

# A UTILIZAÇÃO DE MAPAS DE TÓPICOS NA COMPATIBILIZAÇÃO DE CONTEÚDOS HIPERTEXTUAIS SEMANTICAMENTE ESTRUTURADOS

Guilherme Baião\*

Gercina Ângela Borém de Oliveira Lima\*\*

## RESUMO

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 facetadas para a representação do conteúdo e melhoria da recuperação da informação durante a navegação. O objetivo principal foi avaliar a possibilidade de aplicação da tecnologia de mapas de tópicos na automatização do processo de compatibilização dessas estruturas. Para isso, foram utilizadas como amostra duas teses do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais que já estavam devidamente analisadas e estruturadas na base de dados do protótipo MHTX. As estruturas facetadas dessas teses, até então representadas em mapas conceituais, foram convertidas em mapas de tópicos. Dessa forma, tornou-se possível a utilização da propriedade de fusão<sup>1</sup> dos mapas de tópicos para promover o inter-relacionamento semântico entre os mapas e, conseqüentemente, entre os recursos de informação hipertextual propriamente ditos. Os resultados da fusão foram analisados à luz das teorias de compatibilização de linguagens desenvolvidas, no âmbito da Ciência da Informação e da Biblioteconomia, a partir da década de 60. Os principais objetivos alcançados foram: (a) a conceitualização detalhada do processo de fusão dos mapas de tópicos, considerando os níveis de compatibilização possíveis e a aplicabilidade dessa tecnologia na integração de estruturas facetadas; (b) a obtenção de uma seqüência detalhada de passos que pode ser utilizada para a implementação de mapas de tópicos a partir de estruturas facetadas.

**Palavras-chave:** Hipertexto. Representação da Informação. Navegação. Mapas de Tópicos. Compatibilização de Vocabulários Controlados.

## 1 INTRODUÇÃO

A crescente utilização dos sistemas de hipertexto através da *Web* desencadeou várias mudanças nos conceitos de organização, armazenamento e recuperação da informação. Com sua estrutura não linear associativa, os documentos hipertextuais apresentam inúmeras inovações em relação aos documentos tradicionais impressos, sobretudo nos processos de pesquisa e aprendizado, uma vez que a leitura hipertextual baseia-se, predominantemente, no processo de navegação.

---

<sup>1</sup> O termo escolhido para tradução do original em inglês “*merge*”.

Em contrapartida, os sistemas de hipertexto apresentam algumas desvantagens a seus usuários, dentre as quais cabe destacar a sobrecarga cognitiva e a sua desorientação durante a navegação.

Com o intuito de reduzir esses problemas, vários sistemas de navegação estruturada em hipertextos estão sendo propostos. Esses novos modelos baseiam-se na aplicação de técnicas e metodologias para organização da informação na rede semântica hipertextual e utilização de interfaces que permitam uma navegação orientada, com base na possibilidade de visualização da estrutura global da rede em diversas formas, tais como diagramas hierárquicos e mapas conceituais. Alguns projetos, como o Mapa Hipertextual (MHTX) proposto por Lima (2004), sugerem a aplicação da Teoria da Análise Facetada (TAF) na análise de documentos tradicionais e conversão dos mesmos em documentos hipertextuais para posterior disponibilização em uma biblioteca digital de teses e dissertações.

Entretanto, quando considerada a coexistência, nesses repositórios, de documentos hipertextuais com estruturas semânticas heterogêneas, a identificação de inter-relacionamentos entre as mesmas torna-se um processo bastante complexo.

Assim, a situação problemática a ser tratada neste estudo é caracterizada pela existência de estruturas conceituais distintas para representação de diferentes documentos hipertextuais em um mesmo repositório (uma biblioteca digital, por exemplo), e que poderão abordar conceitos correlatos entre si, ou até mesmo abordar o mesmo conceito através de termos e contextos diferentes. Por esse motivo, a criação de mecanismos que possibilitem o inter-relacionamento entre essas estruturas tende a facilitar a recuperação de informações por parte dos usuários.

Considerando que a análise de assunto de cada um desses documentos foi realizada com base na TAF e com respaldo de garantia literária em relação aos documentos analisados, ao final desse processo teremos diferentes vocabulários controlados sendo utilizados para organização dos documentos. Essas diferenças ocorrem principalmente devido à possibilidade de designação de termos diferentes para um mesmo conceito e à possibilidade de um mesmo conceito estar associado a outros de formas distintas em cada uma dessas estruturas.

A proposta deste estudo foi avaliar, à luz de teorias de compatibilização de linguagens e vocabulários, o padrão proposto nos mapas de tópicos como ferramenta para automatização da integração de estruturas conceituais facetadas heterogêneas de representação e recuperação de informações em conteúdos hipertextuais. Esses mapas podem ser utilizados para representação da informação a partir de instrumentos das áreas de Ciência da Informação e Biblioteconomia, tais como tesouros, taxonomias, mapas conceituais e classificações

facetadas (GARSHOL, 2004) e são considerados ferramenta eficiente para integração de diferentes estruturas conceituais (PARK, 2003; PEPPER, 2000).

## 2 OS MAPAS DE TÓPICOS

Os mapas de tópicos, ou *Topic Maps*, formam um padrão que permite a implementação de instrumentos de representação da informação, utilizado para descrever e navegar por objetos informacionais em sistemas digitais de forma contextualizada.

De acordo com Park (2003), os mapas de tópicos e a organização do conhecimento são parceiros naturais, pois a invenção do primeiro foi direcionada para atender aos requisitos e problemas do segundo, sobretudo no que diz respeito à integração semântica (ou compatibilização de esquemas de assunto heterogêneos) e construção compartilhada de conceitualizações (ou modelos mentais). Para esse mesmo autor, a construção de mapas de tópicos deve sempre ser embasada nas teorias, técnicas e metodologias desenvolvidas pela organização do conhecimento para a construção de estruturas intelectuais, tais como a classificação facetada.

Assim como em outros instrumentos de representação da informação (como os tesouros e os mapas conceituais, por exemplo), os mapas de tópicos baseiam-se na representação de conceitos (tópicos) através de relações (associações). A partir desse mapeamento de conceitos e associações são construídos “caminhos” que apontam para os recursos de informação relevantes para esses conceitos. Este paradigma requer que o autor do mapa de tópicos pense em termos de tópicos (assuntos, tópicos de conversação, noções específicas, idéias ou conceitos) e associe vários tipos de informação a um tópico específico.

Librelotto (2005) aponta como principais objetivos dos mapas de tópicos: (a) estruturar recursos de informação através de mecanismos externos a estes recursos, (b) permitir buscas que recuperem a informação desejada e (c) criar visões diferentes para usuários ou aplicações específicas, através de filtros sobre as informações.

