

VIII ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informação  
28 a 31 de outubro de 2007 • Salvador • Bahia • Brasil

GT 2 – Organização e Representação do Conhecimento  
Pôster

**ORGANIZAÇÃO DA INFORMAÇÃO EM REPOSITÓRIOS  
TEMÁTICOS: o uso da modelagem conceitual**

***ORGANIZATION OF INFORMATION IN THEMATIC REPOSITORY:  
use of the modelling knowledge***

Fernanda Monteiro (PPGCIInf/UnB, fernanda.s.monteiro@gmail.com)  
Marisa Bräscher (PPGCIInf/UnB, marisab@unb.br)

**Resumo:** Trata a perspectiva cada vez mais dinâmica da informação no meio digital, favorecida pelo acesso livre e arquivos abertos que, dispostos em repositórios, vislumbram a possibilidade de reunir, divulgar e disponibilizar informações científicas. Neste contexto, essas informações encontram alternativa aos mecanismos tradicionais de comunicação, mas esbarra na problemática ocasionada pelo amplo universo de informação disponível: a dificuldade de organizá-la coerentemente para facilitar sua recuperação. A partir destas considerações, este estudo demonstra como a modelagem dos conceitos compreendidos pela área de conhecimento abordada por um repositório temático pode contribuir para a organização da sua informação. Esta proposta prevê a representação do conhecimento com respaldo da Classificação Facetada, Teoria do Conceito e Terminologia, para definir relações entre os conceitos do domínio coberto pelo repositório, permitindo assim, que a modelagem oriente semanticamente sua estrutura de informação e a disposição dos documentos a serem depositados.

**Palavras-chave:** acesso aberto; arquivos abertos; modelagem conceitual; organização da informação; repositório.

**Abstract:** Each more dynamic time deals with the perspective the scientific publications in the digital way, favored for the open access and the open archives. In this context, these scientific publications find an alternative to its traditional mechanisms of communication, but, it finds the problematic caused by the ample universe of available information: the difficulty to organize it coherently to information retrieval. To leave of these considerations, this study presents the modeling of knowledge for the contribute information organization.

**Keywords:** open access movement; open archives; modelling of knowledge; knowledge representation.

## 1 Introdução

A informação contempla uma perspectiva cada vez mais dinâmica com o crescimento da sua disseminação em meio digital. O tônio desta perspectiva é favorecido por inovações tecnológicas como os arquivos de publicações digitais livres, ou arquivos abertos<sup>i</sup>, que associam importantes características, como a acessibilidade, à promoção de seu uso. Os arquivos abertos, dispostos em repositórios, vislumbram a possibilidade de reunir, divulgar e disponibilizar uma vasta quantidade de informações. De acordo com Suber (2005), os repositórios podem ser organizados por tema, ou seja, conjuntos de documentos sobre uma área de conhecimento, ou agrupados com caráter multidisciplinar, constituindo repositórios institucionais. Neste contexto, a informação científica encontra uma alternativa aos seus mecanismos de comunicação tradicionalmente intermediados por periódicos científicos impressos (DAY, 1999), mas em contrapartida, esbarra na problemática ocasionada por esse amplo universo de informação disponível: a dificuldade de organizá-la coerentemente para facilitar sua recuperação.

A partir destas considerações, este estudo evidencia como a modelagem dos conceitos compreendidos pela área de conhecimento abordada por um Repositório Temático (RT) pode contribuir para a organização da sua informação. Esta proposta prevê o uso da representação do conhecimento com respaldo da Classificação Facetada (RANGANATHAN, 1967), da Teoria do Conceito (DAHLBERG, 1978) e da Terminologia (NEDOBITY, 1985; apud Campos, 1994), para definir relações entre os conceitos do domínio coberto pelo RT, permitindo assim, que a modelagem oriente semanticamente a estrutura da informação no repositório e a disposição dos documentos depositados.

Sendo esta uma pesquisa em andamento, são apresentados a seguir importantes aspectos que devem ser observados para sua elucidação, tais como, o levantamento de informações sobre repositórios, suas características e estrutura; além da abordagem sobre modelagem conceitual e os resultados esperados da sua utilização na organização da informação em RTs.

## 2 O contexto de desenvolvimento dos repositórios

Resultados de pesquisa são intermediados tradicionalmente por grandes editoras e suas publicações científicas que, a partir da avaliação, aceite e posterior publicação de artigos, mediam a comunicação científica. Diante de nítidos intervenientes vislumbrados neste panorama, como por exemplo, o custo da assinatura de periódicos científicos, o dispendioso processo de submissão e publicação e a inacessibilidade aos documentos que estão fora dos circuitos editoriais<sup>ii</sup>, a Internet e seus recursos de comunicação foram identificados como um meio favorável para minimizar, ou ainda eliminar, tais problemas.

Este contexto é referendado pela nomeada Filosofia Aberta que define o movimento observado nos últimos anos em direção ao uso de ferramentas, estratégias e metodologias que denotam o novo modelo de comunicação científica (COSTA; MOREIRA, 2003). Nas premissas deste modelo releva-se o acesso aberto<sup>iii</sup> para disseminação ampla e irrestrita de resultados da pesquisa científica; os arquivos abertos, com padrões específicos que viabilizam sua utilização e acesso global; e o uso de software livre, para o desenvolvimento de aplicações em computador que intermedeiem este tipo de comunicação (COSTA, 2006).

