

Especificação e Processamento de Ontologias para a *Semantic Web*

Giovani Rubert Librelotto e Pedro Rangel Henriques

¹ UFMSM – Universidade Federal de Santa Maria
DELIC – Departamento de Eletrônica e Computação
Av. Roraima, 1000, Bairro Camobi, 97105-900, Santa Maria - RS, Brasil

²UM – Universidade do Minho
Departamento de Informática – Campus de Gualtar
Braga, Portugal, 4710-057

librelotto@inf.ufsm.br, prh@di.uminho.pt

Resumo. *Este minicurso consiste de uma visão geral sobre a Semantic Web. Serão abordadas as maneiras de representação de conhecimento (dicionários, índices, taxonomias, thesauri e ontologias), dando uma ênfase maior nesta última. Com o objetivo de permitir a especificação e o processamento de ontologias, introduzir-se-á as linguagens para representação de ontologias para a Semantic Web (RDF, RDF Schema, XOL, SHOE, OWL, Topic Maps e TMCL), assim como uma visão geral das ferramentas que suportam o desenvolvimento de aplicações e ontologias para a Semantic Web. Neste momento, será apresentado uma metodologia para o projeto de ontologias.*

1. Introdução

A *Semantic Web* é uma área emergente que tem como objetivo tornar a Web mais útil e fácil de usar. É uma evolução da Web atual na qual os recursos de informação estão associados a uma interpretação possibilitando que, além do tradicional utilizador humano, também o computador nela pode navegar.

A ideia de ter na Web dados definidos e relacionados para serem usados por máquinas, não apenas para apresentação a humanos, com vista à sua integração e partilha entre aplicações, ainda é um anseio. Para isto, os dados devem ser descritos de forma que o homem e a máquina entendam seu significado.

Basicamente, os mecanismos a serem desenvolvidos para o estabelecimento da *Semantic Web* compreendem duas vertentes: a disponibilização de uma coleção de dados estruturados; e regras de inferência associadas que permitirão a criação de ferramentas capazes de percorrer a Web realizando tarefas complexas com base nessas estruturas de conhecimento. Estas duas vertentes serão abordadas, recorrendo-se a ontologias para estruturar as fontes de informação e a linguagens de anotação para descrevê-las.

Este minicurso apresenta, portanto, as metodologias para a representação do conhecimento, seguindo-se de linguagens que permitam a sua representação de maneira formal. De modo a finalizar o minicurso, abordam-se ferramentas que permitam o processamento destas linguagens, de forma a construir aplicações para a *Semantic Web*.

2. Estruturando o Conhecimento

A *Semantic Web* é uma extensão da Web atual, que permitirá aos computadores e humanos trabalharem cooperativamente. Ela tem como finalidade conseguir atribuir um significado aos conteúdos publicados na Internet, de modo que seja perceptível tanto pelo humano como pelo computador [Shadbolt et al. 2000]. Idealizada em 2001, seu principal objetivo é desenvolver tecnologias e linguagens que tornem a informação legível para as máquinas. A finalidade passa pelo desenvolvimento de um modelo tecnológico que permita o compartilhamento global de conhecimento assistido por máquinas. A integração das linguagens XML, arquiteturas de metadados, ontologias, agentes computacionais, entre outras, favorecerá o aparecimento de serviços Web que garantam a interoperabilidade e cooperação [Ahmed et al. 2001].

Esses sistemas de organização de conhecimento são mecanismos que permitem uma estruturação da informação e podem ser empregados em casos onde exista uma grande coleção de dados, como museus, bibliotecas e arquivos. Nesta etapa do minicurso serão discutidas as principais formas de organização de conhecimento, os quais podem ser divididos nas seguintes classes:

Sistemas baseados em listas de termos: normalmente este tipo de lista tem uma estrutura simples para a representação de conhecimento. Como exemplos, tem-se:

- Dicionários;
- Índices.

Sistemas baseados em grafos: determinam associações entre os termos através de um conjunto de relações semânticas. Os exemplos mais conhecidos são:

- Taxonomias;
- Thesaurus;
- Ontologias.

As ontologias, os thesauri, as taxonomias, os índices e os dicionários são similares nos seguintes aspectos:

- São abordagens para estruturar, classificar, modelar e representar conceitos e relacionamentos pertencentes a algum tema de interesse significativo para alguma comunidade;
- Permitem que uma comunidade adote e use o mesmo conjunto de termos de um modo uniforme;
- O significado dos termos são especificados de alguma maneira a um certo nível.

