

1

Introdução

Atualmente na *World Wide Web* tem-se um grande volume de informações disponibilizado sem uma forma estruturada de representação de conhecimento. Desta maneira, seu conteúdo é processado apenas por humanos; máquinas não obtêm suporte explícito para este tipo de tarefa.

Para fornecer suporte ao processamento por máquinas das informações disponíveis na Web – *World Wide Web*, ontologias vêm sendo utilizadas. Nas ontologias, informações são estruturadas em um vocabulário livre de ambigüidades e com um formalismo passível de processamento automático.

No entanto, além do processamento é desejável também que máquinas troquem informações. Para satisfazer esta necessidade, o problema de interoperabilidade semântica precisa ser resolvido. Parte deste problema, a interoperabilidade semântica na Web, é tratada ao longo do texto.

Questões relativas à evolução da Web para a Web Semântica, à interoperabilidade semântica e às ontologias são tratadas nos próximos tópicos deste capítulo. Por fim, um guia do leitor também é apresentado.

1.1. A Web Semântica

Pesquisadores da indústria e da academia vêm explorando a possibilidade de criar uma Web Semântica. Nesta nova Web, informações estarão organizadas de forma que máquinas processem e integrem seus recursos de maneira inteligente, possibilitando, por exemplo, buscas de informações mais rápidas e precisas, e facilitando a comunicação entre seus dispositivos heterogêneos. Além disso, através da estruturação e conjuntos de regras de inferência, informações poderão ser deduzidas automaticamente. Desta maneira, ao contrário da Web atual, o conteúdo da Web Semântica não será processado apenas por humanos, mas também por máquinas (Berners-Lee et al., 2001).

A Web Semântica não é uma nova Web desconectada da Web atual, mas sim, sua extensão. Seu desenvolvimento é um esforço colaborativo da *W3C* –

World Wide Web Consortium – (SemanticWeb, 2004) com a participação de um grande número de pesquisadores e parceiros da indústria.

A arquitetura da Web Semântica é definida como ilustrado na Figura 1 (Berners-Lee, 2000b). Nesta arquitetura, cada camada estende a funcionalidade e expressividade de suas camadas inferiores. A camada base refere-se ao uso de identificadores únicos para nomeação dos recursos Web, os *URIs* – *Uniform Resource Identifiers*. As camadas referentes às linguagens *XML* – *Extensible Markup Language* – (XML, 2004) e *RDF* – *Resource Description Framework* – (Miller et al., 2004) desempenham papéis fundamentais (Decker et al., 2000). A camada *XML* é responsável pela estruturação de documentos e a *RDF* pelo modelo de seus dados. No entanto, essas camadas são elementares e, por isso, são necessárias outras camadas de suporte às ontologias e à lógica para fornecer a semântica esperada para a evolução da Web atual.

Ontologias disponibilizam um vocabulário estruturado para explicar as relações entre seus diferentes termos e permitir as interpretações destes, livres de ambigüidades. A camada de ontologias na arquitetura da Web Semântica fornece um padrão bem definido para a estruturação de informação. A camada lógica é necessária para prover o formalismo passível de processamento automático (Fensel, 2001) como, por exemplo, permitir serviços de raciocínio por máquinas. Por fim, a última camada trata-se da “Web da Confiança”, onde as assinaturas digitais funcionam como um mecanismo de prevenção de inconsistências nessa Web Semântica (Berners-Lee, 2000a).

