

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

MARÍLIA TERRA DE MELLO

**Integração de Dados e Processos Disponíveis
em Portais Web Baseada em Ontologias de
Domínio**

Dissertação apresentada como requisito parcial
para a obtenção do grau de Mestre em Ciência
da Computação

Prof^a. Dr^a. Mara Abel
Orientadora

Porto Alegre, maio de 2008.

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Mello, Marília Terra de

Integração de Dados e Processos Disponíveis em Portais Web Baseada em Ontologias de Domínio / Marília Terra de Mello – Porto Alegre: Programa de Pós-Graduação em Computação, 2008.

113 f.:il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação. Porto Alegre, BR – RS, 2008. Orientadora: Mara Abel.

1.Integração de Informação. 2.Ontologias de Domínio 3.Serviços Web Semânticos. 4. Biomedicina. 5. Geologia. I. Abel, Mara. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. José Carlos Ferraz Hennemann

Vice-Reitor: Prof. Pedro Cezar Dutra Fonseca

Pró-Reitora Adjunta de Pós-Graduação: Profa. Valquiria Linck Bassani

Diretor do Instituto de Informática: Prof. Flávio Rech Wagner

Coordenador do PPGC: Profa. Luciana Porcher Nedel

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

*O rio atinge seus objetivos,
porque aprendeu a contornar obstáculos.*

Agradeço primeiramente ao Instituto de Informática da UFRGS e ao Programa de Pós-Graduação em Computação por proporcionar a infra-estrutura necessária à pesquisa. Aos funcionários por serem sempre atenciosos e prestativos, e aos professores pelo incentivo à busca pelo conhecimento. Agradeço também a Universidad de Murcia pela oportunidade de intercâmbio e por todo o suporte oferecido durante a minha estadia na Espanha.

Agradeço todo o apoio e amizade oferecidos pelos colegas do grupo de pesquisa de Banco de Dados Inteligentes: Carlos Santin, Sandro Rama Fiorini, William Gonçalves, Henrique Monteiro e Rogério Dutra Soares. Tanto as conversas descontraídas em horários de almoço e churrascos, como as sérias discussões realizadas nas reuniões de grupo, contribuíram significativamente em todo o andamento do trabalho. Aos colaboradores Laura Mastella, Eduardo Castro, Felipe Victoreti e Luiz Hermes Svoboda pelas sugestões pertinentes ao trabalho desenvolvido no mestrado. Obrigada por estarem sempre dispostos a ajudar!

Aos membros do grupo de Tecnologías de Modelado, Procesamiento y Gestión Del Conocimiento por me receberem em seu grupo e por toda a ajuda no período de intercâmbio. Ao professor Rodrigo Martínez Bejar pela oportunidade oferecida e pelas sugestões de pesquisa. A Francisco Gracia-Sánchez por todo conhecimento compartilhado, pelas idéias sugeridas e por ter desempenhado o papel de meu tutor durante minha pesquisa em Murcia.

Agradeço especialmente a professora Mara Abel, minha orientadora, por todo o estímulo a este trabalho. Agradeço as oportunidades oferecidas, o incentivo à pesquisa e à minha carreira, as sugestões valiosas e a confiança em mim depositada. Muito obrigada por desempenhar não apenas o papel de orientadora, mas também de amiga, conselheira, além de um exemplo de profissional a ser seguido.

Aos meus amigos sem os quais todo e qualquer trabalho seria bem mais difícil. Gostaria de agradecer a Bruna Bettega por me acolher como irmã em Porto Alegre. Obrigada por estar presente sempre que precisei, por compreender minhas ausências e me acompanhar tanto nos momentos difíceis como nos momentos de descontração. Aos colegas de mestrado, em especial aos amigos: Rafael Borges e Carla Lopes, pelos grupos de estudo, pelo incentivo, pelos conselhos valiosos e pelas pausas para o mate

nos intervalos das aulas. Aos ex-colegas da UFPel agradeço pelos encontros que ainda realizamos e por todas as conversas repletas de conselhos e incentivos. Agradeço também aos grandes amigos que ficaram em Pelotas, na torcida pela minha realização aqui em Porto Alegre.

A toda minha família por acreditar sempre no meu potencial. Agradeço principalmente aos meus pais Celso e Aglaé por estarem presentes em todo o meu caminho, apoiando incondicionalmente todas as minhas decisões. Obrigada por todo o amor, a confiança, os conselhos, o suporte e o incentivo. Obrigada por estarem sempre interessados em meus projetos, me ajudando a tomar as decisões mais difíceis de minha vida. Tudo isso sem contar a ajuda financeira sem a qual eu não estaria em Porto Alegre terminando este curso. Meus pais são os principais exemplos que eu procuro seguir na vida.

Agradeço com carinho a Ricardo Rochedo por estar sempre ao meu lado, desempenhando um papel fundamental em minha vida. Obrigada pela confiança, pelo apoio constante, pela motivação e pela compreensão com as minhas ausências devido aos momentos de estudo, mesmo quando a ausência se estendeu por seis meses.

Por fim, agradeço a Deus, pela força de seguir em frente e pelas oportunidades oferecidas nos momentos certos. Muito obrigada!

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	2
LISTA DE FIGURAS	4
LISTA DE TABELAS	6
RESUMO	8
ABSTRACT	9
1 INTRODUÇÃO	15
1.1 Objetivo.....	17
1.2 Abordagem de Pesquisa.....	18
1.3 Organização dos Capítulos.....	19
2 ABORDAGENS ATUAIS PARA INTEGRAÇÃO DE DADOS E PROCESSOS NA WEB	21
2.1 Ontologias de Domínio.....	31
2.1.1 Linguagens Ontológicas.....	32
2.2 Serviços Web Semânticos – Uma Revisão.....	33
2.2.1 Serviços Web.....	33
2.2.2 Web Semântica.....	35
2.2.3 Serviços Web Semânticos.....	36
2.2.4 A Abordagem OWL-S.....	38
3 O SISTEMA DE INTEGRAÇÃO W-KIS	41
3.1 Domínios de Aplicação.....	43
3.1.1 Biomedicina.....	43
3.1.2 Geologia.....	44
3.2 Processo de Desenvolvimento.....	45
4 DESCRIÇÃO DA APLICAÇÃO	49
4.1 Modelos de Domínio.....	49
4.2 Fontes de Informação.....	51
4.3 Descrições Semânticas.....	54
4.4 Arquitetura do sistema W-KIS.....	57
4.4.1 Interface de Usuário.....	58
4.4.2 Componente de Consulta.....	59
4.4.3 Componente de Execução.....	60

4.4.4	Componente <i>Core</i>	60
5	AVALIAÇÃO DO SISTEMA W-KIS.....	63
5.1	Estudos de Caso	63
5.1.1	Domínio da Biomedicina.....	63
5.1.2	Domínio da Geologia	65
5.1.3	Avaliação nos Dois Domínios	67
5.2	Comparação com Outras Abordagens	68
5.3	Manutenção e Evolução	77
6	CONCLUSÃO.....	79
6.1	Sugestões para Trabalhos Futuros.....	80
	REFERÊNCIAS.....	83
	ANEXO A ONTOLOGIA <i>ONCOGEN</i>.....	87
	ANEXO B ONTOLOGIA <i>PETROGRAPHER</i>.....	93
	APÊNDICE A DESCRIÇÃO WSDL DO SERVIÇO <i>MINDAT</i>.....	105
	APÊNDICE B DESCRIÇÃO OWL-S DA OPERAÇÃO <i>GETMINERAL</i> DO SERVIÇO <i>MINDAT</i>	107
	APÊNDICE C EXEMPLOS DE CONSULTAS REALIZADAS COM O SISTEMA W-KIS.....	111

LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
HTTP	Hypertext Transfer Protocol
IRS-II	Internet Reasoning Service
OWL	Ontology Web Language
OWL-S	Ontology Web Language for Services
RDF	Resource Description Framework
SAWSDL	Semantic Annotations for WSDL and XML Schema
SOAP	Simple Object Access Protocol
SW	Serviço Web
SWS	Serviço Web Semântico
UDDI	Universal Description, Discovery and Integration
UPML	Unified Problem Solving Method Development Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
WSDL	Web Service Description Language
WSMF	Web Service Modeling Framework
WSMO	Web Service Modeling Ontology
W3C	World Wide Web Consortium
XML	Extensible Markup Language

LISTA DE FIGURAS

Figura 2.1: Interação de Serviço Web	34
Figura 2.2: A ontologia de serviço OWL-S.....	38
Figura 3.1: Abordagem de integração	42
Figura 3.2: Processo de desenvolvimento da aplicação	46
Figura 4.1: Principais conceitos da ontologia <i>Oncogen</i>	50
Figura 4.2: Principais conceitos da ontologia <i>PetroGrapher</i>	50
Figura 4.3: Portal Web <i>Mindat</i>	53
Figura 4.4: Parte do arquivo WSDL do serviço <i>Mindat</i>	54
Figura 4.5: Operações oferecidas pelo serviço <i>Geologynet</i>	54
Figura 4.6: Anotação semântica através do <i>OWL-S Editor</i>	55
Figura 4.7: Mapeamento do parâmetro de entrada	56
Figura 4.8: Parte do arquivo OWL-S do serviço <i>Mindat</i>	56
Figura 4.9: Parte do arquivo OWL-S demonstrando a operação <i>getMineral</i> e seus parâmetros.....	57
Figura 4.10: Arquitetura do sistema W-KIS.....	58
Figura 4.11: Modelo de processo do sistema W-KIS	58
Figura 4.12: Interface para o domínio da Biomedicina.....	59
Figura 4.13: Interface para o domínio da Geologia.....	59
Figura 5.1: Exemplo de consulta	64
Figura 5.2: Correspondência do termo de entrada com o tipo definido como “Gene” ..	65
Figura 5.3: Resultado da busca pelo termo “A4”	65
Figura 5.4: Exemplo de consulta	66
Figura 5.5: Correspondência do parâmetro de entrada.....	67
Figura 5.6: Resultado da busca pelo termo “matrix”.....	67
Figura 5.7: Primeiros resultados da busca pelo termo “A4” utilizando <i>Google</i>	70
Figura 5.8: Primeiros resultados da busca pelo termo “matrix” utilizando <i>Google</i>	71
Figura 5.9: Primeiros resultados da busca por “A4+gene” utilizando <i>Google</i>	72
Figura 5.10: Primeiros resultados da busca por “matrix+constituent” utilizando <i>Google</i>	72
Figura 5.11: Primeiros resultados da busca pelo termo “A4” utilizando <i>MetaCrawler</i>	74
Figura 5.12: Primeiros resultados da busca pelo termo “matrix” utilizando <i>MetaCrawler</i>	75
Figura 5.13: Primeiros resultados da busca pelo termo “A4” utilizando <i>About.com</i>	76
Figura 5.14: Primeiros resultados da busca pelo termo “matrix” utilizando <i>About.com</i>	77

LISTA DE TABELAS

Tabela 2.1: Comparação entre as técnicas utilizadas para integração na Web.....	26
Tabela 2.2: Comparação entre abordagens para integração na Web.....	27
Tabela 5.1: Tabela contendo os serviços no domínio da Biomedicina	64
Tabela 5.2: Tabela contendo os serviços no domínio da Geologia	66
Tabela 5.3: Avaliação das consultas realizadas com motores de busca	76

RESUMO

A integração e a troca de informação na Web é uma necessidade reconhecida no mundo atual nos mais diversos domínios de aplicação. A Web constitui hoje o principal meio de acesso a informações, disponibilizando várias formas de acesso a esse conteúdo. No entanto, disponibilizar essa informação de forma integrada para o usuário ainda é um desafio. O Serviço Web Semântico é uma tecnologia de Inteligência Artificial promissora a qual pode contribuir significativamente com o processo de integração tanto de dados como de processos disponíveis em portais Web.

