

## **Evolução ontológica das linguagens documentárias: relato de uma experiência de curso organizado conjuntamente para o DT/SIBi-USP e o PPGCI/ECA.**

*Ontological evolution of documentary language: report of an organized course experience  
jointly to DT / SIBi-USP and PPGCI/ECA*

**José Antonio Moreiro González**

Profesor Catedrático de Biblioteconomía y Documentación e Decano da Facultad de Humanidades,  
Comunicación y Documentación da Universidad Carlos III de Madrid.

E-mail: [jamore@bib.uc3m.es](mailto:jamore@bib.uc3m.es)

### **Resumo**

Aborda-se o contexto no que se produziu a revisão do conceito de tesouro como linguagem documental prototípico. O contexto vem determinado pela revalorização das Taxonomias, o aparecimento das Ontologias como KOS; e a evolução dos próprios tesouros para o seu enriquecimento conceitual e a sua transformação em redes semânticas. Atende-se também ao novo modelo de visualização e representação conceitual contribuído pelos Topic maps, bem como a relevância atingida pelas Folksonomias à hora de reinterpretar o sistema das palavras-chave. Finalmente, descreve-se uma experiência de curso de atualização dado conjuntamente a bibliotecárias do DT/SIBi-USP e a alunas do PPGCI/ECA.

**Palavras-chave:** Conceitos; Contexto; Revisão conceitual; Tesouros; Taxonomias; Ontologias; KOS; Tesouros conceituais; Redes semânticas; Topic maps; Folksonomias; Relato da experiência; curso de atualização; DT/SIBi-USP; PPGCI/ECA.

### **Abstract**

It addresses the context in which it has produced the review of the thesaurus concept like prototypic documentary language. The context is determined by the Taxonomies revalorization, the apparition of the Ontologies like KOS; and the evolution of the own thesauri to his conceptual enrichment and his transformation in semantic networks. It attends also to the new model of visualization and conceptual representation contributed by the Topic maps, as well as to importance achieved by the Folksonomies to reinterpret the key-words system. Finally, it describes an experience of update course given jointly to librarians of the DT/SIBi-USP and to students of the PPGCI/ECA.

**Keywords:** Concepts and context; Conceptual Review; Thesaurus; Taxonomies; Ontologies; KOS; Conceptual Thesaurus; Semantic Networks; Topic maps; Folksonomies; Relate of experience; Update Course; DT/SIBi-USP; PPGCI/ECA.

## Os conceitos e o contexto<sup>1</sup>

A estabilidade terminológica, de origem racionalista, foi a proposta que as ciências naturais procuraram para determinar o sistema de comunicação científica entre os séculos XVI e XX. Desde uma perspectiva lingüística, a história da ciência está associada com a constituição de uma terminologia própria (BENVENISTE, 1989. p. 221). Na perseguição de instrumentos simbólicos e conceituais chegaram a metalinguagens de representação amplamente compartilhadas, muito formalizadas e independentes das línguas naturais, procurando especialmente a precisão, objetividade e clareza expressivas (BARROS, 2006, p. 22-26).

Neste marco empírico e utilitarista procuraram-se categorias universais, como formas *a - priori* de representação. A pretensão de uma comunicação inequívoca teve como resultado uma linguagem muito normalizada que relacionava os conceitos dentro de classes e que determinava entre eles, então, uma sintaxe dependente das relações hierárquicas. Esse modelo, proveniente das ciências mais formalizadas, tinha o substantivo como forma de representação privilegiada (CABRÉ, 1993).

Na organização destas linguagens é representativa a classificação decimal idealizada por Dewey (1979), que analisava a totalidade do conhecimento humano e incluía-lo dentro de umas divisões decimais, seguindo a idéia de aspirar a uma expressão universalmente válida que desse uma visão geral dos conhecimentos de todo o mundo. Tentava graduar as disciplinas científicas mediante uma seqüência que partia do nível mais amplo da classificação do saber humano de Francis Bacon (1980), passando pelos níveis meios, segundo os conceitos do século XIX<sup>2</sup>, até chegar aos níveis mais específicos (CUTTER, 1972, p. 128). Refletia-se assim que, até as primeiras décadas do século XX, as ciências tivessem se constituído por ramificação, por subdivisão sucessiva.

Coincidentemente apareciam nos Estados Unidos, o mesmo ano de 1876, as *Rules for a dictionary catalog de Cutter* (1962), pré-amostra das linguagens controladas, com caráter

---

<sup>1</sup> Os conceitos aqui contribuídos foram atendidos com maior profundidade e rigor no livro "*Linguagens documentárias e vocabulários semânticos para a Web. Elementos conceituais*" que está para aparecer graças à Editora da UFBA. A origem mais concreta deste artigo são os esquemas mostrados num curso ministrado conjuntamente para o DT/SIBI-USP e o PPGCI/ECA na primavera do ano 2010.

<sup>2</sup> A ciência moderna se constituiu a partir da metodologia racionalista proposta por Descartes. "A ideia base é que o todo pode ser reconstituído a partir das partes, ou seja, de que o todo é igual à soma das partes". Pombo, O. Dispersão e unidade: Para uma poética da simpatia, em Ginez de Lara, M. L.; Smit, J. W. (2010) *Temas de Pesquisa em Ciência da Informação no Brasil*. São Paulo: Escola de Comunicações e Artes/USP. p. 31. <http://www.pos.eca.usp.br/sites/default/files/enancibdigital.pdf>

pré-coordenado, estrutura associativa e controle de vocabulário de aplicação específica. Sem dúvida a facilidade de uso para o usuário, frente à rigidez arbórea dos sistemas classificatórios, deu aos Cabeçalhos de assunto uma projeção para as linguagens controladas superior à qual teriam as classificações. Uma variação desta proposta nomeadora seriam os tesouros.

