

1

Introdução

Os¹ sonhos de software que poderiam “entender” dados (na *Web*) têm sido tratados por diferentes abordagens por pesquisadores de diversas áreas, tais como banco de dados, dados semi estruturados, gerência do conhecimento, lógica, linguagens formais e sistemas para a *Web*. Esses sonhos não são novos e, além disso, cada vez mais dados estão disponíveis na rede (W3C, 2005a) e “claramente” relacionados por *hiperlinks* (Rossi *et al.*, 1999). A *Web* é distribuída, dinâmica, um mundo aberto (*open world*) e possui uma grande quantidade de dados (Heflin *et al.*, 2003). Esses fatos e características trazem à tona novos requisitos, alguns desses já foram tratados por uma organização (*World Wide Web Consortium* - W3C) no intuito² de levar a *Web* ao seu potencial máximo, através do desenvolvimento de protocolos, especificações e sugestões incluindo o desenvolvimento da *Web Semântica*.

A *Web Semântica* tenta resolver problemas como interoperabilidade, melhoria de técnicas de busca e confiança nos dados, entre outros. Isso é feito através da explicitação da semântica dos dados. Essa explicitação da semântica possibilitará às aplicações a realização de inferências e o provimento de serviços mais personalizados para os usuários (Berners-Lee, 1998) (Berners-Lee *et al.*, 2001). De acordo com o W3C *Semantic Web Activity Statement*: “A meta da iniciativa de *Web Semântica* é tão abrangente quanto a da *Web*: criar um meio universal para a troca de dados. Prevê-se a conexão sutil do gerenciamento de informações pessoais, da integração de aplicações corporativas e do compartilhamento global de dados comerciais, científicos e culturais” (W3C, 2005b). Contudo, se isso é possível, como pode ser feito?

No caso da *Web Semântica*, Fensel *et alli* (Fensel *et al.*, 2003) identificaram que os seguintes elementos são necessários (Figura 1):

- Linguagens formais para expressar e representar ontologias, que são de maneira grosseira, os artefatos responsáveis pela explicitação da semântica dos dados;
- editores para construir, unir (*merge*) e reutilizar ontologias;

¹ Esta tese foi redigida e defendida em inglês. Recomenda-se fortemente a leitura da sua versão original, publicada como o relatório técnico (Cunha, 2007).

² About the World Wide Web Consortium (W3C) - <http://www.w3.org/Consortium/> - acesso em: 26/09/2006.

- serviços de inferência (*reasoning*) que possibilitem consultas avançadas e ajudem a mapear diferentes terminologias;
- ferramentas de anotação que possibilitem relacionar fontes de informação desestruturadas ou semi estruturadas a metadados;
- ferramentas para acesso e navegação a informação que possibilitem um acesso “inteligente” às fontes de informação por humanos; e
- serviços de tradução e integração entre diferentes ontologias para possibilitar a troca de dados em diferentes padrões.

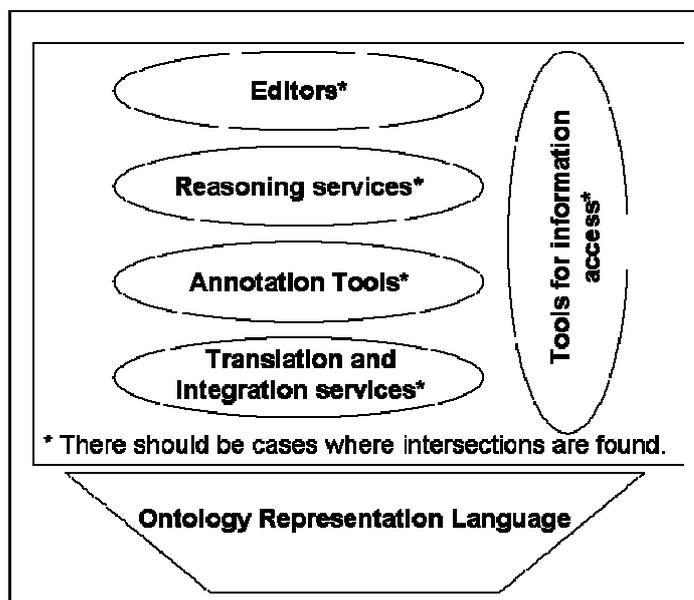


Figura 1 – Uma “interpretação” da lista de ferramentas ou tecnologias da *Web Semântica* de Fensel *et alli* (Fensel *et al.*, 2003)

Muitos dos elementos (ferramentas ou tecnologias) da Figura 1 foram e são alvo de estudo de vários pesquisadores. Entretanto uma questão ainda chama a atenção. Como são as “ferramentas para acesso e navegação a informação que possibilitam um acesso inteligente por humanos”? E também, como desenvolvê-las? Conforme Alavi e Leider (Alavi & Leider, 1999), sistemas de gerência de conhecimento (*knowledge management systems* - KMS), como as ferramentas em questão, têm de lidar com diferentes aspectos tais como aspectos de informação, de tecnologia e culturais. E ainda, segundo eles, em sua pesquisa não foi encontrada uma tecnologia, ferramenta (navegador, ferramenta de vídeo conferência etc.) ou produto KMS que lidasse com todos os aspectos.