O objetivo do presente trabalho é realizar a integração das diferentes funcionalidades oferecidas por portais disponibilizados na Web, fazendo uso de ontologias de domínio através de Serviços Web Semânticos. Para alcançar o objetivo proposto, um sistema de busca e integração, denominado W-KIS, foi projetado e desenvolvido, fornecendo uma visão unificada e transparente ao usuário de toda informação disponível em um determinado número de portais Web, contextualizados a um domínio de aplicação específico.

Dois domínios de aplicação foram explorados neste trabalho através de dois estudos de caso: a Biomedicina e a Geologia, cada um fornecendo sua própria ontologia de domínio. Serviços Web Semânticos foram desenvolvidos para encapsular portais Web em ambos os domínios, permitindo ao sistema realizar a integração em cada domínio, separadamente.

A aplicação proposta foi validada através da avaliação de sua atuação em ambos os domínios, além de testes com usuários finais e comparações com outras abordagens para integração na Web e motores de busca tradicionais. Através da utilização de Serviços Web Semânticos, W-KIS é capaz de localizar e executar os serviços de forma automática, além de lidar com a heterogeneidade semântica entre as fontes de informação por meio das ontologias de domínio. Desta forma, o sistema fornece apenas informação contextualizada ao interesse do usuário.

Palavras-Chave: Integração de Informação, Ontologias de Domínio, Serviços Web Semânticos, Biomedicina, Geologia.

Web Portal Data and Processes Integration Based on Domain Ontologies

ABSTRACT

Integration and exchange of information across the Internet is a universally recognized need, in a wide variety of domains. Nowadays, the Web is the main source of information, providing many ways to access this content. However, making this information available to the user in an integrated way is still a challenge. New promising application fields such as Semantic Web Services can improve the integration process of data and processes available in Web portals.

The goal of the present research is to accomplish the integration of the different functionalities offered by portals available in the Web, making use of domain ontologies through Semantic Web Services. In order to achieve the proposed goal, a search and integration system, named W-KIS, was designed and developed. The system provides the user with a unified and transparent view of the whole information available in a number of Web portals, related to a specific application domain.

Two application domains were explored in this work through two case studies: the Biomedicine and the Geology, each one providing its own domain ontology. Semantic Web Services were designed to encapsulate Web portals in both domains, allowing the system to accomplish the integration in each domain, separately.

The application proposed in this project was validated through the evaluation of its behavior in both domains, besides other tests with final users and comparisons with other approaches for integration in the Web, and traditional search engines. Making use of Semantic Web Services, W-KIS is able to find and execute the services in an automatic way, besides dealing with the semantic heterogeneity in the information sources by means of domain ontologies. This way, the system provides the user with information contextualized to its interest.

Keywords: Information Integration, Domain Ontologies, Semantic Web Services, Biomedicine, Geology.

1 INTRODUÇÃO

Atualmente, é possível encontrar uma vasta quantidade de fontes de informação independentes e heterogêneas, disponíveis através da Internet, contendo informações relacionadas aos mais variados domínios de aplicação. Entretanto, disponibilizar essa informação de forma integrada ainda é um desafio. O rápido aumento na quantidade de informação que é fornecida na Web reflete diretamente nas ferramentas de manipulação deste conteúdo, exigindo ferramentas mais robustas que permitam o acesso à informação de forma fácil e ao mesmo tempo eficiente. A integração da informação torna-se extremamente importante para esta tarefa. A comunidade da Inteligência Artificial dedica correntemente grandes esforços na exploração de abordagens para melhorar o processo de descoberta, integração e acesso à informação na Web, tendo como principal motivação a crescente demanda exigida por seus usuários.

As ferramentas disponibilizadas hoje na Web são voltadas para o usuário comum que realiza buscas por informações genéricas. Estas ferramentas não levam em conta quem é o usuário, o que ele está buscando e por que ele está realizando uma determinada busca; ou seja, elas ignoram o contexto no qual a informação é utilizada. Entretanto, atualmente, inúmeras pessoas realizam buscas rotineiras na Web como suporte a suas tarefas e estas tarefas definem o contexto de suas consultas. Existe, portanto, a necessidade de ferramentas capazes de realizar buscas contextualizadas, isto é, ferramentas que utilizem conhecimento de domínio para dirigir a busca do usuário. Quanto mais conhecimento a ferramenta de busca possui sobre o contexto da consulta, melhor é sua capacidade de fornecer a informação correta para o usuário apropriado no momento certo.

A identificação e integração da informação disponível na Web é um processo essencial para atender consultas de usuários que buscam informação dentro do seu contexto de trabalho. Porém, a integração não é um processo simples. Suas principais dificuldades são: (1) a heterogeneidade na apresentação das informações, (2) a heterogeneidade de contexto e (3) a heterogeneidade nas formas de acesso às fontes, onde estão disponíveis não apenas dados estáticos, mas também serviços através de portais Web.

O problema da heterogeneidade na apresentação das informações engloba as heterogeneidades sintática e estrutural. A heterogeneidade sintática refere-se às diferenças na representação do formato dos dados. Os portais Web representam o seu conteúdo de formas bem diferentes. Alguns portais, por exemplo, apresentam a resposta para a consulta do usuário utilizando texto simples, enquanto outros utilizam tipos de dados estruturados. Além disso, os portais armazenam seu conteúdo em diferentes níveis de detalhamento, seguindo diferentes paradigmas de modelagem de dados, o que caracteriza a heterogeneidade estrutural. Alguns portais armazenam sua informação em

forma de glossários *online*, enquanto outros utilizam uma ou mais bases de dados para armazenar as informações.

A heterogeneidade de contexto está relacionada às diferentes áreas de conhecimento abordadas nas fontes de informação e aos diferentes contextos nos quais os usuários realizam suas consultas. Enquanto os portais Web disponibilizam informação sobre os mais variados domínios, o usuário realiza cada busca dentro de um contexto do seu interesse. Para lidar com esta heterogeneidade, a ferramenta deve integrar apenas as fontes com informações sobre o domínio de interesse do usuário, fornecendo, como resultado da consulta, apenas respostas contextualizadas a este domínio.

Este problema, que está relacionado ao problema da heterogeneidade semântica, representa hoje uma grande dificuldade para o processo de integração de informações na Web. Termos diferentes presentes em diversos portais Web podem ter o mesmo significado, isto é, podem denotar um mesmo conceito, ao mesmo tempo em que um mesmo termo pode denotar conceitos diferentes em fontes de informação distintas. Devido a esta dificuldade, somente extrair dados de páginas Web normalmente não é suficiente para dar suporte à integração. Um sistema de integração deve ser capaz de aprender ou ter acesso às descrições semânticas das fontes, para então realizar um mapeamento através de um modelo de conhecimento de domínio. Em um exemplo simples, se o problema de um usuário está ligado ao automobilismo, uma consulta buscando pelo termo “jaguar” deve retornar os automóveis deste tipo. Por outro lado, se o contexto é a zoologia, a mesma consulta deve retornar descrições de animais.

Já a heterogeneidade nas formas de acesso às fontes refere-se ao fato de que os portais Web não incluem apenas dados, mas também serviços, tais como motores de busca, disponibilizando a informação de diversas fontes distribuídas de uma maneira unificada. Por estarem focadas apenas na integração de dados estáticos, as tecnologias existentes atualmente não permitem que o usuário também tenha acesso à informação fornecida pelos serviços disponíveis nos portais Web. Porém, o resultado obtido com a integração apenas de dados estáticos não é mais suficiente para suprir as necessidades do usuário da Web atual. Além disso, acessar um único serviço também não é suficiente, visto que não existe um serviço que sozinho seja capaz de cobrir todo o conteúdo referente a uma determinada área de conhecimento. Como hoje existe um grande número de portais oferecendo serviços na Web, disponibilizar um acesso integrado a estes serviços é a melhor forma de prover informação contextualizada ao usuário, sem a necessidade de armazenamento de grandes bases de dados.

Os Serviços Web podem ser vistos como uma maneira de expor a funcionalidade de um sistema de informação e torná-la disponível através de tecnologias padronizadas para a Web (ALONSO et al., 2004). A vantagem da tecnologia de Serviços Web está no fato de que ela estabelece uma plataforma comum para a integração de aplicações computacionais distribuídas, tanto em intranets como na própria Internet. Já o Serviço Web Semântico é um novo paradigma de pesquisa que pode ser definido como o aumento das descrições dos Serviços Web através de anotações semânticas. Através do uso de Serviços Web Semânticos é possível automatizar o processo de descoberta e execução dos serviços fontes, automatizando, portanto, o processo de integração. Além disso, eles fazem uso de ontologias de domínio com as quais é possível lidar com o problema da heterogeneidade das fontes de informação.

As ontologias de domínio fornecem um modelo comum dos conceitos que são importantes e seus relacionamentos permitidos, dentro de um domínio específico. A

utilização destas ontologias facilita o compartilhamento de informações em geral, dentro de um mesmo contexto, e o processo de integração de bases de dados. Através desta facilidade, os usuários formulam consultas sobre a ontologia e o sistema tem a responsabilidade de gerenciar a heterogeneidade e a distribuição nas diferentes bases.

No presente trabalho, um sistema de busca e integração, denominado W-KIS (*Web Knowledge Integration System*), inicialmente proposto pela autora deste trabalho em (DE MELLO et al., 2007) foi projetado e desenvolvido, utilizando a tecnologia de Serviços Web Semânticos. Este sistema integra dados e processos oferecidos através de diferentes portais Web conectados ao sistema, com o objetivo de atender a consultas contextualizadas em domínios específicos.

1.1 Objetivo

O principal objetivo do presente trabalho é realizar a integração das diferentes funcionalidades oferecidas por diversos portais Web para o atendimento a consultas contextualizadas em domínios de aplicação específicos. Para validar o objetivo proposto, um sistema de conhecimento foi desenvolvido e aplicado em dois domínios diferentes, utilizando ontologias previamente definidas. Um sistema de conhecimento tem como objetivo reduzir o nível de perícia do usuário necessário para utilizar o sistema. A intenção é que o usuário não necessite ter conhecimento das fontes de informação assim como da estrutura semântica do conteúdo a ser buscado. Este conhecimento será dominado e disponibilizado pelo sistema, reduzindo a necessidade de treinamento do usuário com guias ou manuais.

Os domínios da Biomedicina e da Geologia foram escolhidos por serem áreas de pesquisa de conhecimento intensivo, nas quais os pesquisadores necessitam ter acesso a toda informação disponível de uma maneira unificada e automatizada, além de disporem de ontologias de domínio formalizadas e disponíveis para acesso.

A Web dispõe hoje de um grande número de fontes de informação tanto no domínio da Biomedicina como no da Geologia. Portanto, torna-se fundamental a interligação entre os dados obtidos pelos diversos projetos de pesquisa ao redor do mundo sobre os tópicos inerentes a cada domínio de aplicação. Além disso, as fontes disponíveis na Web são, em sua maioria, otimizadas para serem utilizadas somente por seres humanos, justificando a necessidade de ferramentas que possibilitem o acesso automatizado a esses recursos.

