pesquisa na base”</p> <p>S4: “Sim, utilizar a biblioteca digital é uma necessidade para o desenvolvimento dos meus trabalhos de pesquisa”</p> <p>S5: (sem resposta)</p>	<p>S3: “Tivesse lido um tutorial ou usado a biblioteca digital antes, pois assim eu conheceria mais recursos sobre a mesma e minha busca seria mais tranquila e fácil”</p> <p>S4: “Sim, a biblioteca digital é interessante, permite que eu filtre meus resultados através da data, linhas de pesquisa, orientador, etc. Isso faz com que a busca fique mais simples e segura, pois estarei explorando toda ou quase toda a coleção que a biblioteca digital contempla”</p> <p>S5: “O sistema é interessante e permite fazer a busca muito personalizada e especificada. Achei a interface simples e fácil de navegar e a busca é muito intuitiva”</p>
Participante 2	<p>S3: “Fosse mais claro o que cada opção poderia retornar”</p> <p>S4: “Sim, utilizo por ser necessário, mas não o considero com boa usabilidade”</p> <p>S5: (sem resposta)</p>	<p>S3: “A parte que utiliza a taxonomia ficasse mais clara. Ex: fosse informada de alguma forma que quando eu selecionasse Coleta de Dados, os arquivos seriam classificados novamente por tipo de instrumento de coleta”</p> <p>S4: “Sim, a proposta é interessante e se mostrou mais fácil de usar que a BDTD”</p> <p>S5: “Exceto pelo exemplo já citado (da Coleta de Dados, que não é muito intuitiva), o sistema foi tranquilo de se usar”</p>
Participante 3	<p>S3: “Tivesse mais opções de campos de busca. Em alguns casos é necessário acessar o PDF do documento para saber determinadas informações”</p> <p>S4: “Sim, é a opção que temos agora. Melhor do que nada. Mas precisa de melhorias”</p> <p>S5: “Acho que faltam campos importantes para a busca. Muito centrado nos campos tradicionais autor, título assunto. Não relaciona os campos, o que facilitaria um melhor refinamento das buscas”</p>	<p>S3: “Haja melhoria da taxonomia”</p> <p>S4: “Sim, esta biblioteca digital traz algumas melhorias em relação às bibliotecas digitais que estão em funcionamento, tais como: maiores possibilidades de combinar as relações entre os campos de pesquisa e mais possibilidades de campos de busca”</p> <p>S5: “Muito boa a iniciativa. Se as bibliotecas digitais caminharem para esta possibilidade de maiores relações entre os campos vai nos dar maior relevância na recuperação da informação”</p>
Participante 4	<p>S3: “Tivesse outras categorias para pesquisa, e não restringisse aos orientadores, título e data de envio”</p> <p>S4: “Não, porque não consegue atender as minhas necessidades de informação”</p> <p>S5: “A página não abre muitas possibilidades para a busca e recuperação da informação. Portanto, a BDTD precisa de uma reavaliação”</p>	<p>S3: “Colocassem uma nova hierarquia dentro das opções das facetas. Exemplo: Bibliotecas e serviços - Biblioteca escolar; Biblioteca pública; Biblioteca Universitária”</p> <p>S4: “Sim, achei bastante pertinente as categorizações das facetas. E auxiliou na busca e recuperação da informação”</p> <p>S5: “O protótipo é bastante interessante, útil e fácil de utilizar, consegui fazer algumas pesquisas com facilidade, e a recuperação da informação teve um índice maior de retorno. Porém, acredito que deve haver uma readequação nas opções dentro das facetas, pois alguns temas relevantes não foram contemplados”</p>
Participante 5	<p>S3: “A biblioteca digital permitisse que eu montasse a minha busca, escolhendo diferentes pontos de acesso e permitisse que eu filtrasse os resultados”</p> <p>S4: “Sim, porque pelo menos posso pesquisar os trabalhos pela coleção da ECI”</p> <p>S5: (sem resposta)</p>	<p>S3: “Buscasse informações precisas”</p> <p>S4: “Sim, possibilidades de busca e combinações legais”</p> <p>S5: (sem resposta)</p>
Participante 6	<p>S3: “A biblioteca digital oferecesse mais possibilidades e categorias para a busca”</p> <p>S4: “Sim, pois é de grande relevância para a minha pesquisa os documentos encontrados no sistema”</p> <p>S5: “O sistema de biblioteca digital poderia ser melhorado com a inclusão de outras categorias para a busca como linha de pesquisa, instrumento para coleta de dados, etc”</p>	<p>S3: “Já tivesse experiência com o sistema de biblioteca digital”</p> <p>S4: “Sim, através da busca dinâmica facetada é mais fácil encontrar os documentos pertinentes”</p> <p>S5: “O sistema demonstrou eficiência para recuperar os documentos e as informações desejadas”</p>

Participante 7	<p>S3: “O mecanismo de busca e filtros fossem aprimorados. Houve dúvida na interpretação de alguns parâmetros”</p> <p>S4: “Sim, o sistema é interessante, porém não atendia a todas as minhas solicitações. O sistema (biblioteca digital) possui algumas funcionalidades relevantes como a busca booleana e outros”</p> <p>S5: “A biblioteca digital apesar de possuir a busca avançada e alguns filtros (data, etc.) não permite que o usuário faça buscas específicas e não permite que o mesmo explore toda a coleção”</p>	<p>S3: “Tive dúvidas nos conteúdos relacionados a 'coleta de dados' e 'métodos”</p> <p>S4: “Sim, bastante útil”</p> <p>S5: “Bem melhor do que o sistema de armazenamento de teses e dissertações usado na UFMG”</p>
-----------------------	---	--

A análise das respostas apresentadas no QUADRO 16 indica a percepção dos participantes sobre as melhorias proporcionadas pela maior capacidade exploratória fornecida pela Busca Dinâmica Facetada. No entanto, algumas participantes indicaram algumas dificuldades bem como aspectos que podem ser melhorados. Essas questões também puderam ser detectadas a partir do relato dos participantes durante os testes e da análise das gravações.

As principais dificuldades apresentadas se referem ao entendimento e ao uso da taxonomia facetada. Esta dificuldade se refere mais especificamente aos aspectos relacionados ao conteúdo dos documentos, modelados pela taxonomia. Os participantes apresentaram dificuldades no entendimento da hierarquia taxonômica, nas diversas categorias/facetadas. Os conceitos utilizados para descrever o primeiro nível da taxonomia se mostraram demasiadamente genéricos e pouco intuitivos em indicar os conceitos subjacentes. Por exemplo, na categoria “Coleta de Dados”, os participantes apresentaram dificuldades em localizar o instrumento “Questionário” devido ao fato de o mesmo estar situado na hierarquia sob o rótulo genérico “Coleta de documentos (impressos e digitais)”. Esta mesma dificuldade se mostrou recorrente em outras categorias da taxonomia. Alguns participantes tiveram dúvidas relacionadas à diferença entre as categorias “Métodos” e “Coleta de Dados”.

Conforme mencionado anteriormente, a taxonomia utilizada como instrumento de organização do conhecimento no contexto desta pesquisa foi desenvolvida inicialmente por Maculan (2011), apenas para a linha de pesquisa Organização e Uso da Informação. A referida autora, que colaborou imensamente com esta pesquisa, revisou e ampliou a taxonomia construída, de modo a abarcar as demais linhas de pesquisa do PPGCI/ECI/UFMG, tendo ainda realizado a classificação do acervo com base na mesma. No entanto, este instrumento deve ser considerado apenas para fins de prototipação, visto que não passou por nenhum tipo de avaliação ou validação formal. Os relatos dos participantes apresentados no QUADRO 16, acima, indicam alguns pontos para a melhoria da taxonomia facetada.

5 CONSIDERAÇÕES FINAIS

Neste capítulo apresentamos algumas considerações sobre a pesquisa e os resultados alcançados frente aos objetivos inicialmente propostos. Ao final são apresentadas e discutidas algumas possibilidades de trabalhos futuros.