1.1. O Problema

Conforme Conallen (Conallen, 1999), as diferenças entre um *site Web* e uma aplicação *Web* envolvem o seu uso. Nas aplicações *Web*, os desenvolvedores devem focar a modelagem da lógica de negócio sem gastar muito esforço com os detalhes de apresentação dos dados. Contudo, essa separação entre os aspectos de negócio e apresentação é o mais importante e se o modelo de apresentação é complexo ele também deve ser modelado mas não como parte do negócio.

Considerando-se as características das aplicações *Web*, as características da *Web* levantadas em (Heflin *et al.*, 2003) tornam-se ainda mais pertinentes. Para eles a *Web* é:

- distribuída: não há uma autoridade centralizadora;
- dinâmica: os dados podem estar, e normalmente estão, desatualizados;
- enorme: uma questão de escalabilidade. Têm-se que restringir a expressividade ou usar algoritmos de inferência (*reasoning*) incompletos com os dados disponíveis;
- um mundo aberto (*open world*): as informações podem estar, e normalmente estão, incompletas.

Voltando a *Web Semântica*, para o projeto e implementação de uma aplicação para a *Web Semântica* (SWAPp - *Semantic Web Application*) é necessário um conjunto de decisões pragmáticas (Tummarello & Morbidoni, 2005). Este trabalho trata da questão de como são as SWAPps e como desenvolvê-las. Portanto são importantes e de interesse o entendimento e a restrição do significado de quais aspectos apóiam aquelas aplicações. A resposta àquelas questões pode levar os usuários finais a entenderem melhor os benefícios das SWAPps e os desenvolvedores a tomarem suas decisões de uma maneira mais consciente.

Para os usuários finais, os benefícios do uso das técnicas ou tecnologias da *Web Semântica* deveriam ser transparentes, já para os desenvolvedores é importante entender como as técnicas e tecnologias se relacionam. Além disso, que decisões pragmáticas devem ser tomadas para alcançar os benefícios oferecidos pela “nova” semântica explícita dos dados.

Se as decisões tomadas pelos desenvolvedores de SWAPps seguirem uma abordagem de Engenharia de Software, essa abordagem pode mostrar o caminho de aplicações que são mais reutilizáveis, portáteis, manuteníveis, tolerantes a falhas e eficientes. Tendo tudo isso em vista, na próxima seção, apresenta-se uma abordagem para responder as questões de como são as SWAPps e como desenvolvê-las.

1.2. Solução Proposta

Para responder as questões apresentadas na seção anterior, este trabalho revisará as aplicações submetidas ao *Semantic Web Challenge*³ (SWC). O SWC é um concurso destinado a demonstrar como as técnicas da *Web Semântica* podem propiciar aplicações atrativas ou com valor agregado para os usuários finais. Como será visto no Capítulo 3, o concurso compartilha de alguns dos mesmos objetivos deste trabalho. A revisão das aplicações apresentará alternativas já implementadas para as decisões pragmáticas que devem ser tomadas por aqueles que desejam desenvolver SWAPps.

Pela revisão dessas aplicações, restringiu-se o domínio deste trabalho ao mesmo do concurso. Portanto a abordagem apresentada aqui é limitada pelo domínio do SWC, isto é, ela não é aplicável a qualquer aplicação para a *Web Semântica* que um desenvolvedor poderia projetar e implementar. Contudo a variedade de aplicações conforme definida pelo concurso já é ampla o bastante. Tratar de todas as aplicações para a *Web Semântica* seria muito difícil ou quase impossível, tanto o é que os organizadores do SWC definem um conjunto, ainda que amplo, de requisitos mínimos e desejáveis para caracterizar uma SWAPp.

A partir da revisão das aplicações, propõe-se uma análise de domínio das submissões para o SWC. Essa análise de domínio define um conjunto de tipos de aplicações e funcionalidades oferecidas por elas. Essa definição servirá como um dos requisitos para a proposta do *framework* para SWAPps. Na próxima seção, a solução proposta é decomposta em objetivos.

³ The Semantic Web Challenge - <http://challenge.semanticweb.org/> - acesso em: 16/06/2006.

1.3. Objetivos

A solução proposta para as questões de como são as SWAPPs e como desenvolvê-las considerando o domínio do SWC levou aos seguintes objetivos:

- revisar as aplicações submetidas ao SWC;
- usar uma formato padrão para registrar o processo de revisão;
- realizar uma análise de domínio das aplicações baseada no processo de revisão;
- propor um *framework* baseado nos tipos de aplicações e funcionalidades descobertas durante a análise de domínio; e
- ilustrar como a arquitetura do *framework* pode ser instanciada.

