

FERNANDO MARIANO DE OLIVEIRA

**UMA PROPOSTA DE MODELO DE ARQUITETURA BASEADA
EM SEMÂNTICA PARA AMBIENTES CORPORATIVOS**

Monografia do curso de pós-graduação lato sensu
MBIS - Master Business Information Systems
apresentado à Pontifícia Universidade Católica de
São Paulo para a obtenção do título de
especialista.

São Paulo

2011

FERNANDO MARIANO DE OLIVEIRA

**UMA PROPOSTA DE MODELO DE ARQUITETURA BASEADA
EM SEMÂNTICA PARA AMBIENTES CORPORATIVOS**

Monografia apresentada à Pontifícia Universidade
Católica de São Paulo para a obtenção do título de
especialista.

Orientador:

Prof. Dr. Gregório Bittar Ivanoff

São Paulo

2011

Agradecimentos

Agradeço primeiramente a Deus que tem me dado saúde, sabedoria e inteligência para que eu consiga completar minhas realizações. Aos meus pais, Waldir Oliveira e Maria Cristina Oliveira que me moldaram e fizeram de mim grande parte do que sou hoje.

A minha esposa, Luciene, pela paciência e incentivo que têm me ajudado tanto no decorrer do curso.

Ao meu Orientador Prof. Dr. Gregório por desde o início do MBIS ter incentivado mudanças profissionais e pessoais e por me acompanhar no desenvolvimento deste.

A todos os professores que passaram por minha vida, que foram muitos, quero agradecer a todos. Todos tiveram sua parcela de contribuição. Sem vocês, mais esta realização, seria impossível, muito obrigado!

A PUC - COGEAE pelo incentivo e apoio a esta capacitação.

FICHA CATALOGRÁFICA

Oliveira, Fernando Mariano de

Uma proposta de modelo de Arquitetura baseada em Semântica para Ambientes Corporativos. São Paulo, 2011. 49p.

Monografia programa de pós-graduação MBIS – Pontifícia Universidade Católica de São Paulo. Departamento de Computação.

1.. 2. 3. . I. Pontifícia Universidade Católica de São Paulo. Departamento de Computação.

Resumo

As organizações já descobriram que o conhecimento é o insumo mais importante de sua cadeia produtiva e iniciaram pesquisas por métodos e ferramentas que sejam capazes de geri-lo.

Movidas pela competitividade do mercado, as organizações perceberam que a qualidade de seus produtos/serviços está diretamente ligada a este conhecimento interno e que grande parte da sabedoria corporativa reside em documentos não estruturados (muitos deles como texto em linguagem natural) que se encontram dispersos por meios diversos. O aumento substancial destes tipos de documentos, neste contexto, representa um desafio extraordinário na extração de “conhecimento assistido por computadores” aos seus colaboradores.

Todo este conteúdo não estruturado é o filão da sabedoria corporativa, mas “minerá-lo” com a arquitetura tradicional baseada apenas na gestão do conhecimento tem se revelado decepcionante para quem precisa de conhecimento preciso dentro de um tempo escasso.

Portanto, esta monografia objetiva apresentar uma proposta de modelo de arquitetura baseada em semântica que trará grandes benefícios ao ambiente corporativo, porém conjuntamente criará grandes desafios técnicos, não menos do que a criação de um mecanismo capaz de extrair metadados de conteúdos e automaticamente aprender ontologias, de maneira que agentes computacionais possam buscar informações relevantes em linguagem natural.

Palavras-chave: Modelo de Arquitetura baseada em Semântica para Ambientes Corporativos. Gestão do conhecimento. Recuperação de Informações.

Abstract

The enterprises have discovered that knowledge is one of the most important inputs in its production chain and started research for methods and tools to manage this knowledge.

Driven by competitive market, corporations have realized that the quality of your products/services is directly linked to this internal knowledge and that, the great part of this wisdom, lies in unstructured documents (many of them as text in natural language) that are spread by different types of media.

The substantial increase in these types of documents in this context represents an extraordinary challenge on extracting "knowledge assisted by computers" to their employees.

All of this unstructured content is the mother lode of corporate wisdom, but "mining it" with the traditional architecture based only on knowledge management has been disappointing for those who need accurate knowledge within a limited time.

Therefore, this monograph aims to present a proposal for a model of semantic-based architecture that will bring great benefits to the corporate environment, but together create huge technical challenges, no less than the creation of a mechanism able to extract content from metadata and automatically learn ontologies so that computational agents could search for relevant information in natural language.

Keywords: Model of Semantic-based Architecture to Enterprise Environments. Knowledge Management. Information Recovery.

Sumário

LISTA DE FIGURAS	I
LISTA DE TABELAS.....	II
LISTA DE PALAVRAS-CHAVE.....	III
1 INTRODUÇÃO	1
1.1 Objetivos	3
1.2 Justificativa	3
1.3 Metodologia	4
1.4 Estrutura do trabalho	5
2 O CENÁRIO ATUAL DOS AMBIENTES CORPORATIVOS.....	6
3 A SEMÂNTICA E SEUS PRINCIPAIS CONCEITOS	9
3.1 A Semântica	9
3.1.1 Metadados	12
3.1.2 Ontologias	13
3.1.2.1 Tipos de ontologias.....	14
3.1.2.2 Exemplos de Ontologias	15
3.1.2.3 Benefícios das Ontologias.....	16
3.2 Busca Semântica	17
4 DESAFIOS: A INTEROPERABILIDADE E O CONFLITO SEMÂNTICO	18
4.1 Uma palavra sobre Interoperabilidade	18
4.2 O conflito semântico	21
5 MODELO DE ARQUITETURA BASEADO NA SEMÂNTICA: OS PILARES	23
5.1 Engenharia Corporativa de Ontologias	23
5.1.1 Tipos de Colaboradores.....	28
5.1.2 Modularização e integração de Ontologias	30
5.2 Colaboração Semântica Corporativa.....	33
5.3 Busca Semântica Corporativa	36
5.3.1 Busca em dados não semânticos	37
5.3.2 Personalização da Busca Semântica Corporativa	39
5.3.3 Cenários de Aplicação.....	40
6 CONSIDERAÇÕES FINAIS	42
6.1 Limitações do Estudo.....	43
6.2 Trabalhos futuros	44
BIBLIOGRAFIA.....	45

Lista de Figuras

Figura 1.1 – O papel da Arquitetura baseada na Semântica no Ambiente Corporativo	1
Figura 2.2 – O posicionamento, evolução e maturidade da pesquisa da semântica no ambiente corporativo, seus pilares e sua colaboração na evolução da pesquisa web semântica.	8
Figura 4.3 – Modelo sugerido de arquitetura baseada em semântica e o conjunto de módulos que compõe o mediador de consultas.....	22
Figura 5.4 – O Ciclo de Vida da semântica em uso no ambiente corporativo e o papel da Engenharia de Ontologias	27
Figura 5.5 – Componentes para modularização e integração de ontologias e decomposição dos módulos do “mediador”	32
Figura 5.6 – Visão global do modelo proposto de arquitetura baseado em semântica	34
Figura 5.7 –Visão global do modelo proposto de arquitetura com foco na Busca Semântica	37
Figura 5.8 – Cenário real de análise realizada atualmente por investidores e o que se poderia obter com a Semântica.	41

Lista de Tabelas

Tabela 1 - Grupos de usuários e suas operações típicas.....	29
---	----

Lista de Palavras-chave

BI	<i>Business Intelligence - Inteligência nos negócios</i>
CMS	<i>Content Management System - Sistema de gestão de conteúdo</i>
DAML+OIL	<i>Linguagem de Ontologia que usa a DAML revisada + as características do OIL</i>
ERP	<i>Enterprise Resource Planning–Sistema Integrado Gestão Empresarial</i>
HTML	<i>HyperText Markup Language – Linguagem usada para páginas Web</i>
KM	<i>Knowledge Management - Gestão do Conhecimento</i>
KB	<i>Knowledge Base - Base de Conhecimento</i>
OIL	<i>Ontology Interchange Language – Linguagem intercâmbio Ontologias</i>
OML	<i>Ontology Markup Language – Linguagem Marcação Ontologias</i>
OWL	<i>Web Ontology Language – Linguagem Ontologia Web</i>
RDF	<i>Resource Description Framework Language</i>
RDFs	<i>Resource Description Framework Schema Language</i>
SC	<i>Semântica Corporativa</i>
TCP/IP	<i>Internet Protocol Suite</i>
URI	<i>Uniform Resources Identifiers</i>
URL	<i>Uniform Resource Locator</i>
XHTML	<i>Extensible Hypertext Markup Language</i>
XML	<i>Extensible Markup Language</i>
XOL	<i>Ontology Exchange Language XML (eXtensible Markup Language) provê uma sintaxe de estruturação de documentos, mas não impõe regras semânticas nesses documentos.</i>

1 INTRODUÇÃO

Um dos grandes problemas que as corporações enfrentam nos dias atuais é o de conseguir manter o conhecimento em movimento e em coordenação com suas demandas estratégicas. Este conhecimento, via de regra, está depositado nos colaboradores que compõem a organização e pode ser perdido, comprometendo os níveis de qualidade e produtividade. A fim de manter seus níveis ótimos, as organizações (empresas e instituições) devem promover meios para que a experiência ou o saber que está nestas pessoas (conhecimento tácito) permaneçam na organização mesmo depois da ausência destes. Esse conhecimento acumulado pelo colaborador sempre é fruto de anos de trabalho em determinada função, tornando-o de difícil identificação e sendo perceptível somente quando da sua ausência.

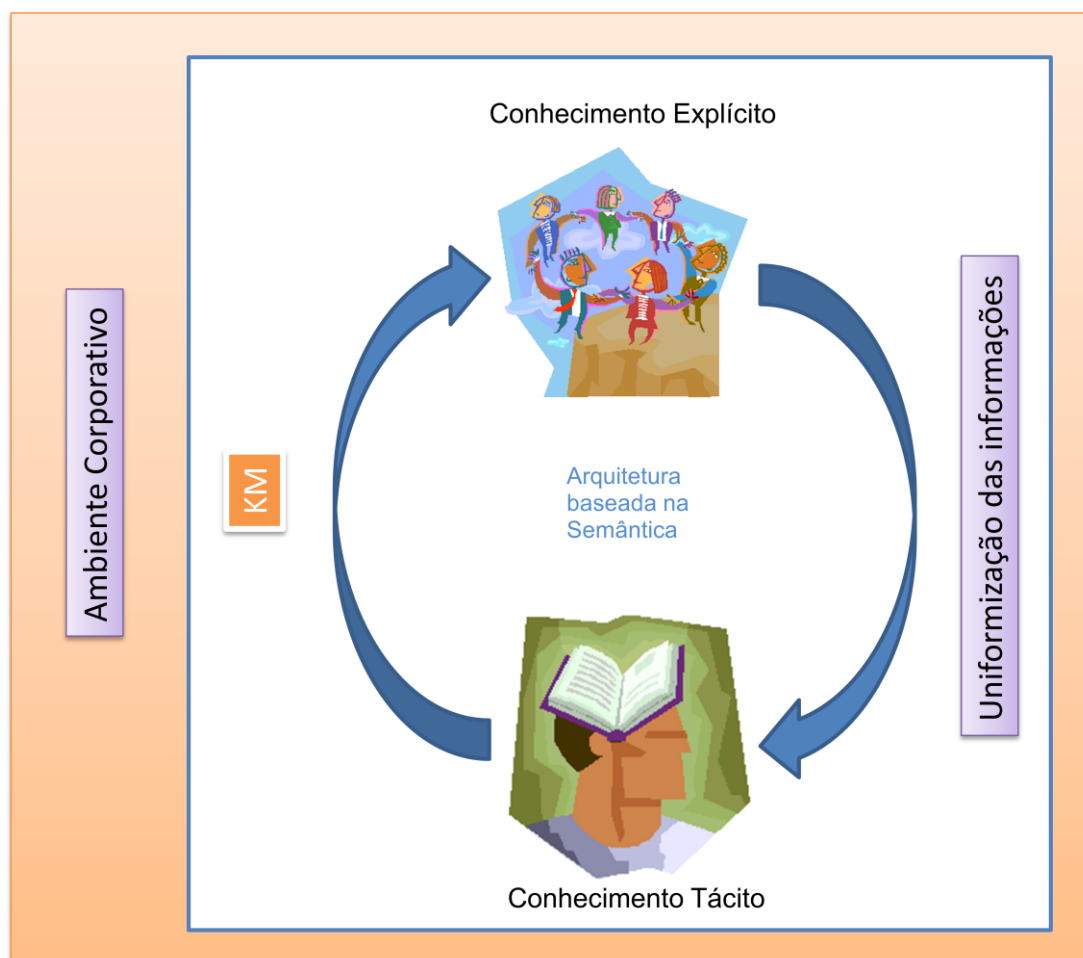


Figura 1.1 – O papel da Arquitetura baseada na Semântica no Ambiente Corporativo

A coordenação dos sistemas de informação torna-se, então, ferramenta importante para aperfeiçoar os processos de geração, difusão e armazenamento de conhecimento nas organizações.