Para alcançar o objetivo proposto, ontologias de domínio são utilizadas através de Serviços Web Semânticos. Uma única ontologia global é utilizada em cada domínio de aplicação, com o objetivo de fornecer o vocabulário compartilhado para a especificação da semântica de todas as fontes conectadas ao sistema. A mesma ontologia serve de base para o usuário fazer sua consulta onde o termo a ser buscado deve ser associado a um conceito do domínio de aplicação, definido na ontologia. Os Serviços Web Semânticos são utilizados para encapsular as fontes de informação e, desta forma, prover a automatização do processo de integração por completo.

Dois estudos de caso foram realizados no processo de validação do sistema proposto – um em cada domínio de aplicação. O primeiro foi feito na Universidad de Murcia - Laboratório KLT, na Espanha, sobre o domínio da Oncologia para o diagnóstico e tratamento do câncer de mama. O segundo foi realizado na Universidade Federal do Rio Grande do Sul - Grupo de Pesquisa de Bancos de Dados Inteligentes, no Brasil, sobre o

domínio da Petrografia Sedimentar para qualificar reservatórios de petróleo e avaliar seu potencial produtor, através de descrições de rochas sedimentares.

Como resultado do trabalho foi obtido um sistema capaz de atuar em domínios de aplicação específicos, baseado em Serviços Web Semânticos e em ontologias de domínio, que encapsule os serviços oferecidos pelos portais Web vinculados aos sistemas, permitindo um acesso integrado a essa informação, de forma automatizada. A integração permite otimizar o processo de consulta e busca em cada domínio específico. O problema de fontes de informação heterogêneas que utilizam representações diferentes é resolvido através do mapeamento de cada esquema de dados fonte para a ontologia de domínio, permitindo que o usuário tenha uma visão unificada e transparente de toda informação fornecida por um determinado número de portais, independente da sua fonte real.

O objetivo de realizar estudos em dois domínios distintos é identificar de que forma ontologias de domínio facilitam a integração de bases de dados e formalizar esse método de forma independente de domínio. Pretende-se disponibilizar o sistema de forma que seus componentes centrais possam ser reutilizados em diferentes áreas de aplicação e, dessa forma, o sistema não seja restrito a apenas um domínio.

1.2 Abordagem de Pesquisa

O objetivo principal desse trabalho é propor uma ferramenta que possibilite realizar a integração tanto de dados como de processos disponibilizados por diferentes portais Web, atendendo a buscas realizadas em domínios de aplicação distintos, separadamente. Esse trabalho foi desenvolvido através das seguintes etapas:

- **Estudo das principais abordagens para integração de informação na Web.** As principais técnicas adotadas atualmente para realizar a integração de dados e processos na Web foram pesquisadas, analisando as vantagens e desvantagens de se utilizar uma técnica em detrimento de outra. Com base na análise realizada, foi possível destacar as deficiências presentes nas soluções propostas e, principalmente, as características fundamentais que uma solução para integração de informação na Web deve fornecer.
- **Estudo do atual estado da arte referente aos principais conceitos relacionados aos Serviços Web Semânticos.** Este levantamento bibliográfico serviu de base para a proposta de solução do presente trabalho.
- **Seleção dos domínios de aplicação a serem utilizados.** Dois domínios a serem utilizados para validar a solução proposta foram escolhidos com base nas áreas de atuação nas quais os projetos de pesquisa envolvidos neste trabalho estão inseridos.
- **Proposta de uma arquitetura para o sistema de integração.** Com base em todos os estudos previamente realizados, foi idealizada uma arquitetura para um sistema de integração, visando a integração de dados e processos disponíveis em portais Web.
- **Implementação do modelo proposto.** Cada componente da arquitetura proposta foi então implementado e, posteriormente, todos os componentes foram interligados, constituindo o sistema de integração.

- **Realização de estudos de caso.** Um estudo de caso foi realizado em cada domínio de aplicação para realizar a validação do sistema. A avaliação do sistema de integração foi feita com base nos resultados obtidos em casa estudo de caso e na comparação com outras abordagens de integração existentes.

1.3 Organização dos Capítulos

No capítulo 2 são apresentadas as principais abordagens existentes na atualidade para a integração de dados e processos da Web. Com base na análise realizada comparando as abordagens, foram ressaltadas as deficiências presentes nas soluções propostas. Por fim, são apresentadas as características fundamentais que uma solução para integração de informação na Web deve fornecer. O capítulo 2 ainda apresenta uma revisão sobre o estado da arte do conteúdo que serviu de embasamento para a solução proposta neste trabalho: Serviços Web, Web Semântica, Serviços Web Semânticos, ontologias de domínio, linguagens ontológicas, além da abordagem OWL-S.

A visão geral da abordagem do presente trabalho é dada no capítulo 3, onde o sistema de integração é apresentado e os domínios de aplicação são descritos. Já a descrição completa da aplicação desenvolvida encontra-se no capítulo 4. Este capítulo apresenta a arquitetura do sistema de integração juntamente com a descrição detalhada de todos os seus componentes, dos modelos de domínio e de cada serviço, bem como a descrição do processo de anotação semântica de cada serviço.

O capítulo 5 apresenta a avaliação do sistema desenvolvido, incluindo a descrição completa de dois estudos de caso, cada estudo em um domínio de aplicação específico; a comparação entre os dois domínios; e a comparação do sistema proposto com outras abordagens de integração. A conclusão do presente trabalho, juntamente com as sugestões para trabalhos futuros, são encontradas no capítulo 6.

2 ABORDAGENS ATUAIS PARA INTEGRAÇÃO DE DADOS E PROCESSOS NA WEB

A integração efetiva de bases de dados e fontes de informações heterogêneas tem sido o foco de um grande número de pesquisas nos últimos anos. Os sistemas de integração de dados têm como objetivo fornecer aos usuários um acesso uniforme às diversas fontes de dados, as quais provavelmente são autônomas, heterogêneas e distribuídas. Ou seja, seu objetivo principal é evitar que o usuário tenha que localizar as fontes de dados, interagir com cada uma isoladamente e combinar manualmente os dados provenientes dessas múltiplas fontes (LEVY, 2000).

Existem na literatura diversos trabalhos aplicando diferentes técnicas para realizar a integração de fontes de informação. Porém, o grande número de soluções propostas não significa que o problema da integração já tenha sido resolvido por completo. Ainda não existe uma solução geral adequada que se ajuste aos diversos problemas de integração.

Os primeiros sistemas de integração foram propostos para integrar dados de diferentes sistemas dentro de uma mesma organização. Em seguida, surgiram as propostas para integrar dados provenientes de organizações distintas. Mais tarde, com o advento da Web, surgiu a necessidade de integrar informação (não apenas dados, mas também processos) disponibilizada na Internet tanto por organizações como por indivíduos.

A Web emergiu como uma tecnologia para compartilhamento de informação através da Internet. Entretanto, ela rapidamente tornou-se o meio de conexão entre clientes remotos e aplicações na Internet e, mais recentemente (com o advento dos Serviços Web), um meio para integrar aplicações na Internet. Pode-se afirmar que a Web é um ambiente dinâmico que cresce rapidamente e está continuamente modificando-se. A tarefa de um sistema de integração de informação na Web é responder consultas que podem exigir a extração e combinação de dados e/ou processos de múltiplos portais Web.

Na Web, estão disponíveis diversos portais os quais oferecem uma grande variedade de recursos e serviços, como por exemplo: e-mails, fóruns, mecanismos de busca, etc. A maioria dos motores de busca transformou-se em portais Web com o objetivo de atrair e manter uma audiência maior. Entretanto, é muito difícil encontrar algum motor de busca que sozinho seja capaz de cobrir todo o conteúdo referente a um domínio de aplicação específico. Além disso, um dos pontos ressaltados como negativo da Web atual é a dificuldade enfrentada pelos usuários na tentativa de encontrar a informação correta.

Atualmente, bases de dados têm proliferado com a disponibilização de informação em rede e, principalmente, na Web. A cada ano que passa, tanto a oferta de novas informações quanto a demanda pela informação já disponível crescem

consideravelmente. A recuperação deste conteúdo, no entanto, não é realizada de forma satisfatória, devido à falta de ferramentas de acesso adequadas, capazes de realizar o tratamento semântico da informação contida nas bases de dados, viabilizando a integração de informações em domínios de conhecimento. A interpretação semântica de conteúdo multimídia é ainda considerada um desafio e pouco progresso tem sido observado nas propostas de solução para este problema.

A grande disponibilidade de informação na Web trouxe à tona a necessidade de prover o usuário com um meio de descobrir, acessar e obter uma visão integrada de tais informações. A integração de informações na Web é um processo bem diferente da integração de informações em bancos de dados. Além do problema da heterogeneidade dos dados também enfrentado na integração de bancos de dados, a integração na Web sofre também as seguintes dificuldades (HSU et al., 2003; IVES et al., 1999 e LEVY, 2000): a mudança dinâmica de conteúdo; a necessidade de automatização; a dificuldade de adaptação ao domínio de aplicação; e a necessidade de incorporar no sistema de integração tanto os dados como os serviços disponíveis na Web.

A heterogeneidade das fontes de informação pode acontecer em nível sintático, estrutural ou semântico (BERGAMASCHI et al., 1998; CRUZ e XIAO, 2005). A heterogeneidade sintática refere-se às diferenças de representação dos tipos de dados. A heterogeneidade estrutural é causada pelas diferenças nos modelos e linguagens utilizados na modelagem das fontes de informação. E a heterogeneidade semântica é a diferença nos significados ou interpretações dos dados em diversos contextos. A integração semântica dos dados, a qual deve ser provida nos sistemas de integração, é definida por Cruz e Xiao (2005) como o uso de uma representação conceitual de tais dados e seus relacionamentos com o objetivo de eliminar possíveis heterogeneidades.

Além de solucionar o problema da heterogeneidade, um sistema de integração, principalmente quando destinado para a Web, deve ser capaz de lidar com a mudança dinâmica de suas fontes, caso contrário, ele torna-se obsoleto cada vez que seu repositório sofre a adição e/ou remoção de fontes, ou quando ocorre alguma alteração no conteúdo das fontes já conectadas ao sistema.

O sistema também deve preocupar-se em prover a automatização do processo de integração de forma a reduzir o envolvimento humano nas tarefas do processo de integração. Além disso, a proposta do sistema não deve se limitar a um único domínio de aplicação. Deve ser prevista a possibilidade de atuação em diferentes domínios.

As fontes de informação na Web incluem portais Web estáticos, bibliotecas digitais, glossários, etc. Porém, além destas fontes, existe um grande número de portais Web dinâmicos os quais oferecem não apenas dados aos usuários, mas também serviços (ACUÑA et al., 2005), tais como os motores de busca. Os serviços (ou processos) oferecidos pelos portais Web, disponibilizam informações de diversas fontes distribuídas de uma maneira unificada através do portal. Apenas disponibilizar a integração de dados estáticos na Web não é mais suficiente para suprir as necessidades dos usuários da Web atual.

Diversas áreas de pesquisa da Ciência da Computação dedicam-se atualmente ao problema da integração de informação na Web. As principais técnicas utilizadas nas propostas para integração de informação na Web são (KAMBHAMPATI e KNOBLOCK, 2003; USCHOLD e GRUNINGER, 2004): a linguagem XML (*eXtensible Markup Language*); ontologias de domínio; Serviços Web; e Serviços Web Semânticos.

A linguagem XML já foi considerada padrão de representação e troca de informação na Web (BERTINO e FERRARI, 2001). Apesar de XML ser capaz de ajudar na tarefa de integração de informação, ela não é suficiente pra lidar com todas as dificuldades inerentes a esta tarefa. XML estabelece uma plataforma sintática para a troca de informação na Web, porém ela não tem a capacidade de lidar com heterogeneidade estrutural entre as fontes, além de não expressar a semântica dos elementos ou de seus relacionamentos explicitamente (as *tags* são compreensíveis apenas por seres humanos). Apesar de XML promover interoperabilidade entre organizações que concordam previamente com os padrões a serem utilizados; o vocabulário divergente diminui sua usabilidade em buscas gerais na Web (HEFLIN e HENDLER, 2000).

