

1 Introdução

1.1. Motivação

A World Wide Web – ou simplesmente Web – foi implementada pela primeira vez como uma solução interna para o compartilhamento de documentos entre aproximadamente mil usuários do CERN (Segal, 1995). Concebida com base em uma série de princípios de projeto como simplicidade de protocolos e interoperabilidade, a Web cresceu a níveis espetaculares. Esses princípios tinham como objetivo provocar a rápida expansão dessa infraestrutura de comunicação, com o aumento do número de usuários e do conteúdo disponível. Com uma quantidade mínima de recursos, qualquer um teria a capacidade de criar e publicar documentos que poderiam ser acessados de qualquer outro ponto desta rede, aumentando o seu valor e utilidade – vale lembrar que, segundo a Lei de Metcalfe, “o valor de uma rede é proporcional ao quadrado do número de usuários” (Reed, 1999).

Além disso, para evitar a fragmentação da Web em várias redes isoladas (Orlean et al., 2001), delimitadas por tecnologias proprietárias, o princípio de interoperabilidade vem regendo substancialmente este crescimento. Independentemente da plataforma e de tecnologia escolhida, os componentes desta rede devem ser capazes de trocar dados e informações entre si. Para isso, foram padronizados diversos protocolos, tecnologias e linguagens para tarefas fundamentais da Web como localização de recursos (URIs, URLs e URNs), transporte de dados (http, ftp, wtp) e marcação de documentos (HTML, XML).

Atualmente, no entanto, há um preço elevado a ser pago por esse sucesso. A crescente disponibilidade de recursos e fontes de informação vêm dificultando a busca, o acesso, a apresentação e a manutenção das informações requeridas por uma grande variedade de usuários (Ding et al, 2002). Se, no princípio, o caos era gerado pelo fato de que cada aplicação utilizava sua própria abordagem de comunicação de dados (Segal, 1995) e, com isso, tornava as informações inacessíveis, hoje este existe justamente pelo motivo inverso, a disponibilidade aparentemente infinita de conteúdo. Benjamins *et alli* (Benjamin et al., 2002) apresentam o problema de

sobrecarga da informação causado pelo contínuo crescimento da web tanto em tamanho, quanto em linguagens e formatos.

As tarefas mais complexas de acesso, extração, interpretação e manutenção são deixadas para o usuário humano (Maedche, 2002), uma vez que a interoperabilidade de dados e informações permanece restrita ao nível sintático, em detrimento do semântico. Entenda-se aqui como semântica a capacidade de se processar e interpretar algum tipo de informação computacionalmente (Uschold, 2001), e não apenas através da mente humana.

1.2.A Web Semântica

A segunda geração da Web – até pouco tempo considerada apenas uma visão de Tim Berners-Lee (Berners-Lee et al., 2001) - vem com o objetivo de solucionar este problema. A Web Semântica, como está sendo chamada, tem como característica principal dotar os documentos ora disponibilizados na Web com informações que possam ser processadas semanticamente, para favorecer o oferecimento de serviços automáticos (Ding et al., 2002), como agentes de software, máquinas de busca inteligentes e web services, por exemplo. Não só a apresentação do conteúdo, mas também a automação de tarefas, integração e reuso dos dados por diversas aplicações devem ser possibilitados através desta infra-estrutura (W3C, 2001). No entanto, apesar do novo nome poder levar a crer que uma nova infra-estrutura independente da Web atual está sendo criada, isso não é verdade. A Web Semântica pode ser encarada como uma extensão da web atual, como mostra a Figura 1 (Semaview, 2002).

A Web Semântica pode ser definida como a teia de informação capaz de ser processada e consumida por aplicações em escala global (Palmer, 2001). Essas aplicações são principalmente descritas por Berners-Lee como agentes de software (Weiss, 1999), mas podem ser caracterizadas também por outros tipos de aplicações, como sistemas de busca, processamento de linguagem natural, integração de bases de informações distribuídas e *web services*, apenas para citar alguns exemplos.

No entanto, já virou lugar-comum falar que a Web como é conhecida hoje é primordialmente desenvolvida para humanos. A grande maioria do conteúdo disponibilizado atualmente encontra-se no formato HTML (HTML), que se preocupa em descrever a forma como este conteúdo deve ser apresentado, em detrimento da sua estrutura.