Entretanto, ao analisarmos os atuais ambientes corporativos típicos, com raras exceções, verificaremos que a arquitetura real da maioria das empresas cresceu organicamente e que estes ainda trabalham essencialmente com informações não estruturadas, ou seja, grande parte das empresas maduras possuem vários bancos de dados e paradigmas de programação múltiplos.

Sistemas mais novos e com características mais modernas mesclam-se com sistemas que estão com pelo menos uma década de idade e muitas vezes em outra plataforma. Em um ambiente tão heterogêneo, podemos encontrar diversos tipos de documentos e informações, e os históricos de aplicação de tecnologias convencionais de buscas (tal como a utilizada pelo Google, baseadas somente em indexação de palavras-chave) são estrategicamente falhas, pois possuem capacidades limitadas para estruturar e interpretar os mesmos, simplesmente porque este processo de busca por palavras-chave podem não agregar o “contexto” da solicitação do usuário, a sua intenção e/ou necessidade.

O resultado desse tipo de busca convencional são pilhas e mais pilhas de documentos irrelevantes, quando na verdade, os ambientes corporativos exigem tomadas de decisões rápidas e acertadas, em respostas a questões críticas.

Deste modo, fica bem claro que atualmente os colaboradores que realizam buscas nos documentos da companhia ainda precisam encontrar “agulhas de saber” em “palheiros de irrelevância”, o que nos leva a deduzir que a relevância dos documentos deve ser levada em conta e de preferência num nível de conhecimento profundo de seu *domínio* e problema a ser solucionado.

Faz-se necessário então a criação de uma camada que interligue as plataformas (definida mais a frente como “mediador”) e diferentes setores combinada com uma *ontologia do domínio*, o que tornará possível a criação de tipos de documentos “inteligentes”. Um documento inteligente “sabe sobre” o seu próprio conteúdo de forma que esteja acessível não somente por seres humanos, mas também por processos automatizados que possam “saber o que fazer” com ele, além de permitir que pessoas e computadores trabalhem em cooperação através da interoperacionalidade.

1.1 Objetivos

O objetivo desse trabalho é explorar a proposta de um possível modelo de arquitetura baseado em semântica para utilização em ambientes corporativos capaz de realizar a integração de ferramentas e plataformas e possibilitar que o conhecimento possa ser agregado com a união das várias fontes de informações. Essa conexão entre diferentes tipos de ambientes computacionais proporcionará maior assertividade nas buscas e utilização do conhecimento dentro da organização, aumentando o universo de conhecimento para seus colaboradores, trazendo conseqüentemente vastos benefícios para a corporação.

1.2 Justificativa

O crescente aumento da competição entre as corporações e a velocidade que se exige destas para se adaptar e buscar novas soluções neste panorama altamente competitivo torna a gestão do conhecimento um fator de diferenciação. A proposta de um modelo de arquitetura baseada em semântica para ambientes corporativos possibilitará o aumento no nível de conhecimento dos colaboradores de forma a expandí-lo na mesma proporção, possibilitando uma recuperação de informações mais eficiente e eficaz que trará benefícios significativos na qualidade dos produtos/serviços das corporações.

1.3 Metodologia

Nesse trabalho foram utilizadas técnicas de pesquisa bibliográfica e exploratória para o embasamento teórico possibilitando maior entendimento dos conceitos abordados, estabelecendo a confrontação das informações obtidas e a consistência destas e de outros dados apresentados.

O planejamento deste trabalho envolveu as seguintes fases:

a) *Levantamento bibliográfico* – foi feito inicialmente uma volumosa pesquisa bibliográfica de diversos livros e trabalhos acadêmicos relevantes à área em questão e que estavam disponíveis para acesso.

Todos os trabalhos, sem exceção foram armazenados.

b) *Seleção* – os trabalhos e livros foram filtrados através de leitura exploratória de material compatível e relevante.

c) *Aprofundamento* - foi realizada leitura profunda e analítica com material definitivo buscando justificativas para o tema.

d) *Interpretação* - realizada leitura interpretativa e anotações que ajudariam a elaborar o trabalho.

e) *Análise de exemplos* – procura por experiências sobre o assunto para fazer análises comparativas.

1.4 Estrutura do trabalho

O **capítulo 2** aborda como foco, o cenário atual do ambiente corporativo, o potencial de um modelo com abordagem semântica e sua possível colaboração nos atuais estudos da Web Semântica.

O **capítulo 3** define os principais conceitos da semântica no qual o modelo de arquitetura proposto está baseado, tais como, seu significado e conceitos filhos interligados que são os metadados e as ontologias. Também é dada uma breve introdução sobre o conceito de busca semântica considerado um dos pilares do modelo de arquitetura demonstrada.

O **capítulo 4** explicita o conceito de interoperabilidade e as barreiras que podem impactar na interligação de diferentes fontes de informação além de tratar o conceito de conflito semântico.

O **capítulo 5** explicita e se aprofunda nos principais pilares do modelo de arquitetura supracitado e define como o mesmo poderá ser utilizado dentro das corporações. Explicita-se também os seus três pilares e principais aspectos, elementos envolvidos e desafios para sua implementação.

No **capítulo 6** finalizamos o trabalho expondo como este mesmo modelo poderá agregar valor ao conhecimento nas corporações, através de suas vantagens com relação ao modelo pragmático utilizado nos dias atuais, conhecido simplesmente como sintático.

2 O CENÁRIO ATUAL DOS AMBIENTES CORPORATIVOS

A transição do ambiente de manufatura para a economia da informação e da globalização progressiva dos mercados colocaram novos desafios às corporações. A quantidade de informação que as mesmas têm para produzir, adquirir, manter, procriar e usar aumentou dramaticamente. Embora as tecnologias e ferramentas que ajudam na colaboração e estruturação de conteúdo, tais como wikis, blogs e plataformas de colaboração estejam atualmente em vigor, às corporações ainda buscam abordagens mais capazes para obter, gerir e utilizar o conhecimento necessário para seus processos de negócio. O montante, a heterogeneidade e a multimodalidade de dados permanecem aspectos problemáticos na interoperabilidade de fontes de dados, apresentação de informações e consequente extração do conhecimento.

Como já citado anteriormente, o modelo de arquitetura que propomos neste trabalho, visa oferecer uma solução promissora, mas que também coloca novos desafios. O aspecto principal deste modelo envolve uma mudança do foco em dados e documentos para o seu próprio conteúdo informativo e uma representação legível por máquina, utilizando para isto ontologias próprias do domínio representado.

Sendo uma especificação formal de uma conceitualização, uma ontologia permite que motores de inferência possam obter novas informações que estejam implicitamente contidas em diversos repositórios. Como resultado da necessidade de um melhor gerenciamento de informações em ambientes corporativos a proposta de um modelo de arquitetura baseada na semântica visa lançar novas luzes na utilização de tecnologias semânticas nas corporações.

Focando-se no ambiente controlado em contrário ao ambiente global da Web Semântica evitamos enfrentar problemas ainda por resolver, tais como a

escalabilidade, uma adoção mais ampla de ontologias partilhadas, e as questões de confiança.

Portanto, o potencial em muitos destes cenários corporativos, bem como sua viabilidade em curto prazo alcançam muitos méritos, embora as tecnologias semânticas sejam relativamente jovens e lacunas nas normas e implementações ainda existam.

O tipo de modelo de arquitetura que apresentaremos deverá conter três pilares que consideramos seus “blocos de construção” e as nomeamos como *Engenharia da Ontologia Corporativa*, *Colaboração Semântica Corporativa* e *BUSCA Semântica Corporativa*.

A evolução das pesquisas neste modelo de arquitetura baseada na semântica e que poderá ser utilizada em ambientes corporativos tem como, por que não, colaborar na evolução da própria Web Semântica pública e agregar ainda mais informações ao seu contexto.