Portanto, há uma forte conexão entre os modos de expressar conhecimento descritos nas seções anteriores. Os conceitos e seus relacionamentos são descritos e definidos em modos diferentes em ontologias, thesauri e taxonomia [Gruber 1993]. Diferentes aspectos de estrutura de conhecimento são implementados em cada um deles.

Uma ontologia descreve os conceitos existentes em um mundo particular; em uma ontologia, todos os conceitos relevantes para um certo domínio são definidos à custa de qualquer relação binária que se julga interessante citar, enquanto que nos thesauri só se usam certas relações específicas [Swartout and Tate 1999]. A taxonomia clarifica os relacionamentos hierárquicos entre os conceitos, criando uma estrutura de classes/subclasses. Os índices fornecem os apontamentos para os locais onde cada conceito é referenciado, enquanto que os dicionários fornecem as definições dos conceitos [Chandrasekaran 1999].

3. Ontologias e Semantic Web

Considere a estrutura de *Semantic Web* apresentada na Figura 1 [Berners-Lee et al. 2001]. Percebe-se que Unicode e URI (*Uniform Resource Identifier*) constituem a base para a legibilidade e o endereçamento na *Semantic Web*. Acima disto, há o XML e os *namespaces*. Esta camada é vista como a camada sintática.

Acima destes dois níveis entra-se no contexto da representação dos dados e dos metadados e o seu esquema. RDF (*Resource Description Framework*), juntamente com RDF Schema (RDFS), fornece uma estrutura bem adaptada para esta necessidade. A sugestão que tem sido feita é que a especificação Topic Maps também pode satisfazer esta necessidade.

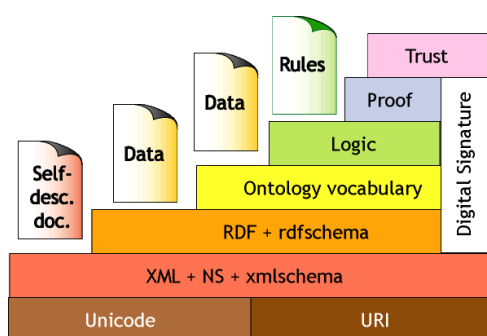


Figura 1. Arquitetura da *Semantic Web*. Fonte: [Berners-Lee et al. 2001]

No nível da camada ontológica, encontram-se linguagens que permitem a especificação de ontologias. O W3C está atualmente envolvido no desenvolvimento de um padrão para representação de ontologias, o *Web Ontology Language* (OWL) [Allemang and Hendler 2008], o qual é amplamente baseado no DAML+OIL [Davies et al. 2006]. OWL adiciona vocabulário para a descrição das classes e propriedades, como relações entre estas classes, cardinalidades, igualdades, tipos e características mais apuradas das propriedades e enumerações das classes.

Finalizando a visão da pirâmide da *Semantic Web*, há ainda as camadas superiores, as quais ainda estão sob desenvolvimento: a camada lógica (*Logic*) expressa conhecimento através de regras, enquanto a camada de prova (*Proof*) utiliza essas regras para inferir conhecimento. A camada confiança (*Trust*) fornece mecanismos para determinar o grau de confiança do conhecimento obtido. As assinaturas digitais (*Digital Signature*) introduzem várias camadas para garantir segurança, através do uso de codificação dos dados (*encryption*) e assinaturas digitais [Antoniou and van Harmelen 2008].

4. Especificação de Ontologias

O projeto de ontologias deve ser estruturado em uma série de passos [Hebeler et al. 2009]. O primeiro passo a ser realizado é a determinação do domínio e o escopo da ontologia desejada. Neste momento, é necessário responder uma série de questões, como: (a) qual será o domínio a ser coberto pela ontologia? (b) qual a finalidade da ontologia? (c) que tipo de questões devem ser respondidas com as informações nela existentes? Cabe ressaltar que as respostas a estas questões podem mudar ao longo do ciclo de vida da ontologia.

Após definido o escopo da ontologia, leva-se em conta a reutilização de outras ontologias. Isto deve ser tratado de forma a poupar esforços, interagindo com as ferramentas que utilizam outras ontologias e empregando ontologias que foram validadas através do uso em outras aplicações [Yu 2007].

O passo seguinte trata da enumeração dos termos importantes. Deve-se dar a devida importância para os termos sobre os quais irá se tratar, sobre as propriedades destes termos e o que espera-se que seja dito sobre cada um deles. Uma vez enumerado os termos, passa-se para a definição de classes e suas hierarquias. As classes são conceitos do domínio, representando uma coleção de elementos com propriedades similares. Tal como na orientação a objetos, os objetos do domínio são instâncias de classes.

