

TOPIC MAPS DIRIGIDOS POR ONTOLOGIAS¹

TOPIC MAPS DRIVEN BY ONTOLOGIES

**Renato P. de Azevedo², Heleno B. Cabral³ e
Giovani Rubert Librelotto⁴**

RESUMO

Topic Maps são usados para adicionar informação ao conceitual e semântica em documentos *Web* e ferramentas. Isso conduz a uma interpretação humana bem próxima do que os projetistas desejam. Ontologias são modelo de dados que representam um conjunto de conceitos dentro de um domínio e os relacionamentos entre eles. Dessa forma, o objetivo deste artigo é apresentar um relacionamento entre ontologias e *Topic Maps*, de forma a demonstrar que os *Topic Maps* podem ser construídos de forma a representar o conhecimento expresso em ontologias.

Palavras-chave: documentos *Web*, modelos de dados.

ABSTRACT

Topic Maps are used to add information to concepts and semantics in *Web* documents and tools. This conducts to a human interpretation pretty close to designers want. Ontologies are data models that represent a set of concepts within one dominion and the relationships among them. In this way, the goal of this article was to present a relationship among ontologies and *Topic Maps*, in order to show that *Topic Maps* can be constructed to represent the knowledge expressed in ontologies.

Keywords: *Web* documents, data models.

¹Trabalho de Iniciação Científica - UNIFRA.

²Acadêmico do Curso de Ciência da Computação - UNIFRA.

³Mestrado em Nanociências - UNIFRA.

⁴Orientador - UNIFRA.

INTRODUÇÃO

Uma ontologia é uma especificação ou *formalização de uma conceitualização* (GRUBER, 1993) de determinado universo de discurso, ou domínio de conhecimento. Uma conceitualização é um conjunto de conceitos e de relações entre si. Alternativamente, uma ontologia é uma teoria lógica, a qual dá uma explicação explícita de uma conceitualização, projetada para ser compartilhada por agentes (humanos ou computadorizados) com objetivos diversos (GUARINO; GIARETTA, 1995).

Uma ontologia difere de outros modelos de dados porque sua preocupação principal é com os conceitos e com os relacionamentos entre si, em que a semântica desses relacionamentos é aplicada uniformemente.

Em uma estrutura típica, os relacionamentos entre dados são específicos e toda interpretação é necessariamente executada por um programa que acessa os dados. No caso de uma base de dados, os relacionamentos entre eles são parcialmente representados por um esquema de dados; todavia, quase toda interpretação é executada por um gestor de base que os acessa. Um humano ou outro programa desprovido de conhecimento da semântica específica da estrutura de dados ou base de dados em particular não tem a mínima ideia do que aquilo representa.

Em uma ontologia, os relacionamentos são definidos formalmente e a semântica de um dado relacionamento é consideravelmente detalhada. Se esses relacionamentos possuem certos nomes apropriados que identificam seu significado, um humano visualizando uma ontologia pode entendê-la diretamente, como um programa pode assumir a semântica de um dado relacionamento e atuar sistematicamente por meio de toda a ontologia.

Resumidamente, ontologias atuam como modelos semânticos conceituais representando um conhecimento comum em um modelo bem definido, consistente, completo, extensível, reutilizável e modular. Dessa forma, este artigo visa à demonstração de como a norma ISO 13250 *Topic Maps* (PARK; HUNTING, 2003) pode ser utilizada para a representação de ontologias.

A estrutura deste artigo é dividida em 5 partes: enquanto a seção 2 tem como função introduzir os conceitos de ontologias e apresentar os passos para a construção de uma ontologia, a seção 3 apresenta a norma ISO 13250 *Topic Maps*. A seção 4 estabelece o que se entende por ontologia em *Topic Maps*. A abordagem *Topic Maps* dirigidos por ontologias é citada na seção 5. O artigo encerra com um sumário na seção 6.

CONSTRUÇÃO DE UMA ONTOLOGIA

Os principais elementos de uma ontologia são: a) conceitos (coisas); b) instância de conceitos; c) classe de conceitos; d) subclasse de conceitos; e) as propriedades desses conceitos; f) os relacionamentos entre os conceitos, que podem ser relações binárias ou de aridade superior a 2, também podem ser de qualquer tipo e g) restrições e regras sobre esses conceitos e as suas relações.