Já no pós-guerra mundial, Vannevar Bush (1945, p. 101-108) percebeu que o problema principal estava na inadequada forma de armazenar e recuperar a informação. Em “*As we may think*” raciocinou que a estrutura seqüencial dos conceitos documentários em classe-subclasse era a causa que as taxonomias alfabéticas ou numéricas fossem incapazes de processar adequadamente grandes acervos de informação, pois os sistemas hierárquicos ficavam superados pela associação de conceitos, imitando o modo em que as pessoas pensam (ROBERT, 1984, p. 273). Parece evidente que as categorias tradicionais da Biblioteconomia resultavam insuficientes para tratar o conteúdo dos documentos, pois as classificações universais contribuíam pouco à circulação efetiva dos conteúdos registrados.

Nas propostas recentes destaca-se a função comunicativa da língua, afastando-se das tentativas anteriores de fixar os termos como referentes ideais e homogêneos ao serviço preferente da comunicação entre expertos (KRIEGER; BOCORNY, 2004). Nesta visão pós-moderna emprega-se um léxico mais próximo da linguagem natural, determinado pelos contextos de uso e a causa de que o usuário tenha-se voltado o sujeito da interpretação. Agora a condição dos termos é mais cambiante. Esta relativização pelo cenário ou situação do discurso, favorece as dúvidas sobre o sentido monossêmico (com o que se reabilita conseqüentemente a polissemia), e sobre a variação vista como perturbação da unidade lingüística. O que levou a que as relações associativas passem a ter predomínio na sintaxe destas novas linguagens documentarias.

A respeito da evolução da web, observam-se duas trajetórias da Web: a Web Semântica e a Web Social (W2.0). A relação entre ambas é confusa, mas as duas se originam de mecanismos para compartilhar informação e recursos. A Web 2.0, proposta por O'Really ao constatar a evolução da Web no ano 2004 (EUZENAT; TRONCY, 2004, p. 157-188), mostra um alto grau de colaboração e vontade de compartilhar recursos. O usuário é o centro de atenção: decide o que quer e como utilizá-lo. A representação semântica é pouco complexa e apresenta dificuldade de criar aplicações para utilizá-las porque não estão normalizadas.

Assim a indexação livre e colaborativa de palavras-chave em etiquetas (tags), denominada folksonomia, com muitos problemas de sinonímia e polissemia.

A proposta de Tim Berners conhecida como Web Semântica tenta trazer as taxonomias e ontologias para o primeiro plano (PALACIOS et al., 2006. p. 313-322). As ontologias, de semântica muito formalizada, estão dirigidas a aplicações de software, mas ainda são pouco legíveis por pessoas (RDF e OWL estreitam as possibilidades) e difíceis de serem desenvolvidas. Mesmo assim, a Web Semântica obrigou a adaptar-se às linguagens preexistentes: Tesouros; Linguagem livre (Folksonomias, agora); e Mapas conceituais (para os *Topic maps*).

### **Revisão do conceito de tesouro**

Tesouro é uma lista de descritores (termos controlados) que representa os conceitos de um domínio do conhecimento. Os termos tem sentido semântico por aparecer contextualizados. A inferência terminológica prove das relações semânticas estabelecidas entre os termos simples ou sintagmáticos da lista. Essas relações organizam a estrutura do tesouro. Origina-se na própria estrutura (classificação + hierarquia), assim como nas associações entre os descritores. A sintaxe dos termos é determinada pelas normas de:

- Ordenação (parte classificatória).
- Subordinação (parte hierárquica).
- Associação (completa a parte alfabética) reflete a concepção mais comunicativa da linguagem (FOSKETT, 1997, p. 111-134.).

Mediante as relações semânticas se representa o pensamento e se estabelecem associações entre os termos do corpus, permitindo aos usuários navegar coerentemente pelos conceitos que representam (DEJEZ, 2000), integrando termos dentro de uma mesma categoria, determinando diferenças de uma espécie com outras, possibilitando a divisão ou análise do gênero pelas diferenças e concedendo definições (*Scope note*) para ajustar o significado dos termos.

O emprego dos tesouros tem sido muito bem-sucedido, mas a ambigüidade, riqueza e capacidade de inovação constante das linguagens nas quais se expressam os documentos

sobre os que trabalham produzem, em ocasiões, uma falta de pertinência na seleção de documentos. Ao que se acrescenta o problema do número de documentos que circulam pela Web e a variedade de seus suportes.

As estruturas definidas nos tesauros só servem para recuperações que contenham aquelas relações estáticas com as quais trabalha e que foram definidas a priori. Por isso colocam dois inconvenientes: A dificuldade para descrever campos com um nível de abstração alto (o da informática, por exemplo), com predomínio das ações (documentos audiovisuais) ou documentos sem nenhum tipo de estrutura (textos de livre formato) (FENTON, 2010, p. 187-205). Ao que se soma o custo elevado na sua criação, manutenção e funcionamento, evitáveis só com a automatização. Existem muitos inconvenientes para fazê-lo, devido à riqueza das estruturas semânticas dos textos.

Se o seu grau de coerência é geralmente alto, a sua eficácia comprovada, e são um bom ponto de partida para criar ontologias, porem não tem muitos mecanismos para compartilhar informação na Web, até a aparição de SKOS (*Simple Knowledge Organization System*). Ao que se acrescenta que tem dificuldades para definir as relações e que são vocabulários que não mostram relação direta com os objetos. Outro inconveniente é a sua criação por consenso prévio dos criadores, não dos usuários.

Nestas condições fazia-se obrigatória a evolução dos Standards sobre tesauros para verdadeiros mapas de representação do conhecimento. Esta transformação leva implícita a criação de relações adaptáveis a cada domínio, assim como a inclusão de grande variedade de descritores cuja categoria gramatical contribui matizes na semântica do mapa conceptual e de novas categorias relacionais que ligam os recursos de informação e as facetas. Sem dúvida alguma, é determinante e definitivo o caso das normas Britânica BS 8723 (2005-2007)<sup>3</sup> e ANSI/NISO Z39.19: 2005<sup>4</sup>. Suas duas primeiras partes são equivalentes as normas ISO 2788 y BS 5723. Mais, a terceira, *Vocabularies other than thesauri*, atende a outras linguagens documentarias: taxonomias, ontologias, cabeçalhos de assunto, esquemas de classificação e redes semânticas. E a quarta, *Interoperability between Vocabularies*, refere se a interoperabilidade na recuperação da informação: *mapping*.