Com o **acesso aberto** (ou livre), a comunidade científica percebeu um meio de aumentar sua visibilidade. Diferentes estudos têm mostrado isto, a citar, Brody; Harnad (2004) que chamam atenção para o fato de que o acesso livre a resultados de pesquisa maximiza o acesso à pesquisa propriamente dita, acelerando e aumentando seu impacto e, conseqüentemente, a produtividade, o progresso e as recompensas da mesma. Essa observação justifica o interesse dos atores<sup>iv</sup> envolvidos com o progresso científico na elaboração de arquivos eletrônicos de

pré-prints e pós-prints, chamados arquivos abertos, como alternativa para publicação (MARCONDES, 2001).

O conceito de **arquivo aberto**, contudo, não inclui o acesso livre como uma prerrogativa e assim já foi destacado por Van de Sompel e outros autores (COSTA, 2005), tal terminologia é pertinente e leva em consideração as características destes, que são arquivos armazenados em um computador disponíveis para o acesso *on line*, a longo prazo, por meio de mecanismos que possibilitem que terceiros utilizem-no diretamente e/ou criem serviços que permitam a busca e utilização por usuários finais. Estes arquivos abertos e automatizados aparecem como parte do modelo mais equitativo e eficiente para a disseminação dos resultados de pesquisa, a alternativa para comunicação científica outrora citada, em que os próprios autores serão responsáveis pela publicação de seus trabalhos (SENA, 2000).

No desenvolvimento destes arquivos abertos, vale citar a iniciativa de Paul Ginsparg ao criar, em 1991, um repositório mantido pelo Laboratório Nacional de Los Alamos (Novo México) (NEVES, 2004). Este precursor, demais personalidades e instituições, somaram-se no intuito de coordenar o desenvolvimento de arquivos abertos e, em 1999 na cidade de Santa Fé (Novo México), realizaram uma reunião com representantes que mantinham ou pretendiam instituir arquivos abertos com serviços de busca ou *links* para o livre acesso a estes documentos. Como resultado deste encontro, a *Santa Fe Convention for the Open Archives Initiative*<sup>v</sup>, foi estabelecido um conjunto de especificações técnicas e princípios organizacionais simples em prol da interoperabilidade e integração desses arquivos (LAGOZE; SOMPEL, 2000).

Outras manifestações que sucedem esta Convenção tiveram crucial importância no cenário dos arquivos abertos e do acesso livre, a saber:

- Budapest Open Access Initiative (BOAI), 2002: movimento planejado pela comunidade científica que define estratégias baseadas no uso do protocolo Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH), e no conceito de acesso aberto como a disponibilidade livre e irrestrita a publicações científicas referenciadas.

- Reunião de Bethesda, 2003: nesta reunião foram definidas condições para o acesso aberto a publicações, prevendo que os autores e detentores de direitos autorais devem assegurar ao usuário o direito livre, irrevogável e perpétuo de acesso a seus trabalhos; licença para copiar, usar, distribuir, transmitir e exibir trabalhos publicamente; e elaborar e distribuir trabalhos derivados em meio digital, para qualquer propósito responsável, sujeito à atribuição apropriada de autoria. Publicações de acesso aberto devem, ainda, ter uma versão completa, o material suplementar e uma cópia da licença, tudo em formato eletrônico apropriado, depositada, imediatamente após a publicação inicial, em pelo menos um repositório on-line. Este repositório deve ser mantido por uma instituição acadêmica ou outra instituição reconhecida de acordo com o acesso livre, a distribuição irrestrita, a interoperabilidade e o arquivamento a longo prazo.

- Declaração de Berlim, 2003-2006: reunião que tem ocorrido anualmente, desde 2003, para discutir assuntos concernentes ao acesso aberto. Seus resultados têm promovido o uso da Internet como instrumento funcional para disponibilização do conhecimento e enunciado medidas a serem seguidas por formuladores de política, instituições de pesquisa, agências de fomento, bibliotecas e museus. A Declaração determina às instituições a necessidade de implementar uma política para o depósito dos artigos publicados por seus pesquisadores em repositórios de acesso aberto e incentiva que estes publiquem seus artigos em periódicos de acesso aberto (2005).

- Apontamentos brasileiros - Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica (13-09-2005); Declaração de Salvador (26-09-2005); Carta de São Paulo (Dez/2005); Declaração de Florianópolis (Mai/2006); Carta aberta à SBPC (18-07-2006); e o Projeto de Lei 1120/2007 (maio-2007): todas estas são iniciativas de atores brasileiros envolvidos com o progresso científico e contribuem para mostrar o interesse e a pré-disposição

destes para incentivar e estabelecer diretrizes que possibilitem o acesso aberto a informação. O Manifesto (2005) faz referência a Declaração de Berlim, 2003-2006 (2005) e estabelece o Paradigma do Acesso Livre à Informação com objetivos de promover o registro e a disseminação da produção científica brasileira, estabelecer uma política nacional de acesso livre à informação científica e buscar o apoio da comunidade científica. A Carta Aberta à SBPC é uma representação de instituições e indivíduos brasileiros para solicitar que as associações científicas, particularmente as filiadas à SBPC, discutam a questão do acesso aberto e se manifestem nos termos dos documentos expedidos a partir das iniciativas nacionais anteriores. E, atualmente, vale comentar a proposição mais ousada no sentido do acesso livre, o Projeto de Lei 1120/2007 que pretende assegurar a construção de repositórios para disseminar, via Internet e de forma livre, a produção técnico-científica brasileira entre outras importantes determinações.