Segundo Ahmed e Moore (2006), a separação entre a estrutura conceitual e os recursos indexados apresenta três grandes benefícios: (a) o mapa serve como uma representação, em mais alto nível, das informações indexadas, facilitando a localização de recursos e a estruturação conceitual dos recursos que estão sendo representados; (b) facilidade de fragmentação de mapas de tópicos para atender a diferentes necessidades; (c) facilidade de combinação entre mapas de tópicos que indexam diferentes conjuntos de recursos.

O padrão de mapas de tópicos (“*topic maps*”) foi concebido inicialmente no começo da década de 90. Desde então, diversas pesquisas têm sido desenvolvidas com o intuito de

criar normas e padrões para sua utilização, bem como avaliar a sua aplicabilidade. No final da década de 90, o grupo ISO/IEC tratou de criar, sob a identificação ISO 13250, um padrão internacional para formalização das estruturas de conhecimento representadas por essa tecnologia. Logo após publicação do padrão ISO 13250, começaram os estudos para sua adaptação à linguagem XML (*eXtensible Markup Language*) que já despontava, em termos de aceitação, como linguagem para comunicação entre sistemas de informação na *Web* (PARK e HUNTING, 2003). Nessa mesma época foi fundada a organização independente *TopicMaps.org*, cujo propósito era criar uma especificação XML para mapas de tópicos. Nesse grupo, destacaram-se nomes como Graham Moore, Steve Pepper e Eric Freese. No início de 2001 surge, então, o XTM 1.0 (*XML Topic Maps*).

Segundo Pepper (2000), o padrão de Mapa de Tópicos foi criado baseando-se em três componentes principais (também presentes nos índices bibliográficos): Tópicos, Associações e Ocorrências. Além destes três principais, esse mesmo autor destaca outros elementos importantes nos mapas de tópicos: a identidade do conceito (*subject identity*), as facetas (*facets*) e o escopo (*scope*). Pela composição dos mnemônicos desses componentes, o autor chamou o primeiro grupo de TAO e o segundo de IFS. Cada um destes conceitos será comentado a seguir.

Um tópico é a representação de um conceito. No seu sentido mais genérico, um conceito pode ser qualquer coisa: uma pessoa, um conceito, uma entidade, etc (PEPPER, 2000). Em uma representação ideal, a relação entre tópicos e conceitos é do tipo “um-para-um”, ou seja, cada tópico representa um único conceito e cada conceito é representado por um único tópico.

Dentro de um mapa de tópicos, qualquer tópico é uma instância de um ou mais “tipos de tópicos” (*topic types*). Por exemplo, pode-se dizer que o tópico “indexador” é do tipo “Pessoa” (sendo o tipo “Pessoa” um tópico em si), que os tópicos “análise de assunto” e “tradução” são do tipo “Processo” e “livro” é um tópico do tipo “material”.

Para cada tópico do mapa, é possível definir um ou mais nomes (“*topic names*”). Librelotto (2005) afirma que a possibilidade de atribuir vários nomes a um tópico pode ter duas finalidades: facilitar o entendimento do significado do tópico através de descrições alternativas e indicar o uso de nomes diferentes em contextos diferentes, como idioma, domínio, área geográfica, período histórico, etc. Segundo Park (2003), a utilização de múltiplos nomes para um mesmo tópico é um requisito essencial para a robustez, escalabilidade e interoperabilidade dos mapas.

Em um mapa de tópicos, as associações têm a função de expressar um relacionamento entre um ou mais tópicos. Uma associação de tópicos é, formalmente, um elemento de ligação que descreve as relações entre dois ou mais tópicos (PEPPER, 2000) e os papéis que cada um deles exerce nesta associação. As associações entre os tópicos também podem ser agrupadas de acordo com seus tipos (*association types*), que por sua vez também são definidos através de tópicos. Ainda nessas associações, conforme citado em Pepper (2000), podemos definir também o papel de cada tópico. Para isso utiliza-se o elemento *association roles*. Por exemplo, quando existe a necessidade de se representar, em um mapa de tópicos, uma relação do tipo todo-parte entre os processos de análise de assunto e extração de conceitos, é possível definir não só o tipo da relação (todo-parte), mas também definir que o todo é a análise de assunto e a parte é a extração de conceitos.

Ao contrário do que acontece nos tesouros e em outros esquemas de classificação tradicionais, nos mapas de tópicos é possível representar qualquer tipo de relação entre os conceitos. A relação do tipo classe-instância é estabelecida por meio da tipificação de tópicos, associações ou ocorrências (PEPPER, 2000). Para os demais tipos de relações, novas associações devem ser definidas.

Nos mapas de tópicos, uma ocorrência é qualquer informação que, de alguma forma, é especificada como sendo relevante para um determinado tópico (conceito). Ou seja, as ocorrências permitem que um tópico seja relacionado a qualquer número de recursos considerados relevantes a ele. Para Pepper (2001), as ocorrências podem ser externas ao tópico (uma página da *Web* sobre aquele tópico, por exemplo) ou interna a ele (definida no próprio mapa). Assim como os tópicos e as associações, as ocorrências também podem ser classificadas por meio de tipos (*occurrence type*). Dessa forma, é possível, por exemplo, distinguir as ocorrências que definem o tópico das ocorrências que exemplificam ou instanciam aquele tópico.

Conforme citado anteriormente, no segundo grupo de componentes dos mapas de tópicos definido por Pepper (2000) temos o elemento responsável pela identificação dos conceitos, facetas e escopo.

O primeiro deles, o identificador de conceitos, tem como objetivo garantir que um conceito será representado por um único tópico no mapa. Para isso, pode-se identificar o conceito em um mapa de tópicos através de: (a) recursos eletrônicos já existentes e endereçáveis, como uma imagem em uma página da *Web*, por exemplo, ou (b) um indicador de conceito criado e disponibilizado para esta finalidade específica, uma vez que o conceito em questão não é endereçável, como um conceito “Brasil”, por exemplo. Nesse segundo caso,

o recurso criado por uma comunidade com o objetivo específico de fornecer uma identificação única para um conceito é chamado de indicador de conceito publicado (*published subject indicator*, ou simplesmente PSI). O site da comunidade Topicmaps.org<sup>2</sup>, por exemplo, define uma série de PSIs referentes a conceitos não endereçáveis eletronicamente, tais como idiomas, países e os próprios conceitos envolvidos na construção de mapas de tópicos.