O processo para a construção de uma ontologia similar ao usado para construir um modelo orientado a objetos (BOOCH et al., 1999) ou um diagrama entidade-relacionamento (TEOREY, 1999) possui uma sequência de passos, ilustrados pelo exemplo que segue:

1. Primeiramente, define-se um universo de discurso, ou seja, escolhe-se qual o domínio que se pretende modelar. Inicia-se esse processo de definição do domínio pela listagem dos conceitos que se deseja incluir, isto é, pelos objetos significativos deste mundo em particular. Muitas vezes, essa tarefa implica em analisar profundamente o domínio, o que significa ler os documentos sobre esse universo em particular, de modo a identificar os temas, os verbos, os objetos e os adjetivos encontrados em suas definições.

Por exemplo, para criar uma ontologia sobre o conceito *Dinamização Científica*, deve-se incluir os temas: *instituição, evento, cidade, escola, universidade e capital* (no conjunto dos conceitos), entre outros, além de relações como *organiza e é organizado por*, e *XATA, Braga e Universidade do Minho* (no conjunto de conceitos concretos, físicos/existentes);

2. Após a fase anterior, identifica-se como os temas podem se relacionar entre si e se percebe, por exemplo, como os verbos identificados acima definem as relações entre os conceitos. Enquanto esse processo é realizado, pode-se descobrir temas que foram esquecidos durante o primeiro passo. Em caso afirmativo, adicionam-se esses novos conceitos e relações.

Nesse exemplo, o termo *organizar* é um relacionamento entre *evento e instituição*. Deseja-se expressar a ideia de que *uma instituição organiza um evento* e de que *um evento é organizado por uma instituição*.

3. Com base na análise realizada, nos passos 1 e 2, constrói-se a ontologia de forma manual ou usando uma ferramenta para gestão de ontologias,

como, por exemplo, o Protégé (PROTÉGÉ, 2005). Além disso, existem processos automatizados para a especificação de extração de *topic maps* baseados em ontologias, possibilitando sua extração automática e evitando a sua edição manual (LIBRELOTTO et al., 2006).

Um guia sobre como projetar uma ontologia está disponível em Gruber (1995).

TOPIC MAPS

A norma ISO 13250 *Topic Maps* (BIEZUNSKY et al., 1999) fornece uma especificação que permite representar conhecimento – em particular, o conhecimento conceitual – por meio do qual se distingue semanticamente os recursos de informação. O processo de criação de ontologias, aplicado à construção de *Topic Maps*, foca precisamente esse aspecto: enfatiza o projeto conceitual e a construção de *topic maps* que refletem corretamente a semântica do conhecimento implícito.

Além de funcionar como um índice para recursos de informação, *Topic Maps* permite expressar conhecimento, o que fornece uma grande vantagem para o projetista de *Topic Maps*, pois uma representação de conhecimento semanticamente rica e correta proporciona mais qualidade a ele. Pelo fato de uma ontologia ser uma representação de conhecimento, este capítulo descreve como ela desempenha um importante papel no projeto de *Topic Maps*.

ONTOLOGIA EM TOPIC MAPS

Ao analisar *Topic Maps*, percebe-se que tópicos e associações podem ser divididos em duas camadas: as que representam conceitos abstratos e as que representam conceitos concretos. A ontologia de um *topic map* basicamente é definida pelos conceitos abstratos, ou seja, que serão definidos como tipos de tópicos, tipos de associações, tipos de papéis de atuação, tipos de nomes e de contextos e tipos de ocorrências. Além disso, considera-se também papéis específicos de atuação relacionados aos tipos de associação, de ocorrência de determinados tópicos e regras implícitas de cardinalidade, etc..

A figura 1 demonstra uma representação esquematizada dessa visão. Essa figura apresenta um *topic map* para o conceito *Dinamização Científica*, que envolve conceitos como *instituição*, *evento* e *cidade*, além das subclasses desses conceitos: *universidade*, *escola*, *conferência*, *summer school* e *capital*.