Estas iniciativas e movimentos pretendem otimizar os custos de disponibilização e acesso à informação promovendo seu fluxo e o desenvolvimento científico e tecnológico da sociedade. Nesta conjectura, o conceito de acesso aberto representa a disponibilidade e acesso às informações em formato eletrônico (COSTA, 2005) que, com os recursos da tecnologia da informação e comunicação, passaram a ser editadas, descritas e depositadas por seus próprios autores em ambientes digitais. Desta forma, os repositórios, apresentados a seguir, são os sistemas implantados a partir de ferramentas específicas para sustentar esta nova forma de produção e divulgação científica.

### 3 Repositórios Temáticos e Institucionais

Dentro do contexto apresentado, que se destaca pelo acesso livre e o uso de arquivos abertos, os repositórios são distinguidos em temáticos e institucionais, além de apresentarem estrutura e características próprias.

O primeiro tipo de repositório, o **repositório temático**, tem a preocupação de armazenar documentos com uma delimitação concisa de sua cobertura designada por um assunto, área do conhecimento ou temática específica. Em outras palavras, Kuramoto (2006, p. 83) afirma esta definição dizendo que repositórios temáticos “são um conjunto de serviços oferecidos por uma sociedade, associação ou organização, para gestão e disseminação da produção técnico-científica em meio digital, de uma área ou subárea específica do conhecimento”.

O êxito destes RTs suscitou discussões sobre o seu funcionamento e a necessidade de um gestor que garantisse seu bom desempenho fazendo surgir a figura de uma instituição responsável e agregadora das iniciativas individuais de desenvolvimento de RTs, neste momento emergem os repositórios institucionais (CAFÉ, 2003).

O **repositório institucional** (RI) é a reunião de RTs sob a responsabilidade técnica e administrativa de uma instituição ou organismo. Por consequência, este tipo de repositório é multidisciplinar e possui uma gama de tipos de documentos ainda maior que os RTs. Além de agregar o conjunto de informações relativas e/ou de interesse para a instituição, dispõem de serviços relativos à organização, tratamento, acesso e disseminação deste conteúdo digital (CAFÉ, 2003).

### 4 Características e estrutura dos repositórios

Os repositórios, sejam eles temáticos ou institucionais, apresentam determinadas características, possuem uma estrutura comum de submissão e acesso às informações e são desenvolvidos segundo padrões de interoperabilidade específicos. Esses assuntos, tratados neste momento, têm o intuito de fazer entender o potencial destes sistemas para agregação e divul-

gação da informação eletrônica. Entretanto, vale dizer que algumas das características ou observações quanto à estrutura destes repositórios são impostas pelos padrões que utilizam.

Uma primeira consideração a ser feita é que de acordo com os objetivos propostos para um repositório este pode contemplar uma infinidade de tipos de documentos, ou seja, ter uma **tipologia variada de documentos**. Mesmo com a concepção de que os repositórios foram desenvolvidos para divulgar documentos já publicados nos meios tradicionais (BUDAPEST..., 2002) não é dada uma delimitação à cerca dos tipos de documentos que atualmente podem compor um repositório (CAFÉ, 2003).

O **processamento automático dos mecanismos de discussão entre os pares** é uma característica que tem por objetivo otimizar a comunicação científica, permitindo que durante o desenvolvimento de um produto de pesquisa (artigo, relatório, etc.) o autor possa se comunicar e trocar experiências com seus pares, assim construindo o conhecimento com a validação e colaboração dos mesmos. Este processo, que muitas vezes é restrito ao universo de colegas próximos do autor, pode ser ampliado com o uso de repositórios, a partir das ferramentas e políticas elaboradas para estes, ocasionando maior transparência ao processo de revisão e crítica pelos pares (CAFÉ, 2003).

A **geração de versões de um mesmo documento**, como consequência da característica anterior, define que diferentes versões de um documento possam ser geradas a partir de sugestões dos pares e alterações elaboradas pelo próprio autor (CAFÉ, 2003). Esta questão implica na manutenção ou substituição destas versões, de acordo com a política do repositório, contudo, é viável optar pela manutenção para que o acesso a um documento não seja se torne indisponível.

No que tange esta manutenção das versões de um documento, é possível identificar uma característica correlacionada, a **preservação digital**. A preocupação em disponibilizar conteúdos a longo prazo dá sustentação a pesquisas sobre preservação digital e isto inclui a migração do conteúdo digital de uma tecnologia em vias de ser desativada por outra, a necessidade de aceitar a responsabilidade pela preservação e controle do conteúdo dos repositórios, demonstrar sua sustentabilidade financeira e viabilidade organizacional, manter a informação sobre os documentos armazenados com o uso de metadados, entre outras considerações (James, et al. 2003; apud HUNTER; DAY, 2005).