Por esse motivo a Web Semântica está deixando de ser uma visão para virar uma realidade. Diversos projetos em todo o mundo já exploram as potencialidades dessa segunda geração da Web para tornar o conteúdo disponibilizado processável por máquinas. No entanto, assim como as pessoas, máquinas não podem compartilhar conhecimento se não possuem uma linguagem em comum (Davenport and Prusak, 1998). Infelizmente (ou felizmente!), os computadores não são capazes de desenvolver por livre e espontânea vontade uma linguagem consensual de comunicação. É nesta etapa que entram em cena as ontologias.

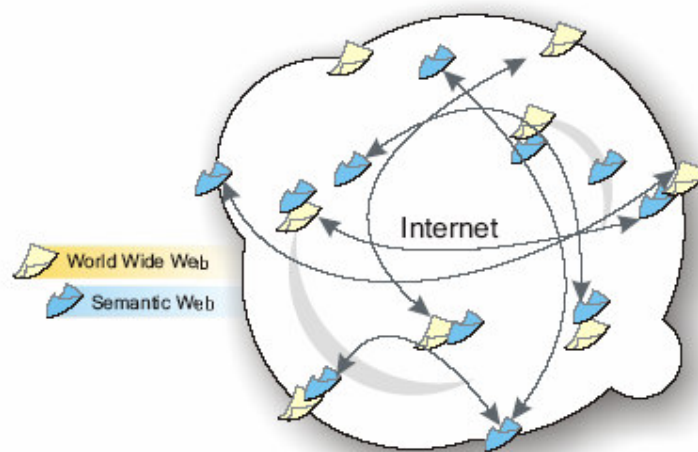


Figura 1 - A Web Semântica como uma extensão da Web atual

1.2.1.A Evolução da Web Semântica

Assim como ocorre com diversas tecnologias, a evolução da Web Semântica ocorrerá de forma gradual, com iniciativas isoladas sendo tomadas antes que ela se torne ubíqua como é a Web atualmente. A Figura 2 (Semaview, 2002) compara os estágios de ambas as tecnologias, e a seguir são descritos os estágios esperados para a Web Semântica, ilustrados na Figura 3.

Estágio 1: *Semantic Islands*

A Web Semântica encontra-se hoje em seu primeiro estágio, no qual conteúdo e aplicações são desenvolvidos no contexto de *intranets* corporativas. São chamadas de ilhas semânticas por estarem isoladas em cenários internos ou entre parceiros de

negócios, experimentando novas potencialidades dessa tecnologia e preparando o caminho para o futuro.

Estágio 2: As Ilhas se conectam

Com o passar do tempo, novas ilhas começarão a se formar a uma velocidade maior, e as ligações entre elas passarão a ser mais frequentes. Novas ferramentas serão desenvolvidas para facilitar a geração de aplicações para a Web Semântica, ao mesmo tempo que as máquinas de busca passarão a levar em conta os meta-dados para realizar suas pesquisas. Agentes de software poderão executar tarefas autônomoamente na web, e uma rede de confiabilidade deverá ser construída para evitar inconsistências.

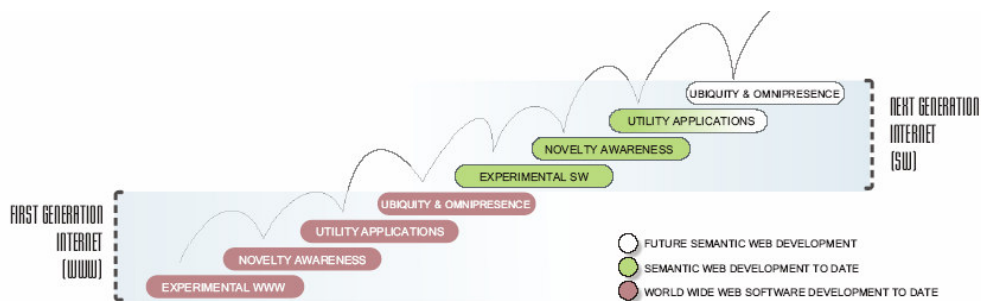


Figura 2 - Os estágios da Web Semântica

Estágio 3: Inferências

Com a disseminação de meta-dados na Web Semântica os agentes de software poderão inferir novos conhecimentos – o que exige também a evolução de recursos da inteligência artificial.