Os Serviços Web, através da padronização da infra-estrutura para a troca de dados, podem simplificar bastante a integração na Web. Sua utilização como proposta de solução para o problema da integração é proposta em diversos trabalhos (HANSEN et al., 2002; VINOSKI, 2003; ZHU et al., 2004) os quais demonstram que os Serviços Web fornecem uma boa infra-estrutura para dar suporte a integração. Porém, estes mesmos trabalhos também apontam as principais limitações dos Serviços Web quando aplicados em sistemas de integração, tais como a falta de suporte para a automatização dos processos de descoberta e execução dos serviços, o que impossibilita a automatização da integração. Além disso, a integração realizada atualmente baseada nesta tecnologia é exclusivamente sintática, não considerando os aspectos semânticos (ACUÑA et al., 2005).

As ontologias de domínio têm sido muito utilizadas em sistemas de integração de informação pelo fato de elas fornecerem um vocabulário comum explícito e processável por computador sobre um domínio de aplicação. Tal vocabulário resolve os conflitos semânticos entre conceitos e termos utilizados nas fontes de informação heterogêneas. Elas assumem um papel fundamental no processo de integração, viabilizando a interoperabilidade semântica de informações heterogêneas. Em ferramentas de integração, as ontologias podem ser aplicadas na identificação e associação de conceitos correspondentes semanticamente. Entretanto, para que a busca baseada em ontologia seja possível, é necessário que a descrição semântica das fontes seja realizada (USCHOLD e GRUNINGER, 2004).

A integração de dados constitui uma das etapas necessárias para conseguir a completa interoperabilidade entre aplicações pretendida no cenário da Web Semântica. Por esta razão, conceitos e técnicas da Web Semântica têm sido aplicados no desenvolvimento de sistemas de integração de dados e processos na Web. Os Serviços Web Semânticos constituem um campo de aplicação novo e promissor que está sendo reconhecido como sendo o suporte adequado para o processo de descoberta e integração de informação na Web.

A adoção de Serviços Web Semânticos como solução para o problema de integração de informação na Web, proposta em trabalhos como (ACUÑA et al., 2005), possibilita usufruir as vantagens proporcionadas pelos Serviços Web e pelas ontologias de domínio. Com uma representação formal das funcionalidades oferecidas pelos Serviços Web, eles tornam-se acessíveis por aplicações automatizadas, tornando viável a automatização do processo de integração.

Existem na literatura inúmeros trabalhos propondo diversas abordagens na tentativa de solucionar o problema da integração de informação na Web. Alguns trabalhos desenvolvidos ao longo dos últimos anos são descritos a seguir.

A ferramenta de busca *SHOE* (HEFLIN e HENDLER, 2000) utiliza uma linguagem de marcação semântica, também denominada SHOE, para realizar buscas mais efetivas na Web. A linguagem faz uso de ontologias para adicionar marcação às páginas Web. A ferramenta SHOE permite que o usuário especifique o contexto para a sua consulta e depois utiliza tal contexto para auxiliar o usuário a formular uma nova consulta.

O sistema *OBSERVER (Ontology Based System Enhanced with Relationships for Vocabulary Heterogeneity Resolution)* (MENA et al., 2000) realiza a integração de repositórios de dados heterogêneos, distribuídos e desenvolvidos independentemente. O conteúdo de cada repositório é descrito por uma ou mais ontologias pré-existentes, as quais são expressas utilizando um sistema baseado em Lógica de Descrição. Um repositório compartilhado contendo os relacionamentos entre as ontologias é utilizado para realizar o mapeamento semântico e disponibilizar tal mapeamento de forma explícita.

Köhler e Schulze-Kremer (2002) propõem o sistema *SEMEDA* para integração semântica inteligente e consulta a bancos de dados federados. O sistema permite lidar com diversas ontologias, entretanto, quando é utilizado para integração, todas as bases devem ser definidas semanticamente com base na mesma ontologia (os atributos das bases de dados são definidos através de referências aos conceitos da ontologia). As bases de dados têm suas informações semânticas e técnicas divulgadas através de um browser Web.

SEWISE (GARDARIN et al., 2003) é um sistema de busca que promove a integração de dados de páginas Web as quais são encapsuladas por *wrappers* (componentes de software cuja função é extrair os elementos de informação úteis presentes nas fontes Web e mapeá-los para um modelo comum de dados acessível para motores de busca e ferramentas de mediação). *SEWISE* utiliza ontologia para descrever um modelo comum de dados em um dado domínio de aplicação. O sistema é capaz de mapear informação textual de diversas fontes Web para uma estrutura XML. Os documentos XML são então enriquecidos semanticamente através de quatro técnicas de mineração de texto: indexação de texto, categorização de texto, extração de sentença, e extração de palavra-chave. Estas técnicas são aplicadas aos elementos do texto com o objetivo de marcar semanticamente o documento XML gerado pelo *wrapper* de cada fonte.

Já a ferramenta *BioMeke* (MARQUET et al., 2003), apesar de ser descrita como baseada em ontologia, faz uso na verdade de terminologias padrões não formais. Esta ferramenta oferece um motor de busca desenvolvido para facilitar a extração de informação biológica e médica, acessível a partir de diversas fontes públicas e repositórios locais de biólogos. Ela utiliza o conhecimento disponível em terminologias padrões existentes (HUGO e UMLS) para realizar a unificação semântica, de forma limitada, de diferentes termos e referências cruzadas.

O sistema *Integra* (BELIAN e SALGADO, 2004) apresenta uma adaptação de um sistema de integração previamente existente para incorporar o tratamento semântico da informação. Ele adota conceitos como metadados, contextos e ontologias em um sistema de integração de informações baseado em mediação. Porém, pelo fato da arquitetura ser baseada em mediação, adotando a abordagem *Global-as-View*, torna-se difícil lidar com alterações nas fontes de informação.

Um exemplo de pesquisa que explora o uso de Serviços Web para realizar integração na Web é o sistema *Application Manifold – AM* (EYAL e MILO, 2001). Ele

tem como objetivo simplificar a tarefa de integração e customização de aplicações de comércio eletrônico disponibilizadas na Web. As aplicações Web são encapsuladas através de Serviços Web, permitindo a integração dos processos disponibilizados por tais aplicações. Porém, o sistema não lida com a heterogeneidade semântica. O valor das variáveis de cada fonte local é fornecido pelo usuário no esquema global.

Atualmente, poucos estudos fazem uso de Serviços Web Semânticos. Um dos estudos mais importantes é o projeto DIP¹ (*Data, Information, and Process Integration with Semantic Web Services*) (VASILIU et al., 2004; DIP, 2006), o qual visa desenvolver soluções para os problemas inerentes a integração através da combinação de tecnologias da Web Semântica e de Serviços Web, fornecendo um ambiente no qual Serviços Web diferentes podem cooperar uns com os outros e serem descobertos automaticamente.

O sistema de integração *BIRD*, para o domínio da Biomedicina (GOMEZ et al., 2007), é a proposta na qual o presente trabalho foi baseado. Seu desenvolvimento fez parte do processo inicial do projeto descrito no presente trabalho, sendo realizado na Universidad de Murcia, na Espanha. Este trabalho propõe uma solução que combina análise de linguagem natural e técnicas que lidam com aspectos semânticos (ontologia de domínio e Serviços Web Semânticos) com o objetivo de satisfazer as necessidades dos usuários buscando informação na Web sobre o domínio de aplicação.

Com o desenvolvimento da Web, os motores de busca globais tornaram-se rapidamente o método preferido pela comunidade em geral para encontrar informação. Estas ferramentas também constituem uma forma de integrar informação na Web, pois combinam dados de múltiplas fontes autônomas disponíveis na Internet para responder consultas de usuários. Motores de busca globais são ferramentas que pesquisam todos os documentos na Web, apresentando resultados de forma aleatória, dependendo do ranking de acessos às páginas Web. Os resultados são retornados na forma de *links* para as páginas onde a resposta realmente se encontra, além disso, as informações retornadas nestes resultados podem estar relacionadas a qualquer assunto. As principais estratégias utilizadas pelas ferramentas de busca é o casamento de *strings* por palavras-chave e a análise de *link*.

O motor de busca mais popular atualmente é a ferramenta do *Google*. Seu principal diferencial é a utilização de um sistema, denominado *PageRank* (GOOGLE, 2008), o qual atribui uma pontuação para as páginas Web de acordo com a quantidade e a qualidade das ligações (através de *links*) que apontam para elas, sendo este um dos fatores de maior peso na definição do ordenamento das páginas apresentadas pelo sistema. Além disso, o *Google* analisa os assuntos mais pesquisados e verifica quais páginas tratam tais temas de maneira significativa (GOOGLE, 2008). Entretanto, tanto o motor do *Google* como os demais motores de busca globais não são capazes de contextualizar a busca do usuário de acordo com suas necessidades. Como consequência disto, o usuário normalmente enfrenta dificuldades em encontrar o resultado desejado para as suas pesquisas.

Além das ferramentas de busca tradicionais, os motores de meta-busca também são utilizados como propostas de solução para o problema da integração. Eles são uma tentativa de prover uma visão integrada de informações disponíveis na Web, disponibilizando uma interface no topo de diversos motores de busca, fornecendo

¹ <http://dip.semanticweb.org>

acesso uniforme a cada mecanismo de busca local. O sistema experimental *WebFusion* (Husek et al., 2007) propõe um acesso integrado a diversos motores de busca tradicionais através de um mecanismo de meta-busca. Ele propõe melhorar o processo de meta-busca de acordo com preferências do usuário, através da análise dos logs da consulta, fornecendo resultados mais relevantes para cada usuário. Os resultados de cada motor de busca são retornados de forma automática, porém o usuário ainda deve clicar no link da(s) página(s) fornecida(s) para obter a resposta concreta.

As propostas atuais para integração na Web, em sua grande maioria, não permitem que o usuário também acesse os serviços oferecidos pelos portais Web. Abordagens tradicionais para integração estão focadas somente na integração de dados, ignorando os serviços disponibilizados pelos portais. Outra desvantagem dos sistemas de integração disponibilizados atualmente é a não automatização do seu processo por completo. Alguns sistemas ainda disponibilizam uma semi-automatização, mas ainda exigindo o envolvimento humano na execução de algumas tarefas. Além disso, apesar de existirem aplicações voltadas para os domínios da Biologia, Genética e Biomedicina, não existem propostas de integração na Web para informações relacionadas ao domínio da Geologia.

A solução adequada para a presente pesquisa deve resolver tanto os aspectos sintáticos, estruturais e semânticos na integração de informações. As ontologias e as tecnologias baseadas em semântica são necessárias para que o sistema seja capaz de lidar com os aspectos semânticos, ou seja, para especificar formalmente o significado da terminologia de cada fonte e definir o mapeamento entre as terminologias. Além disso, a solução deve ser capaz de integrar não apenas dados, mas também processos os quais são os serviços oferecidos pelos portais aos usuários.

Nas tabelas 2.1 e 2.2 são apresentados comparativos entre as principais técnicas para a integração de informação na Web e as principais abordagens existentes, respectivamente.