O segundo componente, as facetas, são, na verdade, metadados utilizados para descrever ocorrências de um mapa de tópicos. Esse recurso foi adotado no padrão ISO 13250 e extinto no padrão XTM, pois, neste último, o recurso e o valor da propriedade devem ser considerados tópicos associados através de associações do tipo “aplica-se a” (PARK, 2003).

O escopo é utilizado em mapas de tópicos para definir o contexto no qual as características de um tópico são válidas, removendo ambigüidades e reduzindo a chance de erros na fusão de mapas de tópicos (PEPPER, 2001). Entende-se como características de um tópico os seus nomes, suas ocorrências e as associações das quais participa. Se nenhum escopo é atribuído ao nome, à ocorrência, ou à associação, significa que todas essas características do tópico são válidas em qualquer situação.

O processo de fusão (*merge*) é um recurso disponibilizado nos mapas de tópicos para integrar dois mapas distintos em um único, de forma a garantir que tópicos que representam o mesmo conceito em mapas distintos serão transformados em um único tópico no mapa resultante. Quando essa fusão ocorre, as características dos tópicos fundidos também devem ser unidas, eliminando duplicidades, e mantidas no tópico resultante (nomes, ocorrências, associações e identificadores).

Esse processo pode ser automatizado em aplicativos específicos para manipulação de mapas de tópicos. Nesses aplicativos, para determinar que dois tópicos representam o mesmo conceito, utiliza-se duas possibilidades distintas. A primeira ocorre quando o identificador único do conceito é o mesmo para dois tópicos distintos, quer o identificador seja um objeto endereçável, quer seja um PSI. O segundo ocorre quando um ou mais nomes de um tópico coincidem com um ou mais nomes de outro.

Para este segundo caso é preciso considerar o problema da homonímia, ou seja, quando um mesmo termo pode representar dois ou mais conceitos diferentes. Para tratar o problema os aplicativos podem utilizar dois caminhos distintos: na primeira o aplicativo apresenta ao usuário as propriedades dos tópicos que foram identificados com mesmo nome

---

<sup>2</sup> Disponível em: <<http://www.topicmaps.org/xtm/>>.

para que ele possa decidir se a fusão deve ou não acontecer. Nesse caso o processo deixa de ser totalmente automatizado. No segundo, o aplicativo considera a equivalência dos nomes somente se estes estiverem definidos para um mesmo escopo, fazendo com que o processo seja totalmente automatizado.

Uma outra possibilidade de fusão entre os elementos dos mapas é a utilização do componente *mergemap* para especificar os tópicos que deverão ser considerados como um único. O *mergemap* referencia um mapa externo através de uma URI, permitindo indicar, nos tópicos de um mapa, tópicos correspondentes em outros mapas.

Como alguns mapas de tópicos podem conter uma grande quantidade de tópicos e associações, atualmente é possível encontrarmos ferramentas que permitem a visualização dos mapas de tópicos através de diferentes interfaces: textuais, geralmente disponibilizadas através de páginas HTML, grafos, árvores ou mapas de diversos tipos, inclusive tridimensionais.

Os estudos mais recentes acerca dos mapas de tópicos estão concentrados na implementação de recursos que viabilizem a definição de restrições em relacionamentos<sup>3</sup> e regras de inferências.

### **3 COMPATIBILIZAÇÃO DE LINGUAGENS E VOCABULÁRIOS CONTROLADOS**

Os estudos acerca da compatibilização de linguagens tiveram início na década de 60. Dahlberg (1981) destaca nessa década os trabalhos desenvolvidos por William Hammond e Staffan Rosenborg, Simon M. Newman, Madeline M. Henderson, John S. Moats e Mary Elizabeth Stevens. Um dos principais resultados alcançados pelos estudos dessa época foi a definição de alguns conceitos centrais para a área, tais como convertibilidade, compatibilidade e suas respectivas aplicações, sobretudo no âmbito da cooperação entre centros de informação. William Hammond, em 1965, definiu compatibilidade como: “a capacidade de um sistema de informação aceitar dados da indexação e resumo de outro sistema sobre qualquer assunto que seja comum a ambos.” (NEWMAN, 1965, p.7).

Coates (1970) comenta que até então a cooperação entre bibliotecas era baseada somente na união de catálogos, ou seja, não havia mecanismos eficientes que permitissem o intercâmbio de descrições de assunto. Porém, segundo Maniez (1997), foi na década de 70 que os estudos sobre compatibilização de linguagens tiveram uma maior produção. Como autores relevantes dessa década, Dahlberg (1981) cita, dentre outros, Hans Wellisch, Verina

---

<sup>3</sup> Alguns exemplos dessas restrições são: “Toda pessoa é nascida em algum lugar” (considerando um mapa que aborda os tópicos “pessoa” e “lugar”); “Um animal é sempre filho de outro animal”.

Horsnell, Wersig, Eric Coates, Elaine Svenonius, Dagobert Soergel, Linda C. Smith e Glushkov. Além desses autores, Lancaster (1986) aponta os trabalhos de H. Neville e R. T. Niehoff.

Em 1971 o relatório UNISIST da Unesco dedicou uma seção para tratar da conversão e compatibilização de instrumentos de indexação. Nele foram apresentadas definições de compatibilização e conversão:

O relatório define ‘compatibilidade’ como a qualidade dos sistemas cujos produtos podem ser usados de maneira intercambiável, apesar das diferenças na notação, estrutura, suporte físico, etc., sem nenhuma conversão especial automatizada e ‘conversão’ como o processo de transformar registros de informação, baseado na transcrição de códigos, estruturas de dados, etc., fazendo com que os mesmos sejam intercambiáveis entre dois ou mais serviços ou sistemas que utilizam diferentes convenções e mídias (DAHLBERG, 1981, p.86).

Na década seguinte, a de 80, a temática foi pouco estudada. Naquela época as tecnologias que permitiam o intercâmbio de dados entre bases diferentes eram bastante restritas. Já na década de 90 os estudos foram retomados em virtude da expansão global das redes de informação, da facilidade de acesso simultâneo a coleções distintas e do aumento do número de intercâmbio entre bases de dados diferentes, sobretudo via *Web* (MANIEZ, 1997). Nesses estudos, novas definições de compatibilidade são apresentadas e, dentre elas, destaca-se a definição dada por Gerhard Riesthuis em um seminário sobre compatibilização e integração de sistemas que ocorreu em 1995, na cidade de Varsóvia: compatibilidade significa que, para cada termo de uma linguagem, existe um termo correspondente, de mesmo significado, na outra linguagem, e por isso é possível converter termos de uma linguagem em outra, sem alterações de significado (MANIEZ, 1997).