Já a **interoperabilidade** entre os repositórios envolve uma série de aspectos, tais como, um conjunto mínimo de metadados, a arquitetura subjacente do sistema, integração com o mecanismo de comunicação já existente no meio científico, possibilidade de uso em contextos interdisciplinares, entre outros (CAFÉ, 2003). Tudo isto, é parte do diálogo que estabelece o uso de protocolos em repositórios e distingue os participantes do seu cenário, a saber, provedores de dados e serviços, fornecendo os pilares da estrutura de submissão, acesso e posterior utilização de suas informações.

#### **4.1 Submissão e acesso à informação**

A OAI com o protocolo OAI-PMH prevê a distinção de dois níveis distintos de participantes no desenvolvimento de coleções em repositórios. Nos termos deste protocolo temos provedores de dados e provedores de serviços, como já mencionado (LAGOZE, et al., 2002). Os provedores de dados permitem a descrição de documentos e seu respectivo armazenamento. Esta descrição é feita com a definição de metadados que ficam armazenados nos provedores de dados e permitem, com o uso das especificações do OAI-PMH, a exposição e colheita destes pelos provedores de serviços. Os provedores de serviço, por sua vez, agregam estes dados e disponibilizam serviços de informação, preocupando-se com a análise da área temática coberta, o tipo de recurso incluído e o nível de qualidade dos conteúdos selecionados. De acordo com Hunter; Day (2005), muitas instituições serão provedoras de dados embora pos-

sam também prover serviços a partir da colheita dos dados de suas diferentes áreas, ou seja, os RIs colhem os dados dos RTs e divulgam-nos. Existe, portanto, a necessidade de cada segmento da instituição compreender não apenas a finalidade do repositório, mas a importância da sua alimentação, sua estrutura de conteúdos e as determinações sobre a submissão e o acesso às informações.

A alimentação dos repositórios prima pelo **auto-arquivamento**, ou seja, requer que os próprios autores depositem seus documentos e gerenciem suas versões e atualizações, bem como, a descrição destes com a atribuição de metadados. Contudo, as diretrizes do repositório devem ser previamente definidas, inclusive o fluxo de trabalho que fornece critérios como as permissões para a submissão dos trabalhos, políticas para a avaliação dos documentos, edição de versões e acesso por meio de pesquisas e *downloads*.

No tocante a **estrutura do conteúdo a ser armazenado**, é possível observar que a formação de um repositório, mais especificamente um provedor de dados, depende da instituição em que este se insere, dos seus objetivos e do seu escopo. Sendo este repositório considerado um arquivo digital de produtos intelectuais criados pela equipe de pesquisadores ou pelos estudantes de uma instituição, é preciso fazer escolhas que delimitem exatamente o conteúdo a ser coberto. Uma primeira decisão diz respeito aos tipos de documentos que o repositório deve abranger (seus *out-puts*), como *pré-prints*, dissertações, teses, registros administrativos, etc. Esta definição, que pode ser modificada ao longo do desenvolvimento do repositório, contribui para a definição dos metadados. Outra decisão mais prática trata do formato destes documentos aceitos pelo repositório, tais como, Html, pdf, postScript, TeX ou tiff, mesmo considerando a existência de outros formatos. Concluindo, a delimitação do tipo de documento a ser coletado, interfere no formato aceito para submissão e no tipo de metadados necessários a sua descrição (HUNTER; DAY 2005).

Considerando a alimentação do repositório por meio do auto-arquivamento e algumas definições estruturais que designam a abrangência de sua coleção, dá-se início ao processo de **submissão dos documentos**. Na maioria das vezes, a submissão esta condicionada a ferramenta escolhida para o desenvolvimento do repositório, embora a própria instituição possa desenvolver seu próprio processo de submissão. Estas ferramentas possuem interface de submissão baseada na *Web*, que permite o depósito (*upload*) do documento e o acréscimo de metadados descritivos. Uma vez depositados, os dados são transferidos para uma área de *buffer* onde serão submetidos ao controle de qualidade, ou seja, a análise antes de serem disponibilizados no repositório. Esse controle varia, porém é possível verificar se o individuo que submeteu o artigo tem autoridade para isso, se o documento está de acordo com o escopo do repositório e se os metadados associados são adequados (HUNTER; DAY 2005).

Depois que os documentos são submetidos, conforme as especificações, tem-se a possibilidade de **acesso ao conteúdo**, aos usuários finais tanto internos quanto externos à instituição, com poucas ou nenhuma barreira para o seu acesso. Mas, enquanto isso está plenamente em conformidade com os princípios de acesso aberto, há razões para os repositórios não tornarem todo seu conteúdo disponível publicamente. Conteúdos potencialmente restritos podem ser os relatórios ou teses que contêm informações comercialmente sensíveis, conjuntos de dados que estão em processo de análise, etc. Alguns desses documentos podem ser distribuídos na íntegra ou em parte dentro da instituição, mas seu conteúdo e os metadados associados, geralmente não estará disponível para usuários externos. Os repositórios, nesse caso podem necessitar da definição de diferenciados níveis de acesso, ou seja, inserir mecanismos de controle de acesso para determinados tipos de documentos restritos, enquanto assegura o acesso para aqueles que podem ser compartilhados de forma mais ampla (CROW, 2002).