Tabela 2.1: Comparação entre as técnicas utilizadas para integração na Web

Tecnologia	Dá suporte para automatização	Atua no nível semântico	Permite integrar processos
XML	Não	Não	Não
Serviços Web	Não	Não	Sim
Ontologias de Domínio	Não	Sim	Não
Serviços Web Semânticos	Sim	Sim	Sim

Tabela 2.2: Comparação entre abordagens para integração na Web

Abordagem	Resumo	Domínio de Aplicação	Automatização	Nível Semântico	Processos	Manutenção	Validação
SHOE (2000)	Utiliza uma linguagem de marcação semântica para realizar a integração da informação em páginas Web e, desta forma, prover buscas mais efetivas na Web.	O domínio é escolhido pelo usuário de acordo com a ontologia selecionada a partir de uma lista.	O processo de integração não é automatizado. Ele requer envolvimento humano em passos intermediários da consulta.	Utiliza ontologias.	Não integra processos.	Permite a adição de novas fontes. Porém, se o conteúdo das fontes é alterado, uma nova descrição semântica deve ser feita.	Realização de testes com consultas simples.
OBSERVER (2000)	Sistema para integração de repositórios de dados heterogêneos, distribuídos e desenvolvidos independentemente, através de múltiplas ontologias pré-existentes.	É específico para o domínio da informação bibliográfica.	Propõe uma solução semi-automatizada.	Utiliza um repositório compartilhado contendo as relações para o mapeamento entre as ontologias.	Integra apenas dados estáticos.	O sistema é independente do número de repositórios de dados e de ontologias.	Desenvolvimento de um protótipo para testes utilizando repositórios do mundo real.
Manifold (2001)	Realiza a integração de aplicações Web previamente existentes através do seu encapsulamento por Serviços Web. Utiliza XML e UML.	É específico para aplicações de comércio eletrônico.	Não dá suporte a automatização do processo de integração.	Não lida com a heterogeneidade e semântica.	Integra processos além de dados.	Permite que novas fontes sejam adicionadas ao modelo.	Experimento com a simulação de venda de livros <i>online</i> .

Tabela 2.2: Comparação entre abordagens para integração na Web

Abordagem	Resumo	Domínio de Aplicação	Automatização	Nível Semântico	Processos	Manutenção	Validação
SEMEDA (2002)	Sistema para integração semântica inteligente, baseado em ontologia.	Implementado para o domínio da biologia molecular. Mas pode ser usado com bases de outros domínios de aplicação.	O processo de consulta não é especificado.	Utiliza ontologia de domínio.	Não integra processos.	Permite a adição de novas bases ao sistema.	O sistema ainda está sob desenvolvimento.
SEWISE (2003)	Sistema de informação para a Web, baseado em ontologia, que dá suporte a descrição e recuperação de informação na Web.	Implementado para o domínio financeiro, mas pode ser aplicado em outros domínios (usando a devida ontologia de domínio).	Propõe uma solução semi-automatizada.	Utiliza ontologia de domínio.	Não integra processos.	Permite que novas páginas Web sejam conectadas ao sistema.	Testes com diversas páginas Web populares que disponibilizam notícias sobre a economia francesa.
BioMeke (2003)	Oferece um motor de busca baseado em terminologias padrões não formais.	É específico para o domínio da Biomedicina.	Não oferece automatização. Muitas tarefas estão sob responsabilidade do usuário.	Lida com a heterogeneidade semântica de forma limitada. Utiliza terminologias padrões existentes.	Não integra processos.	Não dá suporte a adição de novas fontes.	Comparação da ferramenta com outros sistemas existentes.

Tabela 2.2: Comparação entre abordagens para integração na Web

Abordagem	Resumo	Domínio de Aplicação	Automatização	Nível Semântico	Processos	Manutenção	Validação
Integra (2004)	Baseia-se em mediação, utilizando metadados, contextos e ontologias para integrar fontes de dados heterogêneas e autônomas na Web.	É específico para a área da saúde.	Propõe uma solução semi-automatizada.	Utiliza ontologia de domínio.	Integra apenas dados estáticos.	Difícil de lidar com alterações nas fontes.	A proposta ainda não foi validada.
DIP (2004-2006)	É um projeto integrado focado no desenvolvimento de soluções para a integração de informação através de Serviços Web Semânticos.	É independente de domínio de aplicação. É aplicado principalmente no comércio eletrônico.	Fornecer uma plataforma onde vários passos envolvidos no desenvolvimento da aplicação são automatizados.	Utiliza ontologia de domínio (através de SWS) para solucionar o problema da heterogeneidade semântica.	Integra processos além de dados.	Suporta adição e remoção de fontes, bem como alterações nas informações de cada fonte.	Diversos estudos de caso foram realizados com o objetivo de validar a proposta.
BIRD (2007)	Utiliza SWS para fornecer uma visão virtual unificada das funcionalidades oferecidas por portais Web.	É específico para o domínio da Biomedicina.	Propõe uma solução automatizada.	Utiliza ontologia de domínio (através de SWS) para solucionar o problema da heterogeneidade semântica.	Integra processos além de dados.	Permite que portais sejam adicionados / removidos. Tem suporte para a modificação do conteúdo das fontes.	Estudo de caso em uma aplicação real.

Tabela 2.2: Comparação entre abordagens para integração na Web

Abordagem	Resumo	Domínio de Aplicação	Automatização	Nível Semântico	Processos	Manutenção	Validação
WebFusion (2007)	Utiliza <i>logs</i> de consulta para definir perfis de usuários com o objetivo de prover um motor de meta-busca mais direcionada a cada usuário.	É independente de domínio de aplicação.	Propõe uma solução semi-automatizada.	Não lida com a heterogeneidade semântica.	Integra resultados de motores de busca.	Permite a adição de novos motores de busca ao sistema.	Estudo de caso com consultas na categoria “Computadores”
Google (2008)	Motor de busca global que combina informações de múltiplas fontes Web autônomas para responder consultas de usuários de forma aleatória.	É independente de domínio de aplicação.	Propõe uma solução automatizada.	Não lida com a heterogeneidade semântica.	Integra processos além de dados. Porém, não retorna as respostas dos serviços diretamente para o usuário.	Tem suporte para qualquer alteração realizada nas fontes Web.	Sistema disponível na Web para a utilização pela comunidade em geral.

A seguir serão apresentados os conceitos e técnicas utilizadas na solução proposta no presente trabalho para o problema da integração de informação na Web.

2.1 Ontologias de Domínio

O termo ‘ontologia’ provém da área da filosofia, onde tem o significado de explicação sistemática da existência. Na ciência da computação, a ontologia é definida como uma especificação formal de uma conceitualização compartilhada (BORST, 1997). Segundo Netches et al. (1991 apud GÓMEZ-PÉREZ, 1999, p. 33), uma ontologia define os termos e as relações abrangendo o vocabulário de uma área específica, assim como as regras para combinar tais termos e relações para definir extensões ao vocabulário. As ontologias são normalmente organizadas em taxonomias, contendo primitivas de modelagem tais como: classes, atributos, relacionamentos, funções, axiomas e instâncias.

A utilização de ontologias possibilita o compartilhamento de conceitos comuns; a especialização destes conceitos e do vocabulário para serem reutilizados em diversas aplicações; o mapeamento de conceitos entre ontologias diferentes; e a composição de novos conceitos a partir de diversas ontologias (MCILRAITH et al., 2001). A principal vantagem proporcionada pelo uso de ontologias é a possibilidade de reusar conhecimento e compartilhá-lo facilmente entre pessoas ou agentes de software. As ontologias têm sido utilizadas principalmente para representação e reuso de conhecimento em aplicações inteligentes; para melhorar a comunicação e o compartilhamento de conhecimento entre pessoas e agentes de software; em máquinas de busca inteligente (Web Semântica); e para facilitar o processo de integração de bases de dados.

Uma possível classificação para as ontologias é dada pelas seguintes categorias (GÓMEZ-PÉREZ et al., 2004): ontologias de representação de conhecimento, ontologias gerais, ontologias *top-level*, ontologias de tarefa, ontologias de domínio-tarefa, meta-ontologias, ontologias de aplicação e ontologias de domínio. Esta última classe engloba as ontologias reusáveis em um domínio específico.

Um domínio é uma área de interesse específica (SCHREIBER et al., 2000). Exemplos de domínios são: Medicina, Geologia, Computação, etc. As ontologias de domínio fornecem vocabulário sobre conceitos em tal domínio e os relacionamentos entre eles, sobre as atividades realizadas no domínio, e sobre as teorias e princípios elementares que o governam. Os conceitos nas ontologias de domínio normalmente são especializações de conceitos já definidos em ontologias *top-level*, o mesmo ocorrendo com os relacionamentos. Uma ontologia de domínio corresponde, portanto, a uma representação estática de um domínio a partir de seus conceitos abstratos e a forma como esses conceitos se relacionam entre si (SCHREIBER et al., 2000). Os conceitos da ontologia de domínio podem ser utilizados para realizar a anotação semântica de recursos da Web, com o objetivo de definir o significado dos termos utilizados nos recursos, em um domínio específico.

O uso de ontologias torna possível definir uma infra-estrutura para integrar sistemas inteligentes no nível de conhecimento. Além disso, outras vantagens do uso de ontologias podem ser destacadas, tais como (GÓMEZ-PÉREZ et al., 2004): a possibilidade de compartilhamento do conhecimento entre indivíduos em um determinado domínio; a facilitação do processo de integração da informação, especialmente em aplicações distribuídas; a disponibilização da informação como fonte

de consulta e de referência do domínio; e a possibilidade de reuso de conceitos em domínios onde o reuso é bastante útil.

2.1.1 Linguagens Ontológicas

Para que as ontologias possam ser utilizadas em aplicações práticas, elas devem ser disponibilizadas utilizando alguma forma de representação concreta. As linguagens ontológicas são linguagens formais utilizadas no desenvolvimento de ontologias, agindo como veículos para que as ontologias possam ser expressas de uma maneira compreensível por máquina. Elas possibilitam que usuários escrevam conceitualizações formais e explícitas de modelos de domínio.

As linguagens ontológicas permitem codificar o conhecimento sobre domínios específicos e frequentemente incluem regras de raciocínio para dar suporte ao processamento de tal conhecimento. Uma grande variedade de linguagens pode ser utilizada para codificar ou representar modelos conceituais. Diferentes linguagens são normalmente baseadas em diferentes paradigmas subjacentes, tais como: lógica de descrição, lógica de primeira ordem, representação baseada em *frames*, taxonomia, rede semântica, etc.

A linguagem escolhida para a especificação das ontologias de domínio do presente trabalho foi a OWL (*Ontology Web Language*), visto que ela é hoje a abordagem mais utilizada para a modelagem de ontologias para uso na Web, sendo bastante expressiva, além de ser recomendada pela W3C (*World Wide Web Consortium*) (MCGUINNESS e VAN HARMELEN, 2004).

OWL pode ser considerada como uma família de três linguagens ontológicas: OWL-Lite, OWL-DL e OWL-Full; desenvolvida pela W3C como parte de seus desenvolvimentos para a Web Semântica (MCGUINNESS e VAN HARMELEN, 2004). OWL-Lite é a sublinguagem mais simples sintaticamente, sendo a menos expressiva. OWL-DL por sua vez é baseada em lógica de descrição, o que a torna uma linguagem bem mais expressiva que a OWL-Lite. OWL-Full é a subontologia OWL mais expressiva, entretanto, não é possível realizar raciocínio automatizado sobre esta ontologia.

A linguagem OWL foi desenvolvida com base na lógica de descrição, apresentando uma semântica bem definida, propriedades formais fáceis de serem compreendidas, além do fato de que existem vários algoritmos desenvolvidos para raciocínio e inúmeros sistemas já implementados, utilizando OWL.