Encontram-se relações entre os conceitos, tal como organizar, que relaciona os conceitos *evento* e *instituição*. Nessa associação, os papéis de atuação “organiza” e “é organizado por” indicam como cada membro da associação relaciona-se com os demais.

Todos esses conceitos (classes, subclasses (tipos de tópico), relações (tipos de associação) e papéis de atuação) formam a ontologia desse *topic map*.

Os tópicos restantes formam a base de conhecimento associada à ontologia, os quais compõem um conjunto de objetos de informação que permite organizar os conceitos e indicar os reais recursos de informação associados a eles (um objeto pode ter múltiplas ocorrências nos recursos de informação).

De acordo com o *topic map* representado na figura 1, os conceitos concretos desse domínio são: *Lisboa*, *Braga*, *XATA* e *UMinho*. É perceptível que nenhum desses conceitos possuem instâncias, ou seja, todos são nodos folhas na árvore hierárquica do *topic map*.

Para clarificar o exemplo, apresentam-se os tópicos *Instituição*, *Universidade*, *Universidade do Minho* e *XATA*, bem como a associação entre os dois últimos chamada organizar:

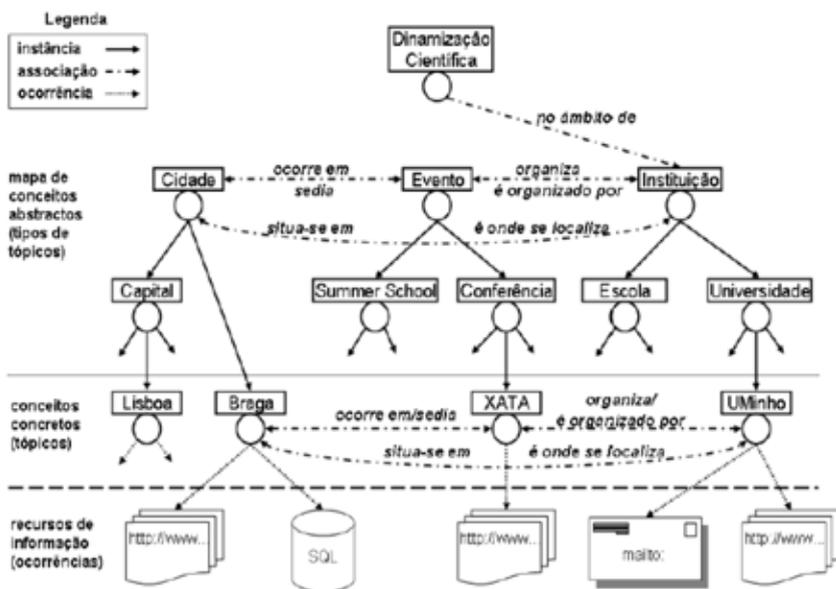


Figura 1 - Um *topic map* para o conceito Dinamização Científica.

```
1 <topicMap>
2   <topic id="instituicao">
3     <baseName>
4       <baseNameString>Instituição</baseNameString>
5     </baseName>
6   </topic>
7   <topic id="universidade">
8     <instanceOf>
9       <topicRef xlink:href="#instituicao"/>
10    </instanceOf>
11    <baseName>
12      <baseNameString>Universidade</baseNameString>
13    </baseName>
14  </topic>
15  <topic id="uminho">
16    <instanceOf>
17      <topicRef xlink:href="#universidade"/>
18    </instanceOf>
19    <baseName>
20      <baseNameString>U. do Minho</baseNameString>
21    </baseName>
22  </topic>
23  <topic id="xata">
24    <instanceOf>
25      <topicRef xlink:href="#conferencia"/>
26    </instanceOf>
27    <baseName>
28      <baseNameString>XATA</baseNameString>
29    </baseName>
30  </topic>
31  <topic id="organiza"/>
32  <topic id="organizado-por"/>
33  <association>
34    <instanceOf>
35      <topicRef xlink:href="#organizar"/>
36    </instanceOf>
```

```
37     <member>
38         <roleSpec>
39             <topicRef xlink:href="#organiza"/>
40         </roleSpec>
41     <topicRef xlink:href="#uminho"/>
42 </member>
43 <member>
44 <roleSpec>
45     <topicRef xlink:href="#organizado-por"/>
46 </roleSpec>
47 <topicRef xlink:href="#xata"/>
48 </member>
49 </association>
50 </topicMap>
```