Nossa pesquisa partiu da observação de que os processos de busca e recuperação de informação, na grande maioria das bibliotecas digitais, assim como ocorre na Internet, baseiam-se, principalmente, nas buscas textuais e na navegação. Neste sentido, podemos dizer que as bibliotecas digitais podem ser vistas, na maioria dos casos, como meros sistemas executores de consultas que devem ser formuladas e objetivadas por seus usuários. Isto acontece nas bibliotecas digitais apesar dos processos de tratamento da informação que ocorrem nas mesmas. Acreditamos que as bibliotecas digitais deveriam possibilitar maior “capacidade exploratória” em seus acervos. O usuário necessita de um sistema exploratório eficiente, além de um sistema executor de consultas. Esta capacidade de exploração pode ser encontrada nas bibliotecas tradicionais no papel do bibliotecário de referência, dialogando com o usuário em torno de uma consulta para melhor defini-la, mas é raramente encontrada nos esquemas de recuperação da informação presentes nas bibliotecas digitais e na Internet. Outro fator importante e que justificou nossa pesquisa tem origem na observação, compartilhada por alguns autores, de que uma série de conhecimentos acumulados sobre como organizar bibliotecas e como tratar informações não está sendo devidamente aproveitada pela instituição correspondente no contexto digital, a biblioteca digital.

Diante do exposto, esta pesquisa teve como proposta a aplicação da teoria da classificação facetada, para a organização do acervo, bem como o uso de taxonomias dinâmicas para o desenvolvimento de novos mecanismos para exploração e recuperação da informação em uma biblioteca digital de teses e dissertações. Com isso, nossa expectativa era poder responder as seguintes questões: como melhorar e facilitar a interação do usuário, em tarefas de exploração e recuperação da informação em bibliotecas digitais de teses e dissertações? O uso de uma estrutura classificatória facetada, para organização do acervo, e taxonomias dinâmicas para compor um novo mecanismo de exploração e a recuperação da informação, pode melhorar a experiência e a satisfação dos usuários da biblioteca?

Foi desenvolvido um protótipo para um sistema de biblioteca digital de teses e dissertações. O análise do mecanismo de busca dinâmica facetada, implementado no protótipo TDF-Bíblia, apresentado na seção 4.1, indicou melhorias na experiência do usuário relacionada à busca e à exploração do acervo da biblioteca digital. Essas melhorias

puderam ser observadas, principalmente, em três cenários de uso: 1) seleção pragmática de documentos; 2) exploração do acervo e 3) análise do acervo.

A busca dinâmica facetada fornece suporte a um novo paradigma de acesso, permitindo a exploração guiada do acervo e fazendo uma ponte entre os processos de busca e a navegação. Além disso, o uso desta abordagem parece ser intuitivo e facilmente utilizado pelos usuários. As taxonomias dinâmicas projetadas de forma facetada requerem um esquema conceitual simples e são capazes de se adaptar a relacionamentos novos e não previstos, permitindo aos usuários descobrir esses relacionamentos nas operações de navegação e recuperação. O controle do usuário e o mecanismo adaptativo claro e consistente, proporcionado pelas taxonomias dinâmicas, encorajam o usuário a experimentar e explorar, proporcionando-lhe a sensação de ter considerado todas as alternativas para chegar ao resultado. Acreditamos que a maior facilidade de interação do usuário é resultado do uso de uma única representação para consulta (definição do foco de interesse) e sumarização (apresentação da taxonomia reduzida). Desta forma, o usuário lida com uma representação conceitual única da base de dados.

Outra questão que merece destaque é a possibilidade, também implementada no protótipo desenvolvido, de integração da busca dinâmica facetada com outros mecanismos de recuperação da informação, como a busca textual. Desta forma, a busca dinâmica facetada pode ser facilmente integrada às bibliotecas digitais, assim como em outros contextos de recuperação da informação em ambientes digitais.

O mecanismo de busca dinâmica facetada deve ser visto como uma alternativa, e não um substituto dos mecanismos de busca textual. Tanto é que o protótipo TDF-Bíblia também disponibiliza, além da busca dinâmica facetada, o mecanismo de busca textual convencional, permitindo inclusive a integração entre ambos. A busca dinâmica facetada é especialmente útil em possibilitar uma maior capacidade exploratória do acervo que, em geral, não pode ser encontrada nas bibliotecas digitais atualmente disponíveis.

A metodologia adotada se mostrou adequada ao alcance dos objetivos traçados. Os novos mecanismos de recuperação da informação foram implementados utilizando-se a plataforma DSpace, que é a mesma utilizada pela BDTD. Este fato permitiu a elaboração de uma avaliação de base comparativa, executada na forma de um experimento e envolvendo testes de uso dos sistemas. No capítulo anterior, foram apresentados os resultados da avaliação comparativa do protótipo TDF-Bíblia, que contou com a participação de usuários. A análise dos dados coletados, relacionados à eficácia da recuperação da informação e

satisfação dos usuários, cumpriu o objetivo de responder às duas perguntas inicialmente formuladas na definição operacional adotada na metodologia (ver seção 3.3.1).

Com relação à recuperação da informação, o protótipo TDF-Bíblia facilitou e melhorou os seus resultados quando considerado o conjunto de questões formuladas e apresentadas aos participantes no contexto do experimento realizado. Na média, o desempenho dos participantes foi melhorado em 42,19%, com probabilidade de erro dentro do nível de significância adotado de 5%. Desta forma, podemos concluir que, no contexto do experimento realizado e considerando a amostra analisada, o protótipo TDF-Bíblia melhorou os resultados médios relacionados à eficácia da recuperação da informação.

Com relação à satisfação do usuário, o grau de satisfação dos participantes foi maior no caso do uso dos mecanismos de busca e acesso implementados no protótipo TDF-Bíblia. Na média, o grau de satisfação dos participantes foi melhorado em 36,75%, com probabilidade de erro dentro do nível de significância adotado de 5%. Desta forma, podemos concluir que, no contexto do experimento realizado e considerando a amostra analisada, o protótipo TDF-Bíblia melhorou o grau médio de satisfação dos participantes.

A avaliação conduzida também permitiu identificar necessidades de melhorias na taxonomia facetada, além de evidenciar a importância deste instrumento para a organização do conhecimento na biblioteca digital.

Chamou a atenção o fato de a maioria dos participantes ter utilizado a busca textual na questão de busca livre (Questão 5), mesmo no caso de uso do protótipo, que oferecia o mecanismo de busca dinâmica facetada. Quanto questionados sobre este fato, os participantes informaram que estavam acostumados com a forma de busca do Google e tinham, instintivamente, a tendência em utilizar a busca textual. Acreditamos que este fato observado possa ser explicado pela nuance do comportamento humano em tentar resolver os problemas com as ferramentas com as quais estamos acostumados. Em outras palavras, “quando a ferramenta que estamos acostumados a utilizar é um martelo, tudo à nossa frente vira um prego”. Este ditado reflete uma das leis fundamentais do comportamento, que é a lei do reforço: se o martelo já foi usado inúmeras vezes para fixar pregos, e prestou-se para isto, é quase fatal que se tente utilizá-lo de novo, quando se quiser fixar alguma coisa, não importando sequer analisar os detalhes ou as alternativas. Acreditamos que na medida em que as bibliotecas digitais incorporarem formas alternativas para busca e exploração do acervo, como a busca dinâmica facetada, estas ferramentas serão também incorporadas no

modelo mental dos usuários, não como substituto da busca textual, mas como uma alternativa a ser utilizada quando for conveniente.

Cabe ainda destacar que, apesar dos esforços no sentido de conseguir um número maior de voluntários para participar do experimento conduzido, a avaliação contou com uma amostra pequena de usuários. Apesar disso, os resultados se mostraram estatisticamente válidos para a amostra utilizada.

Os resultados obtidos devem ser avaliados no contexto restrito das questões que foram apresentadas aos participantes nas tarefas de busca e recuperação da informação, tendo em vista o número reduzido de documentos que compõem o acervo utilizado. Outros conjuntos de questões poderiam gerar resultados diferentes, principalmente aqueles relacionados à recuperação da informação. Diante do exposto, os resultados obtidos indicam melhorias significativas proporcionadas pela abordagem proposta por esta pesquisa, mas que devem ser confirmadas por outras avaliações envolvendo um maior número de participantes e, principalmente, com a possibilidade de utilização de acervos maiores, que possibilitem a utilização das métricas tradicionais da área de recuperação da informação.

É importante destacar ainda a importância em se resgatar conhecimentos desenvolvidos no campo da biblioteconomia e ciência da informação, como a Teoria da Classificação Facetada. É neste ponto que acreditamos estar a principal contribuição da nossa pesquisa. Esses conhecimentos, utilizados em conjunto com técnicas desenvolvidas no contexto da computação, e potencializados com as tecnologias da Internet, podem trazer importantes avanços nas pesquisas. Tais avanços são certamente mais fáceis de serem alcançados em contextos nos quais exista o adequado tratamento da informação, como no caso das bibliotecas digitais, e podem favorecer o surgimento de bibliotecas digitais mais eficientes no que diz respeito ao processo de comunicação do qual são mediadoras.