---

<sup>3</sup> BSI Group. **Structured vocabularies for information retrieval**: guide. London: BSI, 2005-2007. (BS 8723/1-4).

<sup>4</sup> Z3919:2005. ISO. ANSI/NISO. Z39.19-2005. **Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies**. Disponível em: <<http://www.niso.org/standards/index.html>>. Acesso em: 11 jan. 2011.

As principais novidades são que no lugar de documentos, falam de objetos de conteúdo: entidades que contêm informação. Para referir-se aos descritores e termos preferentes, usa termo: palavra que representa um conceito.

As linguagens controladas mantêm finalidades herdadas, junto a outras de nova aparição. Assim, continuam traduzindo do vocabulário natural aos termos, ajudam na busca dos objetos de conteúdo, e garantem a consistência ou uniformidade na assinação dos termos. Mas agora fazem indicação explícita das relações semânticas entre os termos; os introduzem mediante etiquetas, e proporcionam hierarquias para visualizar os objetos de conteúdo desejados.

Distinguem quatro tipos de linguagens, segundo a sua complexidade:

- Listas.
- Anéis de sinônimos.
- Taxonomias.
- Tesouros.

Lista: É um grupo simples de termos preferentes sem estrutura e acostumam apresentar-se em ordem alfabética ou outra seqüência lógica.

Anéis de sinônimos: Sua utilização supõe acabar com a finalidade das relações de equivalência dos tesouros. Ao primar o conceito (o significado) sobre a expressão (o significante), não importa as diferentes possibilidades de expressão dum conceito. Só interessa que esse conceito seja bem entendido num determinado contexto de uso.

Taxonomias: Definem categorias temáticas gerais que agrupam termos ordenados numérica ou alfabeticamente. Classificam ou categorizam um conjunto de coisas de forma hierárquica. Por definição, taxonomia é classificação, sistemática. Uma hierarquia é basicamente uma estrutura em forma de árvore que estabelece uma relação entre os objetos de generalização-especialização.

Tesouros: abrange os grupos de relações (Tesouros, Ontologias e *Topic maps*), de equivalência, hierarquia e associação: agora estas relações têm um incremento pragmático, sobre tudo as de associação (TUDHOPE; ALANI; JONES, 2001). Assim aceitam-se conceitos associados pela aparição simultânea dentro de uma mesma definição, o quando

mostram contigüidade na expressão dum texto, e chegasse até a considerar relações associativas não especificadas ou associações sem identificar.

### **A revalorização das Taxonomias.**

Neste contexto voltou a ter destacada importância o emprego de Taxonomias, mesmo que agora mostrem características diferentes. As taxonomias estão presentes nos Esquemas de classificação, Tesouros, Modelos conceituais e Ontologias. Podemos considerar que uma Taxonomia é uma lista de termos preferenciais com estrutura hierárquica que ordena a informação utilizando a relação de generalização ou “tipo de”. Trata-se de uma estrutura em forma de árvore que estabelece uma semântica simples entre os objetos de generalização-especialização (ISO, 2005, p. 9).

O principal valor atual é servir como instrumento para a organização e recuperação de informação nas empresas e instituições. Pois organizam e administram os recursos de informação digitais que, como organizações complexas, alojam nos seus servidores Web, procurando categorizá-los e navegar não só nos conteúdos próprios delas, mas também nos seus serviços, produtos e recursos humanos (DACONTA; OBRST; SMITH, 2003, p. 145). Compõem uma rede semântica de conceitos inter-relacionados para cobrir as necessidades empresariais e a forma com que os trabalhadores relacionam-se com a informação.

Devem valorar-se como meios de acesso, pois atuam como mapas conceituais dos assuntos solicitados em um serviço de recuperação: buscando categorizá-las, folheá-las e navegar por elas. Sua primeira função tem, então, mais de visualização do que recuperação (MOREIRO, J. A. et al., 2006, p. 17).

As taxonomias mostram uma estrutura classificatória hierárquica (contexto conceitual), com determinados níveis de profundidade. Dispõem de umas entradas etiquetadas orientadas ao usuário que dão nome aos conceitos (PETERS; STOCK, 2010, p. 81-93). São os elementos da estrutura, agrupados em blocos temáticos. Admitem facetas não hierárquicas que dão dinamismo e pluralidade de enfoques aos assuntos (CAMPOS; GOMES, 2008).

Nos tesouros os termos têm justificação documentária. Nas taxonomias o conteúdo adapta-se aos usos e à situação dos usuários. Permitem aos próprios usuários classificar as matérias usando hierarquias. Nelas, os termos cobrem o contexto de aplicação e são fáceis de

modificar, pois facilitam acrescentar dados, além de explicitar um modelo conceitual do domínio. Atendem às linguagens particulares de uma instituição que não possui tesouros e classificações e conectam a experiência e características do pessoal próprio de cada instituição.

As taxonomias mostram uma desvantagem, são restritas nas suas possibilidades de exploração, ao conter apenas relações hierárquicas e partitivas. Assim, têm que ser complementadas com mecanismos de busca, se quer acessar a outras relações.

### **As Ontologias como KOS.**

Nos anos 90, Internet e seu enlace hipertextual de documentos levaram para uma maior diferenciação na representação dos conteúdos. Fez aos tesouros evoluir para características mais conceptuais ou até a precisar a utilização de verbos. Teve como consequência a imediata dinamização do tesouro com maior número de relações, agora explicitamente identificadas. Melhorava assim a representação e aumentava a precisão e eficácia. Voltavam-se instrumentos de navegação e visualização, a través dos grafos explícitos de conexões pela informação. Os tesouros caminhavam para a sua transformação em ontologias e *Topic maps*.

As três primeiras camadas da Web Semântica (sintática, semântica e ontológica) dão as bases para poder representar a informação de maneira semanticamente acessível. As outras três (lógica, de prova e de confiança) definem os elementos que permitem aos agentes de software autenticar e comprovar a confiabilidade dos diferentes elementos do modelo (recursos, agentes, inferências obtidas, etc.).

Para que a camada ontológica possa inferir conhecimento é necessário contextualizar a informação dentro de um cenário concreto. Nesta camada é onde se pode definir uma ontologia.

