

2

Web Semântica

De acordo com Berners-Lee (Berners-Lee, 1998) (Berners-Lee *et al.*, 2001), uma definição da *Web Semântica* é: uma extensão da *Web* obtida através da adição de semântica ao atual formato de representação de dados”. O principal propósito de haver uma *Web Semântica* é tornar a *Web* compreensível tanto por humanos quanto por agentes de software (Silva *et al.*, 2003) ou componentes (Szyperski, 1998). Neste sentido, se o conteúdo da *Web* fosse de fácil processamento por computadores (*machine processable*), as aplicações poderiam ter acesso a uma grande variedade de recursos, que poderiam ser compartilhados, integrados e processados para produzir resultados com mais valor agregado para o usuário.

A “base” da *Web* atual é a HyperText Markup Language (HTML), que possibilita a comunicação entre humanos já que estes conseguem entender o conteúdo de páginas criadas com essa linguagem. Benjamins *et alli* (Benjamins *et al.*, 2002), apresenta a *Web Semântica* como um meio de tratar o problema de sobrecarga (*overload*) de informação causado pelo crescimento contínuo da *Web* em tamanho, linguagens e formatos. Na *Web Semântica*, as páginas não apresentam somente um conjunto de palavras, figuras, tabelas e outros elementos, mas também o código e a estrutura dos seus significados, possibilitando assim o seu processamento eletrônico.

A representação formal de significado pode ser feita de várias maneiras. Uma dos mais antigos formalismos é o de *semantic networks*. Uma *semantic network* representa o conhecimento como um conjunto de nós (*nodes*) ligados por arcos nomeados (*labeled links*). O significado é inferido pelo modo como um conceito está conectado a outros. Outra abordagem são *frames systems* que são isomórficos a *semantic networks* (Heflin *et al.*, 2003). Uma outra forma de facilitar a expressão e justificativa de argumentos é através do uso de lógica formal. Em muitos ramos da lógica, um sistema consiste de:

- uma linguagem bem definida para a representação do conhecimento; e
- métodos bem definidos para inferência (*reasoning*).

Esses sistemas são limitados pelo tipo de conhecimento que podem representar e pelo tipo de inferência que podem realizar (Frost, 1986). Conseqüentemente, lógicos desenvolveram outros tipos de lógica para contornar estas restrições. Exemplos de tais ramos da lógica são lógica de predicado,

lógica de primeira ordem, lógica não-monotônica, lógica de descrição (*description logic*) entre outras. No caso da *Web*, restrições computacionais são uma das mais importantes e por essa razão houve a necessidade de escolher um formalismo de representação do conhecimento específico, por exemplo, um ramo da lógica, para implementar a *Web Semântica*.

Uma vez que este formalismo está escolhido, algum artefato precisa ser definido para conter o código e a estrutura do significado dos elementos na *Web Semântica*. Este papel é, de forma geral, o de uma ontologia. Na próxima seção, definições de ontologia são apresentadas com mais detalhes e nas seções seguintes é mostrado o relacionamento entre o W3C e a *Web Semântica* e também uma controvérsia que existe sobre uma das bases arquiteturais da *Web Semântica*: a *Semantic Web stack*.

2.1. Ontologias

Uma das definições de ontologia mais referenciadas em Ciência da Computação é a de Gruber (Gruber, 1993). Para ele, uma ontologia é uma especificação explícita de uma conceitualização. Nessa definição, por conceitualização entendem-se os conceitos, objetos e outras entidades que existem numa área de interesse e os relacionamentos entre elas. Borst (Borst, 1997) fez uma modificação sutil na definição de Gruber o que, parece, tornou-a mais apropriada: ontologias são definidas como especificações formais de conceitualizações compartilhadas.

Seguindo a definição de Borst, infere-se que ontologias são importantes para sistemas de software que almejam buscar, combinar ou integrar informação de diferentes fontes. E esse é exatamente o caso das informações na *Web*, onde ontologias podem possibilitar a representação semântica dos dados.

O intuito desta seção é oferecer definições de ontologia. Contudo não existe uma definição única e exata deste termo em Ciência da Computação. Uma das razões é uma série contínua (*spectrum*) de possíveis usos para ontologias (Breitman & Casanova, 2006). Essa série contínua é apresentada na Figura 2. Para mais detalhes sobre cada uso de ontologias apresentado na figura vide (McGuinness, 2003) ou (Breitman & Casanova, 2006).

Catalog/ ID	Thesauruses "Narrower term" relation	Formal is-a	Frames (Properties)	General Logic Constraints
Terms/ Glossary	Informal is-a	Formal Instance	Value Restrict.	Disjointness, Inverse, part- of ...

Figura 2 – *An ontology spectrum* (McGuinness, 2003)

Outra forma de entender os diversos usos de ontologias é através do entendimento do termo “semântica” em *Web Semântica*. Uschold (Uschold, 2003) oferece uma abordagem para tal através de um continuum semântico apresentado na Figura 3.