Os 3 primeiros tópicos desse *topic map* na sintaxe XTM apresentam uma relação classe/subclasse de 3 níveis: *Instituição* (linha 2 a 6) é super-tipo de *Universidade* (linha 7 a 14), que por sua vez é super-tipo de *Universidade do Minho* (linha 15 a 22). O tópico *XATA* (linha 23 a 30) é subclasse de *Conferência* – o qual não se encontra nesse *topic map* para simplificá-lo. Os tópicos *organiza* (linha 31) e *organizado-por* (linha 32) são definidos como papéis de atuação em associações, pois os mesmos são referenciados na associação *organizar* (linha 33 a 49), a qual envolve os tópicos *UMinho* e *XATA*. Essa associação expõe que:

- UMinho organiza a XATA
- XATA é organizado pela UMinho

TOPIC MAPS DIRIGIDOS POR ONTOLOGIA

Esta seção introduz a noção de *Topic Maps dirigidos por Ontologia* (a partir de agora, referenciado apenas como *TMdO*) (PARK; HUNTING, 2003). Essa abordagem faz parte da tendência em posicionar ontologias no coração dos sistemas informação (GUARINO, 1998) e em determinar que o conhecimento é a entidade de principal importância e que as ontologias desempenham o papel central no projeto e na operação de sistemas de informação.

COMO ONTOLOGIAS SE RELACIONAM COM *TOPIC MAPS*

Um formato de anotação comum para designação de índices – objetivo inicial de *Topic Maps* – é um passo crucial em direção ao objetivo de se conquistar a interoperabilidade entre esquemas de índices; com isso, resta ainda obter-se a interoperabilidade semântica. Enquanto que a especificação *Topic Maps* garante interoperabilidade sintática, ontologias asseguram interoperabilidade semântica. Portanto, se *Topic Maps* forem construídos a partir de uma ontologia consistente, eles poderão oferecer interoperabilidade semântica.

Topic Maps podem ser gerados a partir de ontologias. Com a abordagem *TMdO*, a ontologia torna-se um excelente ponto de partida para a geração de *topic maps*.

VANTAGENS DA ABORDAGEM *TMdO*

Um *topic map* dirigido por uma ontologia torna mais simples o seu processo de criação e manutenção. O projeto de uma ontologia da qual pretende-se gerar um *topic map* deve ser realizado em separado do processo de construção de um *topic map*, pois o *topic map* expresso em XTM pode crescer gradualmente, visando a atender todos os temas relevantes para a representação do conhecimento de um domínio em particular. Dessa forma, se a ontologia de um dado *topic map* mantém-se inalterada, somente o mapeamento do domínio para o *topic map* deve ser realizado.

Uma mudança no mapeamento pode ser ocasionada por uma modificação nos requisitos do processo. Ao separar a ontologia e o *topic map*, permite-se que as mudanças conceituais possam ser efetivadas independente das demais mudanças. Nesse caso, a abordagem *TMdO* oferece as vantagens de uma típica abordagem *loose coupling*¹ (MARCH; OLSEN, 1976).

A abordagem *TMdO* disponibiliza o uso de uma grande quantidade de ontologias existentes, que são resultados de um investimento significativo, como são os *Topic Maps*. Tal abordagem poupa esforços na construção de um domínio para o qual um trabalho de representação de conhecimento tenha sido realizado previamente.

¹ O conceito de *loose coupling* é um conceito fundamental da programação. O código de qualidade deve manter os seus diversos componentes *loosely coupled*, ou seja, independentes entre si, mas com uma grande coesão interna. Por componentes, entende-se classes, assemblies, tudo o que possa ser considerada uma unidades lógica de funcionalidade.