## **5 A modelagem conceitual e a organização da informação em repositórios temáticos**

Primeiramente, vale ressaltar que a preocupação desta pesquisa fundamenta-se na utilização da modelagem conceitual, com o estabelecimento de relações entre os conceitos, que sirva para orientar a distribuição dos conteúdos em repositórios temáticos. Desta forma, a organização, respaldada pela modelagem, será focada na categorização dos assuntos, facilitando o entendimento de sua arquitetura, a compreensão do seu escopo e, até mesmo, a atribuição de metadados que descrevem o assunto dos documentos. Diante disto, apresentado o panorama em que se desenvolvem os repositórios e as observações sobre suas características e estrutura, cabe delimitar o objeto deste estudo, os repositórios temáticos, relevando que estes devem ser desenvolvidos diante da perspectiva do acesso livre a informação, disponibilizar arquivos abertos conforme as especificações da OAI e demais manifestações, preservar o conteúdo armazenado, garantir sua interoperabilidade e primar pelo auto-arquivamento.

A organização da informação em um RT, que assume o papel de provedor de dados e tem as especificações apresentadas, pode ser arbitrária, pode seguir os critérios de um grupo de especialista, ou ser definida conforme as permissões administrativas concedidas aos responsáveis por este repositório. A submissão de conteúdos em repositórios e a sua disposição conforme as estruturas organizacionais deste, implicam no posterior acesso a seu conteúdo e na avaliação por parte de provedores de serviço sobre a pertinência de determinada coleção aos seus serviços de informação. Portanto, uma organização coerente e que ressalte o assunto abordado pelo RT é uma proposta que favorece a utilidade dos repositórios.

Para tanto, um modelo, oportunamente utilizado na organização da informação, é a “representação simplificada e abstrata de fenômeno ou situação concreta, e que serve de referência para a observação, estudo ou análise” (FERREIRA, 2004, p. 1345). Logo, a modelagem conceitual, possibilita a representação do conhecimento<sup>vi</sup> contido nas definições de forma sistêmica, apoiando-se em alguns princípios comuns encontrados em instrumentos para a organização do conhecimento<sup>vii</sup>. Tais princípios advêm da Teoria do Conceito, Teoria da Classificação Facetada e Teoria Geral da Terminologia (CAMPOS, 1994). Esta base teórica respalda a definição dos conceitos, a organização do domínio e o estabelecimento de relações conceituais, tornando necessário sua investigação.

## 5.1 Teorias que respaldam a modelagem conceitual e as suas etapas

A Teoria do Conceito (DAHLBERG, 1978) esquematiza a **definição de conceitos** ao passo que define, estabelece sua composição e seus relacionamentos. Os princípios desta teoria que permeiam a modelagem conceitual procuram responder o que é o conceito, quais suas partes e para que este serve. Segundo esta autora, existem conceitos individuais e gerais, relativos aos diferentes tipos de objetos. “Os elementos contidos nos conceitos gerais encontram-se também nos conceitos individuais, sendo, portanto, possível reduzir os conceitos individuais aos gerais e ordená-los de acordo com os conceitos gerais” (DAHLBERG, 1978, p. 102).

A Lingüística traz outras soluções para a formação e definição dos termos em uma estrutura conceitual, especificamente para classificações e tesouros, mas estas ficam apenas no campo da língua e não da representação dos conceitos (CAMPOS, 1995). Leska (1981, apud CAMPOS, 1995), aponta na Terminologia princípios para expandir as referências lexicais e tornar os conceitos mais precisos e bem definidos, qualificando-os de acordo com seu relacionamento com outros descritores. Este terminólogo afirma ainda, que “a atividade de desenvolvimento e aperfeiçoamento do sistema de conceitos não fica fora da influência das regras gramaticais que governam os nomes que representam esses conceitos (...)” (Leska, 1981, apud CAMPOS, 1995, p. 4). Segundo Wüester, autor da Teoria Geral da Terminologia, (1981, apud CAMPOS, 1995; LARA, 2004) a normalização terminológica é pertinente à elaboração de sistemas de conceitos.

Já a **organização do domínio de conhecimento a ser modelado** prevê que a estrutura do domínio deve ser entendida como a exteriorização do conhecimento organizado conforme as inferências e abstrações pretendidas. Ranganathan elabora a Teoria da Classificação Facetada (RANGANATHAN, 1967), potencialmente auxiliando a organização de conceitos hierarquicamente estruturados. A teoria do autor encontra-se no âmbito da Ciência da Informação e propõe um conjunto de categorias capazes de identificar os conceitos de acordo com sua natureza, predeterminando a perspectiva a ser analisada.

Frente à Teoria Geral da Terminologia (Wüester, 1981, apud CAMPOS, 2003) a principal contribuição para a estrutura dos conceitos está na sua organização alfabética, não linear e associativa que, insuficiente para uma organização sistemática, retorna à classificação. E a terminologia, por causa de sua natureza sistemática, ao lado da classificação, tem contribuído para a representação da informação.