Apesar de a linguagem OWL ter sido inicialmente desenvolvida para ser utilizada na Web Semântica e no seu desenvolvimento, ela tornou-se rapidamente um padrão para o desenvolvimento de ontologias (GRAU et al., 2006). A linguagem OWL é uma revisão da linguagem ontológica DAML+OIL, também desenvolvida para a Web. Ela foi desenvolvida como uma extensão de vocabulário da linguagem RDF (*Resource Description Framework*), proporcionando uma representação de conhecimento mais expressiva. OWL é uma linguagem semântica utilizada para publicação e compartilhamento de ontologias na Web (DEAN e SCHREIBER, 2004).

OWL provê uma linguagem que utiliza a conexão fornecida por RDF para adicionar as seguintes capacidades às ontologias: possibilidade de estar distribuída através de vários sistemas; ser compatível com os padrões da Web para acessibilidade e internacionalização; além de ser estendível. OWL foi desenvolvida para ser mais

flexível e expressiva que XML, RDF(s) e outras estruturas semânticas desenvolvidas para a Web. Ela tem mais facilidades para expressar significados e semântica do que a linguagem RDF(s), indo além desta linguagem na sua habilidade de representar conteúdo interpretável por máquina na Web.

2.2 Serviços Web Semânticos – Uma Revisão

A Web Semântica é uma evolução promissora da Web atual que pretende fornecer meios para tornar o processo de automação fácil e efetivo. Um de seus componentes fundamentais será a marcação de Serviços Web para torná-los processáveis por computador; aparentes ao usuário; e executáveis de forma automatizada (MCILRAITH et al., 2001). Nesta seção são apresentados os principais conceitos relacionados aos Serviços Web Semânticos. Primeiramente, os Serviços Web são definidos, assim como seus objetivos. Em seguida, uma visão geral da Web Semântica é apresentada. Na sequência, o conceito de Serviços Web Semânticos é introduzido. Por fim, uma atenção especial é dada para OWL-S, uma abordagem para o desenvolvimento de Serviços Web Semânticos.

2.2.1 Serviços Web

Serviços Web são componentes de software bem definidos e reutilizáveis que realizam tarefas encapsuladas específicas através de mecanismos padrões orientados para a Web (CABRAL et al., 2004). Eles podem ser descobertos, invocados, e a composição de diversos serviços pode ser coreografada utilizando estruturas de modelagem de *workflow* bem definidas.

O foco principal de um Serviço Web é a comunicação entre aplicativos, abstraindo as plataformas nas quais estes foram desenvolvidos. Aplicações baseadas em serviço devem ser capazes de interoperar apesar de terem sido desenvolvidas em diferentes linguagens de programação, em diferentes períodos de tempo, por pessoas diferentes, com design baseado em suposições diferentes.

De acordo com Antoniou e Van Harmelen (2004), um simples Serviço Web envolve um único programa, sensor ou mecanismo acessível através da Web que não conta com outros Serviços Web, nem requer interação adicional com o usuário além de uma resposta simples. Já os Serviços Web compostos, por sua vez, são formados por serviços mais simples e normalmente requerem interação contínua com o usuário na qual ele pode fazer escolhas ou fornecer informações de forma condicional.

Correntemente, um Serviço Web é um sistema de software identificado por uma URI (*Uniform Resource Identifier*), cujas interfaces e ligações públicas são definidas e descritas utilizando XML (ALESSO, 2005). Um Serviço Web pode ser acessado na Internet através da sua interface exposta, onde suas definições e funções podem ser descobertas por outros sistemas de software de uma forma controlada (ALONSO et al., 2004). Estes sistemas interagem então com o Serviço Web utilizando mensagens baseadas em XML sobre protocolos da Internet. Basicamente, a interface do Serviço Web é apresentada em WSDL (*Web Services Description Language*), que descreve como outros sistemas podem interagir com ele através da utilização de mensagens SOAP (*Simple Object Access Protocol*). Essas mensagens são tipicamente transportadas utilizando HTTP (*Hypertext Transfer Protocol*) com uma formatação XML e em conjunto com outras convenções relacionadas a Web.

Uma interação de Serviço Web, como demonstrada na figura 2.1, envolve dois ou mais agentes de software (clientes e provedores) trocando informações na forma de mensagens. Primeiro, o provedor registra o serviço que ele está oferecendo, utilizando WSDL. O registro é baseado na especificação UDDI (*Universal Description, Discovery and Integration*), a qual define as estruturas de dados e APIs (*Application Programming Interface*) para a publicação (ou descrição) e descoberta de Serviços Web (ALONSO et al., 2004). Uma vez que o serviço esteja registrado, o cliente pode consultar o registro para buscar pelos serviços nos quais ele tem interesse. Após receber a resposta, o cliente invoca o serviço que ele deseja, utilizando SOAP.

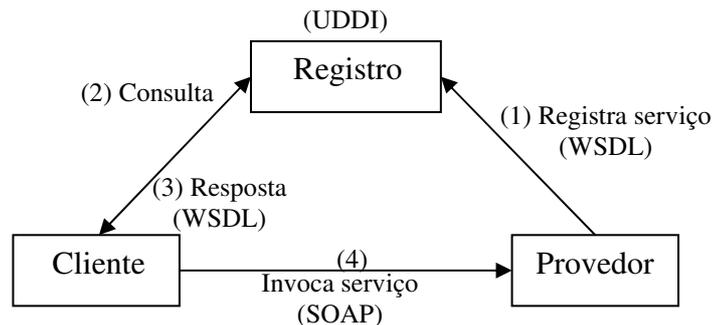


Figura 2.1: Interação de Serviço Web

As tecnologias de Serviços Web citadas anteriormente são descritas resumidamente a seguir (ALONSO et al., 2004):

- *eXtensible Markup Language (XML)*: linguagem baseada em *tags* que tem como principal função a descrição de informações para troca de dados em ambiente Web. Amplamente aceita pela indústria de software e suportada pelas principais ferramentas do mercado, XML fornece uma sintaxe comum para todas as especificações de Serviços Web.
- *Universal Description, Discovery and Integration (UDDI)*: proporciona uma plataforma padrão que permite que companhias e aplicações encontrem Serviços Web de forma rápida, fácil e dinâmica, através da Internet. É utilizado como repositório para publicação de Serviços Web, onde os fornecedores cadastram seus serviços e os clientes podem buscá-los. O registro apresenta três partes fundamentais: (1) páginas brancas, que registram informação sobre o endereço, contato e outros identificadores; (2) páginas amarelas as quais indicam uma categorização industrial baseada em taxonomias; (3) páginas verdes, que oferecem informação técnica sobre os serviços.
- *Simple Object Access Protocol (SOAP)*: protocolo leve para troca de informações em ambiente distribuído, utilizando mensagens, desenvolvido para ser independente de qualquer modelo de programação ou outra semântica específica de implementação. É baseado em definições XML e utilizado para acessar Serviços Web. Esse protocolo encapsula as chamadas e retornos aos Serviços Web, sendo utilizado, principalmente, sobre HTTP.
- *Web Services Description Language (WSDL)*: linguagem de descrição de serviços baseada em definições XML, utilizada para descrever Serviços Web. É

responsável por prover as informações necessárias para se invocar um Serviço Web como sua localização, operações disponíveis e suas assinaturas.

WSDL é a linguagem recomendada pela W3C (CHINNICI et al., 2007) para descrever a interface do serviço, utilizando notação XML. Ela possibilita a especificação de tipos de dados, assinaturas de operações, formatos de mensagens e detalhes de protocolos de comunicação, além do agrupamento de diferentes portas em um Serviço Web (TOSIC et al., 2002). Entretanto, WSDL não possibilita a especificação de diversas restrições relacionadas a operações em um Serviço Web. As operações são definidas em termos de suas mensagens de entrada e de saída.

WSDL descreve as interfaces públicas e as ligações dos Serviços Web, enquanto os Serviços Web por sua vez podem ser vistos como implementações remotas e independentes de plataforma de tais interfaces. WSDL define as operações suportadas por um Serviço Web assim como a forma da troca de dados através de mensagens.

A definição WSDL apresenta duas partes: uma parte lógica, denominada *port-type* ou *interface*, que define os tipos e mensagens de entrada e saída independentes de protocolo; e uma parte física que define as ligações específicas com algum protocolo. O atual padrão WSDL opera somente no nível sintático, não apresentando a expressividade semântica necessária para representar as requisições e as capacidades dos Serviços Web.

À medida que a Web cresce em tamanho e diversidade, a automatização de aspectos relacionados aos Serviços Web como a descoberta, a seleção, a composição e a execução, se faz cada vez mais necessária. O problema é que a tecnologia atual, baseada em UDDI, WSDL e SOAP, não proporciona meios para conseguir este dinamismo. Atualmente, existem poucas especificações disponíveis para Serviços Web capazes de fornecer algum tipo de definição formal para o possível significado das descrições sintáticas das funcionalidades dos Serviços Web. A falta de semântica, de uma forma interpretável por máquina, requer intervenção humana para a descoberta automatizada de um serviço e para a composição em sistemas abertos. Isto dificulta a sua utilização em contextos complexos de negócio (CABRAL et al., 2004).

2.2.2 Web Semântica

A Web Semântica fornece uma estrutura comum a qual permite que dados sejam compartilhados e reusados além das aplicações, empresas e dos limites de comunidades (W3C, 2003). Ela é baseada em RDF e faz parte de um esforço colaborativo encabeçado pela W3C com a participação de um grande número de pesquisadores e parceiros industriais.

A Web Semântica é uma extensão da Web atual onde a informação tem um significado bem definido (W3C, 2003). O projeto da Semântica Web pretende criar uma forma universal de troca de informação, fornecendo conhecimento aos conteúdos dos documentos na Web de forma que eles possam ser compreendidos por computadores. Seu principal objetivo é transformar a Web em um meio através do qual dados possam ser compartilhados, compreendidos e processados por ferramentas automatizadas; isto é, permitir que os recursos na Web sejam acessíveis não só por seres humanos, mas também para processos automatizados (agentes automáticos navegando na Web em busca de informação de forma eficiente e precisa).

De forma resumida, a principal limitação da Web atual é o fato de que suas tecnologias não são capazes de capturar ou representar formalmente a semântica do conteúdo apresentado. Uma grande proporção dos dados atualmente presentes na Web é compreensível apenas por seres-humanos ou aplicações desenvolvidas “sob medida”. A Web Semântica irá proporcionar estrutura para o conteúdo significativo de páginas Web, criando um ambiente no qual agentes de software vagando de página em página podem realizar imediatamente as tarefas sofisticadas para os usuários (BERNERS-LEE et al., 2001). Para que isto seja possível, há a necessidade de enriquecer a Web com ontologias as quais capturam o conhecimento do domínio. A Web Semântica juntamente com suas tecnologias emergentes – linguagens de descrição de ontologias e a noção de Serviços Web – possibilitam uma modelagem de domínio rica semanticamente, permitindo um acesso inteligente às suas informações.

Os Serviços Web podem ter contribuições cruciais em sua estrutura através das tecnologias da Web Semântica (ELENIUS et al., 2005). Os Serviços Web Semânticos baseiam-se nesta idéia, introduzindo ontologias para descrever, por um lado, os conceitos nos domínios dos serviços e, por outro lado, as características dos serviços e seus relacionamentos com as ontologias de domínio. Essas descrições, ricas em significado, possibilitam o raciocínio automático sobre o serviço e descrições do domínio, dando suporte à automação da descoberta e execução do serviço, além de reduzir configurações manuais e esforços de programação.