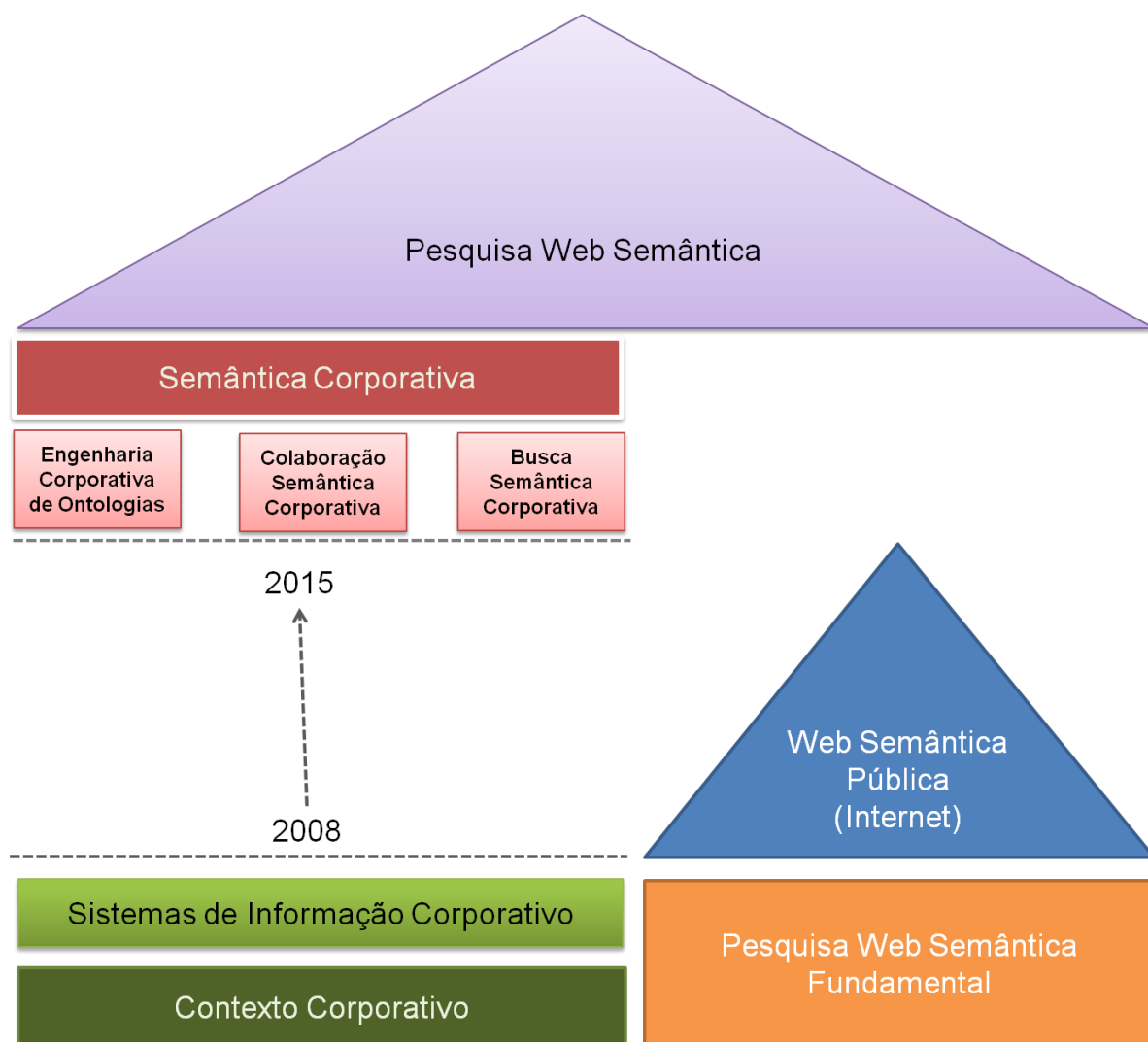


Figura 2.2 – O posicionamento, evolução e maturidade da pesquisa da semântica no ambiente corporativo, seus pilares e sua colaboração na evolução da pesquisa web semântica.

No capítulo 3, detalhamos alguns conceitos básicos da Web Semântica tais como seu significado e objetivos, como a mesma é apoiada pelos metadados e ontologias, e o papel que estes conceitos exercerão dentro do ambiente corporativo.

3 A SEMÂNTICA E SEUS PRINCIPAIS CONCEITOS

3.1 A Semântica

A forma como encontramos dados na Internet, e nas corporações, em Intranets, possui diversas limitações e Breitman (2005) já dizia que a atual maneira de obtê-las pode ser sinonimamente chamada de Sintática, já que os equipamentos computacionais apenas apresentam informações, deixando o processo de interpretação sob a inteira responsabilidade dos usuários que estão utilizando os serviços.

Desta forma, o esforço exigido para classificar e selecionar as informações mais relevantes é gigantesco, segundo Moura (2004) e Bonifacio e Heuser (2002) que complementam dizendo que a Internet atual apresenta grandes problemas para localizar, acessar, apresentar e gerir seus dados e isso acontece pelo simples fato de seu compartilhamento de recursos ser distribuído, seus usuários serem autônomos e heterogêneos, e pela falta de mínimos padrões.

Nos anos 90, Tim Berners-Lee, Fischetti e Dertouzos (2000), propuseram um modelo de Internet que ainda hoje está abaixo das expectativas. Segundo Breitman (2005), a Internet é caótica, do ponto de vista das informações e necessita de uma estrutura organizacional que possibilite a interconexão entre máquinas e pessoas de forma mais simplificada e natural.

Embora a Internet tenha sido criada com a finalidade de possibilitar a troca e recuperação de informações de forma mais fácil e objetiva, Souza e Alvarenga (2004) afirmam que a mesma foi implementada de forma descentralizada e anárquica, vem crescendo exponencialmente e atualmente possui um repositório

inimaginável de documentos que inviabilizam processos de pesquisas eficazes, que recuperem informações realmente relevantes aos usuários.

Atualmente a recuperação de informações desses documentos é executada através de mecanismos de busca, entre os quais, os mais relevantes são o Bing, o Yahoo, e o mais conhecido deles, o Google. Todos estes mecanismos de busca utilizam basicamente o conceito de armazenamento e indexação de palavras-chaves contidas nos textos de documentos, o que demonstra que praticamente ainda não existem estratégias satisfatórias para a indexação correta destes documentos, que desta forma, acabam tornando-se ineficazes (SOUZA; ALVARENGA, 2004).

Estes mecanismos de busca apesar de recuperarem uma grande parte das informações, as recuperam apenas utilizando uma pequena parte da web, enquanto a acessibilidade de parte considerável deste conteúdo fica totalmente bloqueada.

O significado em inglês da palavra Semântica é “o estudo do significado, ou uma interpretação do significado, de uma palavra, sinal ou frase”, ou seja, é um termo muito mais profundo e que se contrapõe diretamente ao conceito sintático considerado bem mais simples. Com base na interpretação direta da palavra surge então, a Web Semântica, através da qual, são estudados mecanismos que busquem dar “significados” as páginas e documentos além de possibilitar que os computadores processem seus conteúdos e os interliguem com outras fontes de dados. Para que isso seja possível Breitman (2005) ressalta a necessidade de utilizar semântica na estrutura dos documentos disponíveis.

Segundo Berners-Lee (2001), a Web Semântica, tem como objetivo criar um ambiente que permita que agentes de software ajudem nas buscas de informações utilizando mecanismos semânticos, que garantam a automatização das pesquisas, baseadas em sua semântica. Souza e Alvarenga (2004) reafirmam este conceito, deixando explícito que a web semântica embute inteligência e contexto nos códigos

de confecção das páginas Web (muitas em XML), de forma a melhorar a interação entre as máquinas e tornar o ambiente mais amigável para seus usuários.

Para exemplificar o conceito da Web semântica, retiramos do livro de Seagaran, Evans e Taylor (2009), o seguinte exemplo:

Avaliando as seguintes frases, verificamos que elas têm a mesma estrutura verbal, sujeito-verbo-objeto:

- Fulano adora aviões.
- Aviões assustam Beltrano.

Se analisarmos seu conteúdo, verificaremos que cada parte da frase representa uma informação diferente. Duas pessoas estão envolvidas na frase, Fulano e Beltrano. As palavras aviões estão se referindo a uma classe de transporte qualquer, e as palavras “adora” e “assustam” relacionam as pessoas com o transporte nessa frase.

Se conhecermos cada palavra individualmente é possível compreender facilmente a sentença. Após a leitura dessas frases estaremos inevitavelmente equipados com mais conhecimento e é exatamente desta forma que a Web semântica se propõe a funcionar, onde símbolos referenciam a conceitos ou objetos, e a seqüência desses símbolos são capazes de construir um significado destinado a frase ou o texto.

Este conceito implementa a idéia de um banco de dados global e uma reestruturação da arquitetura sintática, dando a todo o ambiente mais colaboração e inteligência, e permitindo que o armazenamento dos dados sejam importantes para a construção das informações que circulam pelo ambiente global (URIARTE, 2010).

Podemos considerar que o desenvolvimento da Web semântica ainda está em seu início e melhorará muito a partir do momento que os metadados e as ontologias

(descritos abaixo) forem mais bem utilizados, porém o modelo de arquitetura semântica para ambientes corporativos se baseará em vários conceitos da Web Semântica.

3.1.1 Metadados

A definição padrão de metadados é “dados que descrevem dados”. Desta forma, metadados são considerados abstrações dos dados, ou seja, dados de mais alto nível que descrevem dados de um nível inferior e são utilizados para ajudar em sua identificação, descrição, localização e gerenciamento. (HORROCKS 2002, DECKER 2000, BERNERS-LEE 2001).

Metadados podem possuir classificação estrutural ou semântica. Um metadado estrutural representa a informação que descreve a organização e estrutura dos dados gravados. Por exemplo, informações sobre o formato, os tipos de dados usados e os relacionamentos sintáticos entre eles.

Em contraste, metadados semânticos fornecem informações sobre o significado dos dados disponíveis e seus relacionamentos semânticos. Por exemplo, dados que descrevem o conteúdo semântico de um valor de dado (como unidades de medida e escala), ou dados que fornecem informações adicionais sobre sua criação (algoritmo de cálculo ou derivação da fórmula usada), linhagem dos dados (fontes) e qualidade (atualidade e precisão).

Então, é desejável uma conceitualização de um domínio específico de problema, ou ontologias que forneçam um acordo comum de vocabulários para assegurar a correta interpretação dos metadados disponíveis. Uma ontologia fornece um entendimento sobre uma conceitualização compartilhada de um determinado domínio de aplicação. Os conceitos específicos numa ontologia fornecem um vocabulário comum para que nenhuma negociação adicional seja necessária.

3.1.2 Ontologias

O termo “ontologia” advém da filosofia, onde uma Ontologia é uma descrição da Existência. Porém, na tecnologia, vem sendo aplicado desde o início da década de 90 na área de inteligência artificial para representar o conhecimento computacional de conhecimento em áreas como a gestão do conhecimento e processamento de linguagem natural. Podemos então entendê-la como uma especificação formal de uma conceitualização que exige, entretanto, uma formalização, visto que ela pode gerar algumas confusões.

A estrutura de um conhecimento representado de alguma maneira normalmente é baseada numa conceitualização: objetos, conceitos interligados e outras entidades que são assumidas para existirem em alguma área de interesse e os relacionamentos que são organizados entre elas. Toda base de conhecimento está comprometida com alguma conceitualização, explícita ou implícita. Uma ontologia é uma especificação explícita de uma conceitualização. (JARRAR, VERVENNE e MAYNARD, 2007).

Usam-se ontologias comuns para descrever a representação de uma base de conhecimento para um conjunto de agentes. Desta forma, estes agentes podem passar informações sobre o domínio em voga, sem necessariamente operar sobre uma teoria globalmente compartilhada. (HUHNS e SINGH, 1997).

O conhecimento é atribuído aos agentes através da observação de suas ações. Um agente “sabe” algo se ele age como se ele tivesse a informação e está agindo racionalmente para alcançar seus objetivos. As “ações” dos agentes, incluindo servidores de base de conhecimento e sistemas baseados em conhecimento, podem ser vistos através de uma ordem e perguntas da interface funcional.

Visto deste ângulo uma ontologia comum define o vocabulário com a qual as pesquisas e afirmações são trocadas entre os agentes. Para representar uma conceitualização é necessária uma linguagem de representação. Muitas linguagens e sistemas de representação são definidos para este fim.

Para aplicações sobre a Web, é importante ter uma linguagem com uma sintaxe padronizada. O que nos leva novamente às linguagens baseadas em XML, definindo-as sobre o topo do XML. Um exemplo é o Resource Description Framework Schema Language (RDFs) entre outras. Todos eles usam a sintaxe do XML, mas com leves diferenças nas *tags* de marcação. (BERNERS-LEE, HENDLER e LASSILA, 2001).

Outra proposta que estende o RDF e o RDF Schema é o OIL (Ontology Inference Layer) que provê uma camada de inteligência entre os dados. Os formatos RDF e RDFs já estão em uso na comunidade bibliotecária e sendo aceitos como um padrão.

Um sucessor do OIL é o DAML+OIL, que cria uma linguagem de manipulação para o OIL e foi desenvolvido conjuntamente por um grupo de cientistas europeus e americanos, porém não é o objetivo deste trabalho aprofundar-se nas técnicas e definições utilizadas pelos processos de mais baixo nível que serão utilizados na Semântica no Ambiente Corporativo.