Originariamente ontologia é um conceito filosófico que estuda o ser em geral, suas propriedades transcendentais e, mais especificamente, a organização da realidade. Na representação de informação pode entender se em quanto soma de conceitos relevantes do conhecimento que compartilham os membros de um domínio dado, junto as Relações que se

estabelecem entre esses conceitos e os Axiomas definidos sobre estes conceitos e relações (NECHES et al., 1991, p. 36-56).

Pelo que trabalham com conceitos relacionados, como os tesauros e redes semânticas, mas permitem definir relações semânticas complexas, regras e axiomas que não estão presentes no resto de KOS (GUARINO, 1995, p. 625-640).

Uma ontologia persegue os objetivos de permitir a compressão comum da estrutura da informação de um domínio tanto por pessoas como por agentes de software (NOY, 2000), assim como permitir a reutilização do conhecimento pertencente a um domínio. Também a de fazer explícitos os pressupostos (axiomas) e o alcance dum domínio. Separar o conhecimento do domínio relativo a seus componentes do conhecimento operacional (relativo aos processos ou tarefas que se desenvolvem nesse domínio). Finalmente, o de analisar o conhecimento num campo, e os termos e relações que o configuram formalmente.

Ontologia é uma conceitualização formal de um domínio ou de uma parcela de realidade, com a qual podem operar diferentes aplicações software:

- Os conceitos ou termos utilizados para a descrição servem como vocabulário comum (sintático e semântico) que favorece a comunicação e a interoperabilidade de recursos.
- Dão sentido pleno à informação ao situá-la dentro de um contexto.

As Ontologias supõem um salto qualitativo, pois de usar motores de busca (com palavras-chave), se passaria a usar agentes de software inteligentes (com conceitos). Possibilitam ir da mera recuperação de informação a obter respostas precisas a consultas concretas (recuperar conhecimento). São utilizadas por pessoas, bases de dados e aplicações que necessitam de partilhar informação sobre um domínio específico da realidade:

- Uma ontologia define os termos utilizados para descrever e representar uma área de conhecimento.
- Incluem definições de conceitos básicos no domínio, e relacionamentos entre eles.

A palavra ontologia usa-se para descrever SKOS com diferentes graus de estrutura. Vão desde simples taxonomias (como hierarquia Yahoo), esquemas de metadados (como o Dublin Core), até as teorias lógicas. A Web Semântica precisa de ontologias com um grau

significativo de estrutura. É necessário especificar descrições para os seguintes tipos de conceitos:

- Classes em muitos domínios de interesse.
- Relações que podem existir entre coisas.
- Propriedades que descrevem as características e atributos (*slots* ou *roles*) que essas coisas ou conceitos podem ter; e restrições sobre os *slots* (facetas-restrições de um rol).

Nos tesauros tradicionais as relações não podem ser nós (descritores), porém nas ontologias uma relação pode ser um conceito (o que leva a considerar aos verbos como conceitos). Este aumento quase ilimitado das relações, si comparado com a situação marcada pelos tesauros, é fruto da maior flexibilidade terminológica. O que faz que uma ontologia, junto com o conjunto de indivíduos de classe que a compõe, constitua uma base de conhecimento. Os componentes que representam o conhecimento do domínio numa ontologia são (GRUBER, 1995, p. 907-928):

- Conceitos: idéias básicas que se intentam formalizar: classes de objetos, métodos, planos, estratégias, processos de raciocínio, etc.
- Relações: representam a interação e enlace entre os conceitos do domínio (Taxonomia e mapa do domínio): subclasse - de, parte - de, etc.
- Funções: um tipo concreto de relação, onde se identifica um elemento mediante o cálculo de uma função que considera vários elementos da ontologia. Por exemplo, podem aparecer funções como categorizar-classe, definir - dado, etc.
- Instâncias: para representar objetos determinados dum conceito.
- Axiomas: assertivas sobre relações que devem cumprir os elementos da ontologia. Por exemplo: “Se A e B são da classe C, então A não é subclasse de B”, etc.

As ontologias oferecem algumas vantagens sobre os tesauros, pois apresentam um nível mais alto de descrição do vocabulário e maior desenvolvimento semântico das relações entre os conceitos (CAMPOS; GOMES, 2006). Baseiam se na lógica descritiva e favorecem o raciocínio inferencial dos agentes de software (GARCÍA JIMÉNEZ, 2004, p. 79-95). São reutilizáveis e podem trabalhar em sistemas heterogêneos e permitem a descrição formal de

objetos, suas propriedades e as relações entre eles. Porém é notório a distancia decrescente entre os novos tesouros para a web e as ontologias, pela maior riqueza semântica dos novos tesouros (GILCHRIST, 2003, p. 7-18.), pela sua maior diversidade de relaciones, e pela aparição de vocabulários de etiquetado com maior capacidade expressiva como *RDF Schema* ou *SKOS Core*.

### **Tesouros conceituais e redes semânticas. Os Topic maps.**

Maniez (1993) determinou o tesouro Conceitual como um modelo de relações associativas que agrupa termos e conceitos reais no contexto específico do usuário. Falava assim de uma rede semântica conceitual na qual cada nó contém um único conceito semântico com uma possível série de descritores associados, que podem identificar-se na rede de descritores relacionados segundo as relações: preferenciais, hierárquicas ou associativas dos tesouros.

Para melhorar a precisão das recuperações os tesouros conceptuais aproveitam ontologias específicas, por domínios, desde as que distinguem os sinônimos, controlam os homônimos e induzem a relações associativas entre os descritores. Uma ontologia para uma base de conhecimentos deve abranger os diferentes tipos de documentos, as descrições conceituais, as relações entre ditos documentos (citações), e as de estes com os diferentes problemas científicos; ademais de índices, descrições bibliográficas, tesouros, códigos classificatórios, formalizações de validade, informação terminológica, etc. Sugerem sempre aos usuários uma ampliação das buscas e supõem uma melhora dos resultados da recuperação, pois permitem buscar mediante uma grande quantidade de termos, entre os próprios dos usuários, os procedentes da base de dados e inclusive de indexações previas. Além disso, os usuários controlam o processo de navegação mediante o tesouro, perante a rigidez existente nos procedimentos automáticos.

