

Criação e aplicação de um modelo baseado em Topic Maps à Documentação Educativa

M. Jesús Colmenero, Miguel Ángel Marzal

Departamento de Biblioteconomía y Documentación
Universidad Carlos III de Madrid
E-mail: {mcolmene, mmarzal}@bib.uc3m.es

José Carlos Ramalho

Departamento de informática
Universidade do Minho
E-mail: jcr@di.uminho.pt

RESUMO

As políticas da União Europeia para a sociedade da informação estão a promover um modelo educativo baseado na aquisição de competências e na aprendizagem ao longo da vida. A eficácia da aprendizagem neste modelo educativo assenta numa educação digital, na qual os recursos digitais estruturam, organizam e representam o seu conteúdo de acordo com as suas potencialidades educativas. Descrevem-se os passos prosseguidos na aplicação da norma ISO/IEC 13250:2000 Topic Maps como escolha para a organização e representação do conhecimento de uso educativo, adaptando a recuperação deste tipo de informação a um modelo de aprendizagem centrado no educando, tal como as actuais políticas europeias promovem. Numa primeira parte apresenta-se a análise realizada sobre as várias metodologias para a representação do conhecimento: thesaurus, mapas conceptuais e ontologias. Discutem-se as características que levaram à selecção dos Topic Maps como a metodologia a ser utilizada neste projecto. Numa segunda parte, apresenta-se a norma ISO/IEC 13250:2000 Topic Maps e discute-se de que forma nos irá auxiliar a criar um ambiente de aprendizagem centrado no educando e utilizando os recursos educativos (objectos de ensino, artigos, monografias, teses, etc) disponíveis na Web. Por fim, numa terceira parte, apresenta-se uma arquitectura funcional capaz de suportar a metodologia seleccionada e o cenário de aplicação: especificação de mapas conceptuais orientados a programas educativos dinâmicos e que reagem às interacções dos educandos; agregação de conteúdos disponíveis na Web com o objectivo de criar um repositório de objectos de ensino (utilização do protocolo OAI-PMH); definição de estratégias para a população do mapa conceptual com os conteúdos recolhidos da Internet; exploração do repositório a partir do mapa conceptual. Em conclusão, iremos demonstrar que já é possível criar um ambiente com estas características utilizando vários componentes e tecnologias existentes mas, também, que é necessário fazer algum trabalho de integração e normalização.

PALAVRAS-CHAVE: Documentação educativa, topic maps, recuperação de informação, representação de conhecimento.

ABSTRACT

Policies of the European Information Society are to promote an educational model based on the acquisition of skills and lifelong learning. The effectiveness of learning in this educational model is based on a digital education, in which digital resources structure, organize and represent its content according to their educational potential. We describe the steps pursued in the implementation of ISO / IEC 13250:2000 Topic Maps as the selected methodology adopted for the organization and representation of knowledge of educational use, adapting the retrieval of such information to a student-centered learning model, as the current European policies promote. The first part presents the analysis on the various methodologies for the representation of knowledge: thesaurus, concept maps and ontologies. In the following we discuss the characteristics that led to the selection of Topic Maps as the methodology to be used in this project. The second part presents the ISO / IEC 13250:2000 Topic Maps, and discusses how it will help to create an environment focused on student learning through the use of educational resources (learning objects, articles, monographs, theses, etc.) available on the Web. Finally, a third part presents a functional architecture capable of supporting the selected methodology and the application scenario: specification of concept maps, oriented educational programs and dynamic interactions that respond to the students; content aggregation available on the Web in order to create a repository of learning objects (using OAI-PMH), strategies for the population of the conceptual map with the contents collected from the Internet, holding the repository from the conceptual map. In conclusion, we demonstrate that it is possible to create an environment with these features using several existing technologies and components, but also that some work for integration and standardization is needed.

KEYWORDS: Educational documentation, topic maps, information retrieval, knowledge organization.

INTRODUÇÃO

As políticas da União Europeia para a sociedade da informação converteram-se em verdadeiras “políticas de informação para a educação” como consequência do impacto derivado da difusão das funcionalidades das tecnologias de informação e comunicação (TICs). O apoio à e-Educação começou com a acção e-Europe 2002 [1], há já 10 anos, tendo-se incrementado exponencialmente com o desenvolvimento posterior das actuações europeias destinadas a impulsionar a dita Sociedade da Informação.

Dois são os aspectos chave que cimentam estas políticas. O primeiro, a mudança em direcção a um modelo educativo onde o elemento activo da aprendizagem é o aluno, baseado nas teorias da aprendizagem significativa, no “aprender a aprender”. Este modelo desenvolve-se num espaço distinto ao tradicional, um novo espaço educativo que transcende o espaço da aula, estendendo-se pelo espaço digital, criando contextos colaborativos, dinâmicos e assíncronos (as “aulas expandidas”). Este espaço foi possibilitado graças às TICs.

O segundo aspecto centra-se em potenciar e impulsionar a “aprendizagem ao longo da vida”, sustentado na necessidade de dotar os cidadãos com as competências adequadas para que possam adaptar-se às mudanças e alcançar as competências adequadas durante toda a sua vida profissional, além da que alcançaram no período inicial de formação.

Ambos os elementos de mudança andam à volta do mesmo eixo: um modelo educativo baseado na aquisição de competências, por oposição à competência baseada num conjunto fechado de conhecimentos e domínio de capacidades e habilidades. Esta realidade reflectiu-se posteriormente nas próprias políticas europeias, juntando os objetivos das políticas de e-learning e aprendizagem permanente num único programa, o Programa de aprendizagem permanente 2007-2013 (Parlamento Europeu, 2008). A própria página web da Direcção Geral de Educação e Cultura da Comissão Europeia dá-lhe relevo no seu título: “*A single umbrella for education and training programmes*”.

A eficácia da aprendizagem neste modelo educativo construído sobre competências, desenvolvido num novo espaço, onde o elemento central é o aluno e projectado para uma aprendizagem contínua, assenta numa “educação digital, na qual os recursos digitais estruturam, organizam e representam o seu conteúdo de acordo com as suas potencialidades educativas. É de destacar, neste sentido, as palavras presentes numa informação produzida para a eEurope2002 (Comissão Europeia, 2000a) “*El tipo de información proporcionada por Internet no corresponde forzosamente a lo que necesita el sistema educativo. En efecto, mientras la enseñanza exige calidad y coherencia informativa, a lo que da acceso Internet es a una cantidad casi ilimitada de información de interés muy diverso y que requiere por lo tanto una selección previa. Los mejores sitios dedicados a la educación y la formación deberán proporcionar un contenido digno de interés pero también servicios que eviten al usuario ahogarse en la masa de información.*”

Convém mesmo assim garantir que estes sítios respeitem uma deontologia que permita evitar o uso abusivo de los dados pessoais, a exposição a mensagens publicitárias ou a utilização puramente comercial”.

O essencial destas palavras ficou registado posteriormente em (Comissão Europeia, 2000b).