3.1.2.1 Tipos de ontologias

Existem alguns tipos de ontologias, de acordo com seu grau de genericidade (adaptado de (GÓMEZ, PEREZ 99)):

- *Ontologias de representação* que definem os primórdios da representação - como atributos e outros.

- *Ontologias gerais (ou de topo)* trazem definições necessárias para a compreensão de aspectos do mundo, como tempo, processos, papéis, espaço, seres, coisas, etc.
- *Ontologias centrais (core ontologies) ou genéricas de domínio* definem os ramos de estudo de uma área e/ou conceitos mais genéricos desta área.
- *Ontologias de domínio* tratam de um domínio mais específico de uma área genérica de conhecimento, como direito tributário, microbiologia, etc. (ARANGO e PRIETO-DÍAZ, 1991).
- *Ontologias de aplicação* procuram solucionar um problema específico de um domínio, como, por exemplo, identificar doenças do coração, a partir de uma ontologia de domínio de cardiologia.

Nem todos os tipos são necessários para a construção de uma aplicação corporativa e temos ainda que nos preocupar em manter as ontologias reusáveis, ou seja, escolher uma ontologia num nível acima da ontologia em uso, ou seja, que esteja ligada a esta, mas evitando relacionamentos restritos com uma ontologia específica. (CROW E SHADBOLT, 2001),

3.1.2.2 Exemplos de Ontologias

As ontologias que a literatura mais referencia são as gerais (ou de topo, como visto acima na definição), pois muitas aplicações usam-nas como base.

Ontologias lingüísticas, como a Wordnet, WordnetPT e a WordnetBR contém sinônimos, adjetivos, advérbios e verbos relacionados por redes semânticas, e

também são bastante referenciadas por sua aplicação em sistemas de processamento de linguagem natural.

3.1.2.3 Benefícios das Ontologias

Ao utilizar ontologias podemos identificar diversos benefícios, entre eles:

- A oportunidade de se reutilizar estas ontologias e bases de conhecimento, mesmo com adaptações e extensões.
- A possibilidade de tradução entre diversas linguagens e formalismos de representação do conhecimento. A tradução concretiza um ideal perseguido por gerações de pesquisadores de Inteligência Artificial. Ela facilita o reuso de conhecimento, e pode vir a permitir comunicação entre agentes em formalismos diferentes.
- O mapeamento entre formalismos de representação de conhecimento, que, utilizando componentes de conectividade para SGBD interliga dois formalismos criando uma interface interoperável de acesso comum para eles, permitindo a um agente acessar o conhecimento de outro agente.

3.2 Busca Semântica

Embora a contribuição de informações seja farta e simples, encontrar informações úteis é sempre bem mais difícil. A busca semântica é uma parte da visão da Web Semântica e que deverá ser aplicada também ao conceito e modelo proposto neste trabalho, porém neste primeiro momento apresentaremos apenas o seu conceito básico já que acima definimos a mesma como um dos pilares e deveremos abordá-la mais tarde em detalhes.

Ao contrário das consultas por palavras-chave, existentes atualmente na maioria dos mecanismos de pesquisa (incluindo o maior deles, o Google), a busca semântica ajuda o usuário encontrar informações de acordo com o seu conceito.

Dois documentos podem descrever o mesmo conceito, mas não compartilhar nenhuma palavra chave. Estes documentos têm proximidade conceitual, ou significado similar, mas não seriam incluídos no resultado de uma simples consulta por palavras-chave.

Uma consulta semântica olhará para o seu significado e incluirá os dois documentos no resultado da busca. No entanto, a construção semântica da base de conhecimento não pode ser alcançada sem o uso de ferramentas correspondentes.

Algumas destas ferramentas são a aplicação dos metadados e ontologias.

4 DESAFIOS: A INTEROPERABILIDADE E O CONFLITO SEMÂNTICO

4.1 Uma palavra sobre Interoperabilidade

Por definição, a interoperabilidade é a habilidade de dois ou mais sistemas ou componentes compartilhar informações e utilizar estas informações compartilhadas (IEEE, 1990). Hoje em dia, o compartilhamento de informações é um fator crítico para o sucesso das organizações, devido à emergente necessidade de comunicação e o compartilhamento de informações. Como citamos anteriormente, a habilidade da organização em adaptar-se rapidamente a novas fontes de informação e modelos de negócios é fator determinante para seu crescimento e sobrevivência no mercado atual.

A história demonstra que o uso de um *único sistema de bancos de dados* era suficiente para organizar, manter e gerenciar os dados destas aplicações. No entanto, as maiorias das empresas atuais estão interessadas em compartilhar o máximo de informações com seus colaboradores.

Neste contexto, o ideal seria que esses mesmos colaboradores pudessem realizar consultas através de múltiplas fontes e encaminhá-las a atualizações diretamente às fontes de onde são provenientes respeitando as restrições de integridade dos múltiplos bancos de dados.

Algumas das barreiras que podemos encontrar na interoperabilidade são:

- *Distribuição* - As fontes de informação podem estar fisicamente localizadas em um único ou em múltiplos computadores, que por sua vez podem estar coligados ou geograficamente distribuídos, mas interconectados por um sistema de comunicação. Devido à evolução das redes de computadores, a

maioria dos computadores encontra-se conectado a algum tipo de rede. Atualmente, o exemplo mais trivial é a Internet. Portanto, é natural pensarmos em interoperar estes sistemas fisicamente dispersos, realizando sua comunicação via rede.

- *Heterogeneidade* - Naturalmente, a heterogeneidade dos sistemas de informação ocorre devido às diferentes soluções adotadas. Cada solução possui uma forma particular de entender o problema, seu próprio ambiente tecnológico e requisitos específicos das aplicações a serem desenvolvidas. Devido a estes e outros fatores, até mesmo os mesmos conceitos do mundo real são modelados de diferentes formas. Dentre as diversas classificações para heterogeneidade apresentadas na literatura, neste trabalho foi adotada a classificação apresentada em (BUSSE et al., 1999) que divide o problema da heterogeneidade dos sistemas de informação em três categorias principais: a sintática, de modelo de dados e a lógica.
 - Sintática: engloba problemas sintáticos relacionados aos aspectos técnicos para comunicação dos sistemas de informação envolvidos. A heterogeneidade técnica aborda diferenças nos aspectos técnicos, tais como: hardware, plataforma, sistema operacional e métodos de acesso (protocolos, etc). Já a heterogeneidade de interface de acesso existe quando o acesso aos componentes é feito de forma diferente, devido a diferenças nas linguagens de consulta (heterogeneidade de linguagem) e a diferenças referentes às restrições das consultas, isto é, quando apenas algumas operações são permitidas.
 - De modelo de dados: demonstra as diferenças entre os modelos de dados que podem ser estruturalmente diferentes. Como exemplo, temos herança e generalização de objetos que não é oferecido pelo modelo relacional.
 - Lógica: envolve as diferenças lógicas relacionadas à comunicação dos sistemas de informação envolvidos. É classificada em três tipos: semântica, esquemática e estrutural. A heterogeneidade semântica preocupa-se com a semântica dos dados e metadados. Ela ocorre

quando existe discordância a respeito do significado, interpretação ou uso pretendido, de um determinado objeto, atributo ou tabela de um banco de dados. Sem dúvida, este é um dos principais problemas a serem solucionados para interligar diversas fontes de dados. Até mesmo esquemas que sejam formulados no mesmo modelo podem ter interpretações diferentes. Os nomes das entidades utilizadas em um esquema carregam uma semântica implícita, representando um conceito que pertence a um determinado contexto. *A interpretação desses nomes não necessariamente irá coincidir quando realizada por diferentes pessoas, caracterizando um conflito semântico.* Esses conflitos podem ocorrer quando nomes iguais representam diferentes conceitos (homônimos) ou quando nomes diferentes representam o mesmo conceito (sinônimos). Além disso, um mesmo atributo pode ter a mesma semântica e mesmo assim apresentar heterogeneidade na sua representação. Como exemplo, temos “preço” como nome de um atributo de uma fonte de dados X, e o mesmo nome para um atributo de uma fonte de dados Y. Na fonte X, o preço está sendo representado em “dólares americanos”, enquanto na fonte Y, em “reais”.

- *Autonomia* - A autonomia das fontes de informação é outro fator que pode influenciar na interoperabilidade de sistemas de informação. A autonomia depende de diversos fatores que podem determinar o grau de independência das fontes que desejamos interligar, como por exemplo: o fato das fontes trocarem ou não informações, de poderem executar transações de forma independente e de ter ou não permissão de modificar informações.

4.2 O conflito semântico

Conflitos semânticos podem ser extremamente problemáticos se não forem tratados corretamente. Se uma corporação que utiliza diversas aplicações e bancos de dados, decidir mudar o significado de algum dado (como por exemplo, mudar a moeda corrente de Euro para Real), vários problemas técnicos e organizacionais podem surgir.

Como, neste caso, poderemos identificar o impacto dessa mudança nas aplicações? Geralmente, esses problemas são resolvidos de forma manual, o que causa reduzida disponibilidade e confiabilidade dos sistemas.

Para que resolvamos estes problemas, é necessário que haja uma integração que seja capaz de garantir interoperabilidade entre os dados. Para isto, devemos levar em consideração não somente a informação, mas também o seu “contexto”. Em termos práticos, esse “contexto” pode ser definido como metadados que sejam relacionados a um significado, às propriedades e à organização da informação. Nele, podem estar contidas informações variadas, como especificações de medidas, de formato de dados (formato de datas, fator de escala), informações sobre qualidade (confiabilidade, fonte) e adoções utilizadas na obtenção ou cálculo de dados, em métodos ou na definição de fórmulas.

Para garantir esta interoperabilidade semântica, os valores devem ser comparados semanticamente. Muitas vezes, porém, o resultado de tal comparação pode ser relativo. Por exemplo, o valor Washington e o valor EUA podem ser semanticamente equivalentes, caso se esteja interessado no país que eles denotam, ou não equivalente, quando se está interessado na cidade representada por eles. Isso demonstra que quando comparamos dois valores semanticamente, devemos fazê-lo em relação a um contexto, que chamamos de contexto alvo. O resultado desta comparação é obtido convertendo ambos os valores semânticos para o contexto alvo, e comparando os valores simples resultantes desta conversão.

Este trabalho tem o objetivo de propor um modelo de arquitetura baseado em semântica que tentará resolver alguns destes conflitos e manter a interoperabilidade. Porém, o tema do conflito semântico envolve determinada complexidade que não será abordada neste trabalho, pois isto demandaria estudo mais profundo de baixo nível apenas em tais conflitos, como dicionários e rotinas de transformação de dados.

O modelo que demonstraremos estará baseado em um “mediador” que receberá e tratará as consultas submetidas ao sistema de descoberta e integração e que será responsável pela reformulação destas consultas em subconsultas que serão enviadas ao repositório de ontologias e fontes de dados. Este “mediador” conterá as rotinas de baixo nível que transformarão informações de forma a minimizar conflitos, porém seus pormenores em “baixo nível” não estão no foco deste trabalho.