2.2.3 Serviços Web Semânticos

Atualmente, a utilização de Serviços Web, em sua grande maioria, exige o envolvimento humano (informações devem ser buscadas ou formulários devem ser preenchidos). As tecnologias existentes para Serviços Web apenas fornecem descrições no nível sintático, tornando difícil para clientes e provedores interpretar ou representarem declarações não triviais tais como o significado de entradas e saídas ou restrições aplicáveis (CABRAL et al., 2004). Assume-se que as diferentes operações, a ordem na qual elas devem ser invocadas e outras propriedades dos serviços são conhecidas antecipadamente pelo programador que está desenvolvendo os clientes.

A visão da Web Semântica, quando aplicada aos Serviços Web, visa automatizar a descoberta, a execução, a composição e o gerenciamento dos Serviços Web, fornecendo descrições dos serviços interpretáveis por máquina (ANTONIOU e VAN HARMELEN, 2004). Um pré-requisito para isto, no entanto, é o surgimento e a evolução da Web Semântica, a qual fornece a infra-estrutura para a interoperabilidade semântica dos Serviços Web.

As anotações semânticas são a base para o processamento automático de portais Web. Para alcançar esta automação, os Serviços Web serão incrementados com descrições formais ricas de suas capacidades de forma que eles possam ser utilizados por aplicações ou outros serviços, sem assistência humana ou acordos muito restritos sobre interfaces e protocolos (CABRAL et al., 2004).

É importante ressaltar que a semântica adicionada ao Serviço Web é sobre a descrição do serviço, e não sobre seu conteúdo. A descrição semântica de um Serviço Web deve conter as seguintes informações: as funcionalidades oferecidas; efeitos ocasionados pela execução do serviço; parâmetros de entrada esperados; parâmetros de saída fornecidos, condições necessárias (pré-requisitos) para a execução do serviço;

onde encontrar o serviço; qual protocolo utilizar; etc. Todo o conteúdo da descrição deve ser disponível de alguma forma interpretável por máquina.

Um Serviço Web Semântico é definido através de uma ontologia de serviço, a qual possibilita que suas capacidades sejam interpretáveis por máquina, além de possibilitar a integração com o conhecimento de domínio. A ontologia de serviço agrega todos os modelos de conceitos relacionados à descrição de um Serviço Web Semântico, constituindo o nível de conhecimento da informação que descreve e dá suporte ao uso do serviço (CABRAL et al., 2004).

Existem quatro propostas que se destacam no desenvolvimento de estruturas para os Serviços Web Semânticos: WSMO (ROMAN et al, 2005), IRS-III (CABRAL et al., 2006), SAWSDL (FARREL e LAUSEN, 2007) e OWL-S (MARTIN et al., 2005). IRS-III (*Internet Reasoning Service*) é uma proposta baseada em conhecimento para os Serviços Web Semânticos que evoluiu a partir de pesquisas em componentes de software reusáveis. WSMO (*Web Service Modeling Ontology*) tem sua base conceitual em WSMF (*Web Service Modeling Framework*), apresentando uma proposta orientada a negócio para possibilitar o comércio eletrônico por completo através da aplicação da tecnologia da Web Semântica nos Serviços Web. SAWSDL (*Semantic Annotations for WSDL and XML Schema*) visa permitir a descrição de semântica adicional dos componentes WSDL. OWL-S (*Ontology Web Language for Services*), por sua vez, é uma proposta orientada a agentes que fornece uma ontologia para descrever as capacidades dos Serviços Web.

WSMO refina WSMF e desenvolve uma ontologia de serviço formal, além de uma linguagem para SWS (ROMAN et al, 2005). A ontologia de serviço WSMO baseia-se em quatro elementos principais: ontologias, objetivos (objetivo de um cliente quando consulta um SW), Serviços Web, e mediadores (elementos utilizados para mapear componentes heterogêneos envolvidos na modelagem do SW). Esta abordagem é mais direcionada para a área do comércio eletrônico. Seu principal ponto negativo é a falta de uma definição de como um serviço WSMO é mapeado para uma especificação concreta.

IRS-III propõe um *framework* para SWS, o qual permite que aplicações executem e descrevam semanticamente os Serviços Web. Ele evoluiu a partir do IRS-II, baseado em UPML (*Unified Problem Solving Method Development Language*), incorporando e estendendo a ontologia WSMO para que a infra-estrutura implementada permita a descrição, publicação e execução de Serviços Web Semânticos (CABRAL et al., 2006). Sua principal característica é o fato da localização e execução do serviço ser dirigida por capacidade.

SAWSDL não especifica uma linguagem para representação de modelos semânticos. Esta abordagem fornece mecanismos através dos quais conceitos de modelos semânticos, normalmente definidos fora do documento WSDL, podem ser referenciados a partir de componentes WSDL e XML Schema, através da utilização de anotações (FARREL e LAUSEN, 2007). Em SAWSDL, as anotações semânticas são atributos XML adicionados a uma descrição WSDL ou associados a um documento XML Schema, no elemento XML sendo descrito.

OWL-S consiste em um conjunto de ontologias desenvolvidas para descrever e raciocinar sobre descrições de serviços. Sua principal vantagem está na combinação da expressividade da linguagem OWL e do pragmatismo encontrado na padronização das tecnologias de Serviços Web. Além disso, sua ontologia de serviço provê o conceito “base do serviço”, o qual define como o modelo de processo na descrição OWL-S é

mapeado para uma descrição WSDL do serviço. A iniciativa OWL-S é a mais madura atualmente, devendo completar-se logo, enquanto as demais propostas ainda estão em fase inicial. A abordagem OWL-S é descrita em maiores detalhes na seção seguinte.

2.2.4 A Abordagem OWL-S

OWL-S é uma ontologia *top level* que utiliza OWL para definir um conjunto de classes e propriedades específicas para a descrição dos Serviços Web (ALESSO, 2005). Ela fornece aos serviços um conjunto *core* de construtos de linguagem de marcação (classes e propriedades específicas) para descrever as propriedades e capacidades dos Serviços Web. Tais descrições facilitam a divulgação automatizada e a descoberta de serviços por agentes inteligentes também de forma automatizada.

A ontologia OWL-S apresenta três subontologias inter-relacionadas: o perfil do serviço (*Service Profile*), o modelo do serviço ou modelo do processo (*Service Model / Process Model*) e a base do serviço (*Service Grounding*). Estas subontologias estão relacionadas ao conceito serviço, como demonstrado na figura 2.2. O conceito serviço serve como ponto de referência organizacional para declarar Serviços Web. Cada serviço é declarado através da criação de uma instância do conceito serviço. O perfil do serviço é a descrição das ofertas e dos requerimentos do serviço. O modelo do serviço descreve como o serviço funciona. Já a base do serviço especifica os detalhes de como um agente pode acessar o serviço. O conceito serviço relaciona-se com o perfil, modelo e base de um dado serviço através das propriedades ‘apresenta’, ‘descrito por’ e ‘dá suporte’, respectivamente.

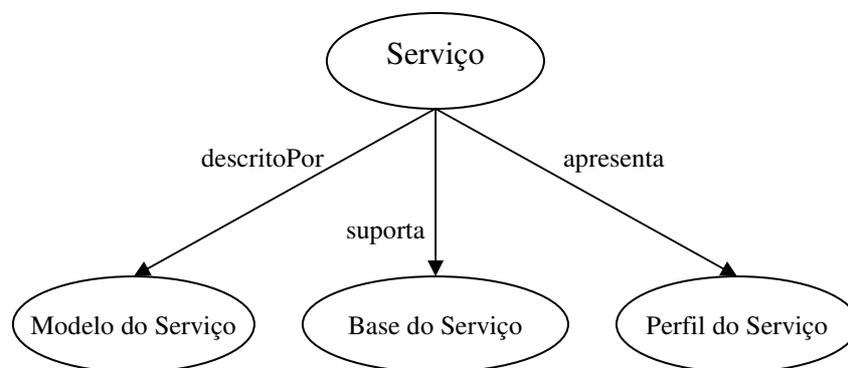


Figura 2.2: A ontologia de serviço OWL-S

OWL-S fornece uma linguagem de descrição importante para Serviços Web Semânticos, tendo como objetivo especificar as funções (pré-condições e efeitos) de uma operação e os tipos semânticos para cada entrada e saída do serviço. Ela propõe estender a capacidade da linguagem OWL para formar a ontologia exigida para os Serviços Web, tornando-os compreensíveis por máquina e, ao mesmo tempo, dando suporte a automação da descoberta, execução, composição, interoperação e monitoramento da execução do Serviço Web.

O perfil do serviço, na especificação OWL-S, fornece uma descrição alto-nível de um serviço e de seu provedor, sendo utilizada por registros de descobertas para requisitar ou divulgar serviços. Ele é descrito através das seguintes propriedades funcionais: entradas, saídas, pré-condições e efeitos. As pré-condições especificam as condições do mundo real que devem ser satisfeitas para que um agente execute um serviço. Os efeitos caracterizam os fatos que se confirmam quando um serviço é

executado com sucesso. Ele ainda inclui os atributos não-funcionais, tais como: uma descrição textual contendo o nome do serviço, etc; e uma especificação das funcionalidades do serviço.

O modelo do serviço permite que ele seja visto como um processo. Através do modelo, o serviço é descrito em termos de suas entradas, saídas, pré-condições, efeitos e composição dos subprocessos componentes. Uma característica importante da abordagem OWL-S é a sua habilidade de modelar processos compostos (ELENIUS et al., 2005). Um processo composto é formado por dois ou mais subprocessos. O controle de fluxo é definido através da utilização de construtos de controle, tais como: *'If-Then-Else'*, Seqüência e *'Repeat-Until'*. O modelo do serviço é utilizado tanto para raciocínio sobre possíveis composições como para o controle da execução de um serviço.

A base OWL-S/WSDL utiliza classes OWL para definir os tipos abstratos das partes de mensagens declaradas em WSDL. A principal função da base OWL-S é especificar como as entradas e saídas de um processo atômico são entendidas como mensagens em um formato transmissível (ALESSO e SMITH, 2005). Um processo atômico em OWL-S corresponde a uma operação WSDL.

Tanto o perfil como o modelo são representações abstratas do serviço. A base do serviço é quem lida com o nível de especificação concreto. Essa classe define os detalhes de como acessar o serviço, incluindo protocolos e formatos de mensagem, serialização, transporte e endereçamento. As descrições OWL-S relacionam-se com arquivos WSDL através desta classe base. Ela pode ser vista como o mapeamento de uma especificação abstrata para uma concreta.

Pelo fato da ontologia de serviço OWL-S ser pública e não exigir uma estrutura de implementação, ela tem sido utilizada como o ponto inicial de esforços individuais nas pesquisas envolvendo Serviços Web Semânticos.

3 O SISTEMA DE INTEGRAÇÃO W-KIS

Atualmente, a Web é o principal meio de acesso a informações nos mais diversos domínios de aplicação. Entretanto, a tarefa de busca de informação na Web ainda não é realizada da melhor forma possível. Sempre que usuários necessitam buscar alguma informação específica de um determinado domínio através da Web, eles precisam acessar um ou mais portais Web, fazendo consultas em cada um, até que uma resposta satisfatória seja fornecida. Uma segunda opção é a utilização de motores de busca tradicionais. Porém, os usuários frequentemente enfrentam um dos seguintes problemas quando realizam sua consulta através destes motores: ou eles não recebem resposta alguma, ou recebem um grande volume de resultados irrelevantes. Além dos usuários da comunidade em geral, usuários que trabalham com áreas de pesquisas específicas, tais como a Biomedicina e a Geologia, necessitam das fontes de informação disponíveis sempre que eles precisam encontrar alguma definição ou descrição específica. Portanto, um acesso integrado aos portais de interesse facilitaria a tarefa deste usuário, fornecendo uma única ferramenta onde toda a informação disponível pode ser acessada com apenas uma única consulta.