Na literatura brasileira, a partir do início da década atual, destacam-se os estudos da Dra. Maria Luiza de Almeida Campos. Campos (2005) destaca como principais teóricos da compatibilização de linguagens na Ciência da Informação: Soergel (1982), Dahlberg (1981), Neville (1970) e Glushkov, Skorokhod’ko e Strongnii (1978).

Segundo Zhang (2006) o objetivo da compatibilização de linguagens de indexação é permitir que os usuários, através de uma única estratégia de busca, recuperem informações armazenadas em qualquer uma das bases ou repositórios que compõem um determinado SRI. Para Rada (1987), a integração<sup>4</sup> de linguagens de indexação controladas pode atender a

---

<sup>4</sup> Segundo Maniez (1997) as palavras “integração”, “harmonização”, “reconciliação” e “concordância” costumam ser utilizadas para indicar o conceito de convergência, ou seja, mesmo conceito de “compatibilização”. Porém, Lancaster (1986) considera a integração de vocabulários como um método específico de compatibilização, no qual não existe a análise conceitual, ou seja, para uma determinada

diferentes propósitos: (a) permitir que o usuário utilize um termo para um determinado conceito e, a partir dele, sejam indicados os descritores apropriados em cada base, (b) permitir o compartilhamento do processo de indexação através da possibilidade de conversão de um tesauro em outro, (c) possibilitar a identificação de relações entre conceitos de vocabulários controlados diferentes e (d) incrementar o conteúdo do vocabulário. Em 1965, William Hammond havia listado os seguintes objetivos da compatibilização: (a) possibilitar a disseminação de informações resultantes de pesquisas em andamento; (b) eliminar a necessidade de duplicação de indexação e resumo de relatórios de pesquisa e (c) prover a recuperação de relatórios armazenados em diferentes repositórios, utilizando como base a indexação original (NEWMAN, 1965).

Segundo Soergel (1982), os principais obstáculos encontrados no processo de compatibilização de vocabulários controlados são: (a) diferentes níveis de pré-combinação dos descritores utilizados em cada um dos sistemas, (b) diferente grau de especificidade dos conceitos que compõem cada um dos sistemas, (c) diferentes formas de se relacionar um termo específico a um único termo genérico no caso de poli-hierarquias<sup>5</sup> (d) diferentes possibilidades de agregação de conceitos específicos formados por termos que, de forma isolada, não representam conceitos, (e) a possibilidade de conotações diferentes para um mesmo termo (homônimos) e (f) ausência de conceitos em algum dos sistemas, sendo que esse mesmo conceito ocorre em outros sistemas envolvidos. Esse último obstáculo tende a ser maior na medida em que a sobreposição dos assuntos tratados pelos vocabulários envolvidos diminui.

Além desses problemas, Lancaster (1986) comenta problemas relacionados ao uso de processamento computacional na tentativa de se automatizar o processo de compatibilização e destaca, como principal problema, a possibilidade de um mesmo termo possuir diferentes significados. Para o mesmo autor, outro problema é a existências de erros na forma como os termos são escritos em alguns sistemas. Para Doerr (2001) as dificuldades encontradas no processo de compatibilização são originadas, principalmente, na subjetividade da escolha de descritores e relações hierárquicas durante a construção de vocabulários controlados.

A compatibilização de vocabulários controlados pode ocorrer em três níveis: terminológico (ou verbal), conceitual e estrutural (SOERGEL, 1982). No primeiro deles, o

---

requisição de um usuário, todas as suas ocorrências (incluindo variações de palavras) e elementos relacionados são mostrados como resultado, sem explicitar a relação ou o contexto em que cada um aparece.

<sup>5</sup> A poli-hierarquia ocorre quando um termo específico pode estar subordinado a mais de um termo genérico. O termo “química geológica”, por exemplo, pode estar subordinado tanto à química quanto à geologia.

nível terminológico, os procedimentos são baseados na comparação dos nomes ou descritores atribuídos aos conceitos. O segundo, conceitual, pode ser considerado o mais complexo de todos, pois é nesse nível que encontramos quase todos os problemas citados anteriormente. Devido a essa complexidade, ao contrário do que ocorre no nível terminológico, a compatibilização no nível conceitual dificilmente acontece de forma automatizada. Ainda no nível conceitual, métodos para validação da etapa anterior podem ser propostos, visto que homônimos podem gerar uma compatibilização terminológica incorreta do ponto de vista conceitual. Já a compatibilização no nível estrutural é responsável por garantir a conformidade das relações estabelecidas entre os elementos.

Glushkov, Skorokhod'ko e Strongnii (1978), por sua vez, dividiu a compatibilização de vocabulários em dois níveis: o nível semântico, ligado à capacidade de ambos representarem o mesmo corpo de conhecimento e o nível estrutural, referente à similaridade de acordo com características internas, o que corresponde aos níveis terminológico, conceitual e estrutural citados por Soergel (1982).

No modelo de compatibilização proposto por Dahlberg (1981) é possível perceber as três etapas da compatibilização: verbal (matriz de comparação alfabética construída no tópico 4 do estudo da autora), conceitual (tópicos 5.1 a 5.4) e estrutural (5.5 a 5.7).

Encontramos na literatura diferentes técnicas e algoritmos para integração de linguagens, dentre os quais podemos destacar o mapeamento, o uso de linguagens intermediárias, microtesauros, macrotesauros e o tesouro universal. Dentre estes, o mapeamento é o mais encontrado em referências sobre o assunto.

O mapeamento consiste em determinar, para cada termo de um vocabulário, o termo de correspondência mais próxima em outro vocabulário. Segundo Doerr (2001) o mapeamento é o processo básico em todas as demais técnicas de compatibilização.

Quando o número de vocabulários envolvidos aumenta, a situação fica extremamente complexa e uma alternativa mais viável passa a ser a utilização de linguagens intermediárias. Nesse caso, cada vocabulário deverá ser mapeado bidirecionalmente com a linguagem intermediária e, dessa forma, obtém-se a correspondência entre todos os vocabulários envolvidos.

A respeito das principais dificuldades encontradas durante o mapeamento (utilizando ou não linguagens intermediárias), Lancaster (1986) aponta: (a) diferenças entre os níveis de pré-coordenação dos vocabulários envolvidos, (b) diferenças entre níveis de especificidade e (c) ausência de elementos correspondentes em um dos vocabulários. Para as duas primeiras situações, a solução proposta pelo autor é mapear um elemento de um vocabulário para vários

elementos do outro. No último caso, a solução seria acrescentar o elemento em um dos vocabulários. Esta situação ocorre em proporções menores à medida que aumenta a sobreposição de assuntos cobertos por ambos os vocabulários.