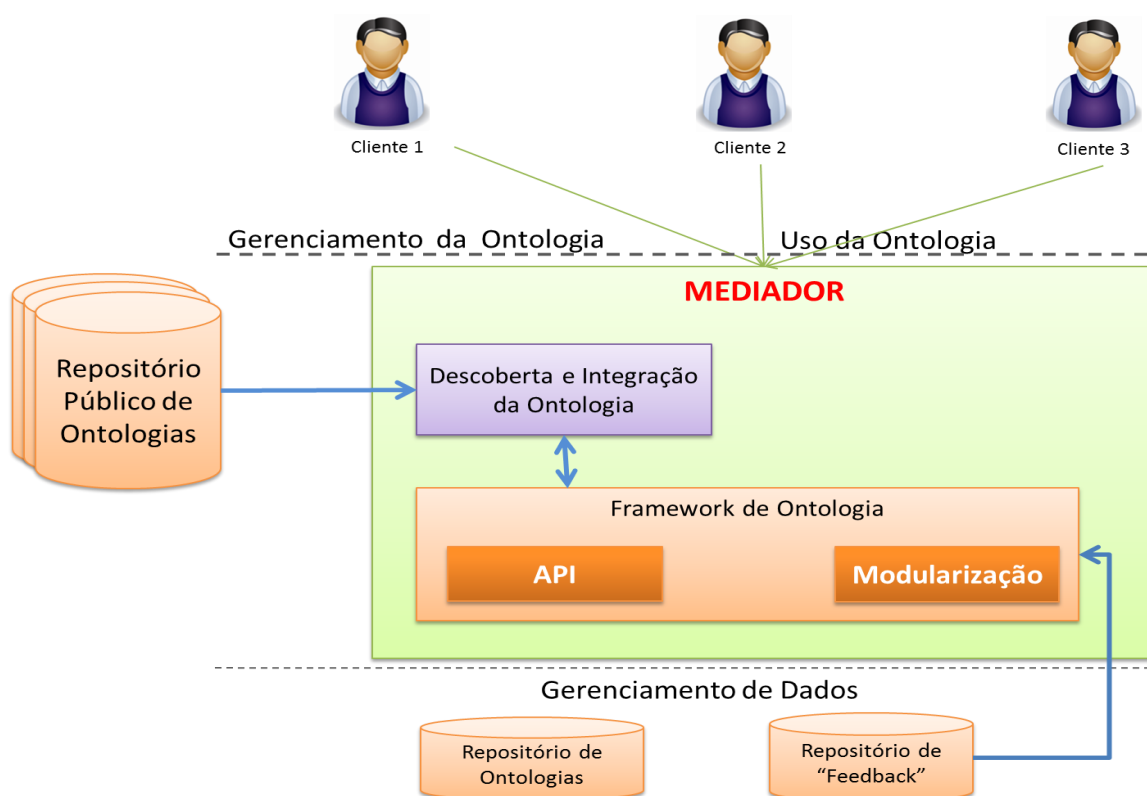


Figura 4.3 – Modelo sugerido de arquitetura baseada em semântica e o conjunto de módulos que compõe o mediador de consultas

5 MODELO DE ARQUITETURA BASEADO NA SEMÂNTICA: OS PILARES

5.1 Engenharia Corporativa de Ontologias

Em um mercado altamente dinâmico, as corporações dependem da rápida integração de seus conhecimentos com seus processos de negócios. Enquanto o conhecimento corporativo multiplica-se e evolui, as empresas buscam ferramentas de apoio à gestão do conhecimento que suportem este processo de forma eficiente e com custo/benefício otimizado.

Na economia moderna, o sucesso das empresas depende fortemente da sua capacidade a rápida adoção de novas conquistas em seus processos de negócio. A criação de valor econômico através do conhecimento é essencial. Mas, ERPs clássicos e sistemas de software são baseados em sistemas de informação estática e soluções de banco de dados. Eles não são capazes de reagir à dinamicidade do conhecimento e não permitem a adoção ágil de novas conquistas para os processos de negócio.

Eles são desenvolvidos principalmente para fins muito especiais visando atender algumas funcionalidades e requisitos da empresa. Aplicações de Gestão do Conhecimento são construídas sobre bases de conhecimento (KB), que permitem gerir o conhecimento declarativo, que é criar, modificar e apagar conhecimento armazenado na KB. (ZACK, 1999).

Tais aplicações são capazes de levar novos conhecimentos imediatamente em conta e, portanto, é uma abordagem muito promissora para deslocamento de dados e sistemas clássicos baseados em software.

Ser capaz de lidar com a dinamicidade e o poder do conhecimento é a chave para processos de negócio flexíveis e ontologias semânticas são um importante meio para a representação de conhecimentos. Elas fornecem um entendimento compartilhado de certo domínio de interesse.

Em um ambiente corporativo, ontologias devem descrever termos e suas inter-relações relevantes ao contexto de negócios. No entanto, existem alguns “*desafios*” (que chamaremos abaixo de **(Dn)**) a vencer entre as necessidades corporativas e o atual status que está a pesquisa de ontologias, são eles:

- **(D1)** o atual nível do estudo acadêmico nesta área – a pesquisa de ontologia e semântica é um campo novo em comparação com outras disciplinas da ciência da computação, como, por exemplo, pesquisas em bancos de dados. Ferramentas e casos de uso são apenas orientados a problemas criados cientificamente. Poucos pioneiros nas áreas de integração de informações empresariais e gestão de conteúdo permitem a criação de casos de uso no mundo real, portanto os objetivos formulados e a evolução duradoura raramente acontecem, ainda.
- **(D2)** a diferença de maturidade das aplicações existentes – Em consequência das lacunas ainda existentes nos estudos acadêmicos, há poucas ferramentas no mercado e menos ainda as que alcançaram grau de maturidade para uso produtivo. Mesmo para essas ferramentas a falta de comparações e “benchmarks” traz um grau de transparência muito baixo o que gera um grau de desconfiança alta nos meios da computação.
- **(D3)** a falta de processos – os desafios referentes à orientação acadêmica da pesquisa em geral e a imaturidade das aplicações e ferramentas são aspectos importantes para a falta de aceitação de ontologias na indústria. A maioria dos pesquisadores desenvolveu metodologias de engenharia de ontologias

somente para a Web. Isso significa que estes pesquisadores estão visando somente o desenvolvimento de ontologias para o mundo todo ou ao menos para domínios interempresariais. O conhecimento individual como uma competência em aplicações intracorporativas e uma demanda de segurança para este conhecimento aparentemente está fora do escopo. A engenharia de software passou a ser utilizada como metodologia no processo de desenvolvimento de ontologias, no entanto, a agilidade de processos de engenharia do conhecimento é muito superior à agilidade dos processos de software.

- **(D4)** dificuldades para estimar a relação custo-benefício – as empresas tem uma forte necessidade de estimar o custo-benefício de investimentos em infraestruturas de informação e sistemas de informação de base ontológica são de difícil estimativa já que devem integrar todas as partes do conhecimento corporativo enquanto as aplicações que utilizam esta ontologia servem apenas para um propósito fixo, ou seja, devido aos múltiplos propósitos e “braços” de uma aplicação ontológica, aplicar regras de estimativas utilizadas em sistemas convencionais únicos é extremamente complicado.

Surge então a necessidade de uma metodologia para criação de ontologias que sejam utilizáveis em ambientes corporativos, ao que denominamos neste trabalho de *Engenharia Corporativa de Ontologias*. Tal metodologia pretende fechar (ou vencer) as lacunas ou desafios citados acima sempre mantendo o foco no ambiente corporativo. Ou seja, ela tem que atender aos seguintes “*requisitos*” (que chamaremos abaixo de **(Rn)**):

- **(R1)** A influência do desenvolvimento da ontologia e o processo de manutenção da mesma pelos profissionais de cada setor têm de ser minimizados para evitar influência negativa sobre a sua produtividade.

- **(R2)** Os sistemas já existentes e em execução não devem ser perturbados.
- **(R3)** A ontologia tem de evoluir com o progresso da empresa.
- **(R4)** A necessidade de engenheiros de ontologia tem de ser minimizada, para reduzir custos.
- Ser focado em ambientes corporativos que estejam dispostos a realizar investimentos em novos estudos que inerentemente “fecham” o desafio **(D1)** e preparam o terreno para novos conhecimentos com base em aplicações **(D2)**.
- Reagir à agilidade dos processos da engenharia do conhecimento **(R3)**.

Para atender aos requisitos acima a Engenharia Corporativa de Ontologias deve seguir alguns princípios básicos na construção de ontologias corporativas (ou de domínio corporativo):

- *Clareza*: manter a objetividade de se definir apenas o que se presume ser útil na resolução do problema definido. Apesar da diminuição de definições, as mesmas devem ser as mais completas possíveis.
- *Legibilidade*: As definições devem corresponder com definições usuais e informais. A ontologia deve usar um vocabulário compartilhável pelos especialistas do domínio.
- *Coerência*: As inferências derivadas da ontologia devem ser corretas e consistentes do ponto de vista formal e informal com as definições.
- *Extensibilidade*: A ontologia deve permitir extensões e especializações e com coerência, sem a necessidade de *revisão de teoria*, que consiste na revisão lógica automática de uma base de conhecimento em busca de contradições.
- *Mínima codificação*: Devem ser especificados conceitos genéricos independente dos padrões estabelecidos para mensuração, notação e codificação.

- *Mínimo compromisso ontológico*: Para maximizar o reuso, apenas o conhecimento essencial deve ser incluído, gerando a menor teoria possível.

Abaixo podemos visualizar como seria o ciclo de vida da semântica em uso no ambiente corporativo, e o papel da Engenharia Corporativa em seu contexto:

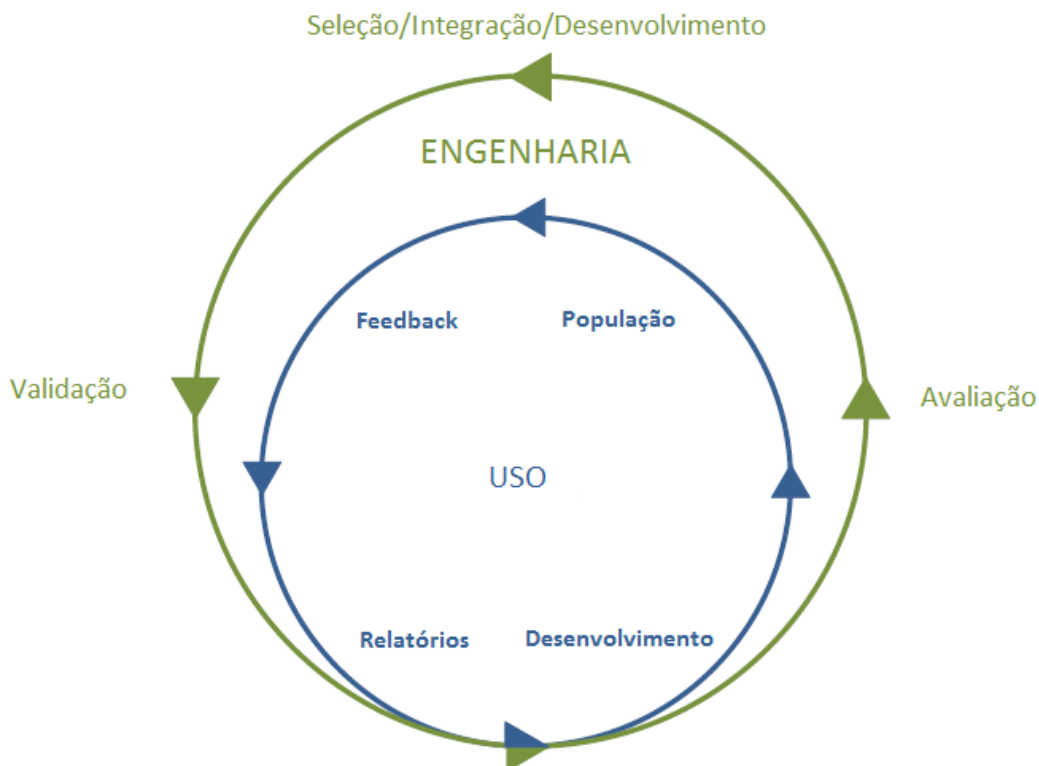


Figura 5.4 – O Ciclo de Vida da semântica em uso no ambiente corporativo e o papel da Engenharia de Ontologias

5.1.1 Tipos de Colaboradores

No uso deste modelo proposto, podemos distinguir diferentes grupos de usuários e que, por sua vez, podem influenciar no desenvolvimento destas ontologias. São os grupos de colaboradores que envolvem desde o usuário engenheiro de ontologias propriamente dito, os usuários de aplicação, os usuários de ontologias e os gestores de conhecimento.

