

CTTO: Semelhanças e Diferenças


Classificações, Tesouros, Taxonomias e Ontologias

-
- ▶ Classificações, tesouros, taxonomias e ontologias estruturam, classificam, modelam e representam conceitos e seus relacionamentos pertinentes num domínio do conhecimento.




-
- ▶ São formados, basicamente, por vocabulários controlados acompanhados de relacionamentos semânticos entre os termos.
 - ▶ Desempenham a função de organizar e recuperar informações. No entanto, como veremos a seguir, os quatro sistemas mantêm pontos em comum, mas diferenciam em alguns aspectos.
-



-
- ▶ Vários tipos de bases de conhecimento:
 - ▶ Esquemas de classificação e tesouro, Dicionários e ontologias desenvolvidos para aplicações de Inteligência Artificial, sistemas linguísticos ou definições de elementos de dados.
 - ▶ Apesar de se desenvolverem com diferentes objetivos, os diferentes tipos de bases de conhecimento sobrepõem-se muito e seguem princípios e métodos muito similares para sua construção.
-
- 

-
- ▶ Uschold (2010) destaca os seguintes aspectos comuns acerca de vocabulários, taxonomias, tesouros, ontologias e meta-modelos:
 - ▶ São diferentes abordagens para ajudar a estruturar, classificar, modelar e/ou representar conceitos e relações de uma área de interesse de uma comunidade.
 - ▶ Estabelecem um conjunto de termos que a comunidade concorda em usar para referir-se a estes conceitos e relações.
 - ▶ Especificam o significado dos termos em algum nível.
 - ▶ São noções não muito bem definidas e usadas de diferentes maneiras por diferentes indivíduos e comunidades.
-



-
- ▶ Ponto comum as seguintes funções fundamentais:
 - ▶ Eliminar ambiguidades
 - ▶ Controlar sinônimos
 - ▶ Estabelecer relacionamentos (hierárquicos e associativos)
 - ▶ Apresentar propriedades (dos conceitos)
-
- 

-
- ▶ Há diferenças entre os SOCs quanto ao
 - ▶ nível de especificação do significado de cada termo e
 - ▶ quanto ao tipo de notação ou linguagem adotada para especificar o significado.
 - Fato que observamos, também, em relação às classificações, tesouros, taxonomias e ontologias.





Aspectos Comuns

Princípios de Classificação

- ▶ CTTOs compartilham os princípios de classificação, que orientam a estrutura sistemática e possibilitam o agrupamento de conceitos que possuem características comuns.



Rótulos ou Etiquetas

- ▶ Os CTTOs utilizam algum tipo de signo para representar os conceitos.



Rótulos ou Etiquetas

- ▶ Na classificação, o uso concomitante do código numérico e do termo (622. Engenharia de Minas).
- ▶ No tesouro e na taxonomia, o uso de termos (azeite, farinha, queijo e revistas, televisão, rádio).
- ▶ Na ontologia, o uso de uma forma de expressão compatível com padrões como XML I 1, XML Namespaces I 2, RDF I 3, e OWL I 4. Ex.: foaf:Group, foaf:Agent.



Rótulos ou Etiquetas

- ▶ Os exemplos não esgotam as possibilidades. Tesouros e taxonomias podem utilizar códigos numéricos ou alfanuméricos para representar os conceitos, mas, o que comumente se observa é que esses dois sistemas tendem a utilizar termos.



Rótulos ou Etiquetas

- ▶ As classificações são as que mais empregam os códigos numéricos. Já as ontologias necessitam expressar o conceito para que o computador possa identificá-lo, daí a necessidade de formalismos definidos pelos padrões da web semântica.



Padronização Terminológica

- ▶ Os CTTOs cumprem a função de padronizar os termos empregados para representar os conceitos. A padronização contempla: aspectos de forma, como, por exemplo, uso do gênero masculino e feminino, plural e singular, sigla e nome por extenso; controle de sinonímias e controle de homografias. A diferença está, mais uma vez, na maneira como cada sistema efetua a padronização.



Padronização Terminológica

- ▶ Para elaboração de tesouros, há normas internacionais que definem os aspectos de controle terminológico (ISO 2788 (1986), ISO 5964 (1985) e ANSI / NISO Z39.19 (2005) e que determinam, entre outras convenções, o uso da relação de equivalência, representada pelos símbolos USE e UP para controle de sinônimos e o uso de qualificadores I5 para controlar homografias.



Padronização Terminológica

- ▶ As ontologias contam com padrões definidos pela OWL, entre os quais destacamos, no caso dos sinônimos, as relações `equivalentClass`, utilizada para indicar `==` classes sinônimas; `equivalentProperty`, que indica a sinonímia entre propriedades de conceitos ou classes e a relação `sameAs`, empregada para criar diferentes nomes que referem ao mesmo indivíduo.

(fonte: <http://www.w3.org/TR/2004/REC-owl-features-20040210/#sameAs>)



Padronização Terminológica

- ▶ Para as classificações e taxonomias, no entanto, não há padrão definido acerca da padronização terminológica. Nesse caso, sobretudo em relação às taxonomias, observamos alguns problemas terminológicos.



Padronização Terminológica

- ▶ Ao analisar taxonomias navegacionais, observamos, por exemplo, a ausência de padrões relativos ao uso de termos em português e de abreviaturas e siglas, como na taxonomia de livrarias virtuais, em que encontramos alguns exemplos: Pocket Books e Franchising; Adm. Financeira / Mat. Finan. ; Dic. Francês/Port., (AQUINO; CARLAN; BRÄSCHER, 2009). Numa rápida consulta a taxonomias de sítios de comércio eletrônico o leitor encontrará diversos problemas relativos à forma dos termos.