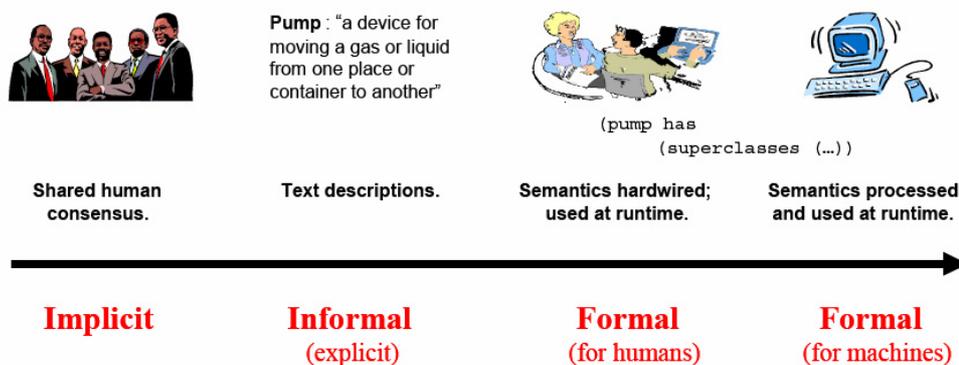


Figura 3 - *Semantic Continuum ...*⁴ (Uschold, 2003).

A partir da discussão apresentada, fica claro que muitas definições de ontologia podem existir e elas podem ser alteradas para melhor se adaptarem a um projeto ou área de pesquisa. Entretanto esse não é o foco deste trabalho, mais definições de ontologias podem ser encontradas em (Guarino, 1998) (Guarino, 1997), (van Heijst *et al.*, 1997), (Guarino, 1995) e (Guarino & Giaretta, 1995).

⁴ Continuação da legenda em (Uschold, 2003): “Semantics may be implicit, existing only in the minds of the humans who communicate and build Web applications. They may also be explicit and informal, or they may be formal. The further we move along the continuum, the less ambiguity there is and the more likely it is to have robust correctly functioning Web applications. For implicit and informal semantics, there is no alternative to hardwiring the semantics into Web application software. In the case of formal semantics, hardwiring remains an option, in which case the formal semantics serve the important role in reducing ambiguity in specifying Web application behavior, compared to implicit or informal semantics. There is also the new possibility of using automated inference to process the semantics at runtime. This would allow for much more robust Web applications, in which agents automatically learn something about the meaning of terms at runtime.”

Ontologias, na *Web Semântica*, são representadas utilizando linguagens de descrição de ontologias para a *Web*. Exemplos dessas linguagens são: Simple HTML Ontology Extensions (SHOE) (Heflin & Hendler, 2000), Resource Description Framework (RDF) (W3C, 2004e), RDF Vocabulary Description Language 1.0 (RDF Schema) (W3C, 2004f), DAML+OIL Language (DAML+OIL) (W3C, 2001), OWL (W3C, 2004b) entre outras. Como essas linguagens são baseadas na eXtensible Markup Language (XML), elas são mais ricas para a descrição de conteúdo do que HTML. As linguagens possibilitam a representação da estrutura dos conteúdos através de sua sintaxe; e a representação da semântica através de ontologias que descrevem propriedades ou relacionamentos dos conceitos. Algumas das linguagens de descrição de ontologias para a *Web* possibilitam o uso de inferências sobre os conceitos e relacionamentos entre eles.

2.2.

O W3C e a *Web Semântica*

O World Wide *Web* Consortium (W3C) é um consórcio internacional com a missão de levar a *Web* ao seu potencial máximo, através do desenvolvimento de protocolos, especificações e sugestões para garantir o crescimento da *Web* no longo prazo.

De acordo com o W3C (W3C, 2004g), algumas das primeiras linguagens usada para representar ontologias, elicitadas anteriormente, e para desenvolver ferramentas para comunidades de usuários não estavam compatíveis com a arquitetura da *Web* em geral, nem, em específico, com a *Web Semântica*. O consórcio então propôs e recomendou o Resource Description Framework (RDF) (W3C, 2004e) que é uma linguagem para representação de informação a respeito de recursos na *Web*. A RDF Vocabulary Description Language 1.0 (RDF Schema) (W3C, 2004f) foi a recomendação seguinte e o próximo passo do W3C para representar ontologias.