Dependendo do conhecimento e da experiência destes usuários podemos explicitar os seguintes grupos:

Usuários de Aplicação: não está ciente do uso de ontologias no sistema, para eles o foco está na funcionalidade da aplicação. Assim, eles não estão diretamente envolvidos na engenharia de ontologias, mas fornecem implicitamente “feedback” aos Engenheiros de Ontologias apenas fazendo uso da aplicação, por exemplo, o sistema observará a interação que existirá entre este usuário e ela mesma.

Usuários de Ontologias: Este grupo usará principalmente a ontologia somente para leitura como se fosse um vocabulário, por exemplo, para anotação de conteúdo. Do seu ponto de vista sobre ontologias são especialmente capazes de julgar a adequação do domínio no modelo. Seus comentários e discussões serão valiosos “feedbacks” para os Gestores de Conhecimentos e Engenheiros de Ontologias.

Gestor de Conhecimento: serão os responsáveis diretos pela manutenção da ontologia, no contexto de uma corporação. Assim, eles não estão apenas usando ou comentando sobre uma ontologia, mas ativamente modificando-a. Um Gestor de Conhecimento pode ser responsável por uma ontologia completa ou apenas por uma parte dela, por exemplo, um módulo de ontologia correspondente a um departamento específico. As mudanças realizadas pelos Gestores de Conhecimento podem ser objeto de revisão pelo Engenheiro de Ontologias.

Engenheiro de Ontologias: tem um profundo entendimento da modelagem de ontologias e não serão necessariamente especialistas do domínio da aplicação. Em contraste com os Gestores de Conhecimento eles têm que considerar a ontologia completa e as interdependências entre os módulos de ontologias, mas não necessariamente precisa conhecer o domínio da aplicação no detalhe. Além disso, Engenheiros de Ontologias tem profundo conhecimento sobre os formatos representativos e suas tecnologias e metodologias.

Tabela 1 - Grupos de usuários e suas operações típicas

Grupo de Usuário	Operações	Nível
Usuário de Aplicação	Nenhuma operação que afete a ontologia	(-)
Usuário de Ontologia	Busca, navegação, comentários, conceitos de “tags” e de uso	(+)
Gestor de Conhecimento	Todas as operações dos usuários de ontologia mais a modificação direta das ontologias e gerenciamento dos padrões do projeto	(++)
Engenheiro de Ontologias	Todas as operações	(+++)

5.1.2 Modularização e integração de Ontologias

Dividir para conquistar é uma idéia muito bem estabelecida no campo da engenharia de software. Desenvolvimentos de grandes e complexos sistemas de software em pequenos blocos baseados em componentes simplificam seu entendimento, assim como seu gerenciamento e permite a reusabilidade e escalabilidade dos módulos por todo o software. Desta forma, o projeto de criar ontologias de uma maneira modular é intuitivamente promissor, a fim de se beneficiar das mesmas vantagens.

Entretanto, modelar conhecimento representado por ontologias é fundamentalmente diferente de desenhar módulos de software, que descrevem principalmente processos. Metodologias e ferramentas de engenharia de software não podem ser facilmente transferidas para a engenharia de ontologias o que exige uma investigação muito mais aprofundada.

Os objetivos gerais da modularização são os seguintes: **(1)** criação de componentes, que possam ser reutilizados de uma forma flexível e dirigidos a **(2)** escalabilidade e **(3)** utilização personalizada; **(4)** compreensibilidade aumentada para os seres humanos por módulos bem dimensionados evitando ontologias muito grandes, de tal forma que a complexidade **(5)** possa ser gerenciada facilmente.

A fim de esclarecer a necessidade da modularização e integração de ontologias e suas vantagens, especialmente para a engenharia de ontologias corporativa é muito importante ter um olhar mais atento sobre o seu impacto sobre as lacunas e as necessidades previamente introduzidas. Mantendo sempre o contexto corporativo em conta, especialmente em **(D4)** (o “desafio” de custo-benefício), a reutilização de ontologias existentes modulares é sempre muito importante para reduzir as despesas de capital e tempo para implementação das mesmas.

Além disso, módulos bem dimensionados aumentam a inteligibilidade dos seres humanos, são de fácil manutenção e permite uma melhor gestão da complexidade, o que leva a uma menor necessidade de Engenheiros de Ontologias (**R2**) e permitem a evolução das ontologias (**R3**).

Ontologias modulares também são benéficas e eficientes, pois possuem escalabilidade e personalização. Durante seu uso apenas os módulos que são necessários no que diz respeito à aplicação e ao usuário podem ser selecionados e integrados para que o carregamento de peças desnecessárias e desperdício de armazenamento, bem como o poder de computação possam ser evitados.

A estrutura e o conteúdo semântico da própria ontologia no ambiente fechado e controlado das corporações permitem a obtenção de informações adicionais benéficas para a modularização e a integração de ontologias. Em contraste com os sistemas abertos e de grande escala como o World Wide Web, as configurações corporativas tornam possível ter informações sobre o uso e evolução de ontologias, bem como levar em consideração a estrutura do ambiente que compreende os departamentos, os fluxos de trabalho e colaboradores.

Ao observar o uso e ter conhecimento sobre os colaboradores, seus fluxos de trabalho e departamentos, fica muito mais fácil concluir quais as partes relevantes de uma ontologia para casos de uso específico. Isso permite a criação de módulos que são otimizados para tais casos de uso e torna as ontologias personalizadas e eficientemente utilizáveis.

Além do mais, observar a evolução da ontologia ajuda a identificar peças confiáveis e peças “vagas”. Conceitos e relações que são usados com frequência, mas não são alterados podem ser marcados como confiáveis, enquanto outros, que são mudados com muita frequência em relação ao seu uso podem ser visto como “vagos” necessitando mais investigação por engenheiros de ontologia e/ou especialistas de domínio.

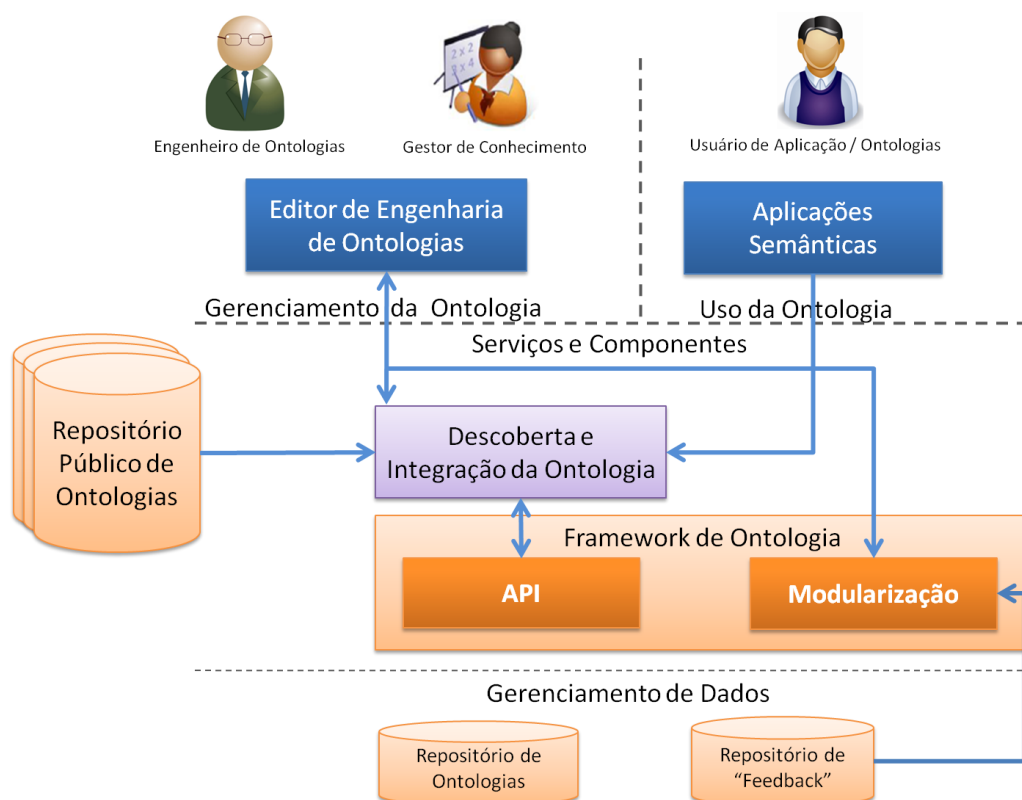


Figura 5.5 – Componentes para modularização e integração de ontologias e decomposição simples dos módulos do “mediador”

A figura 5.5 demonstra os componentes funcionais necessários em um ambiente corporativo para realizar a modularização e a integração de ontologias e assim satisfazer as metas supracitadas. Podemos observar partes relevantes da arquitetura global, os principais módulos do “mediador” e visualizar a interrelação entre eles. O componente que chamamos de Editor de Engenharia de Ontologias é um aplicativo front-end para engenheiros de ontologia e gestores de conhecimento. É uma ferramenta para criar, carregar, visualizar e modificar ontologias diretamente.

O componente de Descoberta e Integração das Ontologias tem acesso a um repositório local de ontologia bem como outros repositórios disponíveis publicamente (tais como a Wordnet, o Swoogle, etc), que seguem requisitos

ontológicos e, baseado em configurações previamente carregadas, é capaz de localizar e integrar módulos de ontologias.

Estes requisitos podem ser definidos manualmente por engenheiros de ontologia usando o Editor de Engenharia de Ontologias ou automaticamente por Aplicações Semânticas. O componente Modularização contido no Framework de Ontologia é capaz de realizar o particionamento de ontologias. De acordo com “feedbacks” do Repositório de “Feedbacks” podem particionar a ontologia automaticamente. Além disso, é possível usar seus algoritmos manuais próprios através do Editor de Engenharia de Ontologias. Também faz parte do componente de modularização a transformação dos dados na tentativa de se evitar conflitos semânticos.

5.2 Colaboração Semântica Corporativa

Enquanto os colaboradores/usuários podem ser considerados os principais atores no processo de evolução do conhecimento nas empresas, as ferramentas atuais de mercado somente oferecem suporte para edição direta da ontologia, exigindo assim, a intervenção manual de Engenheiros de Ontologias.

Isso dificulta o processo de evolução das ontologias e pode tornar o processo extremamente oneroso. Seria desejável que os próprios colaboradores/usuários pudessem modificar ontologias sem ter que ser adepto das formalidades necessárias para tais alterações.

No figura 5.6, propomos um modelo de arquitetura de uma ontologia baseado em um editor web simples e que permite que colaboradores ou especialistas de domínio sem conhecimento profundo sobre ontologias possam contribuir para o “corpus” do conhecimento corporativo.

O modelo proposto na figura abaixo permite acesso às ontologias (o framework de ontologias), rastrear interações de colaboradores/usuários (tracking dos feedbacks) e ajudar o colaborador na modelagem dos conceitos da ontologia.