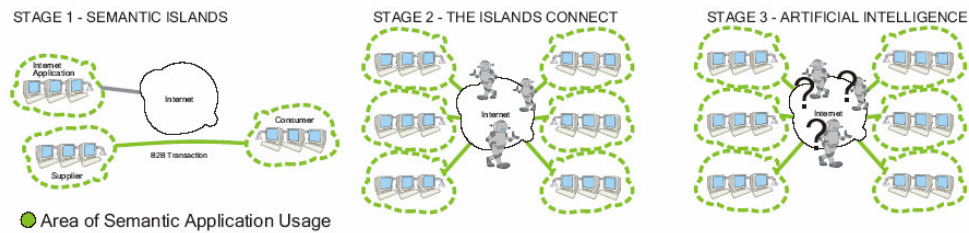


Figura 3 - Os estágios da Web Semântica 2

A abordagem defendida pelo HorizOnto para a Gestão Distribuída de Conhecimento na Web Semântica segue uma abordagem análoga a evolução da própria Web Semântica. As ontologias bases de conhecimento individuais fazem o papel das ilhas semânticas. Aos poucos essas bases serão integradas ou alinhadas em outras ontologias que descrevem os horizontes de conhecimento, em múltiplos níveis. Agentes de software serão responsáveis por integrar essas bases e horizontes de conhecimento descritos por ontologias.

1.3. Ontologias

Conhecida no ramo da filosofia como uma “teoria sobre a natureza da existência”, as ontologias têm sido encaradas de maneira consideravelmente diferente no universo computacional. Na comunidade de Ciências da Computação, uma das primeiras referências a esse termo pode ser vista em (Neches et al., 1991). Objetivando alcançar um alto grau de reuso na construção de sistemas baseados em conhecimento, Neches *et alli* usaram a conceituação de ontologias para definir componentes de conhecimento reutilizáveis.

Desde então, diversas comunidades vêm demonstrando interesse nesta área de pesquisa, passando por processamento de linguagem natural, sistemas de informação, integração de sistemas e inteligência artificial, apenas para citar algumas.

Segundo (Gruber, 1993), responsável por uma das definições mais referenciadas sobre ontologias, estas representam um entendimento comum e compartilhado sobre um domínio específico, e têm como o objetivo principal permitir a comunicação entre organizações, pessoas e aplicações. Mais especificamente, para Gruber, “uma ontologia é uma especificação explícita de uma

conceitualização”. Conceitualização pode ser entendida como a identificação dos conceitos, objetos e outras entidades existentes no domínio de interesse da ontologia, bem como as relações entre estes elementos (Genesereth et al., 1987). Borst (Borst, 1997), por sua vez, complementou esta definição de Gruber: “ontologias são definidas como especificações formais de entendimentos compartilhados”.

Alcançar este entendimento compartilhado exige o comprometimento de grupos de pessoas e o uso de metodologias, processos e/ou métodos bem definidos. Com o aumento de sua importância, o projeto e desenvolvimento de ontologias vêm deixando de ser uma arte para se transformar em um processo de engenharia.

Ontologias consistem tipicamente em definições de conceitos, suas relações e axiomas (Staab et al., 2000) - estas últimas premissas que devem ser consideradas levando-se em conta o domínio que as ontologias pretendem descrever. As relações entre os conceitos da ontologia podem ser tanto taxonômicas (“um professor é um (tipo de) funcionário”) como não taxonômicas (“um professor orienta um aluno”), o que permite um mapeamento bastante próximo da realidade do domínio em questão, e, conseqüentemente, uma linguagem mais adequada para a comunicação e o entendimento compartilhado.

De acordo com (Maedche,2002) uma ontologia pode ser descrita através de uma estrutura (O) composta pela seguinte quintupla:

$O := \{ C, R, Hc, rel, Ao \}$, onde:

1. C e R são dois conjuntos disjuntos formados por conceitos e relações, respectivamente;
2. Hc é a hierarquia de conceitos, que contém as relações taxonômicas da ontologia;
3. a Função rel que relaciona conceitos não taxonômicos;
4. e Ao, o conjunto de axiomas da ontologia.