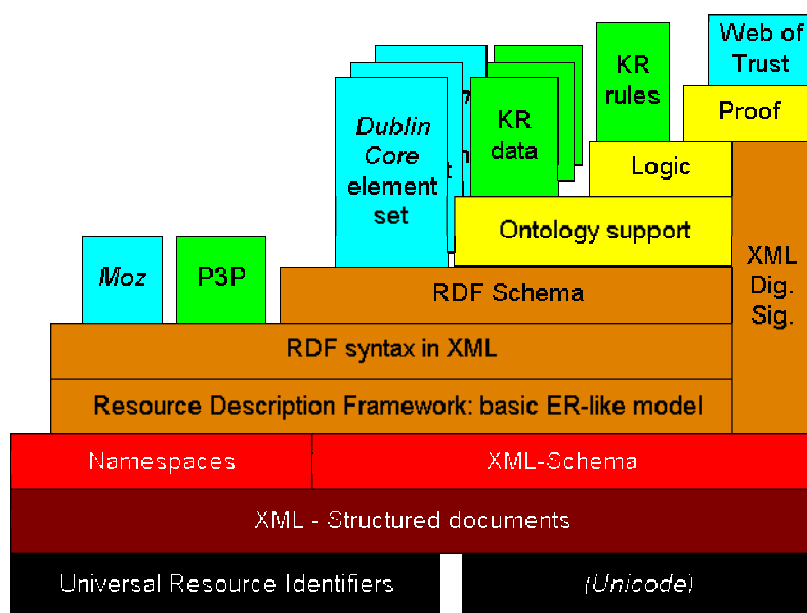


Figura 1 – Arquitetura definida para a Web Semântica em (Berners-Lee, 2000b)

De acordo com esta arquitetura definida, a Web Semântica pode ser entendida como uma reengenharia da Web atual, onde não só a linguagem *HTML – HyperText Markup Language* – (HTML, 2004) é utilizada para dar a formatação da informação disponível, mas também, haverá o uso de outras linguagens que garantam o entendimento comum de tal informação, agora estruturada de acordo com um padrão formal e bem definido. Desta maneira, não só humanos, mas principalmente as máquinas, poderão processar a informação de forma muito mais eficaz e eficiente. Essa evolução possibilitará tanto uma maior comunicação, i.e., troca de informações, entre as aplicações dessa nova Web quanto, conseqüentemente, a interoperabilidade semântica entre elas.

1.2. Interoperabilidade Semântica

Entende-se por interoperabilidade semântica a capacidade de dois ou mais sistemas heterogêneos e distribuídos trabalharem em conjunto, compartilhando as informações entre eles com entendimento comum de seu significado (Buranarach, 2001).

Para que as informações disponíveis sejam utilizadas pelos diferentes sistemas é necessário, em um primeiro momento, que estas sejam localizadas, acessadas e processadas por tais sistemas. Em um segundo momento, devido a sua heterogeneidade estrutural e semântica, a compatibilidade de seu conteúdo deve ser realizada (Moreira, 2003). Para auxiliar a satisfazer essas necessidades, com um esforço computacional reduzido, acredita-se que uma possível solução seja o uso de padrões, i.e., convenções (Buranarach, 2001).

Para convenções serem adotadas na solução de problemas é necessário, sobretudo, que estas sejam expressivas, inequívocas (sem ambigüidade) e bem aceitas, além de extensíveis. Mesmo satisfazendo esses requisitos, o processo de aceitação de uma convenção é demorado e, conseqüentemente, custoso. Desta maneira, as novas extensões procuram ser as extensões de convenções já adotadas. Estas extensões devem ser sucessivas e progressivas, fazendo uso, por exemplo, de soluções com arquitetura em camadas.

Soluções tradicionais com arquitetura em camadas, dadas para tentar garantir a compatibilidade de informações, são conhecidas na área de banco de dados como, por exemplo, soluções que fazem uso de mediadores e conversores.

No entanto, essas soluções funcionam bem apenas em um universo onde existam os mesmos tipos de estruturas para o armazenamento de informações e estas sejam previamente conhecidas. Assim, novas soluções com arquitetura em camadas para a compatibilidade de informações são desejadas.

Limitando os sistemas heterogêneos e distribuídos às diversas aplicações existentes na Web Semântica, a interoperabilidade semântica limita-se à interoperabilidade semântica de aplicações desta Web Semântica.