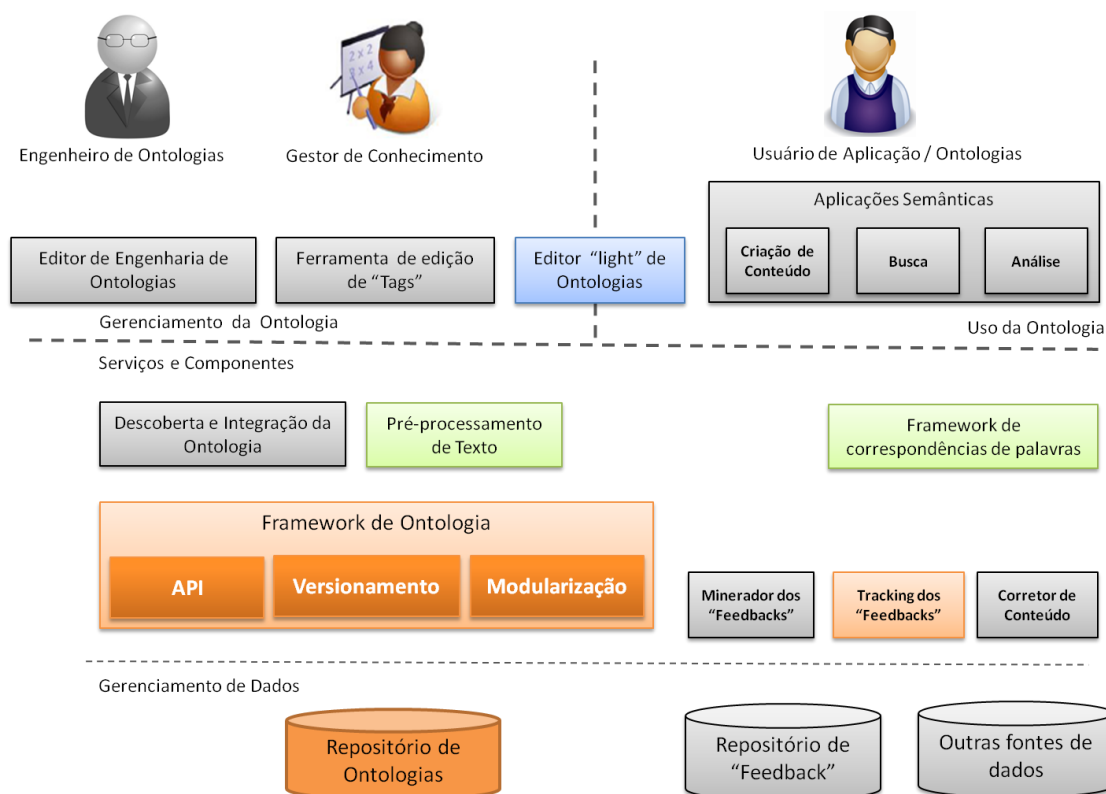


Figura 5.6 – Visão global do modelo proposto de arquitetura baseado em semântica

Abaixo citamos como funcionaria esta arquitetura e as operações que cada usuário exerceria. Algumas operações podem requerer um conhecimento mais profundo da ontologia, portanto na lista de operações indicaremos o grau requerido utilizando os níveis citados na tabela 1.

Qualquer alteração que seja realizada a ontologia deve ser imediatamente permanente e visível para todos ou para ser revista pelo Gestor de Conhecimento.

As principais operações na sequência seriam:

- *Busca de Conceitos* – um usuário busca por um conceito na ontologia que corresponde a um determinado termo de busca. Se a informação estiver disponível na ontologia e retornar um determinado resultado o módulo de buscas utilizará este termo para classificar os próximos resultados com base em um mecanismo de “ranking”. (+)
- *Navegação* – Partindo de algum conceito o usuário poderá navegar por todos os seus relacionamentos. Informações contextuais poderá ser usada para indicar ao usuário os melhores caminhos para navegação. (+)
- *Adicionar conceitos ou relacionamentos* – um usuário adiciona um novo conceito para uma ontologia entrando com uma palavra em um campo do formulário de adição ou simplesmente adicionando frases quando está marcando um documento qualquer que esteja redigindo ou pode aceitar sugestões automáticas do sistema. (++)
- *Comentar e discutir* – Usuários podem adicionar comentários a partes de uma ontologia. Usando esta funcionalidade os usuários podem colaborativamente esclarecer o significado de conceitos e relacionamentos para a modificação consensual de ontologias. (+)
- *Remover conceitos ou relacionamentos* – apagar conceitos e relacionamentos de uma ontologia. (++)
- *Mover subgrafos* – ao realizar esta operação, um usuário muda o termo pai de um conceito e muda os filhos também para outro contexto. (+++)

Para realizar modificações em uma ontologia o sistema tem de garantir que todos os dados estão consistentes após a aplicação das modificações. Para adicionar um conceito, por exemplo, tem de verificar se esse conceito já não existe na ontologia alvo.

Se isso acontecer o sistema deve executar procedimentos para resolver o conflito, por exemplo, automaticamente remover ambiguidades com ajuda de informações sensíveis ao contexto. A situação mais difícil surgirá quando um usuário mover um subgrafo da ontologia porque isto poderá afetar muitos outros conceitos.

5.3 Busca Semântica Corporativa

A Busca Semântica Corporativa deve investigar e descobrir informações em ambientes semânticos e principalmente em não semânticos, utilizando técnicas inovadoras de busca para facilitar a análise profunda das informações disponíveis, buscando relacionar os dados não semânticos com os semânticos, e fornecer aos usuários acesso personalizado às informações corporativas.

O modelo de arquitetura proposto depende fortemente das ontologias como meios para a representação formal do conhecimento, enquanto a maioria dos documentos não possui anotação semântica formal.

Como uma solução para a busca em dados não semânticos o mecanismo de indexação deverá utilizar algoritmos de análise de tendência decorrentes de fontes de informação não estruturada e combinar métodos estatísticos com abordagens de marcações realizadas manualmente por seres humanos para o reconhecimento da semântica.

Outro aspecto importante no domínio da pesquisa é a personalização. Empresas com grande escopo de produtos e alta diversidade de clientes buscam por soluções de pesquisas personalizadas.

5.3.1 Busca em dados não semânticos

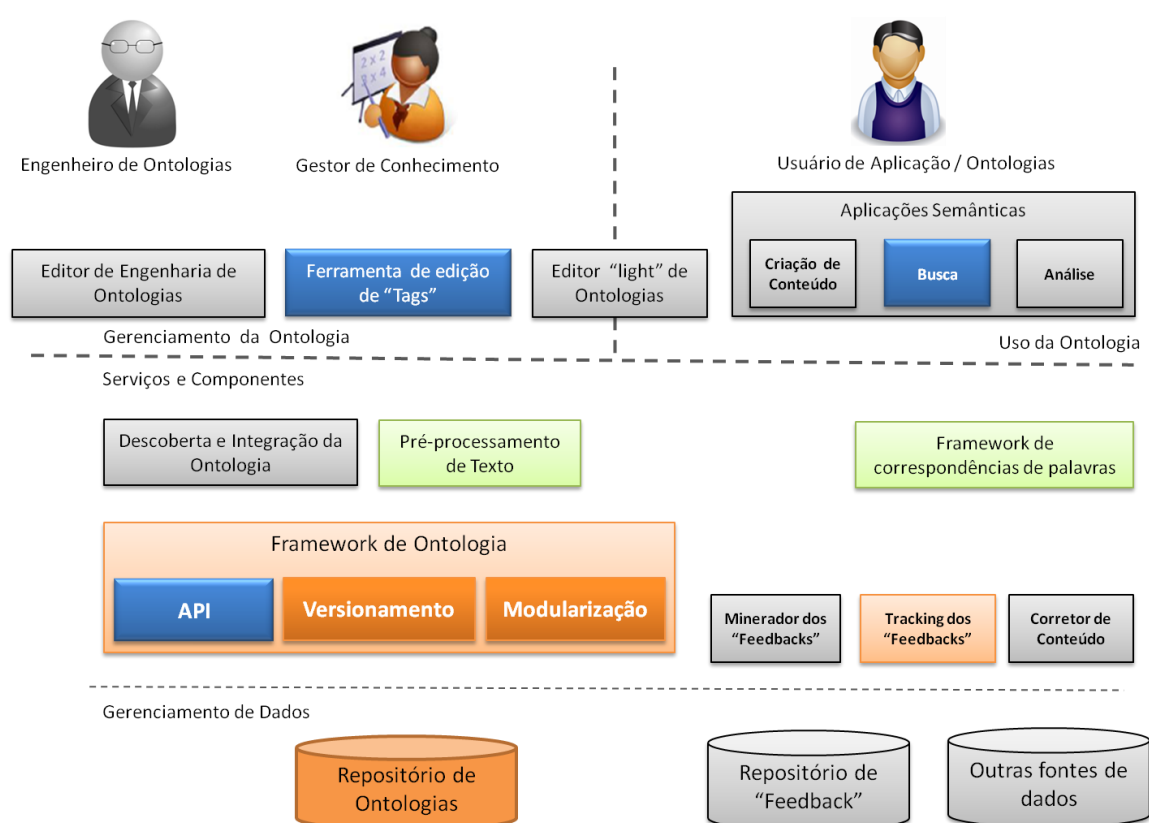


Figura 5.7 – Visão global do modelo proposto de arquitetura com foco na Busca Semântica

A Semântica no Ambiente Corporativo deve prezar pela Engenharia das Ontologias e pela Colaboração Semântica Corporativa concentrando seus esforços na preservação e no avanço de tais ontologias e no conhecimento que será agregado.

Portanto, este trabalho crucial de busca em dados não semânticos deve ser cumprido por mecanismos semiautomáticos que realizarão a pesquisa por conceitos e estabelecerão as relações complexas entre eles, de forma a incorporá-los na ontologia de domínio. Deste ponto de vista, estes mecanismos praticamente extrairão a “semântica” das coleções de textos corporativos.

No entanto, esta "semântica" atribuída aos documentos em coleções de textos corporativos não está explícita. Textos são apresentados em formato eletrônico, no melhor dos casos formatados em XML. Desta forma, um mecanismo de extração inteligente será necessário para realizar tal trabalho.

Outro aspecto que se fará necessário é a intervenção manual para a utilização de uma técnica de marcação, que já é a forma mais comum usada por muitos usuários, principalmente os de corporações, para anotar o conhecimento e a informação posterior que é a *marcação de ajuda* utilizada para melhorar a compreensão da informação que está sendo marcada. Combinando as duas coisas: pesquisa em coleções de texto e a marcação manual citada acima, podemos satisfazer a necessidade de extração de "semântica" a partir de textos e criar uma abordagem automática para a "compreensão" de um determinado texto dentro de uma coleção de texto.

5.3.2 Personalização da Busca Semântica Corporativa

No contexto da Web, define-se o acesso personalizado aos dados da Web como o processo de apoio ao utilizador individual na busca, seleção, acesso e recuperação de recursos da Web (ou subconjuntos desse processo).

Serviços ou aplicações adaptadas para o usuário exigem um perfil de usuário ou uma descrição de suas preferências, a fim de ser capaz de selecionar o conjunto de informações relevantes.

Além disso, sistemas personalizados requerem alguma forma de representação de seu domínio de aplicativo. Ontologias têm o potencial para cumprir esse papel, fornecendo dados formalizados que máquinas conseguirão ler e prover representações significantes para ambos, usuários e recursos do domínio.