A Web Semântica proposta por Berners-Lee aponta na direção de padrões - ou seja, normas e modelos amplamente aceitas pela comunidade - para representar o conhecimento em forma de ontologias, através de linguagens criadas para este fim. A linguagem HTML padronizou a forma de se representar e apresentar documentos hipertexto. O XML (XML) padronizou a forma de se descrever dados. No contexto

da Web Semântica padronizam-se, por exemplo, linguagens de representação de ontologias que permitam descrever conhecimento semanticamente – possibilitando, então, que este conhecimento seja automaticamente processado por máquinas. Como já foi dito, no entanto, que a Web Semântica não vem substituir a Web como esta é conhecida atualmente, mas sim representar uma evolução na forma como seres humanos e entidades autônomas interagem através dela. A Web Semântica não é nada além da web atual dotada de mecanismos para representação formal de informações semânticas através de ontologias (Berners-Lee et al., 2001), (Berners-Lee, 2002).

1.4. Engenharia de Ontologias

Como a própria definição adotada denota, uma ontologia deve ser capaz de capturar o conhecimento de um domínio e de promover um entendimento compartilhado, o que tipicamente exige o comprometimento de grupos de pessoas (Sure et al., 2002a) e o uso de metodologias, processos e/ou métodos bem definidos. Para cumprir esse tipo de objetivo, uma ontologia deve ser explicitada, o que geralmente incorre no desenvolvimento de um artefato computacional, o qual deve ser primeiro criado e estruturado de acordo com os requisitos apontados para esse artefato, para ser então utilizado e manipulado (Ontolingua, 1997) por pessoas e aplicações.

Construir uma ontologia pode parecer uma tarefa similar, a princípio, ao desenvolvimento de um programa orientado a objetos, mas existem diversas diferenças sutis e significativas (Ontolingua, 1997) que devem ser consideradas. O desenvolvimento de um programa OO está geralmente focado na solução de um problema específico, com escopo bem definido. Mesmo um framework está intimamente relacionado a uma família de aplicações em um mesmo domínio. Isso exige a definição de classes e objetos que descrevem não só elementos do mundo real mas também estruturas de dados e elementos funcionais. Uma ontologia, no entanto, deve conter apenas classes e objetos do mundo real, uma vez que, como já foi dito, representa a especificação explícita de uma conceitualização (Gruber, 1993). O desenvolvimento de uma ontologia pode até ser motivada pela necessidade específica de uma aplicação, mas, no contexto da web semântica, esta deve permitir a integração de diversas fontes de informação ou mesmo ser acessível a aplicações desconhecidas *a priori*. Por outro lado, o reuso de código usual – e incentivado – em OO pode

violam princípios fundamentais das ontologias, que devem refletir a estrutura do domínio, e não a estrutura de dados.

Diversas propostas já foram apresentadas para o problema de projeto de ontologias ((Uschold and King, 1995), (Uschold and Gruninger, 1996), (Guarino and Welty, 2000), (Staab et al., 2000), (Fernandez-Lopez et al., 1997), etc.), muitas delas apoiadas por estudos de caso acadêmicos e industriais. No entanto, é importante notar que nenhuma dessas metodologias – que em muitos casos resumem-se apenas a heurísticas extraídas da experiência de seus autores ou a orientações sobre como alcançar bons resultados – atende por completo os requisitos potenciais do projeto de uma ontologia.

Por ter como um de seus objetivos permitir o processamento semântico de informações, uma ontologia deve ter um grau de formalismo que permita com que esta seja computada. Deve, além disso, representar a teoria compartilhada e consensual sobre um determinado domínio, permitindo a comunicação entre pessoas e aplicações, levando em consideração questões antagônicas como expressividade e complexidade (toda a informação necessária – hoje ou em um futuro próximo - deve estar disponível) versus facilidade de implementação, uso e manutenção (nenhuma informação desnecessária deve estar disponível). Desta forma, uma ontologia pode apresentar diferentes graus de abstração e detalhamento que permitam o cumprimento dos diferentes objetivos para os quais está sendo desenvolvida, seja permitir a comunicação entre agentes de software ou mesmo a geração de um portal de conhecimento dotado com informações semânticas sobre seu conteúdo. Por fim, deve levar em conta as evoluções que uma ontologia, por ser, em última instância, a representação de um domínio do mundo real, está propensa a sofrer com o passar tempo, exigindo sempre seu refinamento e manutenção.