Outro benefício das ontologias existentes é que muitas têm sido testadas e utilizadas com sucesso por várias aplicações. Exemplos na área médica incluem as ontologias encontradas nos projetos *Unified Medical Language System (UMLS)*² e *Generalised Architecture for Languages, Encyclopaedias and Nomenclatures in medicine (GALEN)*³. Nessa perspectiva, faz sentido a reutilização de porções relevantes dessas ontologias para construir *topic maps* sobre medicina, por exemplo.

A abordagem *TMdO* também tem como vantagem o fato de que as ontologias construídas para a criação de *Topic Maps* podem ser utilizadas para outros fins. Elas são construídas usando linguagens específicas para a representação do conhecimento, enquanto que *Topic Maps* foram criados especificamente para organizar recursos na *internet* (isto é, criando uma camada semântica sobre os recursos que permanecem inalterados). Uma ontologia simples, então, pode ser usada em várias aplicações além da geração de *Topic Maps*, como, por exemplo: motores de inferência, aplicações de linguagem natural, etc.. Dessa forma, é recomendado manter a conceitualização do domínio em uma ontologia, deixando a implementação real do domínio para um *topic map*.

MAPEAMENTO DE ONTOLOGIAS PARA *TOPIC MAPS*

Depois de relembrar a definição de ontologia, após uma breve apresentação sobre como construir ontologias, considerando a definição de *Topic Maps*, pôde-se determinar um mapeamento entre uma Ontologia e um *Topic Maps*. Assim, definiu-se o mapeamento:

- Conceito -> Tópico
- Instância de um Conceito -> Tópico
- Classe -> Tipo de tópico
- Subclasse -> Instância de tópico
- Propriedades das Classes -> Características dos tópicos
- Relacionamento -> Associação
- Axiomas que representam as condições -> Não existente na norma ISO 13250

² UMLS: <http://www.nlm.nih.gov/research/umls/>

³ GALEN: <http://www.cs.man.ac.uk/mig/projets/old/galen/brochure.html>

A partir desse mapeamento, percebe-se que quase todos os principais elementos de uma ontologia podem ser mapeados para elementos de *Topic Maps*. Entretanto, a norma *Topic Maps* não possui mecanismos para definir os axiomas que representam as condições a serem impostas sobre uma ontologia, pois ela necessita de um mecanismo que possibilite a validação de sua representação por meio de restrições semânticas. Essas restrições podem garantir que uma ontologia realmente representa o que o projetista tinha em mente.

Para resolver esse problema, definiu-se a linguagem TMCL (*Topic Maps Constraint Language*) (LIBRELOTTO et al., 2005). Essa linguagem permite a representação de restrições em ontologias expressas em *Topic Maps*.

AS TENDÊNCIAS PARA A ABORDAGEM *TMdO*

Obviamente, é possível criar *topic maps* de uma forma direta, especialmente se ele é pequeno e expressa uma quantidade limitada de conhecimento. Além disso, alguns *topic maps* não necessitam oferecer muito mais do que uma simples classificação de recursos de informação.

Contudo, ao se construir um índice simples para um documento ou um mapa de navegação sofisticado para um *web site*, implicitamente se expressa conhecimento. Assim, criando-se uma camada de conhecimento rica semanticamente (a ontologia), será adicionada qualidade significativa para o *topic map*.

A tendência aponta para a geração de *topic maps*, a partir de ontologias em que a edição mantém e gera o conhecimento. Elas poderão ser publicadas em conjunto com seus *topic maps*. Para aqueles *topic maps* que não forem gerados a partir de ontologias, deverão ser projetados com princípios baseados em ontologia para garantir consistência semântica dos seus termos.

Poderão ser encontradas bibliotecas de *topic maps*, assim como bibliotecas de ontologias para sua geração. Para uma dada ontologia, poderão haver múltiplos *topic maps*, em que cada um seria a imagem de um mapeamento distinto. Espera-se que ontologias normalizadas sejam disponibilizadas para vários domínios de conhecimento.