Um grande esforço vem sendo aplicado pelo IBICT na construção de bibliotecas de teses e dissertações nas universidades brasileiras. Os resultados obtidos têm sido bastante satisfatórios no que se refere à construção dos acervos em cada instituição, assim como na interoperabilidade entre estes acervos de modo a possibilitar buscas integradas. As teses e dissertações representam o produto final das atividades de pesquisa em nível de pós-graduação. Estes documentos precisam ser efetivamente acessados e utilizados como insumo em novas pesquisas. Acreditamos que a abordagem proposta neste trabalho de pesquisa pode trazer grande contribuição no sentido de melhorar e facilitar a recuperação, e conseqüentemente contribuir para o uso mais efetivo destes acervos no âmbito dos programas de pós-graduação no Brasil.

Com relação a trabalhos futuros, destacamos a necessidade de aplicar e avaliar o mecanismo de busca dinâmica facetada em outras bibliotecas digitais, preferencialmente envolvendo acervos maiores. Outra possibilidade para trabalhos futuros se refere ao desenvolvimento de taxonomias facetadas a serem aplicadas em outras áreas de conhecimento, para a classificação de acervos digitais, bem como o estudo de princípios genéricos para o desenvolvimento dessas taxonomias com o objetivo de facilitar a interação do usuário no contexto da abordagem proposta pela presente pesquisa.

Finalmente, a abordagem proposta na presente pesquisa utilizou uma estrutura semântica construída na forma de uma taxonomia facetada que, desta forma, possibilitou a busca e exploração em todo o acervo. Não se trabalhou a possibilidade de navegação interna nos documentos. Outra possibilidade para trabalhos futuros se refere à integração com a abordagem proposta por Lima (2004), que consiste na representação semântica do conteúdo de cada documento individual, visando possibilitar também a navegação em contexto dentro do documento.

REFERÊNCIAS

ADKISSON, H. P. **Web design practices**: use of faceted classification. 2005. Disponível em: <<http://www.webdesignpractices.com/navigation/facets.html>>. Acesso em 02 set. 2010.

ALVARENGA, L. Representação do conhecimento na perspectiva da Ciência da Informação em tempo e espaço digitais. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n.15, 2003.

ANDERSON, J. D. Organization of knowledge. IN: **International Encyclopedia of Information and Library Science**. 2nd. ed. Ed. by John Feather & Paul Sturges. London: Routledge (pp. 471-490). 2003.

ATKINS, A.; FOX, E.; FRANCE, R.; SULEMAN, H. (Eds). **ETD-MS**: an Interoperability Metadata Standard for Electronic Theses and Dissertations. Version 1.00, revision 2, 25 jun. 2008. Disponível em: <<http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html>>. Acesso em 29 out. 2010.

BAINBRIDGE, D.; BUCHANAN, G.; MCPHERSON, J.; JONES, S. et al. Greenstone: A Platform for Distributed Digital Library Applications. In ECDL'01: Proceedings of the 5th **European Conference on Research and Advanced Technology for Digital Libraries**, pp. 137–148. Springer-Verlag, London, UK, 2001.

BAINBRIDGE, D.; DON, K. J.; BUCHANAN, G.; WITTEN, I. H. et al. Dynamic Digital Library Construction and Configuration. In HEERY, R.; LYON, L. (eds.) ECDL, vol. 3232 of **Lecture Notes in Computer Science**, pp. 1–13. Springer, 2004.

BARBETTA, Pedro Alberto. **Estatística Aplicada às Ciências Sociais**. 11. ed. Florianópolis: Editora da UFSC, 2011.

BLANDFORD, A., Buchanan, G., Jones, M.: Usability of digital libraries. **Int. J. Digit. Libr.** 4, 69–70 (2004).

BOHMERWALD, Paula. Uma proposta metodológica para avaliação de bibliotecas digitais: usabilidade e comportamento de busca por informação na Biblioteca Digital da PUC/Minas. **Ciência da Informação**, Brasília, v. 34, n. 1, p. 95-103, 2005.

BORBINHA, J. L.; KUNZE, J.; SPINAZZÈ, A.; MUTSCHKE, P. et al. Reference models for digital libraries: actors and roles. **Int. J. on Digital Libraries**, 5(4):325–330, 2005.

BRÄSCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...** São Paulo: Ancib, 2008.

BRESLIN, J. G.; HARTH, A.; BOJARS, U.; DECKER, S. Towards Semantically-Interlinked Online Communities. vol. 3532, pp. 500–514. 2005. Proceedings of the 2nd **European Semantic Web Conference** (ESWC '05), Heraklion, Greece.

BRODY, Tim; HARNAD, Stevan. **The research impact cycle**. Disponível em: <<http://opcit.eprints.org/feb190a/harnad-cycle.ppt>>. Acesso em: 17 set. 2009.

BROWN, Mary E. **History and definition of digital libraries**. New Haven, C.T. : Southern Connecticut State University, 2005. Disponível em: <http://www.southernct.edu/~brownm/dl_history.html>. Acesso em: 30 mar. 2009.

BROUGHTON, Vanda. Facet analytical theory as a basis for a knowledge organization tool in a subject portal. In: López-Huertas, María J. and Muñoz-Fernández, Francisco J. (Eds). Würzburg Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries: proceedings of the the **Seventh International ISKO Conference**, 10-13 July 2002, Granada, Spain). Ergon Verlag, 2002. (Advances in Knowledge Organization; Vol 8). pp. 135-142.

BROUGHTON, Vanda. The need for a faceted classification as the basis of all methods of information retrieval. **Aslib Proceedings**: New Information Perspectives Vol. 58 No. 1/2, pp. 49-72. 2006.

BROUGHTON, V. A faceted classification as the basis of a faceted terminology. **Axiomathes** 18(2) 193-210. 2008.

BUCHANAN, G.; HINZE, A. A generic alerting service for digital libraries. In JCDL '05: Proceedings of the 5th **ACM/IEEE-CS joint conference on Digital libraries**, pp. 131–140. ACM, New York, NY, USA, 2005.

BUCHANAN, G.; BAINBRIDGE, D.; DON, K. J.; WITTEN, I. H. A new framework for building digital library collections. In **JCDL '05**: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, pp. 23–31. ACM, New York, NY, USA, 2005.

BUSH, V. As we may think. **Atlantic Monthly**, v.176, n.1, p.101-108, July 1945. Disponível em < <http://theatlantic.com/doc/194507/bush> >. Acesso em: 14 abr. 2010.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Taxonomia e classificação: a categorização como princípio. In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. Anais... Salvador: Ancib, 2007.

CANDELA, Leonardo et al. Setting the foundation of digital libraries. **D-Lib Magazine**, v.13, n. 3/4, Mar./Apr. 2007a. Disponível em: <<http://www.dlib.org/dlib/march07/castelli03castelli.html>>. Acesso em: 30 mai. 2010.

CANDELA, L.; CASTELLI, D.; FERRO, N.; IOANNIDIS, Y.; et al. **The DELOS Digital Library Reference Model** - Foundations for Digital Libraries. Tech. rep., DELOS, 2007b.

CHAN, Lois Mai, Eric Childress, Rebecca Dean, Edward T. O'Neill, and Diane Vizine-Goetz. A Faceted Approach to Subject Data in the Dublin Core Metadata Record. **Journal of Internet Cataloging** 4 no. 1/2 (2001): 35-47. 2001.

CHIN, J. P.; DIEHL, V. A.; NORMAN, K. L. Development of an instrument measuring user satisfaction of the human-computer interface. In **CHI '88**: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 213–218. ACM, New York, NY, USA, 1988.

CLEVELAND, Gary. **Digital libraries**: definitions, issues e challenges. UDT Occasional Paper, n.8, March 1998. Disponível em: < <http://www.ifla.org/VI/5/op/udtpo8/udtop8.htm>> Acesso em: 21/09/2010.

CUNHA, Murilo Bastos. Das bibliotecas convencionais às digitais: diferenças e convergências. **Perspectivas em Ciência da Informação**, v.13, n.1, p.2-17, jan./abr. 2008.

CUNHA, Murilo Bastos. Bibliografia sobre o fluxo do documento na biblioteca digital. DataGramZero - Revista de Ciência da Informação. Rio de Janeiro, v. 10, n. 5, out. 2009. Disponível em: <http://www.dgz.org.br/out09/Art_01.htm>. Acesso em: 26 abr. 2011.