Neville (1970) apresenta um exemplo de passo-a-passo para o mapeamento de tesouros, que ele chamou de reconciliação de tesouros. Considerando os níveis terminológico e conceitual, o autor identificou seis tipos de correspondência possíveis entre elementos de vocabulários distintos: (1) correspondência exata, (2) sinônimos, (3) genérico para específico, (4) mapeável a um nível de pré-coordenação diferente, (5) fatoração semântica e (6) antônimos.

Soergel (1974) apresenta uma tabela para mapeamento de vocabulários controlados na qual ele sugere cinco possíveis níveis de equivalência conceitual entre elementos de linguagens distintas e, para cada um desses níveis, cinco possíveis níveis de equivalência terminológica: mesmo termo, termos distintos, combinações de descritores utilizando operadores lógicos “OU”, combinações de descritores utilizando operadores lógicos “E” e combinações de descritores utilizando operadores lógicos “OU” e “E”.

O mapeamento de um vocabulário em outro é uma tarefa intelectual cansativa e demorada (LANCASTER, 1986). Por esse motivo, alguns estudos foram realizados no final da década de 60 e início da década de 70 com o intuito de desenvolver algoritmos que automatizassem pelo menos uma parte do processo. Nesses algoritmos foram identificados alguns tipos de mapeamento possíveis de serem realizados automaticamente.

Lancaster (1986) apresenta, também, um outro tipo específico de técnica para compatibilização de vocabulários que ele chama de abordagem baseada na integração de vocabulários<sup>6</sup>, cujo propósito é um pouco diferente das linguagens intermediárias, pois as equivalências entre os elementos dos vocabulários envolvidos são definidas de acordo com a estratégia de busca do usuário e, conseqüentemente, nenhuma nova linguagem neutra é criada. Nesses casos, o próprio usuário é que indica a estratégia de comparação: equivalência exata, sinônimos, termos relacionados, termos adjacentes, vocabulários a serem considerados, número de termos a serem exibidos como resultado, etc. No resultado da busca são exibidas as ocorrências nas bases de dados selecionadas, cada uma com seus termos específicos e o motivo pelo qual a ocorrência foi considerada. Para Lancaster (1986), a matriz de compatibilidade proposta por Dahlberg (1981) é um outro exemplo de abordagem baseada na integração de vocabulários.

---

<sup>6</sup> Termo traduzido do inglês: *Integrated Vocabulary Approach*.

#### 4 PROCEDIMENTOS ADOTADOS

Foram utilizadas como objeto empírico para o experimento duas estruturas conceituais facetadas. A primeira foi desenvolvida por Lima (2004) durante a implementação do protótipo do MHTX, apresentado em sua tese de doutorado. Trata-se do resultado da análise facetada da tese de doutorado de Madalena Martins Lopes Naves<sup>7</sup>, do ano 2000. Em tal análise, foi utilizado o conjunto de princípios sugerido pelo Modelo Simplificado de Spiteri.

O segundo foi construído pela professora Gercina Ângela Borém de Oliveira Lima, em parceria com a professora Maria Luiza de Almeida Campos<sup>8</sup>, e foi resultado da análise facetada da própria tese de doutorado de Lima (2004). Nesse caso, a categorização dos conceitos e facetas teve como ponto de partida as categorias e subcategorias formais propostas por Dahlberg (1978).

É importante ressaltar que, além das diferenças entre os modelos de categorização adotados na construção de cada um dos dois esquemas, o segundo foi construído desconsiderando o cenário do primeiro, ou seja, em nenhum momento houve preocupação com a compatibilidade entre os dois. Dessa forma, eles podem ser considerados uma amostra ideal para a elaboração do cenário problemático descrito anteriormente.

Segundo Garshol (2004) os três passos para a criação de um mapa de tópicos a partir de um sistema de classificação facetada são: (a) para cada faceta, criar um tópico do tipo “faceta” e dar a ele o nome da faceta; (b) para cada termo de mais alto nível de cada faceta, criar um tópico do tipo “termo”, cujo nome será o próprio termo, e associá-lo à faceta correspondente utilizando uma associação do tipo “pertence à faceta”; (c) para cada termo abaixo do termo de nível mais alto, criar um tópico do tipo “termo”, cujo nome será o próprio termo, e associá-lo a seu pai utilizando uma associação do tipo “genérico/específico”.

Para manter a representatividade dos mapas conceituais desenvolvidos por Lima (2004) e evitar perda semântica na sua estrutura facetada, o roteiro sugerido por Garshol (2004) precisou sofrer adaptações: (a) ao invés de criar um único tópico do tipo “faceta”, foram criados oito tópicos, cada um representando um nível diferente de faceta ou subfaceta, uma vez que no esquema proposto por Lima (2004) a cada nível era atribuído uma cor diferente no mapa conceitual; (b) em cada associação foram definidos também papéis

---

<sup>7</sup> Madalena Martins Lopes Naves é doutora em Ciência da Informação pela Escola de Ciência da Informação da UFMG e professora da área de tratamento e organização da informação na mesma escola.

<sup>8</sup> Ambas graduadas em Biblioteconomia e doutoras em Ciência da Informação.

“roles”) para os tópicos envolvidos. Numa associação do tipo “pertence à faceta”, por exemplo, um dos tópicos envolvidos assume o papel de “faceta” e o outro o papel de “termo”; (c) nos casos em que foram mantidos os termos sinônimos (Conceito/Idéia/Pensamento, por exemplo) utilizou-se a propriedade “topic names” para indicar que todos os termos se referem a um mesmo conceito; (d) para a indicação dos critérios de divisão foram criados tópicos do tipo “critério de divisão” que, em seguida, foram associados à faceta correspondente utilizando uma associação do tipo “divide a faceta”. Na associação “pertence à faceta” entre os termos subordinados a cada um desses critérios com a faceta imediatamente superior, foi necessário adicionar um terceiro papel indicando o critério de divisão, com o objetivo de explicitar quais termos subordinados à faceta pertencem a qual critério. Isto significa que, no caso de subordinações para as quais havia critério de divisão explicitado no esquema original, foi gerada uma relação ternária envolvendo a faceta, a subfaceta (cujo papel foi definido como “termo”) e o critério de divisão em questão.

Em seguida foram criadas as ocorrências para cada um dos tópicos. Cada ocorrência apontando para a página HTML correspondente ao trecho da tese de Naves (2000) que melhor define o seu conteúdo semântico. Dessa forma, todas as ligações entre elementos do mapa e os respectivos recursos de informação que haviam sido criadas por Lima (2004) foram mantidas. Tornou-se possível, então, a navegação em contexto através do mapa de tópicos.