Especificação do Conceito

- ▶ CTTO apresentam recursos para delimitar o significado que compreendem, desde o simples agrupamento de termos em classes ou categorias adotado pelas classificações, até as complexas definições de propriedades/atributos e formalismos empregados pelas ontologias.



Especificação do Conceito

- ▶ As classificações e taxonomias, de maneira geral, não incluem definições precisas dos termos que arrolam, como nos exemplos do quadro I e da figura 4. No entanto, o posicionamento dos termos em determinadas classes ou cadeias hierárquicas delimita o significado. A inclusão de gol na classe de empresas aéreas, por exemplo, restringe o seu significado e exclui outras possibilidades, tais como marca de automóvel e evento do futebol.



Especificação do Conceito

- ▶ Nos tesouros os termos costumam ser acompanhados de definições expressas em linguagem natural. A norma ANSI/NISO Z39.19 determina que o escopo dos termos é restrito aos significados que possuem no domínio que o tesouro abrange, e, ainda, que os termos devem ser formulados de maneira a tornar claro, a qualquer usuário do tesouro, qual o escopo pretendido.



Especificação do Conceito

- ▶ Para tanto, o recurso utilizado em tesouros são as notas explicativas ou de escopo, que —devem exprimir o significado selecionado para o termo e, também, indicar outros significados reconhecidos na linguagem natural, mas que foram deliberadamente excluídos do vocabulário controlado (ANSI/NISO Z39.19, 2005, p. 22). A inclusão dos termos em categorias e seus relacionamentos são outros recursos que auxiliam na delimitação dos conceitos em tesouros, como no exemplo da Figura 2.



Especificação do Conceito

- ▶ O termo indexação é empregado em sentidos diferentes na Ciência da Informação e na Economia, portanto, é um termo homógrafo. O qualificador (economia) é utilizado para solucionar a homografia e já restringe o significado do termo. Mesmo que observássemos apenas a subclasse 330 – Economia – Finanças públicas ou as relações ‘termo específico’ e ‘termo relacionado’ e não atentássemos para o qualificador, já poderíamos inferir que se trata do conceito de indexação usado na economia.



Especificação do Conceito

- ▶ No entanto, a definição que acompanha o termo é o recurso que melhor cumpre a função de delimitar o significado. Assim como nos tesauros, os recursos de categorização, relacionamentos conceituais e definições estão presentes nas ontologias. No entanto, as ontologias dispõem de mecanismos mais complexos que permitem elencar todos os atributos ou propriedades de determinada classe ou conceito, restringir propriedades e definir sua cardinalidade. A figura 8 exemplifica os recursos mencionados no desenvolvimento de uma ontologia na ferramenta Protégé.



Relacionamento conceitual

- ▶ CTTO estabelecem relações entre conceitos. Os tipos de relacionamentos e a forma de representá-los variam segundo o sistema, sendo que os quatro contemplam a noção de hierarquia entre conceitos. Em tesouros e ontologias, as hierarquias seguem princípios rígidos, por outro lado, as classificações e taxonomias não possuem recursos para a diferenciação precisa dos tipos de relacionamentos hierárquicos.



Relacionamento Conceitual

- ▶ As diretrizes e normas de elaboração de tesouros definem os seguintes relacionamentos hierárquicos:
 - ▶ a) relações gênero/espécie (TG – termo genérico / TE – termo específico). Ex.: instituição de ensino superior / universidade
 - ▶ b) relações todo/parte (TGP - termo genérico partitivo / TEP – termo específico partitivo). Ex.:
universidade/faculdade/departamento
 - ▶ c) relação de instanciação (é um, ou IsA) ¹⁷. Ex.: universidade /Universidade de Brasília.



Relacionamento Conceitual

- ▶ Nas ontologias, a relação gênero/espécie orienta a hierarquia de classes, como vimos no exemplo da ontologia de vinho. Em tesouro teríamos que vinho é um termo específico de bebida, na ontologia temos que vinho é uma subclasse de bebida.



Relacionamento Conceitual

- ▶ Nas ontologias, a relação de instanciação ocorre pela ligação de cada indivíduo como membro de determinada classe. Assim, ainda no exemplo do domínio vinho, Casillero del Diablo, Cabernet Sauvignon, 2007 18 é uma instância da classe vinho tinto. A relação todo/parte, assim como outros tipos de relacionamentos entre conceitos podem ser especificados nas ontologias por meio das propriedades (rdf:Property, na OWL), as quais são definidas segundo os objetivos da ontologia.



Relacionamento Conceitual

- ▶ A identificação do relacionamento hierárquico em classificações e taxonomias é feita, de maneira geral, pelo posicionamento, na estrutura desses sistemas, de um conceito abaixo de outro, como nos exemplos a seguir, extraídos da taxonomia navegacional da loja virtual submarino (<http://www.submarino.com.br/>)



Referência

- ▶ BRÄSCHER, Marisa; CARLAN, Eliana. Sistemas de organização do conhecimento: antigas e novas linguagens. In: Jaime Robredo; Marisa Bräscher (Orgs.). Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento. Brasília DF: IBICT, 2010, 335 p. ISBN: 978-85-7013-072-3. Capítulo 8, p. 147-176 Edição eletrônica. Disponível em: <http://www.ibict.br/publicacoes/eroic.pdf>. (Edição comemorativa dos 10 anos do Grupo de Pesquisa EROIC).