CUNHA, Murilo Bastos; MCCARTHY, Cavan. Estado atual das bibliotecas digitais no Brasil. In: MARCONDES, Carlos Henrique; KURAMOTO, Hélio; TOUTAIN, Lidia Brandão; SAYÃO, Luis Fernando (Org.). **Bibliotecas digitais**: saberes e práticas. Salvador/Brasília: UFBA/IBICT, 2 ed. 2006, p. 25-54.

CROFTS, N.; DOERR, M.; GILL, T.; STEAD, S; et al. Current Official Version of the CIDOC CRM version 4.2 of the reference document. **Definition of the CIDOC Conceptual Reference Model**. 2005.

CUNNINGHAM, H.; MAYNARD, D.; BONTCHEVA, K.; TABLAN, V. GATE: A framework and graphical development environment for robust NLP tools and applications. In **Proceedings of the 40th Annual Meeting of the ACL**. 2002. Disponível em: <<http://citeseer.ist.psu.edu/context/2035358/0>>. Acesso em: 30 jun. 2010.

D'ARCUS, B. Citation Oriented Bibliographic Vocabulary. **Tech. rep.**, 2006. Disponível em: <<http://vocab.org/biblio/schema>>. Acesso em 21 jun. 2010.

D'ARCUS B.; GIASSON F. Bibliographic Ontology Specification. **Tech. rep.**, Zitgist, 2007. Disponível em: <<http://bibliontology.com/>>. Acesso em: 21 jun. 2010.

DANIEL, R.; LAGOZE C.; PAYETTE, S. D. **A metadata architecture for digital libraries**. pp. 276–288. 1998. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=670428>. Acesso em 29 jun. 2010.

DAVIS, I.; NEWMAN, R. **Expression of Core FRBR Concepts in RDF**. Working draft, 2005. Disponível em <<http://vocab.org/frbr/core>>. Acesso em 21 jun. 2010.

DENTON, William. **How to Make a Faceted Classification and Put It On the Web**. Nov. 2003. Disponível em: <<http://www.miskatonic.org/library/facetweb-howto.html>>. Acesso em 30 ago. 2010.

DIAS, Eduardo Wense. Contexto digital e tratamento da informação. **DataGramZero** – Revista de Ciência da Informação. Rio de Janeiro, v.2, n.5, out/2001. Disponível em: <http://www.dgz.org.br/out01/Art_01.htm>. Acesso em 19 out. 2009.

DODDS, L. **An Introduction to FOAF**. 2004. Disponível em: <<http://www.xml.com/pub/a/2004/02/04/foaf.html>> Acesso em: 21 jun. 2010.

DSPACE FOUNDATION. 2009. **DSpace 1.5.2 Manual**. Disponível em: <<http://www.dspace.org>>. Acesso em 4 abr. 2010.

DCMI - DUBLIN CORE METADATA INITIATIVE. **Dublin Core Metadata Element Set, Version 1.1: Reference Description**. 2003. Disponível em: <<http://dublincore.org/documents/dces/>> Acesso em 21 jun. 2010.

ELLIS, David. **Progress and problems in information retrieval**. London: Library Association, 1996. p.1-22.

ELLIS, David; VASCONCELOS, Ana. The Relevance of Facet Analysis for World Wide Web Subject Organization and Searching. **Journal of Internet Cataloging** 2 no. 3/4: 97-114. 2000.

FEDORA DEVELOPMENT TEAM. Fedora Open Source Repository Software: White Paper. **Tech. rep.**, Cornell University, 2005.

FOX, E.A.; EATON, J. L.; McMILLAN, G. et al. Networked Digital Library of Theses and Dissertations: An International Effort Unlocking University Resources. **D-Lib Magazine**, September, 1997.

FREW, J.; FREESTON, M.; FREITAS, N.; HILL, L. L.; et al. The Alexandria Digital Library Architecture. In **Proceedings of ECDL'1998**. 1998.

FREYNE, J.; FARZAN, R.; BRUSILOVSKY, P.; SMYTH, B. et al. Collecting community wisdom: integrating social search & social navigation. In IUI '07: **Proceedings of the 12th international conference on Intelligent user interfaces**, pp. 52–61. ACM, New York, NY, USA, 2007.

FUHR, N.; HANSEN, P.; MABE, M.; MICSIK, A. e SOLVBERG, I. (2001). Digital libraries: A generic classification and evaluation scheme. In **Proceedings of the 5th European Conference on Digital Libraries**, pp. 187-199, Darmstadt, Germany. Springer. 2001.

FUHR, N.; TSAKONAS, G.; AALBERG, T.; AGOSTI, M.; et al. Evaluation of Digital Libraries. **International Journal on Digital Libraries (IJDL)**, 2007. Disponível em : <<http://www.springerlink.com/content/w2w0j4h272k26812/>>. Acesso em 6 ago. 2010.

GARCEZ, Maria Stuart; RADOS, Gregório J. Varvakis. Biblioteca híbrida: um novo enfoque no suporte à educação à distância. **Ciência da Informação**, Brasília, v.31, n.2, p.44-51, maio/ago. 2002.

GARFIELD, Eugene. **A Tribute to S.R. Ranganathan: Part 1. Life and Works**. Disponível em: <<http://www.garfield.library.upenn.edu/essays/v7p037y1984.pdf>>. Acesso em 28 out. 2010.

GIL, A. C. **Como elaborar projetos de pesquisa**. 3. ed. São Paulo: Atlas, 1991. 159p.

GODERT, W. Facet classification in online retrieval, **International Classification**, Vol. 18 No. 2, pp. 98-109. 1991.

GONÇALVES, M. A.; FOX, E. A.; WATSON, L. T.; KIPP, N. A. Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. **ACM Transactions on Information Systems**, 22(2):270–312, 2004.

GONÇALVES, M.A.; MOREIRA, B.L.; FOX, E.A.; WATSON, L.T. What is a good digital library? – A quality model for digital libraries. **Information Processing & Management**, 43(5), 1416–1437, 2007.

GREENSTEIN, Daniel; THORIN, Suzanne Elizabeth. **The Digital Library: A Biography**. Washington, DC: Digital Library Federation. 2002. Disponível em: <<http://www.clir.org/PUBS/reports/pub109/pub109.pdf>>. Acesso em 19 out. 2009.

GROZA, T.; HANDSCHUH, S.; MUELLER, K.; DECKER, S. SALT – Semantically Annotated LTEX for Scientific Publications. In **The 4th European SemanticWeb Conference**. Innsbruck, Austria, 2007. Disponível em: <<http://www.eswc2007.org/>>. Acesso em 01 jul. 2010.

GRUBER, T.R. Towards Principles for the Design of Ontologies used for Knowledge Sharing. **International Journal of Human-Computer Studies**, v. 43, p. 907- 928, 1995.

GUARINO; GIARETTA. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: N. Mars (Editor). **Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing**. Amsterdam : IOS Press, 1995. p. 25-32.

HARTER, Stephen. Scholarly communication and the digital library: problem and issues. **Journal of Digital Information**, v.1, n.1, 1997. Disponível em: <<http://jodi.tamu.edu/Articles/v01/i01/Harter/>>. Acesso em: 29 maio 2008.

HILBERT, D.M.; REDMILES, D.F. Extracting usability information from user interface events. **ACM Comput. Surv.** 32, 384–421. 2000.

HJORLAND, B.; ALBRECHTSEN, H. Toward a New Horizon in Information Science: Domain-Analysis. **Journal of the American Society for Information Science**, 46(6), 400-425. 1995.

HJORLAND, B.; ALBRECHTSEN, H. An analysis of some trends in classification research. **Knowledge Organization**, 26(3), 131–139. 1999.

HJORLAND, Birger. Library and information science: practice, theory, and philosophical basis. **Information Processing and Management**. v.36, p.500-531, 2000.

HJORLAND, B. The method of constructing classification schemes: a discussion of the state-of-the-art. In Lopez-Huertas, M.J. and Munoz-Fernandez, F.J. (Eds), **Challenges in Knowledge Representation and Organization for the 21st Century. Integration of Knowledge across Boundaries, Proceedings of the the 7th International ISKO Conference**, Granada, Spain, 10-13 July 2002, **Advances in Knowledge Organization**, Vol. 8, Ergon Verlag, Würzburg, p. 451.

HJORLAND, B. **Approaches to knowledge organization**. Aula dada na University of Rome La Sapienza, 20 Abr. 2007. Disponível em: <<http://www.w3.uniroma1.it/ssab/new/ApproachesToKO.zip>>. Acesso em 28 out. 2010.