O passo seguinte foi a criação de um repositório de PSIs. Conforme já mencionado anteriormente, os PSIs são os principais responsáveis pela portabilidade dos mapas de tópicos, pois têm como objetivo a identificação única e sem ambigüidades de um conceito. Por esse motivo, apesar de um PSI poder ser representado por qualquer tipo de dado, o ideal é que estejam em um repositório que possa ser disponibilizado na *Web* e que cada um dos elementos seja endereçável através de uma URI específica. Isso permite a reutilização e a construção colaborativa de repositórios de PSIs cada vez mais completos do ponto de vista conceitual e mais abrangentes.

Para a obtenção do segundo mapa de tópico, foram aplicados os mesmos procedimentos, porém sobre uma estrutura facetada distinta, obtida a partir da análise da tese de Lima (2004). Entretanto, nesse segundo mapa, não foram adicionados os PSIs, pois a criação de um PSI para cada novo tópico, assim como foi feito no primeiro mapa, poderia gerar mais de um PSI para um mesmo conceito, caso esse estivesse sendo tratado nos dois mapas. Ou seja, a reutilização dos assuntos já publicados exige uma análise conceitual prévia e manual por parte do profissional que está construindo o mapa de tópicos. Como este estudo

trata da compatibilização por meio de recursos automatizados presentes nos mapas de tópicos, tal análise para inserção dos PSIs no segundo mapa não se fez justificável.

O software utilizado para auxiliar na construção, edição e navegação nos mapas de tópicos foi o Topincs <<http://www.cerny-online.com/topincs/>>, versão 1.3.3. Trata-se de um software gratuito e de código aberto para fins acadêmicos e que permite a exportação dos mapas para arquivos no padrão XTM.

Para visualização gráfica dos mapas de tópicos, a partir do arquivo XTM gerado pelo Topincs, foi utilizado o TMNav, software que permite diferentes formas gráficas de visualização, tais como hierárquica (em árvore) e através de mapa conceitual hiperbólico, que é uma interface bastante indicada para a navegação em recursos hipertextuais.

Vários outros softwares foram analisados antes da seleção destes. Porém alguns apresentaram alto nível de complexidade para instalação e manipulação, exigindo inclusive domínio da linguagem de programação ou arquitetura do código (o TM4J<sup>9</sup>, por exemplo, exige amplo conhecimento em Java para instalação e compilação de seus componentes). Outros, como o TProc<sup>10</sup>, não suportam o padrão XTM, além de possuir como requisito a instalação prévia de vários outros softwares pouco conhecidos.

Na fusão automatizada dos dois mapas de tópicos obtidos nas etapas descritas anteriormente, foi utilizado o *Ontopia OKS Samplers*.

## 5 ANÁLISE DOS RESULTADOS

A análise dos resultados teve como passo inicial a unificação manual das estruturas facetadas envolvidas. A estrutura unificada resultante desse procedimento serviu como parâmetro para a comparação com o mapa de tópico obtido após a fusão automatizada dos dois mapas gerados anteriormente a partir dessas mesmas estruturas.

Para a confecção do esquema unificado manualmente, a estrutura referente à tese de Lima (2004) foi considerada como básica. Em seguida, procurou-se alocar, nesta estrutura básica, cada um dos descritores do segundo esquema, referente à tese de Naves (2000), considerando o conceito ao qual cada um se refere em cada um dos trabalhos.

Durante esse processo, diferentes tipos de relações entre os elementos das duas estruturas puderam ser identificados. Essas relações foram classificadas considerando os níveis três níveis de compatibilização (terminológico, conceitual e estrutural) apresentados

---

<sup>9</sup> <http://www.techquila.com/tm4j.html>

<sup>10</sup> O TProc foi desenvolvido pelo grupo Ontopia (<http://www.ontopia.net:80/software/tmproc/>) e foi um dos primeiros softwares desenvolvidos para manipulação de mapas de tópicos.

por Soergel (1982). O primeiro tipo de relação identificado foi a equivalência terminológica, conceitual e estrutural, ou seja, o mesmo descritor presente nas duas estruturas, denotando o mesmo conceito e em hierarquias semelhantes. Um segundo tipo de relação foi identificado entre elementos de mesma carga conceitual e terminologia com pequenas variações, como, por exemplo, variação de número. Outro tipo foi identificado entre elementos equivalentes sob o ponto de vista conceitual e terminológico, porém apresentado em subordinação diferente nos dois esquemas, o que representa uma divergência no nível estrutural. O quarto tipo é uma variação do anterior, devido a variações terminológicas entre os descritores. O quinto tipo foi identificado entre elementos com descritores diferentes e conceitualmente divergentes apenas em relação à especificidade dos mesmos. O último tipo de relação identificada ocorreu nos casos em que os elementos não possuíam nenhuma das relações anteriores, mas que tiveram a sua colocação definida para fins de compatibilidade estrutural entre os esquemas envolvidos e o esquema resultante.

Para facilitar a identificação dos seis tipos de relação no esquema resultante, utilizou-se um mapeamento de cores para facilitar a identificação visual desses tipos, ou seja, cada elemento do segundo esquema, ao ser inserido no esquema escolhido como base, recebeu uma cor que correspondia ao tipo de relação em questão.

A identificação das relações, além de permitir uma melhor comparação dos resultados, serve como justificativa para a inserção de cada termo do segundo esquema no primeiro. Um exemplo disso é o termo “ciências exatas” que, por ter sido inserido na cor laranja, nos permite deduzir que se trata de uma divergência somente em nível de especificidade. Considerando a cadeia hierárquica da qual o termo passou a fazer parte, pode-se deduzir que se trata de um conceito cujo nível de especificidade é maior do que “Ciências Reais Naturais” e menor que “Ciência da Computação”.

Para facilitar a comparação dos resultados obtidos nos dois processos (a unificação manual dos dois esquemas e a fusão automatizada dos mapas de tópicos), foi necessário navegar por todo o mapa de tópico resultante, utilizando o próprio *OKS Samplers*, a fim de extrair uma estrutura hierárquica no mesmo formato do esquema gerado através da unificação manual.

Neste ponto, detectou-se a existência, no mapa de tópico resultante, de tópicos subordinados a mais de um tópico “pai”, ou seja, a ocorrência de poli-hierarquias. Isso ocorreu em todos os casos nos quais dois descritores idênticos eram tratados em níveis diferentes nos dois esquemas originais. Um exemplo dessa duplicidade de subordinações é o elemento “indexador” e toda a estrutura a ele subordinada. No mapa resultante, devido à

ocorrência aninhada de poli-hierarquias, esses elementos puderam ser acessados a partir de quatro caminhos distintos:

- > *Entidade > Profissionais > Profissional da Informação > Bibliotecário > Indexador*
- > *Entidade > Profissionais > Profissional da Informação > Indexador*
- > *Entidade > Profissional da Informação > Bibliotecário > Indexador*
- > *Entidade > Profissional da Informação > Indexador*

A ocorrência de poli-hierarquias evidenciou que, a nível estrutural, a fusão dos mapas faz uma soma ou união, isto é, para os casos em que o mesmo descritor ocorre nos dois mapas, as duas hierarquias nas quais ocorrem são mantidas após a fusão.