Outra proposta para melhorar tesouros é incluir verbos que complementem aos tradicionais tesouros estáticos de substantivos (LEVIN, 1993). Originada na necessidade de melhorar a precisão quando se trata, por exemplo, de reutilizar software ou de recuperar imagens em movimento, o uso de descritores verbais oferece vantagens como a possibilidade de indexar imagens de vídeo mediante gerúndios, a de identificar mediante verbos associações funcionais muito más adaptáveis a domínios concretos, a de mostrar a relação existente entre

dois conceitos usando as possibilidades inúmeras da linguagem natural (categorias verbais a modo de relaciones facetáveis) e, finalmente, contribuem na determinação da não ambigüidade conceitual.

Relacionar conceitos mediante verbos em trabalhos de recuperação oferece mais possibilidades semânticas que nos tesouros tradicionais, pois podem se mostrar os agentes que interagem com o sistema e como o fazem (casos de uso); ao tempo que permitem também mostrar os diagramas de seqüência e atividade.

A integração verbal procede da área pedagógica, dos mapas conceituais. Contam com um antecedente direto, a base de dados de referencia léxica *WordNet*, que dispõe de relações de equivalência e de hierarquia entre diferentes categorias gramaticais, para tirar a ambigüidade dos conceitos, em especial mediante o uso de verbos. A justificativa para utilizar verbos na procura e recuperação documentaria prove da necessidade de navegar de modo hipertextual mediante estruturas cognitivas numa Web e Bases de Dados a texto completo, com problemas associados a procura da informação e a subsequente sobrecarga informativa e desorientação dos usuários.

A convergência destes métodos desembocou numa evolução dos padrões sobre tesouros para verdadeiros mapas de representação do conhecimento. Esta transformação leva implícita a criação de relações adaptáveis a cada domínio, assim como a inclusão de grande variedade de descritores cuja categoria gramatical contribui matizes na semântica do mapa conceptual e de novas categorias relacionais que ligam os recursos de informação e as facetas. A diferença dos tesouros e das ontologias, as Redes semânticas não definem a estrutura dos conceitos mediante hierarquias.

Estas redes representam-se graficamente mediante nós orientados, com um nó inicial e um nó final. A relação existente entre os conceitos se define numa direção determinada. Assim podem-se visualizar os conceitos macroestruturais<sup>5</sup>. Sua base é que o conhecimento representa-se por frases simples e estruturadas:

- Sujeito - verbo - predicado;
- Como uma associação: conceito → relação → conceito

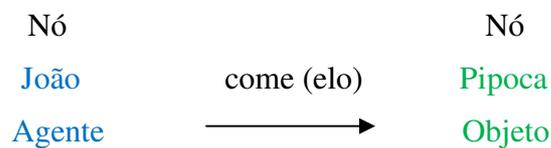
---

<sup>5</sup> Os *Mapas conceituais* são uma técnica desenvolvida por Joseph Novak nos anos 60 para ajudar a ensinar e aprender nas aulas Ausubel, D.P.; Novak, J.D. e Hanesian, H. (1980) *Psicologia Educacional*. Rio de Janeiro: Interamericana, 1980.

Os Mapas conceituais representam o conhecimento em gráficas cognitivas ou redes de conceitos, sem nenhum controle de termos (FARIA, 1995). Compõem-se de:

- Nós (pontos / vértices) que representam conceitos.
- Elos de união (*arcs*: arcos / *edges*: extremos, satélites) que representam as relações entre os conceitos.

Os nós de uma Rede Semântica estão unidos por elos, que indicam a relação que existe entre eles, assim a expressão “João come pipoca” tem dois nós: João e pipoca. “João” é um nó agente, “pipoca” é um nó objeto. Agente e objeto estão enlaçados pelo elo verbal “come”. De forma que estas relações podem se mapear assim:



Um documento ou domínio do conhecimento pode representar-se por um conjunto de conceitos inter-relacionados e dispostos como um grafo. Os conceitos podem representar entidades concretas ou abstratas, eventos ou coisas; e distinguir entre processos, procedimentos ou produtos.

Os mapas conceituais oferecem uma rede de relações mais rica que os tesouros. Na rede, os conceitos organizam-se do mais geral para o mais específico. Cada par de conceitos fica enlaçado mediante linhas etiquetadas o que faz explícito o tipo de associação. As associações entre conceitos não estão normalizadas, mas podem referir-se a:

- Relações hierárquicas.
  - Conjunto – Membro do conjunto;
  - Todo - Parte;
- Relações específicas (x influencia y; é perto de; leva a; etc.);
- Relações físicas (sobre, debaixo, perto...);
- Relações temporais (antes, depois);
- Relações lógicas (causa - efeito);

A evolução mais concreta das redes semânticas concretizou-se nos *Topic Maps*, onde os conceitos têm atingido a cima da representação. Trata-se de um sistema padronizado para descrever estruturas de informação e melhorar a localização dos conceitos e objetos de informação na Web semântica. Num vasto mar de informação, precisam-se sistemas capazes de navegar dentro de um contexto de busca e recuperação de informações. Os motores que movem esse navio são a precisão e facilidade de busca.

Um *Topic map* é um documento, ou um conjunto de documentos SGML ou XML inter-relacionados em um espaço multidimensional no qual as localizações são *topic*<sup>6</sup>. Seus antecedentes são os índices e tesouros, aos que incorpora um modelo de dados flexível e uma sintaxe padronizada que permite acessar aos conceitos mediante aplicações que a fazem funcional. Os *Topic maps* são uma norma para descrever estruturas de informação representadas por estes elementos fundamentais:

- *Topics*
- Relações entre *Topics*
- Enlaces com os documentos que os contem (ocorrências) (PEPPER, 2002).

Os *Topics* identificam-se pelas suas propriedades, não pela sua denominação terminológica; tem controle de vocabulário ad hoc; as relações são feitas usualmente mediante verbos; a tipologia de relações tem um número potencialmente infinito; uma relação pode ser um *topic* (verbos como *topic*); as variações ortográficas são normalizadas *a posteriori*; os *topics* se determinam mediante suas ocorrências e associações.