Por fim, o estabelecimento de **relações conceituais** em um dado contexto formam a estrutura conceitual deste contexto e são de natureza diversa (CAMPOS, 2004, p. 27). A forma como o conhecimento é disposto e organizado no modelo constitui um todo coeso com as definições ligadas umas as outras até o ponto em que podem ser estabelecidos exemplos. Campos (2004), enumera alguns tipos de relações estabelecidas para associar conceitos, tornando possível observar algumas contribuições do uso da modelagem conceitual na organização da informação em RTs:

- Relação categorial – reúne, em um primeiro grande agrupamento, os objetos por sua natureza. Constitui uma classe de maior amplitude que possibilita uma classificação generalista. Na Teoria do Conceito e na Teoria da Classificação Facetada se coloca como um elemento agregador que reúne os conceitos em um nível mais alto de uma dada taxonomia (Ranganathan, 1967; Dalhberg, 1978). A Teoria da Terminologia não possui esta relação, pois seu maior nível de agrupamento de conceitos, o sistema de conceitos, possui significado próprio representando uma classe de conceitos, e não a reunião de todas as classes.

A modelagem da área do conhecimento abordada por um RT permite que, com a identificação de assuntos principais, ou categorias, possam ser agregados, sob um ponto de vista comum, documentos que fazem referência a um mesmo assunto. Desta forma, o tema do repositório pode ser classificado e compreendido a partir de demais relações dispostas no modelo conceitual.

- Relação hierárquica – permite relacionar objetos de uma mesma natureza ou não, dependendo da abordagem realizada. É uma das principais relações em qualquer estrutura classificatória formando a espinha dorsal da mesma. Na Teoria do Conceito, esta relação pode ser de abstração de gênero e espécie, que forma cadeias de conceitos, ou relação lateral, que forma renques de conceitos. Para Dalhberg (1978), a relação hierárquica baseia-se em uma relação lógica de implicação, ou seja, nela os conceitos devem ser da mesma natureza. Entretanto, para Ranganathan (1967), uma relação hierárquica possui conceitos que não estão somente em uma relação de generalização e especialização, mas também em uma relação partetodo. Para o todo e suas partes, em muitos casos, os conceitos são de natureza diferentes. Na Teoria da Terminologia, esta denomina-se *relação lógica* e especifica muito mais os seus tipos. Para a Terminologia, com o objetivo de evidenciar as relações entre os conceitos e não de especificar uma estrutura sistemática, esta relação é fundamental na elaboração de definições, pois permite a compreensão do conceito.

Este tipo de relação, com a abordagem de qualquer das teorias expostas, permite interligar os assuntos que compreendem a coleção de um RT. A generalização e especialização facilita a disposição ideal de um documento no repositório organizado segundo a modelagem, considerando a predefinição de assuntos. A relação lateral permite visualizar as diferentes abordagens da temática de um repositório.

- Relação partitiva – recai sobre a constituição do objeto, ou seja, quais são suas partes e elementos. Esta relação, na maioria das teorias, é tratada simplesmente como a parte de um todo, não especificando o que é o todo ou o que é à parte. Porém, a Terminologia apresenta uma tipologia das relações partitivas em que as possibilidades de coordenação e interseção entre os objetos auxiliam nesta distinção. A relação partitiva para a Teoria da Classificação, é um tipo de relação hierárquica.

Esta relação facilita a organização do repositório tornando possível, por exemplo, considerações sobre os tipos de documentos.

- Relação entre categorias – dá-se a partir da ligação prescritiva entre objetos de natureza distinta. Esta relação pode ser reconhecida por tornar evidente uma determinada demanda, ou função, entre os objetos no mundo fenomenal, não objetivando explicitar o objeto e suas propriedades. É a interseção daquilo que se deseja inferir. Apresentada, na Teoria do Conceito, como relação funcional sintagmática, ou seja, relações que se estabelecem entre categorias, esta relação, é mais flexível que as demais. A Teoria da Terminologia, por sua vez, procura disciplinar este tipo de relação classificando-a como relação ontológica de encadeamento, que relaciona contigüidades no tempo, e relação ontológica de causalidade, que estabelece um elo sucessivo de causas. A Teoria da Classificação Facetada não apresenta esta relação devido a sua estrutura evidentemente hierárquica com cadeias e renques especificando uma temática determinada.

O uso deste tipo de relação deixa a possibilidade de adequação dos documentos de um repositório que não, necessariamente, tem relação com sua temática. É possível, por exemplo, dispor informações meramente institucionais tendo em vista que este conteúdo tem relação com os assuntos já agregados pelas categorias do repositório.

- Relação de equivalência – encontrada na forma de expressar os conceitos, no âmbito da denominação. Faz parte do plano da representação gráfica e é pouco apresentada nos modelos que compõem as relações entre conceitos. Representada, na maioria das vezes, nas teorias do Conceito e da Terminologia, diante da premissa comunicacional das mesmas. Entretanto, na Teoria da Classificação facetada já foi citada por Ranganathan (1967) como pertencente ao plano verbal, irrelevante [ou pouco relevante] na representação de conhecimento.

Aqui, podem ser estabelecidos assuntos referenciados com outro termo ou denominação.