Na seqüência, o W3C propôs e recomendou a *Web Ontology Language* (OWL) (W3C, 2004b) que “estende” RDF e RDF Schema fornecendo algumas outras possibilidades para ontologias como escalabilidade, distribuição, compatibilidade com os padrões (*standards*) *Web* para acessibilidade e internacionalização, *openness* e extensibilidade. Conforme dito antes, apenas alguns ramos da lógica são computáveis. Portanto OWL foi projetada para oferecer 3 sub-linguagens de crescente expressividade (W3C, 2004a):

- OWL Lite: oferece basicamente suporte a hierarquias de classificação e restrições simples (*simple constraint features*);
- OWL DL: oferece o máximo de expressividade sem perda de completude computacional⁵ e decidibilidade⁶ de sistemas de inferência (*reasoning systems*). OWL DL é assim chamada devido a sua correspondência com lógica de descrição (*description logics*). Lógica de descrição (Baader *et al.*, 2003) é um área de pesquisa que estuda um fragmento específico e decidível de lógica de primeira ordem;
- OWL Full: oferece a máxima expressividade e a liberdade sintática de RDF sem garantias computacionais.

De acordo com (W3C, 2004a): “Desenvolvedores de ontologia que adotarem OWL devem considerar qual sub-linguagem melhor atende as suas necessidades. A escolha entre OWL Lite e OWL DL depende da necessidade de construção de restrições mais expressivas oferecida por OWL DL. *Reasoners* para OWL Lite terão propriedades computacionais desejáveis. Contudo *reasoners* para OWL DL, que lidam com uma sub-linguagem decidível, estarão sujeitos a mais complexidades de pior caso. A escolha entre OWL DL e OWL Full depende principalmente da necessidade dos usuários de utilizarem as facilidades de meta-modelagem oferecidas por RDF Schema, isto é, definir classes de classes. Ao utilizar OWL Full em comparação com OWL DL, o suporte a *reasoning* é menos predizível”. Mais informações sobre essas diferenças podem ser encontradas no OWL semantics document (W3C, 2004c).

Uma vez que é escolhida qual sub-linguagem será utilizada, uma questão que emerge é quais são as vantagens de tal escolha. De fato, o “uso de ontologias por aplicações *Web*” ou a “compreensão (*understanding*) e o processamento de ontologias por agentes de software” podem ser vistos como uma “forma de construir aplicações mais inteligentes em um futuro próximo e de executar tarefas em um nível conceitual mais próximo do nível conceitual humano” (W3C, 2004d). Essa última afirmação é muito parecida com um dos objetivos da área de Inteligência Artificial. Contudo, conforme destacado por Breitman e Casanova (Breitman & Casanova, 2006), há uma distinção entre Inteligência Artificial e a *Web Semântica*.

⁵ Todos os *entailments* são garantidos de serem computados.

⁶ Todas as computações terminarão em tempo finito.

Inteligência Artificial almeja a construção de software que é capaz de mostrar um nível de inteligência que é próximo ou similar (ou superior) ao da inteligência humana. Já uma das metas da *Web Semântica* é desenvolver software que é capaz de auxiliar os humanos em suas tomadas de decisão. Além disso como dito por Uschold (Uschold, 2003), a semântica implícita, ou o consenso humano compartilhado, está “conceitualmente” longe da semântica formal processada e usada em tempo de execução pelas máquinas conforme mostra a Figura 3.

Além dessas discussões, deseja-se que as aplicações sejam mais seguras e confiáveis baseadas em ontologias certificadas e em informações inferidas. A *Web Semântica* poderia possibilitar isso e outras funcionalidades interessantes através do uso de lógica complexa (ou mais expressiva) e através da troca de provas para estabelecer relacionamentos de confiança (Hendler, 2001).

A recomendação de OWL e algumas das discussões anteriores sobre o uso de ontologias e o que é esperado das aplicações que as utilizam são ilustradas em uma das bases arquiteturais da *Web Semântica*, a “*Semantic Web stack*” (apresentada no contexto da Figura 4). A *Semantic Web stack* foi apresentada pela primeira vez por Berners-Lee numa palestra na conferência XML 2000 Event (Berners-Lee, 2000). Mais informações sobre cada uma das camadas que a compõe podem ser encontradas em (Fensel *et al.*, 2003). Contudo há uma controvérsia sobre a *Semantic Web stack* que é, brevemente, apresentada na próxima seção.

2.3. A Controvérsia Sobre a *Semantic Web Stack*

Projetar e implementar uma aplicação para a *Web Semântica* (SWAPP) requer muitas decisões pragmáticas (Tummarello & Morbidoni, 2005). A Figura 4 mostra um exemplo dessas decisões baseando-se na *Semantic Web stack* de Berners-Lee (Berners-Lee, 2000).