Os elementos dos *Topic maps* estruturam a informação ao redor dos *topics* (uma matéria = um *topic*) ou de referências digitais para uma matéria. Um *Topic* pode ter um ou mais nomes (*Topic name*): formais, simbólicos, apelidos, referências cruzadas, etimologias, traduzidas, etc

Os *Topic* mostram estes *roles*: Relações entre *topics*: *Associations*. As relações estão caracterizadas e classificadas. Existem classes de relações, a diferença de tesouros e taxonomias. Grupos de objetos de informação (ocorrências das matérias que os *topics* cosificam). *Ocurrence role*: função do fato. Contribui informação a uma matéria. *Ocurrence*

---

<sup>6</sup> ISO, 2000. *ISO/IEC 13250: 2000. SGML-Topic Maps*.

*type*: relevância de um fato em relação com sua matéria são os fatos do *Topic*, os recursos aos que está ligado.

Finalmente, determinam um espaço multidimensional (rede) no qual as localizações (nodos, vértices,...) são *Topic*. A distância entre *topics* se mede pelo número de *topics* visitáveis para passar desde um *Topic* até outro.

### **Folksonomias. A renovada vigência das palavras**

O neologismo Folk + sonomy = gente + taxonomia foi proposto por Thomas Vander Wal como modo de indexação cooperativa que adota palavras-chave para representar a informação de maneira espontânea e livre (VANDER WAL, 2007). O contexto vem determinado pela arquitetura da informação social (Web 2.0), em que cada usuário escolhe o que mais se ajusta a seu entendimento.

Nas folksonomias não é essencial construir uma hierarquia, e sim contar com um instrumento que permita o emprego livre de etiquetas semânticas para efetuar as indexações. O fato de abandonar as taxonomias por listas de palavras-chave não é algo novo. Agora, os bloggers preferem descrever seus textos com palavras-chaves livres. Beneficiam-se do vocabulário pessoal e da contribuição social para ajudar a encontrar e controlar na web os assuntos mais interessantes. As folksonomias incluem cada palavra proposta. Não deixam nenhuma palavra de fora e não há uma autoridade central que imponha sua visão, de forma que cada voz tem seu lugar. Marcar com etiquetas de caráter social pode ser considerado “uma ponte de baixa inversão entre a indexação pessoal e a indexação compartilhada” (PORTER, 2005).

Então, Para que servem as folksonomías? As folksonomías refletem exatamente o modelo de conceitos que a gente tem. É um sistema simples e emergente, que se afasta da maneira tradicional de classificar desde em cima para abaixo. Sua vantagem prove da capacidade de associar as necessidades verdadeiras dos usuários e a língua, não de buscar sua precisão. Mesmo si fosse uma moda, tanto no que respeita ao significado como ao próprio termo, as folksonomías permitem estudar que termos são os mais empregados e como evoluem de acordo com as tendências e segundo as comunidades que os empreguem.

Os vocabulários controlados não são aplicáveis à maioria dos casos onde se pode usar metadados. Sua construção, manutenção e aplicação têm custo alto tanto em tempo de desenvolvimento como de preparação dos usuários.

### **Relato da experiência dum curso ministrado à bibliotecárias do DT/SIBi-USP e aos alunas do PPGCI/ECA.**

O curso titulado Bases e modelos da evolução dos vocabulários documentários: dos tesouros aos mapas semânticos foi organizado pelo Programa de Pós-Graduação em Ciência da Informação e pelo SIBI-USP, e foi ministrado na Escola de Comunicações e Artes da Universidade de São Paulo entre os dias 27 de setembro e 01 de outubro de 2010, e teve um valor de 2 cr. Tem sido os responsáveis pelo curso a Profa. Marilda Lopes G. Lara (USP) e o Prof. José Antonio Moreira (*Universidad Carlos III de Madrid*). O curso contou com 15 funcionárias do SIBI inscritas, mais 05 alunas da Pós-graduação em Ciência da Informação. Acho que foi proveitosa a integração dos docentes e alunos do doutorado ECA com os profissionais das bibliotecas (SIBi-USP) pois permitiu a troca de experiências e a atualização profissional de pessoas interessadas nesta temática, sempre num ambiente harmonioso e o mais distendido possível.

A abordagem do curso efetivou-se em cinco sessões, cada uma delas com uma atividade pratica que facilitasse a compreensão dos conceitos atendidos. Procurou-se dar a oportunidade de discutir os assuntos mostrados e de elucidar as dúvidas surgidas.

O curso atendeu a os seguintes itens:

1. Bases epistemológicas da representação da informação. As estruturas simbolizáveis a partir da convergência do modelo retórico e o modelo textual.
2. Modelos lingüísticos de recuperação de informação: tesouros e terminologia.
3. As novas normas sobre tesouros. Usos e condicionantes da geração de tesouros conceituais. Seu relacionamento com as ontologias.
4. Os tesouros de ações, os mapas de conceitos e os *Topic Maps*. E, finalmente,
5. Volta à indexação livre: as folksonomias.

A proposta quis alcançar os objetivos de:

- Conhecer a evolução e os princípios que regem os métodos da representação semântica dos documentos;
- Atingir elementos de julgamento que permitam apreciar a qualidade na representação semântica;
- Mostrar perspectivas na investigação aplicada à análise do conteúdo dos documentos.

A mais de perseguir estes objetivos específicos de aprendizagem:

- Distinguir as vantagens e inconvenientes que proporciona o uso das linguagens documentais e da linguagem livre em Documentação;
- Aproximar aos métodos próprios do tratamento semântico dos documentos;
- Mostrar perspectivas na investigação aplicada à análise automática do conteúdo documental;
- Conhecer as técnicas sobre as que se estabelecem os novos modelos de linguagem documental associativo e visual;
- Adquirir os conhecimentos que lhes permitam compreender como se constroem e mantêm as Taxonomias, as Ontologias e os *Topic Maps*.