É precisamente este o trabalho que os Centros de Documentação com função educativa (bibliotecas educativas de todos os níveis formativos, museus e arquivos) devem assumir na sua dimensão actual de Centros de Recursos para a Aprendizagem e na qual as Ciências Documentais podem contribuir decisivamente para a concretização deste cenário: constituir-se em espaços educativos organizados que favoreçam o trabalho docente e de aprendizagem, mediante o control, identificação, busca, selecção e organização de recursos informativos digitais e tradicionais, o suporte das comunidades virtuais e a edição e gestão de materiais educativos, implementando, além disso, sistemas de recuperação eficazes destes recursos postos à disposição dos seus usuários mediante a actualização dos seus instrumentos de organização de conteúdos, a partir do seu desenho conceptual na análise documental do conteúdo (ADC).

Este novo modelo educativo desenhado requiere uns instrumentos novos capazes de suportar o processo inferencial na direcção do conhecimento, que deve realizar o educando, a partir dos dados e blocos informativos da rede. Estes têm certas características distintas face aos tradicionais que condicionam o desenho e aplicação de instrumentos que permitam alcançar a adequada efectividade didáctica dos recursos digitais.

Fundamentalmente há que ter em conta que os recursos digitais são hipertextuais e interactivos. A hipertextualidade modifica o modo de compreensão e aprendizagem da informação. Os investigadores do campo da cultura escrita, como Roger Chartier, realçaram que as mudanças produzidas historicamente na apresentação dos textos tiveram influência tanto nos modos de lê-los como na sua compreensão, o qual está a acontecer também frente ao hipertexto. Ainda que em grande parte e todavia está imitando a disposição do texto tradicional, permite uma leitura fragmentada e “à deriva” distinta da do texto linear.

A interactividade, que permite a relação entre o recurso e o utilizador, favorece a possibilidade de adaptação entre ambos em função das capacidades do utilizador. A isto há que somar a possibilidade de interacção e comunicação entre utilizadores no mesmo recurso, favorecendo a aprendizagem colaborativa.

Assim, estes dois condicionantes, hipertextualidade e interactividade, devem ser considerados para intentar conseguir una optimização da eficácia didáctica da navegação, elemento inerente ao recurso digital [2], pela sua capacidade de suportar um processo intelectual complexo. Um bom sistema de navegação deve interligar-se com uma adequada estruturação da informação, dado que os utilizadores não reproduzem a forma física na qual a informação está estruturada, mas sim as relações conceptuais subjacentes para construir as representações mentais do seu conteúdo (Schawn 2002). Os aspectos cognitivos envolvidos num processo de

aprendizagem, então, precisam de fórmulas de navegação baseadas em conceitos acomodados às associações semânticas horizontais.

INSTRUMENTOS DE ORGANIZAÇÃO DO CONHECIMENTO: ANÁLISE DA SUA POTENCIALIDADE FACE AOS DOCUMENTOS DIGITAIS

A Análise Documental de Conteúdo (ADC) constitui a parte do processamento técnico da informação que pretende estabelecer, com a menor ambiguidade possível, as ideias relevantes de um documento com a finalidade de o recuperar posteriormente em função do seu conteúdo temático.

Este trabalho tradicional da AC deve sofrer uma transformação que a habilite a ser aplicada adequadamente no meio digital quando enfrentada à singularidade deste tipo de documentos. Segundo Moreiro y Marzal (2001), três são os condicionantes principais que, no processo que vai desde a informação ao conhecimento, obrigam a que esta mudança se produza.

O primeiro é consequência da função de leitor do utilizador no espaço digital, tomando as suas decisões quanto ao caminho a percorrer pelo hipertexto, guiado pelo seu interesse. Num ambiente de aprendizagem, a principal necessidade deste utilizador-leitor é chegar da informação ao conhecimento, fazendo-o através da selecção. A informação produz-se em blocos e é o utilizador quem deve escolher quais utilizará, de entre a multiplicidade de opções que se lhe apresentam. Assim, os utilizadores seleccionam e geram a sua própria colecção documental digital, que irá sendo mais pessoal segundo o seu nível cognitivo.

A ADC deve, portanto, converter as suas linguagens documentais em instrumentos mais próximos da linguagem dos utilizadores, e não à dos documentos, favorecendo a manipulação correcta de los sistemas informativos. Dar-lhes uma orientação “centrada no utilizador”, com a flexibilidade suficiente para responder às questões deste com a suficiente rapidez e eficácia, de forma que prevaleçam as competências do utilizador quanto à selecção e avaliação dos recursos pertinentes para as suas necessidades informativas.

O segundo condicionante vem da diluição da noção de “matéria” nos conteúdos. A tradicional separação académica em disciplinas foi-se transformando num emaranhado de relações múltiplas entre elas, levando à multidisciplinariedade e interdisciplinariedade. No meio digital, esta situação está especialmente presente, pelo que a ADC deve tratá-la como conceitos ou centros de interesse, em relação às necessidades do utilizador quanto à navegação e busca de informação. As representações baseadas em conceitos assim obtidos melhoram a aprendizagem quando o material se organiza em torno delas (Smith 2002).

O terceiro e último condicionante vem da necessária Alfabetização em Informação (*information literacy*) ou Alfabetização informacional. Em 1989 a American Library Association (ALA) entendeu-a como meio de formação permanente, mediante a que se dota de competência para encontrar informação, organizar o seu conhecimento e usá-lo para aprender. Também se

entende como o novo modo de aprender em contextos tecnológicos, uma vez adquiridas as competências necessárias por parte do educando para poder abordá-lo em óptimas condições (Bawden 2002). A ADC deve ajudar a organizar os conteúdos desde a perspectiva de geração de bibliotecas virtuais “educativas” (MARZAL 2003a, b), em consonância com o já exposto na introdução.

O instrumento da ADC tradicional mais adequado para este espaço digital, pela sua capacidade associativa e relacional, é o thesaurus, o qual está a sofrer uma evolução dos seus postulados teóricos que permitirão adaptá-los às exigências actuais.

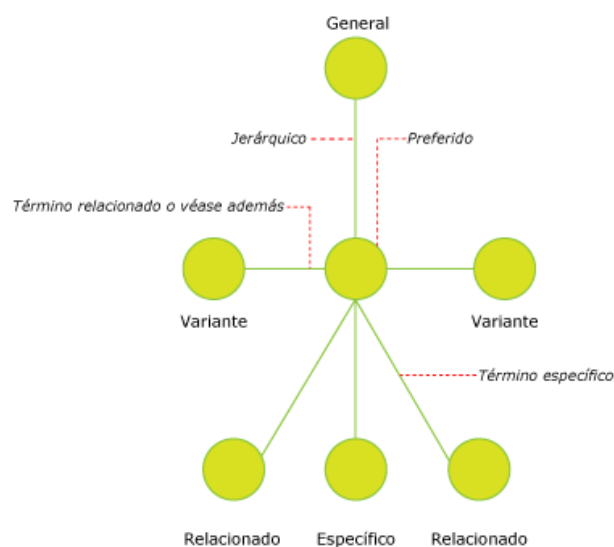


Figura 1: Relações em um thesaurus [extraído <http://www.infoxicacion.cl/vocabulario-controlado-tesauros/>]

ANÁLISE COMPARATIVA DE SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

No contexto descrito, realizou-se uma análise comparativa de thesaurus, mapas conceituais e ontologias, que constituem os sistemas mais usuais de representação de conhecimento, quanto à sua eficácia na organização de conteúdos de educação no espaço digital. Aunque proceden de diferentes disciplinas poseen características adecuadas para la representación del conocimiento con una estructura asociativa

Um exame comparativo dos instrumentos analisados leva-nos a destacar as suas semelhanças e potencialidades, que recaem fundamentalmente na inclusão de elementos de associatividade entre conceitos, acordos no interesse que nos ocupa, ou o que é igual, a busca de ferramentas documentais que permitam a organização dos recursos educativos no meio digital adaptados às características do processo de aprendizagem significativo.