## **5.2 Instrumentos ou métodos para representar a modelagem conceitual**

Existem diferentes formas de organizar o conhecimento resultante da abstração e representação de um domínio específico por meio da modelagem conceitual. A representação possui funcionalidades distintas que demandam instrumentos capazes de contemplar o objetivo de favorecer a organização da informação em RTs. Os esquemas de representação do conhecimento, largamente utilizados, são processos que envolvem um objeto, alguma coisa que o representa e o efeito da representação, na ausência do objeto, na mente de um usuário (PIERCE, 1977). Sendo assim, alguns exemplos de instrumentos e modelos são indicados: Terminologia, Classificação, Tesouro, Taxonomia e Ontologia.

## **6 Metodologia da pesquisa**

No intuito de favorecer a organização de informação para RTs, optou-se por desenvolver uma modelagem conceitual. São observados importantes aspectos para tornar isto viável

e, em decorrência, é definida a metodologia capaz de se adequar ao contexto, prescrevendo etapas para a conclusão bem sucedida do estudo. Desta forma, apresentam-se a seguir, as indicações metodológicas para o andamento e realização da pesquisa:

- Revisão bibliográfica - etapa designada ao levantamento bibliográfico dos assuntos que respaldam o estudo;
- Definição do instrumento ou métodos para representar a modelagem conceitual que, no contexto apresentado, deve possibilitar a representação de domínios de conhecimento estruturado e formalizado, objetivando a modelagem do conhecimento através da definição dos conceitos e de seus relacionamentos;
- Definição de uma arquitetura que contemple a utilização de uma modelagem conceitual em favor da organização da informação em RTs.

## 7 Resultados esperados

Compor um vocabulário de conceitos e relações com os quais os domínios de conhecimento abrangidos por um repositório podem ser formalmente representados, como proposto na modelagem conceitual, é uma proposta para melhorar o desempenho destes repositórios no que diz respeito à organização das informações e a sua interoperabilidade semântica, ou seja, a disposição e interação dos conteúdos armazenados.

Um expectativa paralela relacionada aos resultados da pesquisa é facilitar e aumentar a qualidade dos metadados atribuídos pelo autor. Pois, a construção de uma modelagem conceitual sobre a área abordada por cada RT não apenas contribui para auxiliar a organização das informações dispostas no repositório, guiando a criação de coleções e a especificação de assuntos como também, estando disponível aos autores, pode facilitar a definição de metadados que exponham o assunto do documento a ser depositado. O uso consensual da modelagem desenvolvida permite o entendimento comum por todos aqueles que compartilham informações depositadas no referido repositório, facilitando a avaliação da pertinência de um documento a ser inserido no repositório.

## 8 Conclusões

Como já observado os repositórios institucionais são compostos por repositórios temáticos que por sua vez, dispõem documentos sobre assuntos específicos. O uso de uma modelagem conceitual para estruturar determinado universo do discurso é capaz de favorecer o estabelecimento das áreas cobertas por RTs. O produto final desta modelagem auxilia a localização e o entendimento da disposição de conteúdo e pode ainda, de acordo com a profundidade desta modelagem, auxiliar a definição de metadados descritivos para os documentos inseridos, como já mencionado. Todavia, a definição do instrumento ou métodos para representar a modelagem conceitual, dentre as opções apresentadas, e a elaboração da arquitetura que contemple sua utilização, etapas metodológicas em desenvolvimento, irão permitir atingir tais expectativas.

## 9 Referências

BRODY, Tim; HARNAD, Stevan. **The research impact cycle**. Disponível em: <<http://opcit.eprints.org/feb19oa/harnad-cycle.ppt>>. Acesso em: fev. 2007.

COSTA, Sely M. S. **O novo papel das tecnologias digitais na comunicação científica**. In: SAYAO, L. F. et al. (Org.). **Bibliotecas digitais: saberes e práticas**. Salvador/Brasília: UFBA/IBICT, 2005. p. 167-183.

COSTA, Sely. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. **Ci. Inf.**, Brasília, v. 35, n. 2, p. 39-50, maio/ago 2006.

COSTA; MOREIRA, Ana Cristina S. The diversity of trends, experiences and approaches in electronic publishing: evidences of a paradigm shift on communication. In: COSTA, S. M. S et al. FROM INFORMATION TO KNOWLEDGE: ICC/IFIP INTERNATIONAL CONFERENCE ON ELECTRONIC PUBLISHING, 7., 2003, Portugal. **Proceedings...** Guimarães: Universidade do Minho, 2003. p. 5-9.

CROW, Raym. **The case for institutional repositories**: a SPARC position paper. 2002. 37p. Disponível em: <[www.arl.org/sparc/bm~doc/ir\\_final\\_release\\_102.pdf](http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf)> Acesso em abril de 2007

CAFÉ, Lígia; et al. Repositórios institucionais: nova estratégia para publicação científica na Rede. **INTERCOM** – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação XXVI Congresso Brasileiro de Ciências da Comunicação BH/MG – 2 a 6 Set 2003.

CAMPOS, Maria Luiza Almeida. **Em busca de Princípios comuns na área de representação da informação**: uma comparação entre o Método de Classificação Facetada, o Método de Tesouro-Baseado-em-Conceito e a Teoria Geral da Terminologia. 1994, 196 f. (Dissertação de Mestrado) – Universidade Federal do Rio de Janeiro, Rio de Janeiro / IBICT, 1994.

CAMPOS, Maria Luiza Almeida. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ci. Inf.**, Brasília, v. 33, n. 1, p. 22-32, jan./abril 2004.