HODJE, G. **Systems of knowledge organization for digital libraries**: beyond traditional authority files. Washington, DC, The Council on Library and Information Resources. 2000. Disponível em: <<http://www.clir.org/pubs/reports/pub91/1knowledge.html>>. Acesso em: 21 maio 2010.

HORNBAEK, K.; LAW, E. L.-C. Meta-analysis of correlations among usability measures. In CHI '07: **Proceedings of the SIGCHI conference on Human factors in computing systems**, pp. 617–626. ACM, New York, NY, USA, 2007.

HUYNH, D.; MAZZOCCHI, S.; KARGER, D. Piggy Bank: Experience the Semantic Web inside your web browser. **Web Semant.**, 5(1):16–27, 2007.

IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. Functional Requirements for Bibliographic Records: Final Report, vol. 19 of **UBCIM Publications-New Series**. K.G.Saur, Muenchen, 1998. Disponível em: <<http://www.ifla.org/VII/s13/frbr/frbr.pdf>>. Acesso em: 17 jun. 2010.

INGWERSEN, P.; WORMELLI, I. Ranganathan in the perspective of advanced information retrieval, **Libri**, Vol. 42, pp. 184-201. 1992.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION – ISO. **ISO 9241-11**: Ergonomic requirements for office work with visual display terminals (VDTs). Part 11: Guidance on usability. 1998. Disponível em: <<http://www.eps.ufsc.br/ergon/disciplinas/EPS5225/normas-.htm>>. Acesso em: 22 ago. 2009.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION - ISO. **ISO/IEC 42010:2007**, Systems and Software Engineering - Recommended practice for architectural description of software-intensive systems. 2007.

KERLINGER, Fred Nichols. **Metodologia da pesquisa em ciências sociais**: um tratamento conceitual. São Paulo: EPU, 1980.

KOCH, T. The role of classification schemes in Internet resource description and discovery; work package 3 of **Telematics for Research project DESIRE**. 1998. Disponível em <<http://www.ukoln.ac.uk/metadata/desire/classification/>>. Acesso em 28 out. 2010.

KOVÁCS, L., Micsik, A. The evaluation computer: a model for structuring evaluation activities. In: Agosti, M., Fuhr, N. (eds.) **DELOS Workshop on the Evaluation of Digital Libraries**, Padova, Italy. 2004. Disponível em: <http://dlib.ionio.gr/wp7/workshop2004_program.html>. Acesso em: 11 out. 2010.

KRUK, S. R.; DECKER, S.; ZIEBORAK, L. JeromeDL - Adding Semantic Web Technologies to Digital Libraries. In ANDERSEN, K. V.; DEBENHAM, J. K.; WAGNER, R. (eds.) DEXA, vol. 3588 of **Lecture Notes in Computer Science**, pp. 716–725. Springer, 2005.

KRUK, S. R.; KRUK, E.; STANKIEWICZ, K. Evaluation of Semantic and Social Technologies for Digital Libraries. In CHRISTENSENDALSGAARD, B.; CASTELLI, D.; JURIK, B. A.; LIPPINCOTT, J. (eds.) ECDL, vol. 5173 of **Lecture Notes in Computer Science**, pp. 74–77. Springer, 2008.

KRUK, Sebastian Ryszart; MCDANIEL, Bill (Eds.) **Semantic digital libraries**; Springer: 2009.

KUNY, Terry; CLEVELAND, Gary. The digital library: myths and challenges. **IFLA Journal**, v. 24, n. 2, 1998. Disponível em: <<http://www.ifla.org/IV/ifla62/62-kuny.pdf>>. Acesso em: 30 mar. 2008.

KURAMOTO, Hélio. Ferramentas de *software* livre para bibliotecas digitais. In: MARCONDES, Carlos Henrique; KURAMOTO, Hélio; TOUTAIN, Lidia Brandão; SAYÃO, Luis Fernando (Org.). **Bibliotecas digitais: saberes e práticas**. Salvador/Brasília: UFBA/IBICT, 2 ed. 2006, p. 145-162.

KURAMOTO, Hélio. **Iniciativas do IBICT para implementações tecnológicas para gestão e acesso à informação**. Palestra proferida no XXII Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação (CBBDD), 2007.

KWASNIK, Barbara H. The role of classification in knowledge representation and discovery. **Library Trends** 48 (1): 22-47. 1999.

LA BARRE, K. **The use of faceted analytico-synthetic theory as revealed in the practice of website construction and design**. Ph.D thesis submitted to the School of Library and Information Science, Indiana University. 2006.

LAGOZE, C. The Warwick Framework. A Container Architecture for Diverse Sets of Metadata. **D-LibMagazine**, 2(7/8), 1996. Disponível em <<http://www.dlib.org/dlib/july96/lagoze07lagoze.html>>. Acesso em 29 jun. 2010.

LAGOZE, C.; HUNTER, J. The ABC Ontology and Model. **Journal of Digital Information**, 2(2), 2001.

LAGOZE, C. Fedora Tutorial 1 – Introduction to Fedora. **Tech. rep.**, Fedora Commons, 2005. Disponível em: <<http://www.fedora.info/download/2.0/userdocs/tutorials/tutorial1.pdf>>. Acesso em 29 jun. 2010.

LAGOZE, C.; VAN de SOMPEL, H. The Santa Fe Convention of the Open Archives Initiative. **Dlib Magazine**, vol. 6. n. 2, february 2000. Disponível em: <<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>>. Acesso em 08 Mar. 2009.

LAGOZE, C.; PAYETTE, S; SHIN, E.; WILPER, C. Fedora: an architecture for complex objects and their relationships. **International Journal on Digital Libraries**, V6(2):124–138, 2006. Disponível em <<http://portal.acm.org/citation.cfm?id=1124652>>. Acesso em 21 Jun. 2010.

LAGOZE, C.; DE SOMPEL, H. V. 2001. The Open Archives Initiative. In **Proceedings of the 1st Joint Conference on Digital Libraries (JCDL'2001)** (Roanoke, Va.). 54–62.

LAKATOS, E. M.; MARCONI, M. de A. **Fundamentos da metodologia científica**. 3. ed. São Paulo: Atlas, 1991.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2 ed. Brasília, DF: Briquet de Lemos, 2004.

LAVILLE, Christian & DIONNE, Jean. **A construção do saber** - Manual de metodologia da pesquisa em ciências humanas. Belo Horizonte: Editora UFMG, 1999.

LEITE, Fernando César Lima. **Como gerenciar e ampliar a visibilidade da informação científica brasileira: Repositórios Institucionais de Acesso Aberto**. Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT. 2009.

LESK, Michael. **Understanding Digital Libraries**. San Francisco, CA: Morgan Kauffmann - Elsevier, 2 ed. 2005.

LEVY, D. M.; MARSHALL C. C. Going digital: a look at assumptions underlying digital libraries. **Commun. ACM**, 38(4):77–84, 1995.

LI, Bin. **The history of digital library**. Disponível em: <<http://www.ils.unc.edu/~lib/digital-library.html>>. Acesso em: 30 mar. 2009.

LIMA, Gercina Ângela Borém. **Mapa hipertextual (MHTX): um modelo para organização hipertextual de documentos**. 2004. 204 f. Tese (Doutorado em Ciência da Informação) – ECI/UFMG, Belo Horizonte, 2004.

LIMA, Izabel França; SOUZA, Renato Rocha; DIAS, Guilherme Ataíde. Abordagens para Avaliar Bibliotecas Digitais. In: ENCONTRO NACIONAL DE PESQUISA EM CIENCIA DA INFORMAÇÃO, 12., 2011, Brasília – Distrito Federal. Anais... Brasília: ENANCIB, 2011.

LOUIE, A. J.; MADDOX, E. L.; and WASHINGTON, W. Using Faceted Classification to Provide Structure for Information Architecture. Paper presented at the **ASIS&T 2003 Information Architecture Summit**, Portland, OR, 21-23 March 2003. Disponível em <http://depts.washington.edu/pettt/presentations/conf_2003/IASummit.pdf>. Acesso em 20 set. 2010.

LOURENÇO, Cíntia de Azevedo. **Modelagem de dados como ferramenta de análise de padrões de metadados em bibliotecas digitais: o padrão de metadados brasileiro para teses e dissertações segundo o modelo entidade-relacionamento**. 2005. 161 f. Tese (Doutorado em Ciência da Informação) – ECI/UFMG, Belo Horizonte, 2005.

MACULAN, Benildes Coura Moreira dos Santos. **Taxonomia facetada navegacional: construção a partir de uma matriz categorial para trabalhos acadêmicos**. Dissertação (Mestrado em Ciência da Informação) – ECI/UFMG, Belo Horizonte, 2011.