Até a escrita deste trabalho, 35 aplicações foram submetidas às 3 primeiras edições do SWC. Como já foi dito, as aplicações não representam todas as possíveis aplicações para a *Web Semântica*. Entretanto elas representam um segmento de aplicações que atendem aos requisitos propostos pelos organizadores do concurso.

O processo de revisão será registrado usando um esquema estendido para a descrição de projetos (veja a seção 3.3 para informações mais detalhadas sobre a escolha do esquema). Com base nas informações colhidas durante a revisão, propõe-se uma análise de domínio das aplicações que servirá como uma fronteira para a proposta do *framework* para SWAPPs.

Com a proposta do *framework*, pretende-se assistir (ou guiar) os desenvolvedores de um conjunto de SWAPPs, que é definido como uma combinação “válida” de funcionalidades oferecida por um tipo de aplicação. A ilustração de como o *framework* pode ser instanciado, através de sua arquitetura, deve demonstrar a adequação e relevância do mesmo.

1.4. Contribuições

Com base nos objetivos definidos, as contribuições deste trabalho são:

- o registro padronizado do processo de revisão, usado neste trabalho, das aplicações submetidas ao SWC;
- a proposta de um conjunto de tipos de SWAPPs e suas funcionalidades; e
- a apresentação de um *framework* para SWAPPs.

1.5. Trabalhos Relacionados

Considerando os elementos necessários para *Web Semântica* como definidos em (Fensel *et al.*, 2003), o *framework* proposto neste trabalho não lida com os fundamentos da *Web Semântica*, isto é, ele não é, por exemplo, uma ferramenta para edição ou armazenamento de ontologias. O *framework* também não é uma aplicação de infra-estrutura que oferece funcionalidades genéricas ou acesso a ferramentas que fornecem suporte para aspectos fundamentais da *Web Semântica* como o Sesame (Broekstra *et al.*, 2002) ou o Jena (Carroll *et al.*, 2004).

O *framework* proposto está portanto num nível intermediário entre as aplicações de infra-estrutura e as SWAPps para usuários finais. Há vários trabalhos neste mesmo nível, contudo não se encontrou nenhum que tratasse do mesmo domínio que o definido neste trabalho. Por exemplo, o *Semantic Hypermedia Design Method (SHDM)* é usado para o desenvolvimento de aplicações hipermídia utilizando tecnologias da *Web Semântica* (Lima, 2003).

O diferencial do trabalho apresentado nesta tese concentra-se na escolha de um domínio específico, ainda que amplo, o domínio do SWC. Além disso, a abordagem adotada de utilizar um *framework* traz os benefícios inerentes a ela como reusabilidade, portabilidade, manutenibilidade e tolerância a falhas. O *framework* proposto também não é uma SWAPp para usuários finais já que representa um conjunto delas que podem ser projetadas e implementadas pela instanciação ou personalização (*customization*) do *framework*.

É importante lembrar o aspecto empírico deste trabalho. Revisou-se 35 aplicações submetidas ao concurso e a análise de domínio foi baseada nesta revisão. O *framework* tem, portanto, a característica de usar uma abordagem *bottom-up* e oferece aos seus usuários a possibilidade de utilizar decisões pragmáticas já utilizadas no desenvolvimento de SWAPps para projetar e implementar as suas próprias.

1.6. Resumo

Neste capítulo, contextualizou-se o problema de como são as aplicações para a *Web Semântica* e como desenvolvê-las. Mostrou-se o quão relevante são essas questões e propôs-se, de forma breve, uma solução à elas. A solução proposta foi decomposta em objetivos que levaram às contribuições deste trabalho. A principal delas é a proposta de um *framework* para SWAPps. Apresentou-se também, brevemente, alguns trabalhos relacionados e como este trabalho se difere deles.

O resto deste trabalho está estruturado da seguinte forma: no próximo capítulo, apresentam-se alguns fundamentos sobre a *Web Semântica*. No capítulo seguinte, o *Semantic Web Challenge (SWC)* é explicado e também como foi feita a extensão de um vocabulário RDF para a revisão das aplicações submetidas ao concurso. O vocabulário original é apresentado no Anexo I – O *DOAP Vocabulary*. A versão estendida, no Apêndice A - O *SWDOAP Vocabulary*.

O capítulo 4 apresenta a análise de domínio das aplicações submetidas ao SWC. Essa análise de domínio é composta de definições e exemplos de tipos de aplicações e suas funcionalidades. Na versão em inglês desta tese (Cunha, 2007) apresenta-se respectivamente, as aplicações submetidas ao SWC em 2003, 2004 e 2005. Naquele trabalho para cada aplicação, além de uma breve descrição, apresentam-se o(s) seu(s) tipo(s) e as funcionalidades que ela oferece com base nas definições do capítulo 4.

O capítulo 5 apresenta e discute o *framework* para aplicações para a *Web Semântica* (SWAPpFW - *Semantic Web application framework*) proposto. O capítulo 6 apresenta as conclusões deste trabalho, suas contribuições e também os trabalhos futuros.