No entanto, o que se observa nas metodologias e processos existentes é que geralmente é dado um foco maior a uma dessas necessidades, em detrimento de outras. Enquanto algumas possuem um processo bem elaborado de levantamento de requisitos e análise do domínio, outras focam em questões referentes ao formalismo da ontologia, em seu projeto ou implementação.

1.5. Objetivo

O objetivo principal desta dissertação é **propor um Processo Unificado para Engenharia de Ontologias através da unificação de abordagens e melhores**

práticas existentes na indústria e na academia. Serão apresentados os critérios utilizados para unificação dessas abordagens e um Framework proposto para este processo, utilizando-se uma notação inspirada na notação adotada pelo SPEM (SPEM, 2002), da OMG (OMG). Para mensurar e identificar omissões e falhas do Framework proposto para o Processo Unificado, intitulado KUP – Knowledge Unified Process, foram conduzidos dois estudos de caso tratando de problemas reais da indústria.

No primeiro estudo de caso foram desenvolvidas uma especificação e uma ontologia que estão sendo usadas na transformação de um sistema de informação que foi iniciado de forma tradicional para uma solução na Web Semântica. Esta solução tem como objetivo monitorar em fontes distribuídas - através de agentes de software - informações sobre vulnerabilidades e códigos maliciosos que possam afetar ativos e componentes de software, alimentando uma base de conhecimento consolidada. Através da atualização dessa base de conhecimento são gerados relatórios personalizados de acordo com a infra-estrutura de TI de cada empresa cadastrada no sistema, de acordo com os ativos e componentes que ela definiu em seu escopo de monitoramento. Este estudo de caso teve como objetivo principal verificar a adequação do KUP em um projeto de solução para a Web Semântica, fornecendo subsídios para seu aprimoramento.

O segundo estudo de caso no contexto do KUP teve como foco a integração de diferentes padrões, especificações e ontologias desenvolvidas por organizações distintas em uma única ontologia para a representação de conhecimento nos domínios de cursos, pessoas e competências. Esta ontologia foi projetada com o intuito de garantir a interoperabilidade entre sistemas de Gestão de Competências, Gestão de Pessoas e de Gerência de Aprendizado Online (LMSs). O objetivo deste estudo de caso foi verificar a utilidade do KUP em situações nas quais as ontologias não são desenvolvidas desde o princípio, mas são geradas a partir da integração de modelos já existentes.

1.6. Estrutura da Dissertação

Esta dissertação está organizada em 7 capítulos. Uma pequena revisão dos conceitos e da fundamentação teórica necessárias para o entendimento deste trabalho podem ser encontradas no capítulo 2. Trabalhos relacionados serão apresentados no capítulo 3. O capítulo 4 faz uma breve introdução ao KUP, descreve o processo de

unificação de metodologias e métodos e os critérios que guiaram esta unificação. O capítulo 5 apresenta o KUP propriamente dito, focando em seu meta-modelo, seus componentes, sua organização e aplicação. O capítulo 6 descreve os dois casos de teste realizados no contexto do KUP e as conclusões obtidas.

Finalmente, o capítulo 7 apresenta as conclusões obtidas com a realização do trabalho e pontos que ainda precisam de estudos mais aprofundados, caracterizando possíveis trabalhos futuros. São apresentados 3 anexos com resultados dos estudos de caso, artefatos, guias e informações adicionais. Ao final do texto, encontram-se as referências bibliográficas utilizadas no desenvolvimento do trabalho.

1.7. Metodologia

A metodologia de trabalho utilizada para esta pesquisa compreende os seguintes passos:

O primeiro passo compreendeu o estudo da fundamentação teórica e de diversos trabalhos relacionados levantados durante o estudo. Um estudo de caso foi realizado em seguida, em conjunto com uma equipe de colaboradores. Este estudo de caso envolveu a construção de uma ontologia de projetos de pesquisa, de uma aplicação hipermídia e a criação de uma camada semântica para a ontologia e os dados da aplicação. A definição do desenvolvimento futuro de agentes de software e serviços de acesso aos dados também fez parte dessa segunda fase.