MARCHIONINI, G. Evaluating digital libraries: a longitudinal and multifaceted view. **Library Trends** 49, (2000).

MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. Brazilian Digital Library of Theses and Dissertations. **Intl. Inform.&Libr. Rev.** 35, 265-279. 2003.

MARCONDES, Carlos Henrique. Metadados: descrição e recuperação de informações na Web. In: MARCONDES, Carlos Henrique; KURAMOTO, Hélio; TOUTAIN, Lidia Brandão; SAYÃO, Luis Fernando (Org.). **Bibliotecas digitais: saberes e práticas**. Salvador/Brasília: UFBA/IBICT, 2 ed. 2006, p. 95-111.

MEGHINI, C.; RISSE, T. BRICKS: A Digital Library Management System for Cultural Heritage. **ERCIM News**, (61), 2005.

METS EDITORIAL BOARD. Metadata Encoding and Transmission Standard: Primer and Reference Manual. **Tech. rep.**, Digital Library Federation, 2007. Disponível em: <<http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf>>. Acesso em 29 jun. 2010.

MEY, Eliane Serrão Alves. **Introdução a Catalogação**. Brasília: Briquet de Lemos, 1995. p.1-11.

MILSTEAD, J. L. (1998). Use of Thesauri in the Full-Text Environment. Baseado em artigo apresentado na 34^a **Clinic on Library Applications of Data Processing**. 1998. Disponível em <<http://www.bayside-indexing.com/Milstead/useof.htm>>. Acesso em 04/09/2010.

MORIGI, V. J.; PAVAN, C. Tecnologias de informação e comunicação: novas sociabilidades nas bibliotecas universitárias. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 117-125, jan./abr. 2004.

NAVATHE, Shamkant B.; ELMASRI, Ramez. **Fundamentals of Database Systems**. 4^a ed Addison-Wesley, 2003.

NISO - National Information Standards Organization. Seção 5.3.4 - Facet Analysis In: **"Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies"** - ANSI/NISO Z39.19-2005, NISO, Bethesda, MD, 2005, p.14.

OCKERBLOOM, J. M. Towards the next generation: Recommendations for the next DSpace Architecture. **Tech. rep.**, DSpace, 2007.

PAYETTE, S. et al. Interoperability for digital objects and repositories: the Cornell/CNRI Experiments. **D-Lib Magazine**, v.5, n.5, 1999. Disponível em: <<http://dlib.org/dlib/may99/payette/05payette.html>>. Acesso em: 7 fev. 2009.

PAYETTE, S.; STAPLES, T. The Mellon Fedora Project. In **ECDL '02: Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries**, pp. 406–421. Springer-Verlag, London, UK, 2002.

PEREIRA, Fernanda. **Avaliação de usabilidade em bibliotecas digitais: um estudo de caso**. 2011. Dissertação (Mestrado em Ciência da Informação) – ECI/UFMG, Belo Horizonte, 2011.

PONTES, Flávio Vieira; BRAGA, José Luís; OLIVEIRA, Dóris Magna Avelar de; PINTO, Márcio Lívio Pereira. SBICafé: uma ferramenta de gestão do conhecimento para a comunidade de pesquisa do café. In: Simpósio Brasileiro de Pesquisa dos Cafés do Brasil (2. : 2001 : Vitória, ES). **Anais**. Brasília, D.F. : Embrapa Café, 2001. (CD-ROM), p. 2606-2612.

PONTES, Flávio Vieira. **Gestão de conhecimento apoiada por ontologias**. Belo Horizonte: UFMG, 2002. (Dissertação - mestrado em Ciência da Computação).

SACCO, Giovanni Maria. Dynamic taxonomies: a model for large information bases. **IEEE Transactions on Knowledge and Data Engineering**, 12(3):468-479, 2000.

SACCO, Giovanni Maria; TZITZICAS, Yannis (Eds.). **Dynamic Taxonomies and Faceted Search: theory, practice, and experience**. Springer-Verlag Berlin Heidelberg, 2009.

SARACEVIC, T. (2000) Digital library evaluation: Toward evolution of concepts. **Library Trends - Special issue on Evaluation of Digital Libraries**, 49(3):350-369, 2000.

SARACEVIC, T. Evaluation of digital libraries: an overview. In: Agosti, M., Fuhr, N. (eds.) Notes of the **DELOS WP7 Workshop on the Evaluation of Digital Libraries**, Padua, Italy (2004). Disponível em: <http://dlib.ionio.gr/wp7/workshop2004_program.html>. Acesso em 11 out. 2010.

SARACEVIC, Tefko. How were digital libraries evaluated? In: CONFERENCE LIBRARIES IN THE DIGITAL AGE (LIDA). Dubrovnik, Croatia, 2005. Disponível em: <<http://comminfo.rutgers.edu/~tefko/articles.htm>> Acesso em: 10 out. 2010.

SAYÃO, Luis Fernando. Preservação digital no contexto das bibliotecas digitais: uma breve introdução. In: MARCONDES, Carlos Henrique; KURAMOTO, Hélio; TOUTAIN, Lidia Brandão; SAYÃO, Luis Fernando (Org.). **Bibliotecas digitais: saberes e práticas**. Salvador/Brasília: UFBA/IBICT, 2 ed. 2006, p. 113-143.

SAYÃO, Luiz Fernando. Bibliotecas digitais e suas utopias. **Ponto de Acesso**, Salvador, v.2, n.2, p. 2-36, ago. /set. 2008.

SAYÃO, Luiz Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **TransInformação**, Campinas, 20(2): 133-148, maio/ago., 2008.

SHANNON, Claude; WEAVER, Warren. **Teoria matemática da comunicação**. São Paulo: Difel. 1975. p.2-29.

SHIRI, Ali. Digital library research: current developments and trends. **Library Review**, v. 52, n. 5, p. 198-2002, 2003. Disponível em: <http://www.cis.strath.ac.uk/cis/research/publications/papers/strath_cis_publication_39.pdf>. Acesso em: 03 abr. 2009.

SHNEIDERMAN, B.; PLAISANT, C. **Designing the user interface**: strategies for effective Human-Computer Interaction. Addison-Wesley, 4th edn., 2004.

SILVA, Guilherme Baião Salgado. **A utilização de mapas de tópicos na compatibilização de conteúdos hipertextuais semanticamente estruturados**. Dissertação (Mestrado em Ciência da Informação) – ECI/UFMG, Belo Horizonte, 2008.

SMEATON A. F.; CALLAN J. Personalisation and recommender systems in digital libraries. **Int. J. on Digital Libraries**, 5(4):299–308, 2005.

SORGEL, Dagobert. Digital Libraries and knowledge organization. In: KRUK, Sebastian Ryszart; MCDANIEL, Bill (Eds.) **Semantic digital libraries**; Springer: 2009. p.9-39.

SPITERI, Louise. A Simplified Model for Facet Analysis: Ranganathan 101. **Canadian Journal of Information and Library Science** 23 no. 1/2 (April-July 1998): 1-30. Disponível em <http://aifia.org/pg/a_simplified_model_for_facet_analysis.php>. Acesso em 30 ago. 2010.

TENNANT, Roy. Of real and digital libraries. **Library Journal**, v.132, n.9, p.29, May 15, 2007.

TOUTAIN, Lídia Maria Batista Brandão. Biblioteca digital: definição de termos. In: MARCONDES, Carlos Henrique; KURAMOTO, Hélio; TOUTAIN, Lidia Brandão; SAYÃO, Luis Fernando (Org.). **Bibliotecas digitais: saberes e práticas**. Salvador/Brasília: UFBA/IBICT, 2 ed. 2006, p. 15-24.

TZITZIKAS, Yannis; SPYRATOS, Nicolas; CONSTANTOPOLOUS, Panos; ANALYTI, Anastasia. Extended Faceted Taxonomies for Web Catalogs. Paper presented at the Third **International Conference on Web Information Systems Engineering**, WISE 2002,

Singapore, December, 2002. Disponível em:
<http://www.csi.forth.gr/~tzitzik/publications/Tzitzikas_WISE_2002.ps.gz> Acesso em 22 set. 2010.

VAN de SOMPEL, H.; NELSON, M.L.; LAGOZE, C.; WARNER, S. Resource Harvesting within the OAI-PMH Framework. **D-LibMagazine**, 10(12), 2004. Disponível em:
<<http://www.dlib.org/dlib/december04/vandesompel/12vandesompel.html>>. Acesso em 17 jun. 2010.