Dois são os denominadores comuns a todos eles: os conceitos e as relações entre eles. Com distintas denominações terminológicas e regras diferentes para o seu uso, um exercício de abstracção equipara-os no plano teórico e conceptual. Esta proximidade levou-os a um plano de interrelação ao fundirem-se desde campos diferentes num espaço comum, o espaço digital. Espaço

que por sua vez se organiza sobre o hipertexto, cuja estrutura baseada em nodos e ligações se sobrepõe igualmente à representação do mesmo.

Os thesaurus, que representam a experiência do campo documental, e os mapas conceptuais, que sintetizam a da educação, surgem neste contexto da web como instrumentos de representação do conhecimento, como ontologias de baixo nível que constituem os primeiros escalões da web semântica. Permitem revelar a estrutura do conhecimento subjacente e, mediante a sua utilização conjunta, aproximar-se terminologicamente do utilizador. Desta forma, permitem suportar de forma multilinguística, factores, todos eles, imprescindíveis para a sua aplicação nas bibliotecas digitais educativas.

Os conceitos são representados em todos eles mediante palavras extraídas da língua natural. Em mapas conceptuais e ontologias, ao contrário dos thesaurus, não há controlo sobre o vocabulário. No entanto, com as ontologias escolhem-se termos representativos dos conceitos, estabelecendo relações e axiomas com o objectivo de especificar o seu significado de forma mais rígida que nos thesaurus. Todos eles, além disso, incluem relações ou ligações entre conceitos, com distintos sistemas de utilização, graus de sistematização e denominações.

As relações de equivalência dos thesaurus fornecem um mecanismo para associar todas as “etiquetas” representativas de um conceito à linguagem do utilizador, tal como é representado nos mapas conceptuais. Este tipo de relação está ausente como tal em mapas conceptuais e ontologias, ainda que se utilize um mecanismo semelhante quando se integram em sistemas automatizados mediante a sua combinação com outras ferramentas de tipo linguístico.

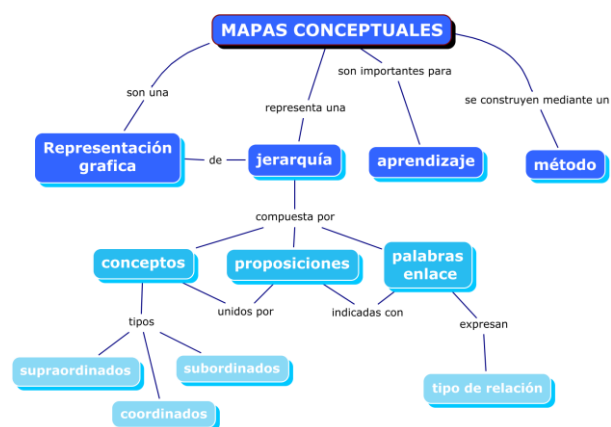


Figura 2. Exemplo de mapa conceptual

A estrutura hierárquica imposta ao esquema de representação do conhecimento do domínio é menos formal nos mapas conceptuais, uma vez que a sua necessidade está subentendida nas palavras de alguns autores e não está sujeita a regras de construção. Tanto em thesaurus como em ontologias há uma parte fundamental, estabelecida a partir de relações hierárquicas que são definidas de forma similar e padecem dos mesmos problemas de imprecisão em ambos os casos, permitindo a inferência a partir de mecanismos de herança.

As relações associativas, por último, representam o maior desafio em qualquer um dos formalismos. Los thesaurus, apesar de terem alguns tipos definidos, deixam a porta aberta à inclusão de outras. Os mapas conceptuais permitem qualquer tipo de relação associativa que o autor queira expressar e as ontologias, onde aquelas se especificam como propriedades, dependem do sistema lógico elegido para as definir.

Uma diferença fundamental dos thesaurus em relação aos outros dois instrumentos é a ausência de especificação do tipo de relação estabelecida. Não possuem mais descrição que um dos três grandes tipos de relação, o que limita a sua expresividade semântica. Como já vimos, contornar estas dificuldades constitui um dos interesses da investigação mais recente, com a intenção de convertê-los em ferramentas mais potentes no âmbito automatizado. Além disso, tanto os mapas conceptuais como as ontologias utilizam principalmente verbos para especificar o tipo de relação estabelecida sendo equiparáveis proposições a propriedades.

Os thesaurus e os mapas conceptuais estão imersos num processo adaptativo ao meio automatizado que os levou, numa primeira etapa, às bases de conhecimento e, numa segunda, às ontologias, fazendo-se acompanhar nesta evolução pelo campo da Inteligência Artificial (IA). No entanto, ambos os instrumentos carecem da suficiente formalização, até ao momento, que permita a sua utilização como ontologias em sistemas que requeiram um grande poder descritivo que permita o raciocínio.

Por outro lado, a aplicação à Web das ontologias muito formalizadas não é tão simples como parece. Os investigadores procedentes da IA encontram-se num meio no qual as suas técnicas tradicionais não podem aplicar-se com eficácia, devido à falta de estruturação que a informação contem e à dificuldade da sua implantação num sistema descentralizado se o nível de aprendizagem necessário para criar conteúdos para este meio aumentar.

Desta forma, sistemas menos formais, mas estruturados, como os thesaurus e os mapas conceptuais parecem ter uma aplicação adequada na Web. Há muitos desenvolvimentos, com óptimos resultados, que são utilizados como base para a construção de ontologias de complexidade crescente e como primeiro escalão em domínios reduzidos. O facto de que um grande número deles tenham uma estruturação dos dados torna-os idóneos para a sua aplicação na Web, que é o espaço sobre o qual se desenvolve a *Educação Digital*. Assim, além de existir a possibilidade de integração de mapas conceptuais e thesaurus como primeiros níveis numa hierarquia de ontologias, ambos os instrumentos permitem a sua utilização como apoio à compreensão da área de conhecimento que descrevem.