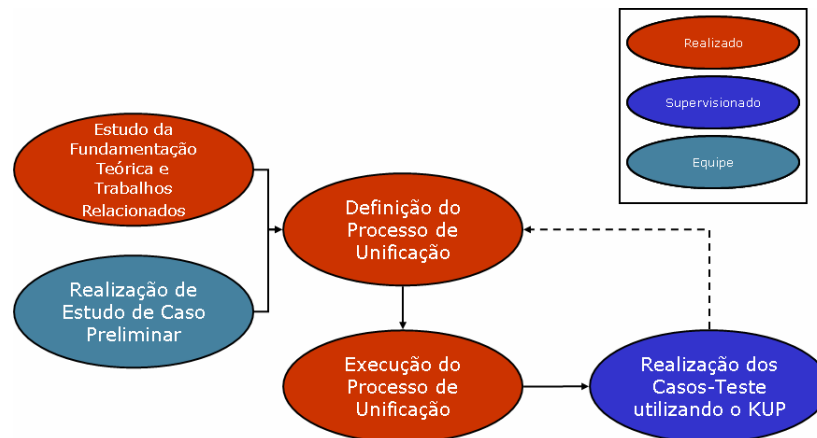


Figura 4 – Metodologia de Pesquisa

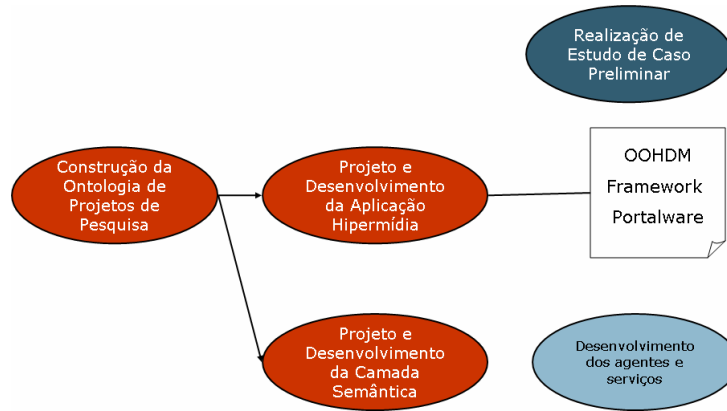


Figura 5 – Metodologia do Estudo de Caso

Com o resultado das duas primeiras fases iniciou-se o processo de definição do processo de unificação. Esta fase compreende as tarefas ilustradas na Figura 6:

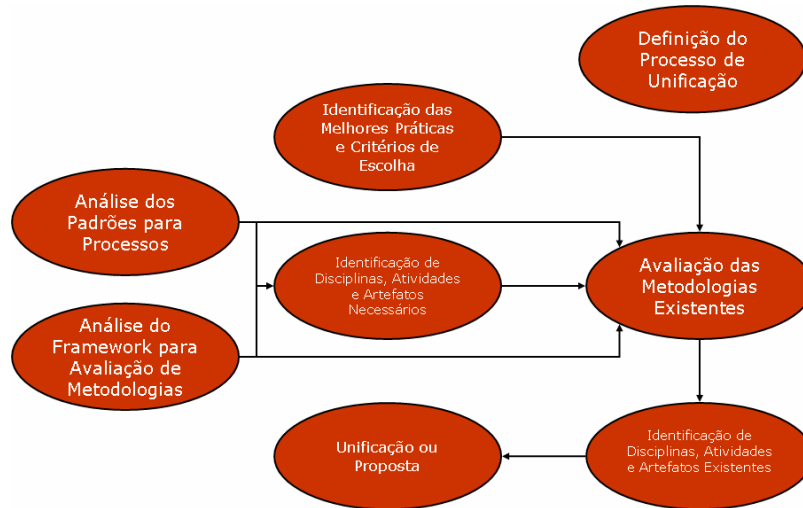


Figura 6 – Metodologia da Definição do Processo de Unificação

Após a definição do processo de unificação foram realizados dois casos teste que forneceram subsídios para o aprimoramento do processo.