Comparando os dois esquemas, observou-se que as variações terminológicas, mesmo quando pequenas, influenciaram negativamente no processo de fusão dos mapas, pois uma das características dos tópicos utilizada para identificação de equivalências são os próprios nomes a eles atribuídos. Os tópicos “autores” e “autor”, por exemplo, que nos dois mapas representavam o mesmo conceito não foram unificados. Já os tópicos que participaram de relações de equivalência terminológica e conceitual foram devidamente unificados.

É relevante comentar que nesses casos em que houve a unificação de dois tópicos em um único, todas as propriedades dos dois, tais como nomes, associações e, principalmente, ocorrências, foram mantidas no tópico resultante. Dessa forma, no momento em que o usuário, ao navegar pelo mapa, atingia algum tópico resultante da unificação de dois tópicos, os dois recursos de informação relevantes eram exibidos.

Os casos de divergência terminológica, divergência conceitual somente em relação ao nível de especificidade e equivalência estrutural, também não receberam nenhum tipo de tratamento no processo automatizado, sendo apresentados no esquema resultante como tópicos não relacionados. Por fim, os casos de equivalência simplesmente estrutural, representados na unificação manual pela cor vermelha, foram tratados devidamente no mapa de tópico resultante da fusão, graças, principalmente, à compatibilidade terminológica entre as grandes categorias tratadas nos dois esquemas.

## **6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS**

A fusão dos mapas de tópicos pode ser definida, no âmbito da interoperabilidade entre vocabulários, como uma compatibilização terminológica automatizada, baseada nos nomes dos tópicos, ou seja, dois tópicos em mapas distintos que possuam um ou mais descritores comuns, são unificados em um único tópico no mapa resultante. Nesse tópico fica depositada

a soma ou união das características dos tópicos originais. Eventuais variações na utilização dos descritores podem influenciar negativamente no resultado do processo, mesmo quando se trata de pequenas variações, tais como divergência na utilização do descritor no singular ou plural. Alguns estudos futuros poderão propor a incorporação de algoritmos de casamento de padrões e verificações ortográficas às ferramentas de manipulação de mapas de tópicos, permitindo o tratamento de pequenas variações terminológicas ou erros de ortografia.

A compatibilização estrutural durante a fusão ocorre como consequência de eventuais unificações no nível terminológico.

Já no nível conceitual, a proposta dos mapas de tópicos é utilizar o conceito de PSIs. Porém, a identificação de equivalências entre um novo tópico e PSIs já existentes é um processo completamente manual e que exige intenso esforço intelectual, a cargo do usuário responsável pelo novo mapa. Ou seja, a utilização de PSIs, que é fundamental para que os mapas de tópicos atinjam seu objetivo de interoperabilidade, é semelhante ao processo de utilização de linguagens intermediárias para compatibilização de vocabulários.

Os estudos acerca da compatibilização de vocabulários podem contribuir no desenvolvimento de novas técnicas de interoperabilidade entre mapas de tópicos. A grande maioria desses estudos, desenvolvidos a partir de década de 60, foi voltada para a sua aplicação em tesouros. Mais recentemente é que estes estudos estão sendo revisitados com o objetivo de integração de outras formas de representação de informações, como as ontologias, por exemplo.

Os mapas de tópicos gerados a partir de esquemas facetados são extremamente simples do ponto de vista dos recursos que os mapas de tópicos oferecem (infinitos tipos de associações, nomes de tópicos, escopo, etc), devido a uma fragilidade de explicitação semântica desses esquemas: tipos de relações limitados, escopo e cardinalidade não são considerados. Além disso, a poli-hierarquia não é permitida, pois os princípios defendidos por Ranganathan estabelecem que as classes formadoras dos renques devem ser mutuamente exclusivas, ou seja, nenhum conceito da estrutura pode pertencer a mais de uma classe no renque. Estudos futuros poderão propor novos modelos de análise de assunto baseada na TAF, que sirvam como base para a construção de mapas de tópicos mais completos sob o ponto de vista semântico e mais preparados para a compatibilização, através da utilização de escopos, por exemplo. Novas propostas de conversão de esquemas de classificação tradicionais da CI em mapas de tópicos poderão ser desenvolvidas, utilizando tipos de tópicos, associações e papéis diferentes dos que foram utilizados neste trabalho.

A ocorrência de poli-hierarquias no mapa resultante da fusão automatizada mostrou-se bastante eficiente no sentido de mostrar diferentes definições conceituais para um mesmo termo, o que é claramente passível de ocorrer quando trabalhamos com recursos de informação de diferentes autorias.

Pelos mesmos motivos, devemos ressaltar a importância da possibilidade de atribuição de vários nomes ao mesmo tópico. Uma vez explicitadas tais equivalências, mesmo quando um conceito qualquer é tratado por diferentes autores através de termos distintos, a garantia literária é mantida. No caso específico de sistemas que priorizam a navegação em contexto, em detrimento de outras técnicas de busca, a garantia literária assume papel de maior importância, do que se comparado a sistemas que priorizam outras técnicas de busca. No caso da busca textual, por exemplo, a garantia de uso passa a ser essencial.

Com relação aos objetivos dos mapas de tópicos de permitir uma navegação contextualizada e oferecer maior flexibilidade para representação de conceitos e relacionamentos, pode-se dizer que a tecnologia cumpre bem o seu papel. Permite diversas interfaces de navegação (textual, hierárquica, em diversas formas de mapas conceituais, ou em diversas outras formas gráficas), sendo que a maioria delas permite ao usuário acessar, de forma eficiente, a localização de um nó qualquer do hipertexto em relação à sua estrutura global. Tal eficiência está fundamentada no mapeamento direto entre a estrutura representada pelo mapa e as fontes de informação às quais essa estrutura se refere. A possibilidade de se estabelecer limites para aplicações de determinados conceitos (escopo) e a possibilidade de representação de um mesmo conceito em diferentes hierarquias (poli-hierarquia) no mesmo mapa contribuem ainda mais para a sua flexibilidade de representação.