A utilização de um instrumento associativo que permita uma aproximação às ontologias para a sua incorporação adequada na Web semântica, servindo além disso de ponte entre estas e os thesaurus, que representam a experiência no campo documental, e os mapas conceptuais, que sintetizam a da educação, seria ideal como ponto intermédio. Isto evitaria uma sobreposição de camadas ou níveis intermédios elevada, com os correspondentes problemas de integração, e por outro lado permitiria conservar a capacidade de revelar a

estrutura do conhecimento subjacente e a sua proximidade terminológica ao utilizador.

Pensamos que um instrumento ideal para este propósito, tal como se descreveu, centrado no documento educativo, é o modelo definido pela norma ISO/IEC 13250: 2003 *Topic Maps*. Esta norma baseou o seu desenvolvimento na semântica implícita dos índices analíticos impressos e na sua adaptação ao processamento automático. Ainda que a sua criação e evolução se tenha produzido em contextos tecnológicos, distantes do documental, não deixa de estender as suas raízes a este. Tanto thesaurus como mapas conceptuais podem ser expressos utilizando esta norma para adaptá-los ao meio digital e aproximá-los a uma maior expressividade semântica ao estilo das ontologias mais formalizadas.

MODELO TOPIC MAPS (ISO/IEC 13250)

Um topic map é um documento estruturado de acordo com a norma ISO/IEC 13250 *Topic Maps*. Define-se como um hiperdocumento que forma uma rede de relações semânticas por cima dos recursos de informação, associando tanto os conceitos com os recursos correspondentes como os conceitos entre si. Assim, realizar uma busca na rede descrita pelo topic map pode assemelhar-se à busca em estruturas de conhecimento (Hemrich, M. y Schäfer, U., 1999).

Esta norma compreende várias versões: a ISO/IEC 13250:2000, que foi a primeira versão na linguagem SGML e que já não é usada; a ISO/IEC 13250:2003, segunda versão que incluía um DTD para a validação de topic maps escritos em XML, conhecida como XTM 1.0 (TOPICMAPS.ORG., 2001) [3]; e uma última versão, ISO/IEC 13250-3:2007, conhecida como XTM 2.0 (ISO/IEC JTC1/SC34, 2006). A decisão de rever XTM 1.0 levou à decisão de converter Topic Maps numa família de normas, da qual estão publicadas a parte 3 (XTM 2.0), a 2 e a 4.

Sem nos centrarmos nos detalhes da sintaxe nem em outras considerações de tipo técnico vamos analisar as características e possibilidades que este modelo oferece.

O núcleo central do modelo definido pela norma ISO/IEC 13250 *Topic Maps* é constituído por três elementos básicos: *Topic*, *Association*, e *Occurrence* (tópicos, associações e ocorrências). Esta tríade de conceitos foi denominada de TAO dos *topic maps* por Steve Pepper (2002), um dos editores originais de XTM. Além destes, ampliando o poder expressivo do modelo e as suas possibilidades, surgem os conceitos de *scope* e *public subject* (contexto e conceito público).

Um *topic map* baseia-se no conceito de **Topic**, o qual constitui a representação material ou concreta do *subject*, percepção humana abstracta de uma realidade. A noção de *subject* é o ponto de partida conceptual sobre o qual assenta o modelo, sendo definido na norma nos seguintes termos: “No sentido mais amplo, um ‘subject’ é qualquer coisa, com independência de se existe ou tem outras características específicas, sobre a qual se pode dizer qualquer coisa com qualquer significado”.

No modelo *topic map* o significado de *subject* está próximo do de representação mental, conceito, noção ou ideia. A nota que acompanha a definição tenta clarificá-lo indicando que “o coração invisível de cada *topic* é o

subject que o seu autor tinha em mente quando foi criado”.

Assim, o termo *topic* indica o objecto ou elemento do *topic map* que representa o *subject* ao qual se está a referir, tornando-o “real” para o sistema. Entre *topic* e *subject* cria-se uma relação biunívoca na qual um *subject* é representado por um único *topic* e viceversa (Pepper, S., 2002). À semelhança do *subject*, um *topic* pode representar qualquer coisa: pessoas, entidades individuais ou colectivas, conceitos, etc.

Mesmo assim, diferencia-se *subjects* direccionáveis e não-direccionáveis. Os direccionáveis são aqueles recursos que têm uma existência no espaço digital; os não-direccionáveis são os que existem fora deste espaço, não podendo portanto ser relacionados directamente, pelo que devem identificar-se de forma indirecta mediante um recurso que funciona como *subject indicator*.

Cada *topic* é uma instância de uma ou mais classes de *topics* que podem ou não declarar-se de forma explícita. Os topics que funcionam como classe de *topic* são *topics* igualmente. Esta relação classe-instância especifica-se com o elemento *instanceOf*.

Cada *topic* pode ter as características seguintes: uma denominação (*topic name*), um ou uns exemplos ou descrições (*topic occurrence*) e um papel como membro de uma associação (*role*). Esta declaração de características considera-se válida para um determinado *scope* ou contexto. Dois *topics* com as mesmas características consideram-se idênticos, produzindo duplicidade, pelo que um deles será eliminado quando o *topic map* seja processado.

O nome, legível para os humanos, de um *topic* define-se através do *topic name*. Dado que um mesmo conceito pode ser designado com uma grande variedade de nomes, e nenhum deles ser válido, o modelo permite definir nomes normalizados para os *topics* que sejam significativos desde o ponto de vista semântico, ao mesmo tempo que concede a possibilidade de associar outros livremente com o objectivo do seu processamento por distintas aplicações (Pepper, S., 2002). Assim, um *topic* pode não ter um nome ou dispôr de vários, mediante a atribuição de múltiplos *base name* (*name* em XTM 2.0). No entanto, cada um dos distintos *base name* é válido para um determinado *scope*, não sendo permitido mais de um para o mesmo contexto. Esta limitação está destinada a facilitar a fusão ou união de dois ou mais *topic maps*. Adicionalmente, o *base name* pode incluir formas alternativas ou variantes do nome (*variant name*) para distintos processos dependentes da aplicação, como a forma em que se mostrará o nome no monitor (com qualquer tipo de dados: caracteres, gráficos, etc.) ou como será ordenado, por exemplo.

A razão principal para que um *topic* possa ter uma multiplicidade de nomes é que podem usar-se para propósitos distintos. A utilização preferencial que se faz desta facilidade, listar sinónimos, variantes dialectais ou idiomáticas, é precisamente a que permite incluir as relações de equivalência e de referência descritor-não descritor (usado por usa-se) de um thesaurus reunindo-as todas à volta do conceito.

Uma **Occurrence** é um recurso externo de informação, ligada por uma referência utilizada para a sua localização, que clarifica ou exemplifica o significado do **topic**. As referências a recursos especificam-se, em XTM, através de URIs (*Uniform Resource Identifiers*).