O projeto de ontologias segue com a definição das propriedades e das restrições das classes. As propriedades são os atributos das instâncias da classe e a sua relação com as outras instâncias, podendo ser propriedades intrínsecas, extrínsecas, simples e complexas. Em relação a definição de restrições, determinam-se as condições para a existência de uma instância, ou seja, regras que todas devem seguir para serem consideradas válidas. A última etapa do projeto de ontologias é a criação das instâncias propriamente ditas [Passin 2004].

Neste minicurso, serão abordados detalhes em todos os passos previamente descritos, fazendo-se uma relação com as ferramentas citadas na seção seguinte.

5. Ferramentas para o Processamento de Ontologias

Após o detalhamento sobre as metodologias de representação do conhecimento e a apresentação das linguagens para representação de ontologias, esta seção discutirá as ferramentas que fornecem suporte a manipulação de ontologias.

Serão vistas ferramentas como o Protégé [Geroimenko and Chen 2006], o JENA, o OntoEdit, o Metamorphosis [Librelotto et al. 2008] e o KAON.

6. Conclusão

A *Semantic Web* é vista como uma nova geração da Web atual, a qual tem como finalidade atribuir um significado aos conteúdos publicados na Web de modo que seja perceptível tanto pelo humano como pelo computador. Para isto, as informações devem estar representadas de alguma forma que se possa extrair o conhecimento nelas inserido.

Desta forma, este minicurso teve por objetivo introduzir os principais conceitos relacionados a *Semantic Web*. Primeiramente, as definições de dado, informação e conhecimento foram clarificadas, de modo a evitar ambiguidade na interpretação de cada conceito. Em seguida, classificou-se e descreveu-se algumas das principais representações de conhecimento utilizadas pela comunidade acadêmica, sempre tentando manter um relacionamento entre estes métodos de representação.

A fim de permitir uma codificação do conhecimento a ser representado na *Semantic Web*, descreveu-se também as principais linguagens para a especificação de ontologias, no âmbito da *Semantic Web*.

Uma metodologia para a construção de ontologias também foi abordada neste minicurso. Propondo uma metodologia composta por sete etapas, parte-se desde a definição do escopo da ontologia desejada até a criação de instâncias da mesma.

A partir desta introdução aos conceitos relacionadas a *Semantic Web*, fornece-se o conhecimento necessário para que os participantes do minicurso possam entender mais detalhadamente o seu funcionamento e possam partir para o desenvolvimento de aplicações para a *Semantic Web*.

Referências

- Ahmed, K., Ayers, D., Birbeck, M., Cousins, J., Dodds, D., Lubell, J., Nic, M., Rivers-Moore, D., Watt, A., Worden, R., and Wrightson, A. (2001). *Professional XML Meta Data*. Wrox Programmer to Programmer Series.
- Allemang, D. and Hendler, J. (2008). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Morgan Kaufmann.
- Antoniou, G. and van Harmelen, F. (2008). *A Semantic Web Primer, 2nd Edition (Cooperative Information Systems)*. The MIT Press; 2 edition.
- Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. In *Scientific American*. <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>.
- Chandrasekaran, B. (1999). What Are Ontologies, and Why do We Need Them? In *IEEE Intelligent Systems and their applications*, volume vl 9, n 1. IEEE.
- Davies, J., Studer, R., and Warren, P. (2006). *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. Wiley.
- Geroimenko, V. and Chen, C. (2006). *Visualizing the Semantic Web*. Springer.
- Gruber, T. R. (1993). 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.
- Hebeler, J., Fisher, M., Blace, R., Perez-Lopez, A., and Dean, M. (2009). *Semantic Web Programming*. Wiley.
- Librelotto, G. R., Ramalho, J. C., Henriques, P. R., Gassen, J. B., and Turchetti, R. C. (2008). A framework to specify, extract and manage Topic Maps driven by ontology. In *SIGDOC 2008*. ACM.
- Passin, T. B. (2004). *Explorer's Guide to the Semantic Web*. Manning Publications.
- Shadbolt, N., Berners-Lee, T., and Hall, W. (2000). The Semantic Web Revisited. In *IEEE Intelligent Systems*, volume 21, pages 96–101. IEEE.
- Swartout, W. and Tate, A. (1999). Ontologies. In *IEEE Intelligent Systems and their applications*, volume vl 14, n 1. IEEE.
- Yu, L. (2007). *Introduction to the Semantic Web and Semantic Web Services*. Chapman & Hall/CRC; 1 edition.