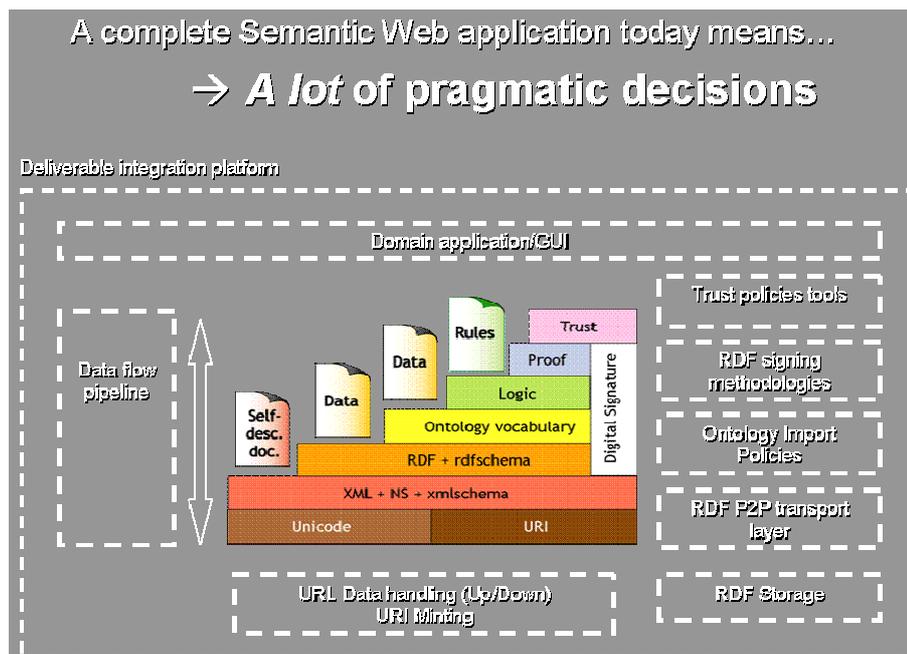


Figura 4 – Decisões pragmáticas baseadas na *Semantic Web stack* para projetar e implementar uma SWAPP (Tummarello & Morbidoni, 2005)

Apesar de decisões pragmáticas terem de ser tomadas, ainda se discute em que base estas decisões são tomadas. Em 2005, Berners-Lee propôs uma atualização (Berners-Lee, 2005) da *Semantic Web stack*, apresentada na Figura 5, que levantou uma controvérsia sobre a mesma em (Horrocks *et al.*, 2005) e (Patel-Schneider, 2005).

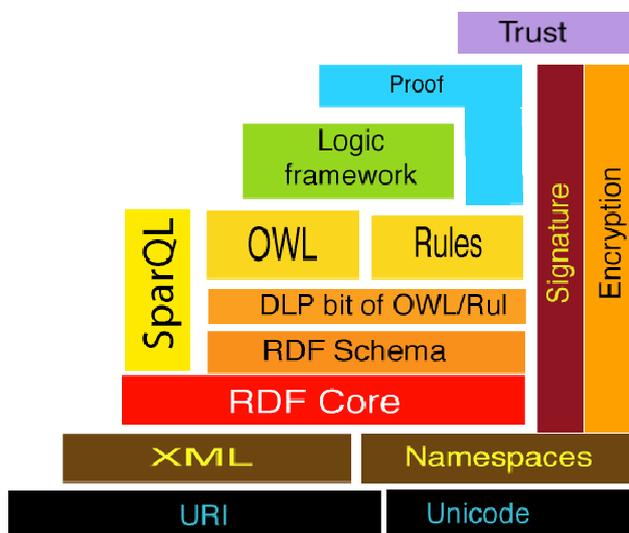


Figura 5 - Uma versão atualizada da *Semantic Web stack* (Berners-Lee, 2005)

Os dois trabalhos, (Horrocks *et al.*, 2005) e (Patel-Schneider, 2005), apresentam alguns mal-entendidos na versão atualizada da *Semantic Web stack*. Eles também discutem e propõem versões revisadas da mesma. A discussão e as revisões propostas estão fora do escopo deste trabalho, mas

como se proporá um *framework* para SWAPps esta controvérsia têm de ser considerada.

2.4. Resumo

Neste capítulo apresentou-se conceitos sobre a *Web Semântica*, ontologias, a relação entre o W3C e a *Web Semântica* através da recomendação da *Web ontology description language* (OWL) e suas sub-linguagens. Introduziu-se, brevemente, a controvérsia sobre a *Semantic Web stack*, que pode ser considerada uma das bases arquiteturais da *Web Semântica*. Grande parte deste capítulo oferece conceitos fundamentais para a contextualização deste trabalho que proporá um *framework* de aplicações para a *Web Semântica*.

Observou-se, que não há uma definição universal de ontologia e que a controvérsia sobre a *Semantic Web stack* e sua decisão sobre sobreposição arquitetural das linguagens serão importantes quando da definição do *framework* de aplicações para a *Web Semântica*.

No próximo capítulo apresenta-se o *Semantic Web Challenge* através de seus requisitos e como foi feita a extensão de um vocabulário RDF existente para a descrição das aplicações submetidas ao concurso.