O uso de ontologias é uma das possibilidades mais promissoras para garantir a interoperabilidade semântica de aplicações Web. Isto porque é a convenção adotada para expressar as informações explícitas e implícitas destas aplicações de forma estruturada, além de fornecer um vocabulário comum com uma semântica bem definida.

1.3. Ontologia

Ontologia é um termo originário da filosofia usado para representar uma visão do mundo em um sistema de categorias. Como descrito em (Sowa, 2003), o assunto ontologias é o estudo das categorias de coisas que existem ou podem existir em algum domínio. O resultado deste estudo, denominado uma ontologia, é um catálogo dos tipos de coisas supostas a existir em um domínio de interesse, na perspectiva de uma pessoa. Esse catálogo é expresso por linguagens para ontologias. Em Ciência da Computação, a definição mais comum de ontologia é a proposta em (Gruber, 1993), onde uma ontologia é definida como a “*especificação explícita e formal de um conceito compartilhado*”.

Neste trabalho, é adotada a estrutura para descrição de ontologias apresentada em (Maedche, 2002), por estar em conformidade com a linguagem padrão atual para ontologias, a linguagem *OWL – OWL Web Ontology Language* – (Dean et al., 2004a). Esta estrutura é representada pela tupla $O := \{C, R, H^C, rel, A^O\}$, onde:

- C (conceitos) e R (relações) são dois conjuntos disjuntos;
- H^C é uma relação direcionada $H^C \subseteq C \times C$ que é chamada hierarquia de conceitos ou taxonomia. Por exemplo, $H^C(C1, C2)$ significa que $C1$ é um subconjunto de $C2$;

- rel é uma função $rel : R \rightarrow C \times C$ que relaciona não taxonomicamente conceitos;
- A^O é um conjunto de axiomas expresso em linguagem lógica apropriada.

Conceitos de um domínio com relacionamentos de especialização, i.e., relacionamentos do tipo *é-um*, são descritos em ontologias utilizando uma organização taxonômica. Aspectos de composição, i.e., relacionamentos do tipo *parte-de/todo*, são ortogonais às ontologias e devem ser representados por meio de funções não taxonômicas, por exemplo, propriedades.

Em um futuro próximo, acredita-se que ontologias serão utilizadas pela maioria das páginas da Web Semântica (Hendler, 2001). As principais razões para esta utilização, como citado em (Noy, 2001b), são:

- Compartilhar o entendimento da estrutura da informação entre pessoas e agentes de software;
- Possibilitar o reuso de conhecimento do domínio;
- Tornar as verdades absolutas do domínio explícitas;
- Separar o conhecimento do domínio do conhecimento operacional; e
- Analisar o conhecimento do domínio.

Ontologias podem possuir descrições mais gerais, utilizadas nas chamadas ontologias genéricas (também conhecidas como as *upper ontologies*) ou mais específicas, utilizadas nas chamadas ontologias de domínio de aplicação Web (também conhecidas como as *web ontologies*).

Como descrito em (SUOWG, 2004), uma ontologia mais genérica (*upper ontology*) limita-se aos conceitos que são meta, genéricos, abstratos e filosóficos. Conceitos específicos de um dado domínio não são incluídos nas ontologias genéricas. Assim, estas ontologias fornecem uma estrutura, i.e., como os conceitos estão organizados, e conceitos genéricos o suficiente para serem utilizados, em um nível elevado, na construção de outras ontologias de várias áreas de domínio.

As ontologias genéricas SUMO – *Suggested Upper Merged Ontology* – (SUMO, 2004) e OpenCyc (OpenCYC, 2004) são os exemplos mais conhecidos de ontologias genéricas (*upper ontologies*). Tais ontologias são padrões do grupo de trabalho de ontologias genéricas da IEEE – *Institute of Electrical and Electronics Engineers* – (SUOWG, 2004).