VELLUCCI, Sherry L. Metadata. **Annual Review of Information Science and Technology (ARIST)**, v.33, p.189-222, 1998.

VICKERY, B.C. **Faceted Classification**: A Guide to Construction and Use of Special Schemes. London: Aslib, 1960.

VICKERY, B. **On 'knowledge organisation'**. Disponível em:
<<http://www.lucis.me.uk/knowlorg.htm#start>>. Acesso: 30 abr. 2010.

ZAFALON, Regina Zaira. Biblioteca em tempo real: o acesso em foco. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.6, n. 1, p. 61-83, jul/dez. 2008 – ISSN: 1678-765X.

APÊNDICE A

E-mail enviado aos estudantes do PPGCI

Data do Envio: 26/12/2012

Destinatários: Relação dos estudantes atualmente matriculados, informada pela Secretaria do PPGCI em 26/11/2012, perfazendo um total de 118 estudantes.

Assunto: Pesquisa do PPGCI/UFMG

Mensagem:

Prezado Colega,

Gostaria de convidá-lo a participar de uma pesquisa do Doutorado em Ciência da Informação do PPGCI/UFMG, do estudante Flávio Pontes, sob a orientação da Profa. Gercina Lima.

Objetivo da pesquisa:

Implementar novos mecanismos para a representação e organização do acervo de uma biblioteca digital de teses e dissertações, com base em uma estrutura classificatória facetada e em taxonomias dinâmicas. Objetiva também avaliar o desempenho desses mecanismos na melhoria da recuperação da informação e na satisfação do usuário.

Orientações:

- Para participar é só responder a um questionário simples, que leva cerca de 10 minutos para ser realizado.
- As respostas são sigilosas e servirão somente para os objetivos listados a seguir.

Objetivos do questionário a ser respondido:

- Obter informações demográficas, hábitos, necessidades de informação e formas de acesso à informação para estudos e pesquisas, entre os estudantes do PPGCI/UFMG.
- Verificar a disponibilidade do estudante para participar das etapas posteriores da pesquisa, que irá envolver um experimento de uso do Protótipo para um Sistema de Biblioteca Digital de Teses e Dissertações, desenvolvido no contexto de nossa pesquisa.

Caso aceite participar clique no link a seguir:

<https://www.onlinepesquisa.com/live.php?code=1b92273>

Sua participação será muito importante para auxiliar a conclusão de minha pesquisa de doutorado.

Muito Obrigado,
Flávio.

APÊNDICE B

Questionário de levantamento inicial - Q1

Questionário de levantamento inicial com o objetivo de obter informações demográficas, hábitos, necessidades de informação e formas de acesso à informação para estudos e pesquisas, entre os estudantes do PPGCI/ECI/UFMG.

Objetiva também verificar a disponibilidade do estudante para participar das etapas posteriores da pesquisa, que irá envolver testes de uso do Protótipo para um sistema de biblioteca digital de teses e dissertações, desenvolvido no contexto de nossa pesquisa.

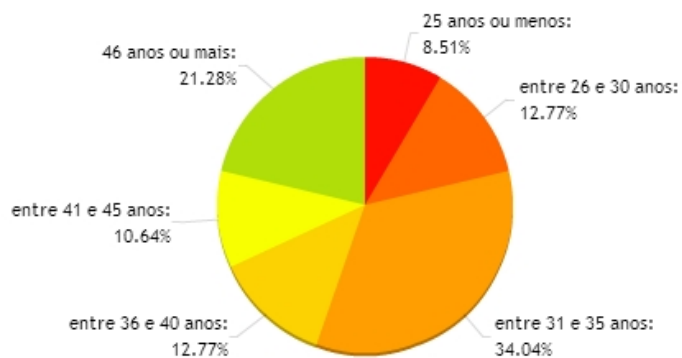
As questões marcadas com * são de preenchimento obrigatório.

Sobre você

1. Sexo: *
- Feminino
 - Masculino

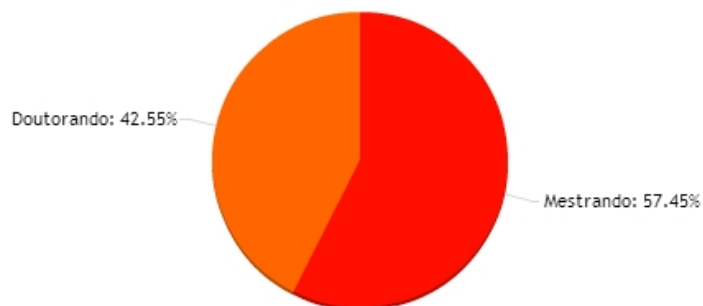


2. Idade: *
- 25 anos ou menos
 - entre 26 e 30 anos
 - entre 31 e 35 anos
 - entre 36 e 40 anos
 - entre 41 e 45 anos
 - 46 anos ou mais



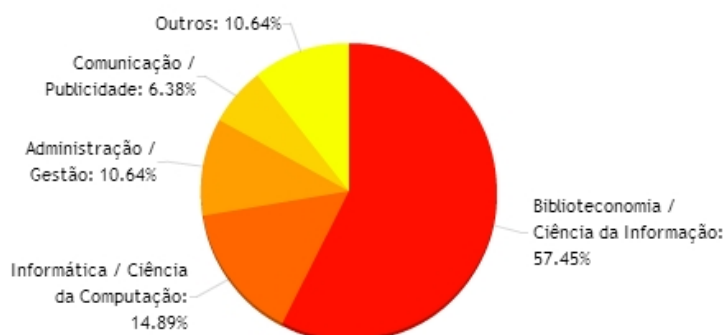
3. Nível: *

- () Mestrando
- () Doutorando



4. Área principal de formação (graduação): *

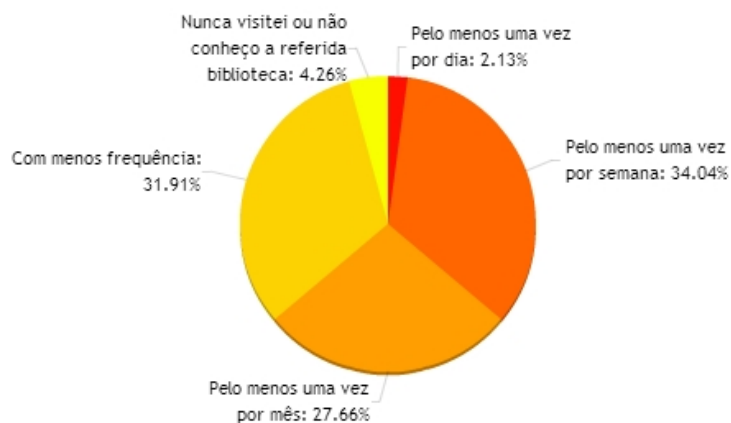
- () Biblioteconomia / Ciência da Informação
- () Informática / Ciência da Computação
- () Administração / Gestão
- () Comunicação / Publicidade
- () Outra:



No que se refere à opção “Outra”, foram mencionadas a área de Engenharia e suas especializações, por 3 respondentes, e Filosofia, por 1 respondente.

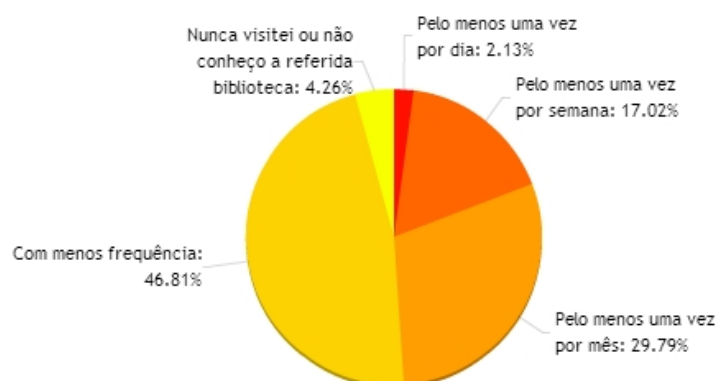
5. Qual a frequência com que visita as instalações da biblioteca da ECI/UFMG? *

- () Pelo menos uma vez por dia
- () Pelo menos uma vez por semana
- () Pelo menos uma vez por mês
- () Com menos frequência
- () Nunca visitei ou não conheço a referida biblioteca



6. Qual a frequência com que visita a biblioteca digital de teses e dissertações da ECI/UFMG? *

- Pelo menos uma vez por dia
- Pelo menos uma vez por semana
- Pelo menos uma vez por mês
- Com menos frequência
- Nunca visitei ou não conheço a referida biblioteca