Estes recursos não se armazenam, habitualmente, no *topic map*, o que implica que o *topic map* e os recursos que indicam as *occurrences* formam camadas separadas; que estes últimos podem constituir um conjunto de informação de qualquer tipo, formato, ou localização; e que os *topic maps* são transportáveis sendo possível aplicá-los a diferentes cenários de recursos informativos. Estas qualidades do modelo, consideradas vantajosas nos contextos tecnológicos, na realidade são as mesmas que têm os thesaurus relativamente aos documentos que enumeram: são um instrumento independente dos recursos, podem-se utilizar em várias colecções documentais e estas colecções podem ser heterogéneas e as suas componentes podem estar em locais distintos. A separação entre recursos e o *topic map* que os referencia e organiza tem a importante vantagem de tornar desnecessária a marcação dos documentos com metadados.

A *occurrence* pode incluir também informação como dados de carácter, o que é especialmente útil quando a sua quantidade é pequena (por exemplo, datas de nascimento, publicação, coordenadas, definições curtas,...).

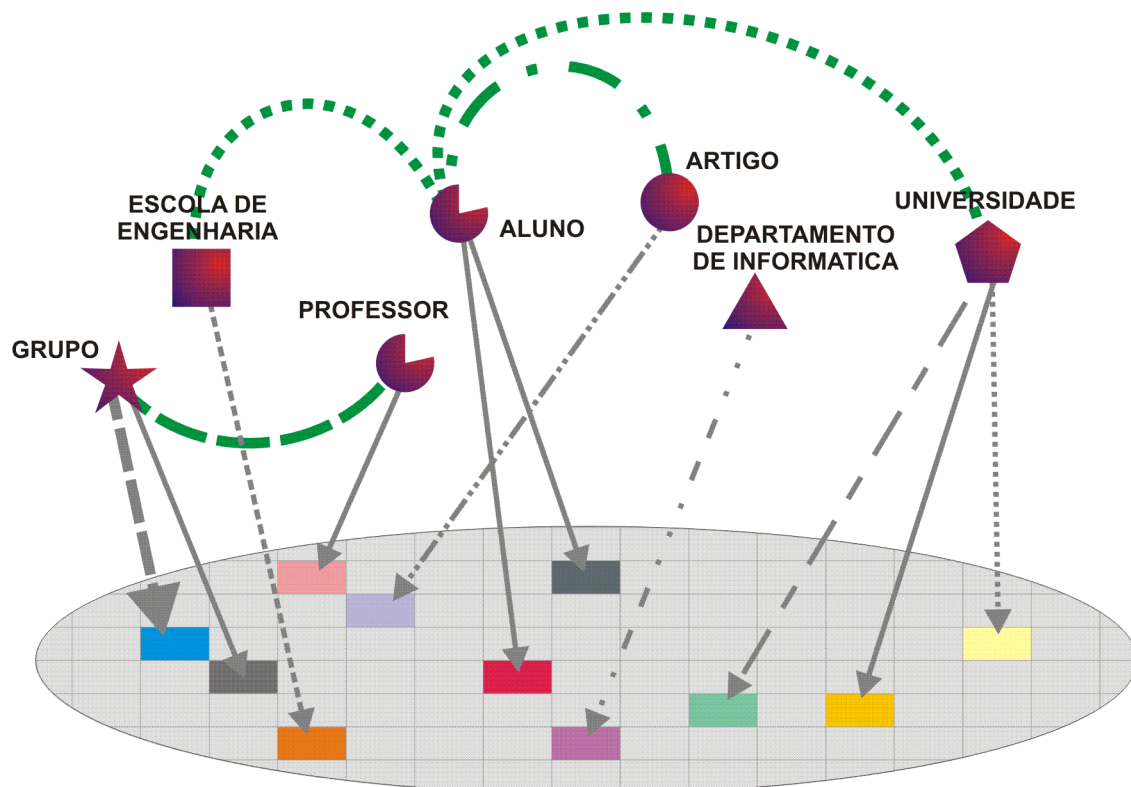
À semelhança dos *topics*, cada *occurrence* é instância de uma classe de *occurrence*, *occurrence type*, ou classe de *occurrence* que assim, é também um *topic*.

O terceiro elemento do núcleo central do modelo *topic maps* é a **Association** (associação). Uma *association* é “uma relação entre um ou mais *topics*, onde cada um deles desempenha um papel [*role*] como membro [*member*] da dita associação”. O papel que um *topic* desempenha numa associação, indica de que maneira participa nela.

Esta relação que pressupõe uma associação vem expressa implicitamente pela expressão verbal que uniria os dois *topics*, assumindo que estes representariam os substantivos da frase assim formada, à semelhança do que se descreveu para os mapas conceptuais. O número de *topics* envolvidos numa associação não está limitado ainda que o mais frequente é que sejam dois (associações binárias) ou, três (associações ternárias).

As associações podem também classificar-se em classes à semelhança dos elementos anteriores. A capacidade de criar classes de associações, i.e., de criar tipos, torna possível agrupar os *topics* que têm a mesma relação com um determinado *topic*, o que facilita o desenvolvimento de interfaces intuitivas e amigáveis para navegar por conjuntos grandes de informação (Pepper, S., 2002a).

A *association* no modelo *Topic Maps* não indica um sentido no qual deva interpretar-se. A relação que estabelece é válida em qualquer caso. Como indica a especificação XTM, depende apenas do tipo de relação estabelecida e do papel que jogam os seus membros; “*Como especificar uma relação é uma questão de denominações, não de direcção*”



Para poder conhecer a função que os *topics* desempenham como membros de uma associação é necessário introduzir o conceito de *association role*. Este conceito deve ser representado por um *topic*, que é o que especifica a dita função ou papel desempenhado.

Na especificação XTM 1.0 estão incluídas duas classes de *association*, que já têm definidos os *subjects* correspondentes através de *published subjects*: a *class-instance association* (associação classe-instância) e a *superclass-subclass association* (associação superclasse-subclasse), com as relações entre *topics* que denotam as suas denominações.

Estas *association* permitem-nos diferenciar aqueles *topics* que estão a funcionar como classes e em que associações, pois pode acontecer que um *topic* seja classe e instância em contextos diferentes: pode ser instância de uma classe e esta por sua vez também ser um *topic*. A *superclass-subclass association* permite também construir hierarquias de classes, aplicadas tradicionalmente para realizar taxonomias ou classificações de matérias. As relações hierárquicas de um thesaurus podem construir-se da mesma forma. Podem criar-se hierarquias de classes de *topics*, de classes de *associations* e de classes de *occurrences*. Em XTM 2.0 estes aspectos foram eliminados e passadas para a parte da norma correspondente ao Modelo de dados (ISO/IEC 13250-2:2006).

Como já referimos, o modelo *Topic Maps* incorpora outros conceitos, que se juntam aos três centrais, conferindo-lhe um poder de especificação muito superior.

O *Scope* é o contexto específico no qual o *topic* se associa para precisar o seu significado, tipicamente o idioma ou dialecto de uma comunidade, os direitos de acesso e outra informação contextual (por exemplo, diferentes níveis de destreza ou interesse do utilizador, ou distintas acepções de uma mesma palavra (homónimos)).