CONCLUSÃO

Este artigo descreveu o relacionamento entre ontologias e *Topic Maps*. A norma ISO/IEC 13250 *Topic Maps* (PARK; HUNTING, 2003) é uma norma para organização e representação de conhecimento sobre um domínio específico, que permite a descrição de temas e de seus relacionamentos. Assim como os *Topic*

Maps, ontologias também podem ser usadas para adicionar informação conceitual e semântica em documentos *Web* e ferramentas. Isso conduz a uma interpretação humana bem próxima do que os projetistas desejam.

O mapeamento expresso na seção 5 entre ontologias e *Topic Maps* é a principal contribuição deste trabalho. Nesse mapeamento, percebe-se que a norma ISO 13250 possui uma estrutura capaz de efetuar a representação de um domínio de informação quase que total. O que não é permitido representar nessa norma, até o momento, são as restrições.

Entretanto, quando a ontologia está representada em um *topic map*, deve-se ter mecanismos de validação do conhecimento nele representado. Para isso, essa validação fornecida por uma linguagem definida por Giovanni Librelotto (LIBRELOTTO et al., 2005). Com isso, a norma ISO 13250 consegue permitir a representação de conhecimento de acordo com a definição de ontologias.

Assim, toda ontologia pode ser mapeada para *Topic Maps*, independente da sua complexidade, pois permite um mapeamento de forma mais direta. Um trabalho futuro seria a criação de uma ferramenta que permitisse a criação automática de *Topic Maps* a partir de ontologias existentes, expressas em outras linguagens de representação de ontologias.

Até o presente momento, não é conhecida nenhuma ferramenta para a geração de *topic maps* a partir de uma ontologia. Por isso, o extrator automático de *topic maps* (escritos em XTM), a partir de um conjunto de recursos de informação heterogêneos, com base na descrição de uma ontologia subjacente, será considerado uma contribuição inovadora, pois esse sistema permite a geração de *topic maps* dirigidos por ontologias.

REFERÊNCIAS

BIEZUNSKY, M.; BRYAN, M.; NEWCOMB, S.; ISO/IEC 13250 - Topic Maps. ISO/IEC JTC 1/SC34. 1999. Disponível em: <http://www.y12.doe.gov/sgml/sc34/document/0129.pdf>.

BOOCH, G.; RUMBAUGH, J.; JACOBSON, I. **The Unified Modeling Language User Guide**. Reading, MA: Addison-Wesley, 1999.

GRUBER, T. R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In Guarino, N. and Poli, R., editors, **Formal Ontology in Conceptual Analysis and Knowledge Representation**, Deventer, The Netherlands. Kluwer Academic Publishers, 1993.

GRUBER, T. R. (1995). How to Design an Ontology. **WEB PAGE AS PART OF THE ONTOLOGIA GUIDED TOUR**. Acesso em: abr., 2002.

GUARINO, N. **Formal Ontology and Information Systems**. In: CONFERENCE ON FORMAL ONTOLOGY (FOIS98), 1998. Disponível em: <http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf>.

GUARINO, N.; GIARETTA, P. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In Mars, N., editor, **Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing**, pages 25–32. Ed. Amsterdam: ISO Press, 1995.

LIBRELOTTO, G. R.; RAMALHO, J. C. ; HENRIQUES, P. R. Constraining topic maps: a TMCL declarative implementation. **Extreme Markup Languages 2005**: Proceedings. IDEAlliance, 2005.

LIBRELOTTO, G. R. ; RAMALHO, J. C.; HENRIQUES, P. R. Metamorphosis - A Topic Maps Based Environment to Handle Heterogeneous Information Resources. 2006. **Lecture Notes in Computer Science**, volume 3873, pages 14–25. Springer-Verlag GmbH.

MARCH, J.; OLSEN, J. **Ambiguity and Choice in Organizations**. Bergen, Norway: Universitetsforlaget, 1976.

PARK, J.; HUNTING, S. **XML Topic Maps: Creating and Using Topic Maps for the Web**, volume ISBN 0-201-74960-2. Addison Wesleyl, 2003.

PROTÉGÉ. **The protégé contology editor and knowledge acquisition system**. 2005. Disponível em: <http://protege.stanford.edu/>.

TEOREY, T. **Database Modeling and Design**. San Francisco, CA: Morgan Kaufman, 3rd edition, 1999.