CAMPOS, Maria Luiza Almeida. Modelos de abstração: perspectivas para o estudo da área de representação da informação. **Ci. Inf.**, Brasília, v. 25, n. 2, p. 22-32, jan./abril 1995.

CAMPOS, Maria Luiza de Almeida; SOUZA, Rosali Fernandez de; CAMPOS, Maria Luiza Machado. Organização de unidades de conhecimento em hiperdocumentos: o modelo conceitual como espaço comunicacional para a realização da autoria. **Ci. Inf.**, Brasília, v. 32, n. 2, p. 7-16, maio/ago. 2003.

DAY, Michael. The scholarly journal in transition and the PubMed Central proposal. **Ariadne**, v. 21, Sept.1999. Disponível em: <<http://www.ariadne.ac.uk/issue21/pubmed/>>. Acesso em: nov. 2005.

DALBERG, I. Teoria do conceito. **Ci. Inf.**, Rio de Janeiro, v.7, n.2, p.101-7,1978.

FERREIRA, Aurélio Buarque de Holanda. **Novo dicionário Aurélio da língua portuguesa**. 3. ed. rev. e amp. Curitiba: Positivo, 2004. 2120 p.

HUNTER, Philip; DAY, Michael. Institutional repositories, aggregator services and collection development. 2005. Disponível: <<http://eprints-uk.rdn.ac.uk/project/docs/studies/coll-development/coll-development.pdf>> Acesso em: abril de 2007

HARNAD, Stevan et al. The **access/impact problem and the green and gold roads to open access**. 2001. Disponível em: <<http://www.ecs.soton.ac.uk/~harnad/Temp/impact.html>>. Acesso em: fev de 2007.

KURAMOTO, Hélio. Informação científica: proposta de um novo modelo para o Brasil. **Ci. Inf.**, Brasília, v. 35, n. 2, p. 91-102, maio/ago 2006.

LAGOZE, Carl; SOMPEL, Herbert van den. The Santa Fe Convention of the Open Archives Initiative. **D-Lib Magazine**, v. 6, n. 2, fev. de 2000. Disponível em: <<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>> Acesso em: nov. 2005.

LARA, Marilda Lopez Ginez de. Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária. **Ci. Inf.**, Brasília, v. 33, n. 2, p. 91-96, maio/ago. 2004

MARCONDES, C. H.; SAYÃO, L. F. Integração e interoperabilidade no acesso a recursos informacionais em C&T: a proposta da Biblioteca Digital Brasileira. **Ci. Inf.**, Brasília, v. 30, n. 3, p. 24-33, set./dez. 2001.

NEVES, Teodora Marly Gama das. Livre acesso à publicação acadêmica. **Ci. Inf.**, Brasília, v. 33, n. 3, p.116-121, set./dez. 2004

PIERCE, Charles. *Semiótica*. São Paulo: Perspectiva, 1977.

RANGANATHAN, S. R. **Prolegomena to Library Classification**. 4th ed. Bombay: Asia Publishing House, 1967.

SENA, N.K. Open archives: caminho alternativo para a comunicação científica. **Ci. Inf.**, Brasília, v.29, n.3, p.71-78, set./dez. 2000.

SOMPEL, H.v.; LAGOZE, C. The Santa Fe convention of the open archives initiative. **D-Lib Magazine**, v.6, n.2, Feb. 2000. Disponível em : <<http://www.dlib.org/dlib/february00/vandesompeloai/vandesompel-oai.html>>. Acesso em: maio 2007.

SUBER, Peter. **Open Access Overview**: Focusing on open access to peer-reviewed research articles and their preprints. Disponível em: <<http://www.earlham.edu/~peters/fos/overview.htm>> Acesso em: fev. de 2007.

TRISTÃO, Ana Maria Delazari; FACHIN, Gleisy Regina Bóries; ALARCON, Orestes Estevam. Sistema de classificação facetada e tesauros: instrumentos para organização do conhecimento. **Ci. Inf.**, Brasília, v. 33, n. 2, p. 161-171, maio / ago. 2004.

---

<sup>i</sup> Do inglês, *open archives*.

<sup>ii</sup> Documentos institucionais, *pré-prints*, relatórios de pesquisa, entre outros.

<sup>iii</sup> Do inglês, *open accsses*.

---

<sup>iv</sup> Pesquisadores, Editores, Agencias de fomento, Unidades de informação, todos aqueles que dependem do acesso ao conhecimento científico, incluindo os bibliotecários, entre outros (COSTA, 2006).

<sup>v</sup> [http://www.openarchives.org/sfc/sfc\\_entry.htm](http://www.openarchives.org/sfc/sfc_entry.htm)

<sup>vi</sup> Para este estudo, o conceito de representação de conhecimento pode ser extraído de Davis et alii (1992, apud CAMPOS, 2004, p. 24): “Uma representação de conhecimento é uma teoria fragmentada de raciocínio que especifica que inferências são válidas e quais são recomendadas. [...] É motivada por alguma percepção de como as pessoas argumentam ou por alguma crença sobre o que significa raciocinar de forma inteligente [...]”.

<sup>vii</sup> A exemplo, dicionário, tesouro, taxonomia e ontologia (TRISTÃO; FACHIN; ALARCON, 2004; CAMPOS, 1995).