Dependendo do elemento ao qual o *scope* está associado as possibilidades de uso são variadas e extensas. Em caso de ser utilizado no nome, os *topics* que desempenham a função de *scope* podem-se agrupar em classes segundo a língua natural que se esteja a utilizar, a norma aplicada, determinar distintas classes de denominações (autoridade, coloquial, mote, pseudónimo) ou incluso espaços de tempo (períodos históricos, calendários, horários). Uma aplicação, interessante em alguns casos, de esta característica é a possibilidade de etiquetar as classes de associações, permitindo dar um “sentido de leitura” à associação. Exemplos de uso sobre as *occurrences* são a gestão de níveis de acesso, subdomínios, localização do recurso, comentários e, de novo, a linguagem que se utiliza. Aplicações semelhantes são as derivadas do seu uso sobre a função (*role*) exercida nas relações (*association*).

Como pode deduzir-se, a principal consequência do uso de *scope* é que facilita a criação de subconjuntos classificados de *topics*, o que resulta numa grande utilidade na hora de aplicar sistemas de consulta (mediante filtros) ou selecção nos interfaces com o utilizador, melhorando a navegação.

O último conceito que presente no modelo é o *Public Subject*. A sua necessidade deriva da possibilidade proporcionada de fundir *topic maps* entre si. A forma de poder estabelecer a identidade entre *topic* aparentemente distintos é através do *public subject* (ou *public subject indicator*; *PSI*), o qual não é mais que um recurso electrónico que identifica o *subject* com a menor ambiguidade possível. Uma vez que a utilização deste elemento é opcional, não existem garantias de que a união seja perfeita e que não resultará em vários *topics* para o mesmo conceito. Se estão identificados e os *topics* são iguais, o resultado da fusão será a união das suas características combinadas.

Um *public subject* é, em resumo, um *subject indicator* que se apresenta como documento publicado e mantido num URI comunicado, com o objectivo de proporcionar-lhe uma garantia de credibilidade, autoridade ou oficialidade. Podem ser criados por qualquer pessoa ou organismo para seu uso público com o objetivo último de facilitar o intercâmbio, a união de *topic maps* e a sua transportabilidade.

Os dicionários da língua e outros produtos terminológicos na sua versão digital, mantidos por instituições reconhecidas poderiam realizar este papel de uma forma magnífica.

Como se deduz da descrição do modelo *Topic Maps*, as suas características são, no que se refere aos conceitos centrais que o sustentam, em tudo parecidas àquelas dos instrumentos de organização do conhecimento antes considerados. Isto faz com que seja possível expressar tanto thesaurus como mapas conceptuais com este modelo e utilizar as vantagens de ambos na sua aplicação à organização de recursos educativos permitindo-nos criar um ambiente de aprendizagem centrado no educando, utilizando os recursos educativos (objectos de ensino, artigos, monografias, teses, etc) disponíveis na Web.

Deve recordar-se que o modelo *Topic maps* nasceu focado no meio digital, ao contrário dos outros dois instrumentos. Ao ser um documento XML, inerentemente digital, permite a sua utilização como estrutura de navegação e, além disso, é possível a sua utilização como instrumento de recuperação através das linguagens de consulta para *Topic Maps* (ainda não normalizadas, mas com algumas propostas operativas).

A possibilidade de usar outras partes do modelo para tornar a estrutura mais ou menos rígida (*Topic Maps Constraint Language* - *TMCL* (ISO 19756)) aproxima-o das ontologias mais formalizadas e às vantagens que estas podem proporcionar (LIBRELOTTO, Giovani. R. *et al.*) (como a geração de novo conhecimento automaticamente a partir da base especificada).

Finalmente, ao assemelhar-se na sua estrutura a um grafo, existem aplicações que os processam e mostram como mapas conceptuais, com as relações /associações “etiquetadas” o que permite utilizar este instrumento de aprendizagem como funcionalidade para a mesma estrutura (LIBRELOTTO, Giovani. R. *et al.*, 2005; RAMALHO, José Carlos *et al.*, 2005).

Deste modo, usando este modelo para a organização de recursos digitais, dispomos de uma ferramenta “três em um” ou multifuncional: estrutura de navegação

conceptual, desenvolvimento simples de um motor de busca sobre a estrutura e visualização em forma de mapa conceptual avançado.

A seguir, discutimos uma arquitectura funcional, baseada na utilização deste modelo, que permitirá a criação de repositórios digitais de recursos educativos que permitem sustentar uma Educação digital óptima, tal como se propôs no início.

ARQUITECTURA DE UM REPOSITÓRIO BASEADO EM TOPIC MAPS

Ainda que o desenvolvimento de um *topic map* de acordo com a norma (seja XTM 1.0 ou XTM 2.0) não requeira maior perícia que um editor de textos e um validador de documentos XML, o que em princípio parece fácil complica-se rapidamente. Isto acontece basicamente por duas razões:

- Um *topic map*, apesar de ser simples, é composto por uma quantidade de tópicos e associações que fazem com que o documento de texto resultante seja bastante grande, o que se agrava quando se começa a juntar as ocorrências seleccionadas, que podem ser de diversos tipos, e outros conceitos do modelo. Isto pressupõe perder-se com facilidade na complexidade crescente, quase exponencial.
- O custo em tempo de desenvolvimento e esforço devido, na maioria das ocasiões, à complexidade conceptual que a área temática e a sua rede de relações possa ter.

A primeira dificuldade pode diluir-se em parte com a utilização de algum dos editores de *topic maps* disponíveis (a complexidade crescente às vezes dificulta a orientação, à semelhança do que ocorre com qualquer outro conjunto grande de informação).

A segunda é, sem dúvida, a mais difícil de contornar. É um problema comum à criação de qualquer estrutura organizada e significativa, tais como thesaurus, mapas conceptuais e ontologias. Este é o ponto que este estudo pretendeu resolver dentro do possível com o objectivo de criar estruturas conceptuais que possam ser usadas como mapas conceptuais orientados a programas educativos dinâmicos e que reajam às interações dos educandos.

Deve fazer-se aqui, de novo, uma sistematização das dificuldades que devem ser resolvidas, a fim de seleccionar a estratégia adequada a cada caso. Basicamente, há que ter duas estratégias distintas, uma para as estruturas de conhecimento a desenvolver para o programa educativo concreto e outra relativa aos recursos seleccionados para ilustrar cada conceito ou parte do programa.

A estrutura de conhecimento pode pensar-se *de novo* ou adaptá-la a partir de fontes já existentes. É razoável, especialmente nos Centros de Documentação com orientação educativa, recorrer aos instrumentos de organização do conhecimento que se estiram analisando na primeira parte: thesaurus relacionados com a temática que a documentação do centro está cobrindo ou que está usando para indexar a sua própria colecção, mapas conceptuais e ontologias.