Considerando os campos de pesquisa de engenharia de ontologia e Colaboração Semântica Corporativa, apresentado em seções anteriores, o foco na formalização do conhecimento de corporações na forma de ontologias, um módulo deverá concentrar-se em fornecer dados aos usuários dentro de seu contexto, tanto internos (funcionários) como externos (clientes, parceiros de negócios, etc), com acesso personalizado as informações.

5.3.3 Cenários de Aplicação

Um cenário de aplicação típico para a mineração de dados através da semântica é o de investimentos no mercado financeiro. Quanto mais informação de boa qualidade um investidor tem, melhor a sua base para tomada de decisões. Uma grande quantidade de informações pode ser encontrada em notícias relacionadas ao mercado e divulgadas por grandes jornais e pela internet, porém como todos os meios de notícias, estes sobrecarregam o investidor também com informações inúteis.

O processo de tomada de decisões torna-se então mais e mais difícil. Informação que é livre também é difícil de filtrar, portanto, o desafio passa a ser encontrar informação de boa qualidade.

Todos os analistas de mercado têm acesso às mesmas informações sobre o mercado e suas tendências, mas os melhores são aqueles capazes de melhorar a informação estatística, trazendo ao processo decisório sua experiência e capacidade de fazer previsões associadas a informações qualitativas.

Porém, estas análises demandam tempo e dinheiro no processo. A fim de apoiar os analistas humanos, novos métodos são necessários. Uma série de métodos e programas de computador tem sido desenvolvida, a fim de apoiar os tomadores de decisão em seus investimentos, mas a maioria deles é baseada apenas na análise de dados numéricos e geração de gráficos com valores de troca de ações. No entanto, os investidores ainda têm de filtrar, ler e analisar informações pertinentes encontradas nas notícias do mercado.

A falta ou perda de uma importante peça de informação pode trazer severas consequências a um investidor. A garantia de seu sucesso reside na sua habilidade de filtrar, entender e reagir a diversos relatórios baseados nas informações coletadas.

A Semântica passa então a ser uma solução extremamente promissora para a integração de conhecimento especializado com o fim de analisar e “entender” informações imprescindíveis em notícias do mercado realizando detecção de tendências de forma automática.



Figura 5.8 – Cenário real de análise realizada atualmente por investidores e o que se poderia obter com a Semântica.

6 CONSIDERAÇÕES FINAIS

A principal contribuição deste trabalho foi oferecer um modelo de arquitetura baseado em semântica para uso em ambientes corporativos ressaltando suas vantagens, benefícios e desafios sem a preocupação de entrar profundamente nos detalhes técnicos que a compõem.

Uma arquitetura baseado em semântica e aplicada ao ambiente corporativo pode dar um novo sopro de vida à utilização das informações nas empresas. Esta prática deve trazer melhorias, tais como, acesso eficiente, automatizado e personalizado a conteúdos semanticamente anotados; suporte a uma autoria mais flexível e automatizada desses conteúdos e ainda a integração de ambientes distintos de repositórios de conhecimento ligados à gestão do conhecimento.

A disponibilização de metadados semânticos (anotações semânticas) e ontologias, acomodados com tecnologias internas pré-existentes (BI, CMS) para a extração e mapeamento do conhecimento, cria um sistema de informação extremamente integrado e capaz de manipular dados de fontes heterogêneas em conjunto com todo o conhecimento corporativo de fundo e que, desta forma, poderá fazer inferências que jamais foram possíveis antes.

Diante de todo o contexto exposto, concluímos que ao se formalizar o modelo proposto – esperamos um impulso exponencial no acesso ao conhecimento dentro das empresas.

6.1 Limitações do Estudo

Como citado em outros pontos deste trabalho, a Semântica é uma área relativamente nova da computação e basicamente seus casos de uso limitam-se as pesquisas acadêmicas. Por outro lado estas mesmas pesquisas acadêmicas atuais, também enfrentam dificuldades e como qualquer área nova de estudo não possui ainda um foco ou horizonte definido. Isto faz com que os cientistas da área exponham cada qual o seu próprio modelo o que leva a diversas discussões dentro da própria comunidade.

Enquanto a maioria dos estudos é focada no ambiente Web e seus entraves, como por exemplo, a velocidade de comunicação das redes atuais, o ambiente corporativo começa a engatinhar no sentido de obter sua própria solução baseada nos conceitos primordiais da Web Semântica. Entretanto, as empresas também enfrentam seus próprios problemas tais como a dificuldade de se demonstrar o custo-benefício interno de tal investimento.

Por isso, esse trabalho limitou-se a documentação escassa e esparsa e por que não dizer, por vezes, incompleta, existente atualmente ao montar um modelo de arquitetura ao menos plausível do ponto de vista corporativo.

6.2 Trabalhos futuros

Deixamos claro que o modelo de arquitetura apresentado neste trabalho não esgota de forma alguma os estudos referentes à semântica para uso corporativo. Melhor seria dizer que estão apenas começando.

Portanto, são apenas indicativos de certa tendência para se fornecer suporte total as consultas e padrões de metadados/ontologias baseados em outra área ainda imatura, o da Web Semântica, que como já citamos, também ainda está em processo de pesquisas, estudos e desenvolvimento.

Por se tratar de um modelo bastante extenso e passível de muitas discussões, a semântica para uso corporativo deve passar, com certeza, por muitos refinamentos. A definição deste modelo, neste trabalho, serve apenas como um ponto inicial de estudos.

Desta forma, cabe-nos a tarefa de complementar estes estudos/pesquisas em detalhes e pormenores, desmembrando cada módulo do “mediador” e buscando criar consistência suficiente para alcançar os principais objetivos e levá-los, a contribuir também para o futuro da Web Semântica.

Bibliografia

ARANGO, G., PRIETO-DÍAZ, R., **Domain Analysis Concepts and Research Directions**, in Domain Analysis and Software Systems Modeling, IEEE Computer Society Press, 1991.

BERNERS-LEE T., LASSILA O., HERNDLER J., **The semantic web**, Scientific American, Disponível em <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>, 2001.

BICK, E., The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Disponível em <http://beta.visl.sdu.dk/visl/pt/parsing/automatic/>, 2000.

BONIFACIO, A. S., HEUSER, C. A., **Ontologias e Consultas Semânticas: uma aplicação ao caso Lattes**. Universidade Estadual de Londrina – Paraná, 2002.

BRACHMAN, J., **On the epistemological status of semantic networks**. In N. Findler, editor, Associative Networks: Representation and use of knowledge by computers. Academic Press, New York, 1979.

BREITMAN, K, **Web Semântica: a internet do futuro**. Rio de Janeiro: LTC, 2005.

BUFFA M., GANDON F, ERETEOB, G., SANDERA P., AND FARONA C. SweetWiki: **A semantic wiki**. KEWI Group, I3S Laboratory, University of Nice, France. Edelweiss Group, INRIA Sophia-Antipolis, France, 2008.

BUSSE, S. et al. **Federated Information Systems**: concepts, terminology and architectures. “Technical Report,” TU Berlin 99-9, 1999. Disponível em: <http://citeseer.ist.psu.edu/busse99federated.html>. Acesso em: 10 ago 2011

CROW, L., SHADBOLT, N., **IMPS – Internet Agents for Knowledge Engineering**. 1999. Disponível em: <http://eprints.ecs.soton.ac.uk/2322>. Acesso em: 15 abr 2011

DECKER, S., **A query and inference service for RDF**. In: The Query Languages Workshops, QL, 1998, Boston. Disponível em: <http://www.w3.org/TandS/QL/QL98/pp/quervservice.html>. Acesso em: 10 jun 2010.

DECKER, S., **The semantic Web: The roles of xml and rdf**. In: IEEE Internet Computing, Los Alamitos, v. 4, n. 5, p. 63-74, Oct., 2000.

FELLBAUM, C., **A semantic network of English verbs**. In: WordNet. An Electronic Lexical Database., pages S. 69–104. MITPRESS Cambridge Mass, London, 1998.

FORTE, M., SOUZA, W. L., PRADO, A. F., **Utilizando Ontologias e Serviços Web na Computação Ubíqua**. In: XX Simposio Brasileiro de Engenharia de Software, 2006.

FREITAS, C., ROCHA, P., BICK, E., **Floresta sintá (c) tica: Bigger, thicker and easier**. In PROPOR '08: Proceedings of the 8th international conference on Computational Processing of the Portuguese Language, pages 216–219, Berlin, Heidelberg, 2008.

GÓMEZ-PEREZ, A., **Tutorial on Ontological Engineering**. In: Internacional Joint Conference on Artificial Intelligence. 1999. Disponível em: <http://amutiara.staff.gunadarma.ac.id/Downloads/files/4679/tutorial-ontology.pdf>, Acesso em 18 abr 2011.

GRUBER, T. R., **Toward principles for the design of ontologies used for knowledge sharing**. Intenational Journal of Human an Computer Studies, v. 43, p.907-928, 1993. Disponível em: <http://tomgruber.org/writing/onto-design.htm>. Acesso em: 20 abr 2011.

CROW, L., SHADBOLT, N., **Extracting focused knowledge from the semantic web**. Int. J. Hum.-Comput. Stud 155~184 Cited By 49, 2001.

HORROCKS, I.; TESSARIS, S., **Querying the Semantic Web: a Formal Approach**. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.14.7901&rep=rep1&type=pdf>. Acesso em: 20 mai 2011.

HUHNS, M. N., SINGH, M. P., **Agents and Multiagent Systems: Themes, Approaches, and Challenges**. In: Huhns, Michael, Singh, Munindar, Eds. 'Readings in Agents', Morgan Kaufman Publishers, 1997.

IEEE STD 610.5-1990, **IEEE Standard Glossary of Data Management Terminology**.

JARRAR, M., VERVENNE, L., MAYNARD, D., HR, **Semantics Roadmap-The Semantic challenges and opportunities in the Human Resources domain**. Technical Report, Bélgica, 2007.

MANNING, C.D., RAGHAVAN, P., SCHUTZE, H. **Introduction to Information Retrieval** in Cambridge University Press, 2008.

MOURA, A. M. de C., **A Web Semântica: fundamentos e tecnologias**. 2004. IME. Disponível em: <http://ipanema.ime.eb.br/~anamoura/publicacoes.html>. Acesso em: 09 Jul 2011.

ORLEAN, D., LUCENA, C., **Um Processo Unificado para Engenharia de Ontologias**. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio de Janeiro, 2003.

SABOU, R. M., **Building Web Services Ontologies**. 187 f. Dissertação (Mestrado) - Departamento de Siks, Vrije Universiteit, Turda, 2006.

SOUSA, L. V., RIOS, C. L., ARAÚJO, A. E. RIOS, R., **Abordagem Semântica para uma Arquitetura Integrada de Suporte à Gestão do Conhecimento**. 2006.

SOUZA, R. R e ALVARENGA, L. **A web semântica e suas contribuições para a ciência da informação**, v.33, n. 1, p.132-141, jan./abr.2004. Disponível em: <http://www.ibict.br/cienciadainformacao/viewarticle.php?id=71> Acesso em: 09 Mar 2011.

URIARTE, R. **Ferramentas de Busca com base Semântica**. 2010.

USCHOLD, M. **Building Ontologies: Towards a Unified Methodology**. Disponível em: <http://citeseer.ist.psu.edu/uschold96building.html>. Acesso em: 09 Mar 2011

ZACK, M. H. **Developing a Knowledge Strategy**. California Management Review, vol. 41, v.3, pg. 125–145, 1999.