7. Quantas horas diárias, em média, você utiliza para acesso a Internet? *

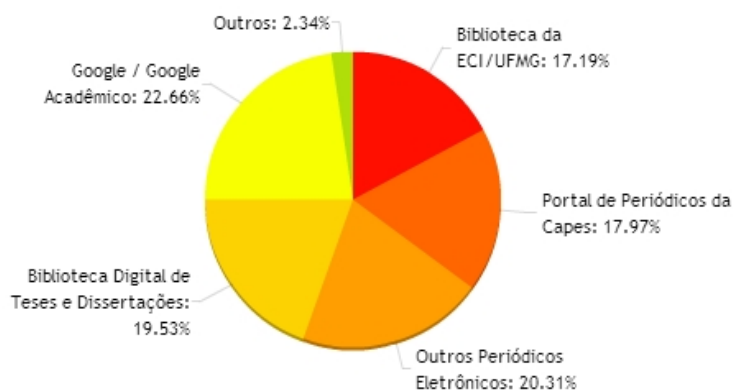
- menos de 1 hora por dia
- entre 1 a 2 horas por dia
- entre 2 a 4 horas por dia
- mais de 4 horas por dia



8. Quais as principais fontes de informação para estudos/pesquisas? (selecione no máximo 3 fontes de informação) *

(selecione as principais fontes de informação utilizadas)

- Biblioteca da ECI/UFMG
- Portal de Periódicos da Capes
- Outros Periódicos Eletrônicos
- Biblioteca Digital de Teses e Dissertações
- Google / Google Acadêmico
- Outra:



No que se refere à opção “Outra”, três respondentes mencionaram, respectivamente: Artigos, revistas e fontes de informação online; Amazon; e Congressos, Seminários, etc.

Sobre suas necessidades de informação

No contexto de seu projeto de pesquisa, elabore livremente uma questão ou pergunta representativa de suas necessidades de informação, que poderia ser respondida a partir de consultas ao acervo de teses e dissertações da BDTD/ECI/UFMG. Descreva a pergunta com suas palavras de modo que possa ser passada a um bibliotecário de referência, responsável pelo atendimento na biblioteca. (aqueles que se dispuserem a participar da segunda etapa da pesquisa serão auxiliados na busca pelas respostas para a questão formulada)

Gostaria de participar da segunda etapa da pesquisa? *

(a segunda etapa da pesquisa irá envolver um experimento de uso do Protótipo para um sistema de biblioteca digital de teses e dissertações, desenvolvido em nossa pesquisa)

Sim

Não

Sobre seu projeto de pesquisa

Forneça algumas informações sobre seu projeto de pesquisa.

Estas informações serão utilizadas para entender melhor suas necessidades de informação.

O sigilo e a confidencialidade das informações serão preservados.

Título do projeto de pesquisa:

Descrição sucinta (em no máximo 5 linhas):

Metodologia utilizada:

Pesquisas relacionadas (já desenvolvidas ou em desenvolvimento) no PPGCI/UFMG:

(relacione qualquer descrição que souber: título, assunto, pesquisador, orientador, das pesquisas que apresentam relação com seu projeto)

Sua identificação

Os dados fornecidos serão utilizados SOMENTE para lhe contactar, no caso de ser selecionado para participar da segunda etapa da pesquisa. Garantimos o sigilo das respostas e a não identificação dos respondentes.

Seu nome completo:

Seu e-mail:

APÊNDICE C

E-mail enviado aos respondentes que se dispuseram a participar do teste de uso dos sistemas

Data do envio: 04/12/2012

Destinatários: Respondentes do Questionário de Levantamento Inicial (Q1) que se dispuseram a participar do teste com usuários, perfazendo um total de 17 estudantes:

Kelly Cristiane Santos Morais kellycristianesantosmorais@gmail.com; Marcel Ferrante Silva marcelf@gmail.com; Janicy Rocha janicy.rocha@gmail.com; Ráisa Mendes Fernandes de Souza raisamendes@gmail.com.br; Rodrigo Moreno Marques rodrigomorenomarques@yahoo.com.br; Marcus Vinicius Rodrigues Martins marcusmartins2005@gmail.com; Edson Marchetti da Silva edson@div.cefetmg.br; Rita Ventura ritakmartins@hotmail.com; Álamo Chaves de Oliveira Pinheiro o.arietino@yahoo.com.br; Alzira Karla Araújo da Silva alzirakarla@gmail.com; Luciana Emirena dos Santos Carneiro lucianaemirena@yahoo.com.br; Ruleandson do Carmo Cruz ruleandson@gmail.com; José Alimatéia de Aquino Ramos alimateia2002@yahoo.com.br; Jasmária de Oliveira josmariadeoliveira@gmail.com; Jacqueline Pawlowski Oliveira jackiepaw@gmail.com; Débora de Carvalho Pereira debcarpe@gmail.com; Gabrielle Francinne de Souza carvalhogfrancinne@gmail.com

Assunto: Participação em Pesquisa de Doutorado do PPGCI

Mensagem:

Prezado(a) <NOME DO(A) ESTUDANTE>,

Em primeiro lugar agradecemos pela colaboração em responder ao questionário da nossa pesquisa. Sua contribuição será muito importante para a conclusão do trabalho.

Conforme mencionado no e-mail anterior, um dos objetivos do referido questionário foi verificar a disponibilidade do estudante para participar da 2ª etapa da pesquisa, que irá envolver um teste de uso do Protótipo para um Sistema de Biblioteca Digital de Teses e Dissertações, desenvolvido no contexto de nossa pesquisa. Diante da sua resposta afirmativa para a questão “Gostaria de participar da segunda etapa da pesquisa?”, estamos entrando em contato novamente para lhe explicar os objetivos e procedimentos previstos, bem como verificar a sua disponibilidade para o teste no período previsto.

No questionário respondido, e considerando o contexto do seu projeto de pesquisa, você elaborou uma questão ou pergunta representativa de suas necessidades de informação, que poderia ser respondida a partir de consultas ao acervo de teses e dissertações da BDTD/ECI/UFMG. No teste a ser realizado na 2ª etapa, você será orientado no uso do Protótipo para um Sistema de Biblioteca Digital de Teses e Dissertações, a fim de encontrar as respostas para a questão formulada. Ao final do teste, você receberá as informações levantadas por uma bibliotecária, profunda conhecedora do acervo de teses e dissertações do BDTD/ECI/UFMG, consideradas relevantes para o desenvolvimento do seu projeto de pesquisa.

Os testes com os participantes selecionados estão previstos para ocorrer na semana de 17 a 21/12/2012. O teste tem duração prevista de 40 minutos, e será realizado nas dependências da ECI. Gostaríamos de saber sobre sua disponibilidade para essas datas, bem como o período (manhã ou tarde) de sua preferência.

Copiamos abaixo suas respostas referentes às suas necessidades de informação, além dos dados informados sobre seu projeto de pesquisa. Fique a vontade para fazer qualquer alteração que considerar pertinente. Aguardamos seu retorno e ficamos a disposição para responder a eventuais dúvidas.

Att.
Flávio

QUESTÃO FORMULADA: <questão formulada pelo participante>

PROJETO DE PESQUISA: <projeto de pesquisa informado pelo participante>

DESCRIÇÃO SUCINTA: <descrição fornecida para o projeto de pesquisa >

METODOLOGIA: <metodologia informada>

PESQUISAS RELACIONADAS: <pesquisas relacionadas informadas pelo participante>

APÊNDICE D

Questionário de satisfação - Q2

Questionário aplicado após as tarefas de busca e recuperação da informação, utilizando cada um dos sistemas (sistema atual da BDTD/ECI/UFMG e protótipo TDF-Bíblia).

1. O que achou da tarefa de encontrar informação na biblioteca digital?

Difícil de entender Fácil de entender

Difícil de executar Fácil de executar

Difícil de aprender, pouco intuitivo Fácil, intuitivo, direto

2. O que achou do mecanismo de busca na biblioteca digital?

Difícil de usar Fácil de usar

Complexo, confuso Simples, organizado

Difícil de aprender, pouco intuitivo Fácil, intuitivo, direto

Desinteressante Interessante

Feio, pouco atraente Atraente

Sem utilidade Útil, conveniente

3. “Eu penso que teria uma melhor experiência caso...” (complete a sentença)

4. Você gostaria de continuar a utilizar o sistema de biblioteca digital avaliado? (justifique)

() Sim

() Não

Justificativa:

5. Comentários gerais sobre a avaliação.