Muitos deles foram-se transferindo para o meio digital e, portanto, podem tratar-se de forma a convertê-los, mediante as transformações adequadas num formato XML que possa transformar-se por sua vez num *topic map* (que também é um documento XML) mediante o seu processamento com XSLT. Se já se encontram em XML o processo é mais directo e rápido pois bastará apenas uma transformação especificada numa folha de estilo. Este seria o caso de estar em formato SKOS, formato que será incorporado na norma ISO para thesaurus (que está em processo de revisão), ou outro formato XML como OWL (“Ontology Web Language”) ou RDF (“Resource Description Format”).

O formato SKOS, e outros parecidos definidos em XML, permitem também incorporar outros vocabulários controlados como esquemas de classificação, listas de tópicos de matéria, taxonomias, ontologias e glossários, se bem que estes últimos carecem de relações.

Os mapas conceptuais são também uma fonte interessante que se pode encontrar no formato XML. Por exemplo, o programa CmapTools, de distribuição gratuita (<http://cmap.ihmc.us/>) permite a exportação dos mapas realizados em XML, inclusive directamente em XTM 1.0, e a busca de mapas por conceitos numa rede própria e na web.

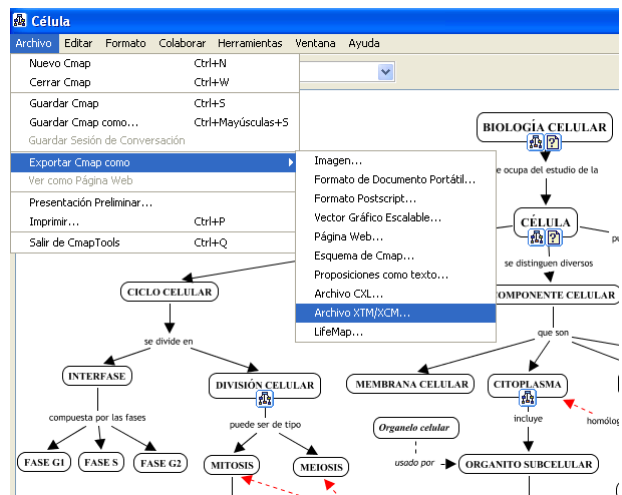


Figura 3. Exportação de CmapTool a XTM

Também é possível usar ontologias escritas em OWL com alguma perda de informação (a resultante dos predicados lógicos e que estão normalmente presentes em OWL-DL e OWL-Full).

Uma vez obtida a estrutura conceptual que representa o domínio de forma quase automática em formato XTM pode-se seguir refinando o modelo à medida dos desejos e necessidades.

O passo seguinte será popular o topic map com os recursos seleccionados para o programa educativo. Como já foi referido deve-se procurar uma possibilidade que evite o custo elevado da realização manual.

Uma primeira aproximação foi a utilização dos repositórios de recursos existentes para criar um repositório de objectos de ensino próprio através da utilização de um colectador (“harvester”) que usa o protocolo OAI-PMH. A partir daqui, os dados dos

recursos são inseridos no *topic map* pré-existente através de folhas de estilo XSLT.

Esta estratégia permite-nos obter recursos já previamente seleccionados e, seleccionando os repositórios onde se vai colher vamos incorporando outros filtros de selecção.

A recolha de recursos directamente da web requiere outros programas de recolha de dados e as transformações pertinentes para a sua incorporação. Uma fonte de recursos fundamental no nosso caso é o próprio catálogo do centro de documentação da instituição, com dados seleccionados e estruturados que podem transformar-se em XML e depois inseridos no *topic map*.

Pode-se optar por armazenar estes recursos (os do próprio catálogo e os recolhidos da web) na base de dados do repositório e extraí-los todos para popular um ou vários *topic maps* à medida das necessidades.

Integrando o repositório com um editor de *topic maps*, como Omnigator ou TM4L, que possibilitam apresentações gráficas interactivas em forma de grafo

torna-se possível fazer a exploração do repositório como se tratasse de um mapa conceptual, expondo a informação existente (os conteúdos educativos que se pretendem transmitir) de uma forma que favorece a aprendizagem e uma navegação dinâmica e interactiva sobre os recursos.

Ainda que seja possível combinar todos estes elementos para obter o resultado exposto existem ainda certos problemas técnicos que é preciso resolver. Os mais importantes centram-se no próprio modelo dos Topic maps pois apesar de ser simples e potente esta mesma simplicidade dá-lhe uma versatilidade que é necessário restringir em várias ocasiões. Assim, a união de mapas distintos nem sempre resulta numa estrutura adequada, para dar um exemplo. Isto está do lado do desenvolvimento dos editores que deveriam ter mecanismos que permitissem ao utilizador tomar a decisão intelectual final.

Para finalizar, se aponta como trabalho futuro a incorporação no sistema de medidas de relevância baseadas na interacção dos utilizadores, que permitam la depuração posterior.