Ao tratar-se de um assunto de atualização contou com a vantagem de ser ministrado para profissionais conhecedoras das linguagens coordenadas. Muitas delas tinham trabalhado com cabeçalhos de assunto ou com tesouros. O que as fazia possuidoras de boas destrezas, pois contavam com uma clara ideia dos conceitos fundamentais, assim como manejavam com bom ar as competências profissionais necessárias. As alunas da Pós-Graduação tinham concluído praticamente as disciplinas do seu currículo, o que as concedia uma visão global e enriquecedora do curso ao possuir um bom nível de conhecimentos com madureza de raciocínio.

O curso quis servir para a atualização, mesmo que os três últimos dias tivessem um nível de especificidade grande. O que dificultou um pouco este objetivo foi o intenso horário noturno, após as jornadas de trabalho. Se o curso tivesse sido realizado noutro período talvez pudessem ter tirado mais proveito. Era cansativo para as bibliotecárias atender a conteúdos densos de uma matéria ampla depois de todo um dia de trabalho. As necessárias atenção e observação exigiam um esforço suplementar.

## Conclusões

O eixo da proposta conceitual baseou-se na diferenciação entre “Conceitos, termos-descriptores, e palavras” tanto na hora de formar os elementos essenciais para cada uma das linguagens estudadas, como na diferenciação de profundidade cognitiva requerida por cada uma delas na hora de representar a informação bem perto da expressão significante ou, melhor, perto do significado a organizar. A riqueza das propostas com maior conteúdo semântico teve como símbolo o estabelecimento de ricas redes semânticas ou associações de conceitos e termos até faz pouco tempo de caráter muito limitado pelas normas surgidas nos três ou mais décadas.

Os novos tesouros constituem redes de conceitos nas que cada conceito semântico relaciona-se com uma possível série de conceitos associados. Aproveitam ontologias específicas para contextualizar os conceitos e incluem verbos que ativam a significação dos descritores.

As novas normas de construção de tesouros abrangem outros muitos tipos de linguagens documentarias, e consideram a interoperabilidade entre eles. Estabelecem a passagem desde os descritores até os termos representativos dos conceitos contidos nos objetos de informação. A mais de incrementar amplamente as possibilidades de relação entre os termos, que agora se fazem explícitas.

Os mapas conceituais utilizam a integração verbal para relacionar os conceitos, ao tempo que constituem verdadeiros mapas de representação do conhecimento.

Nos *Topic Maps* os conceitos identificam-se pelas suas propriedades. A denominação terminológica não tem importância; em tanto as relações são feitas usualmente mediante verbos.

Não é fácil construir e manter uma taxonomia ou uma linguagem associativa. Neste sentido as folksonomias são uma solução quando a indexação tradicional não é viável.

## Referencias

- BACON, F. **Instauratio Magna. Novum Organum. Nueva Atlántida**. México: Porrúa, 1980.
- BARROS, L. A. Aspectos epistemológicos e perspectivas científicas da terminologia. **Ciência e Cultura**, v. 58, n. 2, p. 22-26, 2006.
- BENVENISTE, E. **Problemas de lingüística general II**. Madrid: Siglo XXI, 1989. p. 221.
- BUSH, V. As we may think. **Atlantic Monthly**, v. 176, p. 101-108, 1945.
- CABRÉ, M. T. **La terminología: teoría, metodología, aplicaciones**. Barcelona: Antártica, 1993.
- CAMPOS, M. L. A.; GOMES, H. E. Taxonomia e classificação: a categorização como princípio. **Datagramazero: revista de ciência da informação**, v. 9, n. 4, 2008.
- CUTTER, B. A. Dewey Decimal Classification. **Encyclopedia of Library and Information Science**, New York, v. VII, p. 128, 1972.
- CUTTER, Ch. **Rules for a dictionary catalog**. 4. ed. London: Chaucer House, Malet Place, 1962.
- DACONTA, M.C.; OBRST, L. J.; SMITH, K. T. **The semantic web**. a guide to the future of XML, web services, and knowledge management. Indianapolis: Wiley, 2003. p. 145.
- DEJEZ, D. **Le thésaurus, concept ancien, usages nouveaux**. Paris: AINTD, 2000.
- DEWEY, M. **Decimal classification and relative index**. 19. ed. Albany: Forest Press, 1979.
- EUZENAT, J. ; TRONCY, R. Web sémantique et pratiques documentaires. In : LE MOAL, J. C.; HIDOINE, B.; CALDERAN, L. (Ed.). **Publier sur internet**. Paris: ADDBS, 2004. p. 157-188.
- FARIA, Wilson de. **Mapas conceituais: aplicações ao ensino, currículo e avaliação**. São Paulo: EPU, 1995.
- FENTON, C. Use of controlled vocabulary and thesauri in UK online finding aids. **Journal of the Society of Archivists**, v. 31, n. 2, p. 187-205, 2010.
- FOSKETT, D. J. Thesaurus. In: SPARCK JONES, K.; WILLET, P. (Ed.). **Readings in Information Retrieval**. San Francisco: Morgan Kaufmann, 1997. p. 111-134.
- GILCHRIST, A. Thesauri, taxonomies and ontologies: an etymological note. **Journal of Documentation**, v. 59, n. 1, p. 7-18, 2003.
- GRUBER, T. Toward principles for the design of ontologies used for knowledge sharing. **International Journal of Human-Computer Studies**, 1995, 43. pp. 907-928.

GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. **International Journal of Human and Computer Studies**, v. 43, n, 5/6, p. 625-640, 1995.

ISO. ANSI/NISO. Z39.19. **Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies**. Bethesda, Maryland: NISO Press, 2005. p. 9.

KRIEGER, M. G.; BOCORNY, M. J. **Introdução à terminologia: teoria e pratica**. São Paulo: Contexto, 2004.

LEVIN, B. **English verb classes and alternations: a preliminary investigation**. Chicago: The University of Chicago Press, 1993.

MANIEZ, J. **Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales**. Madrid: Pirámide, Fundación Germán Sánchez Rupérez, 1993.

MOREIRO, J. A. et al. Categorización de los conceptos en el análisis de contenido: su señalamiento desde la Retórica clásica hasta los Topic Maps. **Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información**, v. 20, n. 40, p. 17, 2006.