Apesar da quantidade de informações conseguida com o uso das ontologias genéricas, este trabalho trata **sempre** de ontologias específicas de domínios de aplicação. Isto porque prioriza o alinhamento das ontologias utilizadas para descrever as aplicações da Web Semântica, na visão da Engenharia de Software. Estas ontologias, parciais e contextualizadas, são as de aplicações Web, i.e., as *web ontologies* (Hendler, 2001).

Acredita-se que ontologias devem ser desenvolvidas localmente por engenheiros de software e não por especialistas em ontologias. Sendo assim, Breitman e Leite em (Breitman e Leite, 2004) defendem que a ontologia é um produto da engenharia de requisitos. Entende-se que é responsabilidade do engenheiro de requisitos modelá-la: primeiro, porque é durante o processo de definição do produto que o conhecimento do Universo de Informação¹ – UdI – é descoberto, i.e., elicitado; e segundo, porque a engenharia de requisitos tem um núcleo de conhecimento sobre os processos para captura, modelagem e análise de informações relevantes que pode auxiliar na tarefa de construção de ontologias.

A tarefa de desenvolvimento ou de reutilização de partes das ontologias existentes deve ser simples o suficiente de modo a permitir que tanto o engenheiro de software quanto as pessoas que não são especialistas em ontologias possam realizá-la.

Com este enfoque, um processo para construção de ontologias, centrado em uma estratégia de elicitação denominada *Léxico Ampliado da Linguagem – LAL* – (Leite e Franco, 1993), é proposto em (Breitman e Leite, 2003). Tal processo automatiza um grande número das tarefas de geração de ontologias, guiando o usuário a executar apenas aquelas em que sua intervenção é necessária. Uma instância deste processo foi implementada e é descrita em (Felicíssimo et al., 2003a).

No entanto, mesmo que uma ontologia seja criada segundo os preceitos da Engenharia de Requisitos ou dos conhecimentos de Modelagem Conceitual, ou tendo sua construção suportada por metodologias e métodos para criação de ontologias, como a metodologia de Grüninger e Fox, a *METHONTOLOGY*, entre tantas outras encontradas, por exemplo, em (Gómez-Pérez et al., 2004), persiste o

¹ Neste trabalho, entende-se por Universo de Informação, o local onde as informações ou fontes de informações sobre uma aplicação, que se deseja elicitar, são encontradas.

problema de como compatibilizar seus termos com os termos de outras diferentes ontologias. Neste cenário, a interoperabilidade de ontologias se torna fundamental.

1.4. Guia do Leitor

O segundo capítulo desta dissertação trata o problema de interoperabilidade de ontologias. Apresenta seus mecanismos, em especial, o mecanismo de alinhamento, que foi escolhido para este trabalho. As identificações priorizadas no alinhamento para a compatibilidade dos termos das ontologias comparadas são numeradas. Ainda neste capítulo, os requisitos para o alinhamento no contexto da Web Semântica e seus resultados considerados satisfatórios são apresentados. Em seguida, alguns trabalhos da literatura revisados sobre interoperabilidade de ontologias são descritos.

O terceiro capítulo apresenta a estratégia elaborada para o alinhamento taxonômico de ontologias, com suas três etapas de execução e sua implementação, o Componente para Alinhamento Taxonômico de Ontologias – CATO. Um exemplo simplificado foi criado para auxiliar o entendimento do alinhamento realizado pela estratégia.

O quarto capítulo apresenta três diferentes estudos de caso realizados para demonstrar o funcionamento da estratégia elaborada. O CATO alinhou as ontologias escolhidas em todos estes estudos de caso. A estratégia de seleção dessas ontologias e os resultados conseguidos com o alinhamento também são apresentados neste capítulo.

No quinto capítulo, as conclusões, as contribuições e algumas avaliações da estratégia são apresentadas. Também neste capítulo, os trabalhos futuros são discutidos.