No que diz respeito à integração semântica de esquemas heterogêneos e construção compartilhada de conceitualizações, que também são considerados objetivos dos mapas de tópicos, observamos a confirmação do que alguns estudiosos da compatibilização de linguagens e vocabulários já haviam observado: a complexidade da compatibilização no nível conceitual, exigindo um alto nível de interferência manual. Sob esse ponto de vista, o padrão de mapas de tópicos passa a desempenhar o papel de simples suporte para as análises e decisões humanas, pois a automatização do processo concentra-se no nível terminológico.

Com a evolução do padrão XTM, o desenvolvimento de linguagens de consulta específicas e a possibilidade de implementar regras de inferência, novos trabalhos poderão ser realizados buscando reavaliar o nível de automatização dos procedimentos de fusão.

## USE OF MAPS TO START THE SUITABILITY OF CONTENTS SEMANTICALLY STRUCTURED HYPERTEXT

### ABSTRACT

Considering the characteristics of hypertext systems and the problems of disorientation and cognitive overload of users of these systems, this project addresses a subject of study hypertext documents that passed through the conceptual design process using faceted structures for the representation of content and improved retrieval of information during navigation. The main objective was to evaluate the possibility of applying the technology of topic maps in automating the process of reconciling these structures. The main objectives were achieved: (a) the conceptualization about the process of merging of topic maps, considering the levels of compliance and the possible applicability of this technology in the integration of faceted structures, (b) to obtain a detailed sequence of steps that can be used for the implementation of topic maps from faceted structures.

**Palavras-chave:** Hipertexto. Representação da Informação. Navegação. Mapas de Tópicos. Compatibilização de Vocabulários Controlados.

### REFERÊNCIAS

AHMED, Kal; MOORE, Graham. Uma Introdução aos Mapas de Tópicos. *Architecture Journal*, [revista on-line], 15 mar. 2006. Disponível em: <<http://msdn.microsoft.com/pt-br/library/aa480048.aspx>>. Acesso em: 20 jun. 2007.

CAMPOS, Maria Luiza Almeida. **A problemática da compatibilização terminológica e a integração de ontologias:** o papel das definições conceituais. In: ENCONTRO NACIONAL DE PESQUISA EM CIENCIA DA INFORMACAO - ENANCIB, 6., 28-30 nov. 2005, Florianópolis. Anais do VI ENANCIB... Florianópolis: Universidade Federal de Santa Catarina, 2005. 12 p.

COATES, E. J. Switching languages for indexing. *Journal of Documentation*, v. 26, n. 2, p. 102-110. 1970.

DAHLBERG, Ingetraut. Uma teoria para o interconcept: teoria analítica do conceito voltada para o referente. Publicado originalmente na revista **International Classification**, v. 5, n. 3, p. 142-151, 1978. [Traduzido por Vânia Teixeira Gonçalves, Bolsista do CNPq, da equipe de Hagar. E. Gomes, Rio de Janeiro, 1990. 34 p.]

DAHLBERG, I. Toward establishment of compatibility between indexing languages. **International Classification**, v. 8, n. 2, p. 86-91, 1981.

DOERR, M. Semantic problems of thesaurus mapping. **Journal of Digital Information**, v. 1, n. 8, art. 52, Mar. 2001. Disponível em: <<http://jodi.tamu.edu/Articles/v01/i08/Doerr/>>. Acesso em: 3 abr. 2007.

GARSHOL, L. M.; Metadata? thesauri? taxonomies? topic maps!. **Journal of Information Science**, v. 30, n. 4, p. 378-391, 2004. Disponível em: <<http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>> Acesso em: 5 jan. 2006.

GLUSHKOV, V.M.; SKOROKHOD'KO, E.F.; STRONGNIL, A.A. Evaluation of the degree of compatibility of information retrieval languages of document retrieval systems. **Autom. Doc. & Math. Ling.**, v. 12, n. 1, p. 18-26, 1978.

LANCASTER, F. W. **Vocabulary Control for Information Retrieval**. Virginia: Info Resources Press, 1986. 270 p.

LIBRELOTTO, Giovani Rubert. **Topic maps: da sintaxe à semântica**. 2005. 333 f. Tese (Doutorado em Informática) – Escola de Engenharia da Universidade do Minho, Braga, Portugal, 2005. Disponível em: <[http://tede.ibict.br/tde\\_busca/arquivo.php?codArquivo=486](http://tede.ibict.br/tde_busca/arquivo.php?codArquivo=486)>. Acesso em 2 jun. 2007.

LIMA, Gercina Angela Borém. **Mapa hipertextual (MHTX): um modelo para organização hipertextual de documentos**. 2004. 204 p. Tese (Doutorado em Ciência da Informação) – ECI/UFMG, Belo Horizonte, 2004.

MANIEZ, Jacques. Database merging and the compatibility of indexing languages. **Knowledge Organization**, v. 24, n. 4. p. 213-224, 1997.

NAVES, M. M. L. **Fatores interferentes no processo de análise de assunto: estudo de caso de indexadores**. 2000. 257 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte.

NEVILLE, H. H. Feasibility study of a scheme for reconciling thesauri covering a common subject. **J. Doc.**, n. 26, v. 4, p. 313-36, Dec. 1970.

NEWMAN, Simon M. **Information systems compatibility**. Washington [D.C.]: Spartan Books; London [Ingl.]: Macmillan, 1965. 150 p.

PARK, J.; HUNTING, S. **XML topic maps: creating end using topic maps for the web**. Boston: Addison Wesley, 2003. 644 p.

PEPPER, S. The TAO of topic Maps: finding the way in the age of infoglut. In: XML EUROPE CONFERENCE, 2000, Paris, França. **Proceedings...** Paris, França: of XML Europe 2000. Disponível em: <<http://www.ontopia.net/topicmaps/materials/tao.html>>. Acesso em: 2 nov. 2005.

PEPPER, S.; MOORE, G. **XML topic maps (XTM) 1.0**. [S/l]: Topicmaps.Org Specification, 2001. [On-line]. Disponível em: <<http://www.topicmaps.org/xtm/1.0/>>. Acesso em: 20 fev. 2007.

SOERGEL, Dagobert. Compatibility of vocabularies. In: THE CONTA CONFERENCE ON CONCEPTUAL AND TERMINOLOGICAL ANALYSIS IN THE SOCIAL SCIENCES, May 24-7, 1981, Bielefeld, FRG. **Proceedings...** Frankfurt: Indeks Verlag, Fred W. Riggs (Ed.), 1982. p. 209-223.

\_\_\_\_\_. **Indexing languages and thesauri:** construction and maintenance. Los Angeles: Melville Publishing Company, 1974. 632 p.

ZHANG, Xueying. Concept integration of document databases using different indexing languages. **Information Processing and Management**, v. 42, n. 1, p. 121-135, Jan. 2006.