NECHES, R. et al. Enabling technology for knowledge sharing. **AI Magazine**, v. 12, n. 3. p. 36-56, 1991.

NOY, N.; MCGUINNESS, D. **Ontology development 101: a guide to creating your first ontology**. Stanford KSL Technical Report KSL-01-05, 2000.

PALACIOS, V. et al. An improved methodology for semantic scheme qualification. In: MoMM2006 & iiWAS2006 Workshops, 2006, Indonesia. **Frontiers in Mobile and Web Computing: Proceedings...** Österreichische Computer Gesellschaft, 2006. p. 313-322.

PEPPER, S. **The TAO of topic maps**. 2002. Disponível em: <<http://www.ontopia.net/topicmaps/materials/tao.html>>. Acesso em: 18 ago. 2005).

PETERS, I.; STOCK, W. "Power tags" in information retrieval. **Library Hi Tech**, v. 28, n. 1, p. 81-93, 2010.

PORTER, J. **Controlled vocabularies cut off the long tail**. 2005. Disponível em: <[http://bokardo.com/archives/controlled\\_vocabularies\\_long\\_tail](http://bokardo.com/archives/controlled_vocabularies_long_tail)>. Acesso em: 18 ago. 2010.

ROBERT, N. The pre history of the information retrieval thesaurus. **Journal of Documentation**, v. 40, n. 4, p. 273, 1984.

TUDHOPE, D.; ALANI, H.; JONES, C. Aumenting thesurus relationships: possibilities for retrieval. **Journal of Digital Information**, 2001. Disponível em: <<http://jodi.ecs.soton.ac.uk/Article/v01/i08/Tudhope>>.

VANDER WAL, Thomas. **Folksonomy coinage and definition**. 2007. Disponível em: <<http://vanderwal.net/folksonomy.html>>. Acesso em: 08 ago. 2009.

## Bibliografia

BOURDON, F. Référentiels, données d'autorité, thésaurus, ontologies, taxonomies... Pour en savoir plus! **Documentaliste: sciences de l'Information**, v. 45, n. 2. p. 14-15, Aug. 2008.

BREITMAN, K. K.; CASANOVA, M. A.; TRUSZKOWSKI, W. **Semantic web: concepts, technologies and applications**. London: Springer-Verlag, 2007.

BUSH, V. As we may think. **Atlantic Monthly**, v. 176, p. 101-108, 1945.

CAMPOS, M. L. A.; GOMES, HAGAR, E. Taxonomia e classificação: a categorização como princípio. **Datagramazero: revista de ciência da informação**, v. 9, n. 4, ago. 2008 Disponível em: <[http://www.dgz.org.br/ago08/Art\\_01.htm](http://www.dgz.org.br/ago08/Art_01.htm)>.

CÔTÉ, J. A. Knowledge taxonomies. **Information outlook: the monthly magazine of the Special Libraries Association**, n. 9, p. 45-52, 2005.

DACONTA, M. C.; OBRST, L. J. Y; SMITH, K. T. **The semantic web: a guide to the future of XML, web services, and knowledge management**. Indianapolis: Wiley, 2003.

DAMME, C.; HEPP, M.; SIORPAES, K. Folksontology: an integrated approach for turning folksonomies into ontologies. In: **ESWC 2007, Workshop Bridging the Gap between Semantic Web and Web 2.0, 2007. Proceedings...** Austria, 2007. p. 71-84.

EÍTO BRUN, R. Topics maps y la indización de recursos electrónicos en la web. **El profesional de la Información**, v. 12, n. 2, p. 141- 148, 2003.

EUZENAT, J. ; TRONCY, R. Web sémantique et pratiques documentaires. In: LE MOAL, J. C. ; HIDOINE, B. ; CALDERAN, L. (Eds.). **Publier sur internet**. Paris: ADBS, 2004. p. 157-188.

GILCHRIST, A. Thesauri, taxonomies and ontologies- an etymological note. **Journal of Documentation**, v. 59, n. 1, p. 7-18, 2003.

GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. **International Journal of Human and Computer Studies**, v. 43, n. 5/6, p. 625-640, 1995.

GRUBER, T. Toward principles for the design of ontologies used for knowledge sharing. **International Journal of Human-Computer Studies**, v. 43, p. 907-928, 1995.

IWANSKA, L. M.; SHAPIRO, S. C. **Natural language processing and knowledge representation: language for knowledge and knowledge for language**. Menlo Park: AAAI Press: MIT, 2000.

JAGERMAN, E. J. **Creating, maintaining and applying quality taxonomies**. Zoetermeer: E. J. Jagerman, 2006.

MAZZOCCHI, F. et al. Relational semantics in thesauri: some remarks at theoretical and practical levels. **Knowledge Organization International Journal**, v. 34, n. 4, p.197-215, 2007.

MOREIRO GONZÁLEZ, J. A. La representación y recuperación de los contenidos digitales: de los tesauros conceptuales a las folksonomías. In: TRAMULLAS, J. (Coord.). **Tendencias en Documentación digital**. Gijón: TREA, 2006. p. 81-109.

PEPPER, S. **The TAO of topic maps**. Disponível em:  
<<http://www.ontopia.net/topicmaps/materials/tao.html>>.

SÁNCHEZ CUADRADO, S. et al. De repente, ¿todos hablamos de ontologías? **El Profesional de la Información**, v. 16, n. 6, p. 562-568, 2007.

TUDHOPE, D., ALANI, H., JONES, C. Aumenting thesaurus relationships: possibilities for retrieval. **Journal of Digital Information**, 2001. Disponível em:  
<<http://jodi.ecs.soton.ac.uk/Article/v01/i08/Tudhope>>.

WAL, T. **Explaining and showing broad and narrow folksonomies**. 2004. Disponível em:  
<<http://www.vanderwal.net/random/entrysel.php?blog=1635>>. Acesso em: 12 jan. 2010.

\_\_\_\_\_. **Folksonomy coinage and definition**. Disponível em:  
<<http://vanderwal.net/folksonomy.html>>. Acesso em: 08 ago. 2009.

Artigo submetido em: 14 dez. 2010

Artigo aceito em: 26 jan. 2011