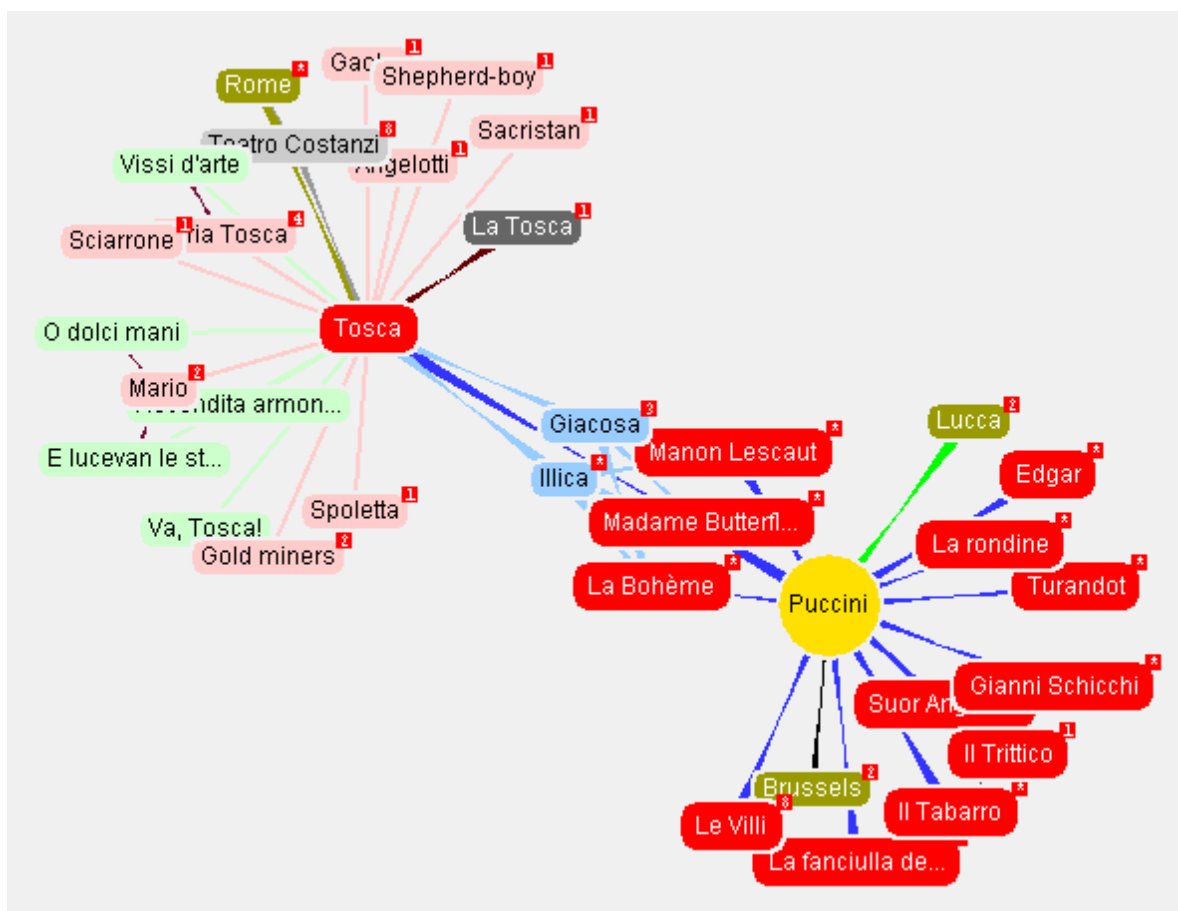


Figura 4: Visualização de *topic maps* em forma de mapa conceptual como um mapa de conceito .mediante Vizigator de Omnigator.

NOTAS

[1] Véase la dirección web del portal EUROPA http://europa.eu/legislation_summaries/education_trainin_g_youth/lifelong_learning/c11046_es.htm, donde se recoge la Comunicación de la Comisión, de 24 de mayo de 2000 “eLearning - Concebir la educación del futuro”.

[2] Los elementos de “navegación” del texto escrito también existen y son por todos conocidos en su forma actual: índices, numeración de páginas, divisiones estructurales, elementos de puntuación.

[3] En <http://doteine.uc3m.es/XTM/index.html> se dispone de una traducción al español de la especificación XTM 1.0.

REFERÊNCIAS

Open Archives Initiative (2002). – The Open Archives Initiative Protocol for Metadata Harvesting. [Em linha] [Consult. 19 de Outubro 2009]. Disponível em: <http://www.openarchives.org/OAI/openarchivesprotocol.html>

COMISSÃO EUROPEIA (2002a). [Em linha]. – Informe de la Comisión al Consejo y al Parlamento Europeo, de 27 de enero de 2000 “Concebir la educación del futuro - Promover la innovación con las nuevas tecnologías” [Consult. 12 de Novembro 2009]. Disponível em: http://europa.eu/legislation_summaries/education_trainin_g_youth/lifelong_learning/c11031b_es.htm

COMISSÃO EUROPEIA (2002b). [Em linha]. – Relatório da Comissão ao Conselho e ao Parlamento Europeu - Pensar o futuro da educação - Promover a inovação através das novas tecnologias. COM(2000) 23 final. [Consult. 12 de Novembro 2009]. Disponível em: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52000DC0023:PT:HTML>

ISO (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION).– Information technology. SGML Applications.Topic Maps. ISO/IEC 13250. 2003. Geneva: ISO.

ISO/IEC JTC1/SC34. 2006. – Information Technology-Document Description and Processing Languages. – Topic Maps — XML Syntax [Em linha]. Garshol, Lars Marius *et al.* (eds.). Disponível em <http://www.isotopicmaps.org/sam/sam-xtm/>. [Consult: 12 de Janeiro de 2010].

GUARINO, N. – Understanding, building, and using ontologies: a commentary to using explicit ontologies in KBS development, by van Heijst, Schreiber, and Wielinga. INTERNATIONAL JOURNAL OF HUMAN AND COMPUTER STUDIES, vol. 46, n. 2/3 (1997), p. 293-310.

LIBRELOTTO, Giovanni. R. *et al.* – Topic maps constraint languages: understanding and comparing. INTERNATIONAL JOURNAL OF REASONING-BASED INTELLIGENT SYSTEMS. Vol. 1, nº 3-4 (2009), p. 173 - 181.

LIBRELOTTO, Giovanni. R. *et al.* – Ulisses: um navegador conceptual para Topic Maps. Em: CLEI, 2005. Cali (Colombia).

MARZAL GARCIA-QUISMONDO, M. A. *et al.* (2003a) – Modelos instrumentales de la documentación para recuperar conocimiento: Metadatos y Topic Maps para educacion. En: SIGEF CONGRESO (10º: 2003: León). Emergent solutions for the information and knowledge economy.

MARZAL GARCIA-QUISMONDO, M. A. *et al.* (2003b) – El recurso educativo como documento: learning objects. En: JORNADAS ANDALUZAS DE DOCUMENTACION. *noviembre 2003*.

MOREIRO GONZÁLEZ, J. A. *et al.* – Modelos teóricos y elementos funcionales para el análisis de contenido documental: definición y tendencias. [Em linha]. Investigación bibliotecológica, vol. 15, n. 31 (2001), p. 125-162. [Consult: 10 de Janeiro de 2010]. Disponível em <http://www.ejournal.unam.mx/ibi/vol15-31/IBI03107.pdf>.

PARLAMENTO EUROPEU. – Decisão 2006/1720/CE do Parlamento Europeu e do Conselho, de 15 de Novembro de 2006, que estabelece um programa de acção no domínio da aprendizagem ao longo da vida. [Em linha]. JO L 327 de 24.11.2006, p. 45–68. [Consult: 12 de Novembro 2009]. Disponível em <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32006D1720:PT:NOT>

PEPPER, Steve. 2002. – The TAO of Topic Maps: Finding the Way in the Age of Infoglut. [Em linha]. [Consult: 10 de Janeiro de 2010]. Disponível em <http://www.ontopia.net/topicmaps/materials/tao.html>. Revisão do original de 2000.

PEPPER, Steve. 2004. – Towards Seamless Knowledge: Integrating Public Sector Portals. [en línea]. En: XML 2004 Conference & Exposition, Washington DC (EEUU), 15-19 Noviembre, 2004. [Consult: 10 de Janeiro de 2010]. Disponível em <http://www.ontopia.net/topicmaps/materials/Towards%20Seamless%20Knowledge.ppt>.

RAMALHO, José Carlos *et al.* – Geração automática de interfaces web para sistemas de informação metamorphosis. Sistemas de Informação. Nº 17 (2005), p. 53-71.

SMITH, T. R. *et al.* – Structured models of scientific concepts for organizing learning materials». Em: LÓPEZ-HUERTAS, M. J. (ed.). Challenges in knowledge representation and Organization for the 21st century. Integration of knowledge across boundaries: proceedings of the seventh international ISKO conference, Granada (Espania) 10-13 Julio 2002. (Alemania: Ergon, 2002), p. 232-239.

TOPICMAPS.ORG. 2001 – XML Topic Maps (XTM) 1.0 Specification. [Em linha]. Pepper, S. *et al.* (eds). [Consult: de Janeiro de 2010]. Disponível em: <http://www.topicmaps.org/xtm/1.0>.