Imaginemos um usuário, no contexto da Medicina, que esteja interessado em adquirir informação sobre alguma proteína específica. Imaginando que existe um portal Web que fornece informação sobre proteínas através de um motor de busca e outro portal Web que fornece informação sobre relações entre proteínas, também através de um serviço de busca. Seria de grande utilidade para este usuário ter a possibilidade de obter uma resposta única integrada, incluindo informação tanto sobre a definição da proteína buscada, como sobre seus relacionamentos. A precisão da busca seria ainda maior se o sistema fosse capaz de focar a busca apenas nos portais com contexto relacionado ao problema do usuário. Com o objetivo de tornar isto possível, o presente trabalho propõe o desenvolvimento de um sistema de integração, denominado W-KIS (*Web Knowledge Integration System*), capaz de atuar em mais de um domínio de aplicação, com o objetivo de oferecer ao usuário um acesso integrado a dados e processos oferecidos pelos portais Web capazes de atender as suas necessidades.

Alguns portais Web apresentam apenas informações estáticas, disponíveis em forma de *thesaurus* ou bases de dados *online*. Porém, existem outros portais que fornecem serviços tais como mecanismos de busca na Web. O sistema W-KIS, inicialmente proposto em (DE MELLO et al., 2007), pretende integrar não só a informação estática presente em alguns portais, mas também os resultados fornecidos por cada serviço de busca individual (quando disponibilizado pelo portal), fornecendo uma interface de consulta uniforme para as diversas fontes de informação. O sistema deve ser capaz de encontrar as fontes que possuem a informação requerida, extrair a informação desejada destas páginas e permitir o acesso aos serviços de busca oferecidos por estas fontes.

A principal funcionalidade esperada é que o sistema de integração atue como um portal de busca baseado na intenção do usuário. A aplicação deve ser capaz de responder a consulta do usuário com informações relevantes para o domínio no qual ela está atuando. Com o objetivo de validação, o sistema foi desenvolvido para atender dois domínios de aplicação: Biomedicina e Geologia, como dois estudos de caso independentes. Na tarefa de integração, foram utilizados portais contendo informações sobre genes, proteínas e suas relações, no domínio da Biomedicina; e sobre minerais e rochas, no domínio da Geologia.

O desenvolvimento completo do sistema fez uso das tecnologias e conceitos detalhados no capítulo anterior. Um Serviço Web foi desenvolvido para encapsular cada portal Web a ser integrado, fornecendo acesso aos seus dados e processos. Entretanto, para que o motor de busca distribuído entenda qual informação está disponível nas fontes, é necessário dispor da descrição do conteúdo dos serviços, de forma significativa. E, além disso, para fazer uso do Serviço Web de forma automatizada, um agente de software requer que a descrição do serviço e dos meios de acessá-lo seja uma descrição formal, isto é, que ela esteja disponível de forma processável por computador.

Para tornar isto possível, os Serviços Web desenvolvidos devem ser enriquecidos com descrições semânticas formais, com base em uma ontologia de domínio. Para atingir este objetivo, a tecnologia de Serviços Web Semânticos foi o recurso utilizado para prover as descrições semânticas aos SW. Eles podem ser vistos como sendo análogos a *wrappers*, encapsulando as fontes de informação (portais Web) e expondo suas funcionalidades de uma forma processável pelo sistema de integração. Os SWS fazem uso de uma ontologia de domínio para padronizar formalmente o vocabulário das diversas fontes conectadas ao sistema. Através do encapsulamento de cada fonte de informação por Serviços Web Semânticos, cada fonte é diretamente relacionada a uma ontologia global compartilhada. Desta forma, o sistema é capaz de fornecer uma interface uniforme para o usuário realizar suas consultas. Esta abordagem de integração pode ser observada na figura 3.1.

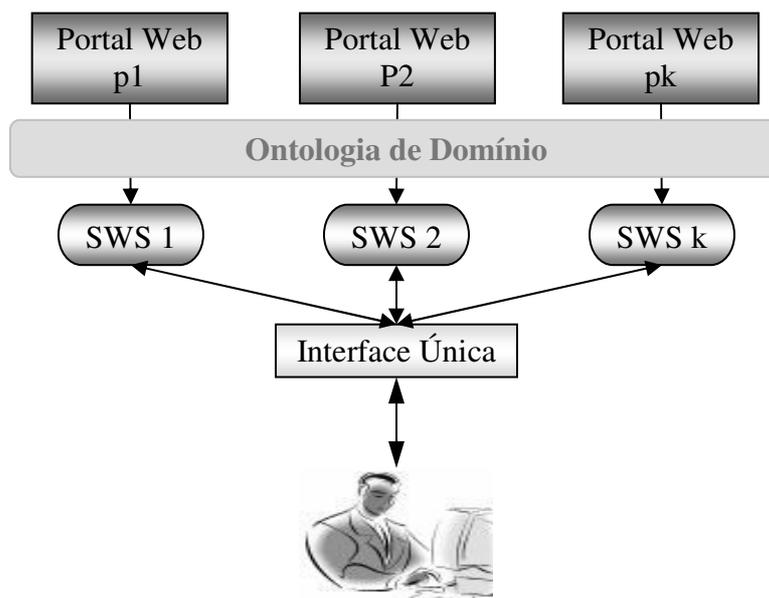


Figura 3.1: Abordagem de integração

Dispondo dos termos utilizados nas descrições definidos formalmente em uma ontologia, o sistema passa a ter acesso ao significado dos termos nas descrições e, dessa forma, torna-se capaz de decidir quais fontes são relevantes para uma consulta particular. Estas tarefas passam então a serem executadas pelo sistema de forma automatizada.

A proposta adota a abordagem onde uma única ontologia é utilizada como um modelo de referência global no sistema. Todas as fontes de informação são semanticamente descritas baseadas na mesma ontologia de domínio. Além disso, a ontologia de domínio é utilizada como esquema de consulta global, ou seja, a consulta do usuário também é mapeada para esta mesma ontologia. A ontologia de domínio auxilia usuários a expressarem seus objetivos e provê a integração da informação através dos Serviços Web Semânticos.

Esse trabalho faz parte do projeto Alfa da Comunidade Européia, que se dedica ao domínio biomédico, e do Projeto *PetroGrapher*, que utiliza sistemas de inteligência artificial no domínio da Petrografia Sedimentar. Ambos os projetos apresentam uma ontologia de domínio como parte integrante de seus sistemas, permitindo que a terminologia (tanto biomédica como geológica) seja padronizada. O processo de integração nos estudos de caso é realizado com base nestas ontologias as quais resolvem o problema da heterogeneidade dos dados nas diferentes fontes de informação pertencentes aos sistemas.

3.1 Domínios de Aplicação

No presente trabalho são abordados dois domínios de aplicação: o domínio da Biomedicina e o domínio da Geologia, cada um provendo sua própria ontologia de domínio. Assume-se que a ontologia seja capaz de capturar não apenas o significado dos principais termos do domínio, mas também o contexto no qual esses termos são aplicados, de certa forma mostrando a intenção do usuário.

O domínio da Biomedicina foi utilizado como um estudo de caso parte de um projeto dedicado a Bioinformática, a qual é uma área de pesquisa que envolve estudos integrando dados e informações pertencentes às áreas da biologia, medicina e de sistemas de informação. Já o domínio da Geologia foi utilizado como um estudo de caso associado ao projeto *PetroGrapher*, o qual se dedica ao suporte de análises petrográficas e interpretações de rochas de reservatórios de petróleo,

O objetivo de realizar estudos em dois domínios distintos é identificar de que forma ontologias de domínio, através dos SWS, facilitam a integração de informação na Web e formalizar esse método de forma independente de domínio. Desta forma, é possível avaliar a capacidade de integração de informação do sistema proposto independentemente do domínio no qual ele é aplicado. Além disso, a possibilidade de integração de diferentes fontes de dados tem grande importância sobre os domínios da Biomedicina e da Petrografia Sedimentar, pois existe uma enorme quantidade de informação armazenada em mais de uma base, o que dificulta a sua interpretação.

3.1.1 Biomedicina

Com os avanços na área da Biologia e Genética, houve um grande aumento na quantidade de informação disponível relacionada ao domínio biológico. Isto fez surgir a necessidade da utilização de tecnologias de informação para coletar, armazenar, analisar

e combinar dados e informações biológicas. A Bioinformática consiste na aplicação de tecnologias de informação para prover os recursos e ferramentas necessárias para promover a pesquisa biomédica. Seu principal objetivo é a pesquisa e o desenvolvimento de sistemas úteis para o entendimento do fluxo de informação dos genes para as estruturas moleculares.

O primeiro estudo está inserido no projeto "*Advanced Methods and Technologies for Bioinformatics*", do programa Alfa da Comunidade Européia, que se dedica ao domínio biomédico. Este projeto conta com a participação das seguintes instituições: Universidad de Murcia (Espanha), University of Ulster (Reino Unido), Université de Rennes I (França), Universidad de Los Andes (Venezuela), Universidad de Talca (Chile), Universidad Nacional de La Plata (Argentina) e Universidade Federal de Rio Grande do Sul (Brasil), através do PPGC.

Biomedicina é a ciência que abrange estudos e pesquisas no campo interdisciplinar entre a Medicina e a Biologia. Esta área está voltada para a pesquisa de doenças humanas, com o objetivo de encontrar suas causas, maneiras de prevenção, diagnóstico e tratamento. O estudo da função de cada gene humano demonstra como genes defeituosos estão relacionados com certas doenças. Uma das áreas de pesquisa da Biomedicina é justamente o estudo dos genes e seus polimorfismos para que, posteriormente, sua relação com as doenças possa ser derivada.

Genes são os elementos fundamentais ligados à hereditariedade. Um gene é um segmento de um cromossomo que corresponde a uma informação para produzir uma determinada proteína ou controlar uma característica. Já as proteínas são os compostos químicos que servem de material para a formação de partes da célula, a transformação de um elemento químico em outro, servir de meio de comunicação entre células, etc.

O grupo de genes implicados no desenvolvimento de câncer são genes ditos danificados, conhecidos como "oncogenes". Oncogenes são os genes que, quando aparecem em células normais, podem contribuir para o desenvolvimento do câncer, dando instruções para que as células produzam proteínas que estimulam a divisão e o crescimento excessivo das células. Logo, os principais conceitos envolvidos neste estudo são: genes e proteínas.

No domínio da Biomedicina, a ontologia *Oncogen* (desenvolvida pelo grupo de pesquisa do laboratório KLT em Murcia, na Espanha) é utilizada para realizar a padronização da terminologia usada pelas fontes a serem integradas. Esta ontologia, que foi desenvolvida como parte do projeto, disponibiliza a terminologia biológica para a oncologia genética. Ela tem como objetivo dar suporte a consulta integrada sobre inúmeras aplicações Web ou fontes de informação biológicas heterogêneas. Esta ontologia foi desenvolvida com base em pesquisa apenas (imersão na literatura), sem o envolvimento de um especialista na área e, portanto, sem ser validada.

3.1.2 Geologia

Já o segundo estudo faz parte do projeto *PetroGrapher* (ABEL, 2001 e 2004) dos Institutos de Informática e Geociências da UFRGS. Este projeto possui como objetivo principal o desenvolvimento de um sistema de banco de dados inteligentes voltado à tarefa de análise de rochas sedimentares com o objetivo de dar suporte à avaliação de reservatórios de petróleo. Este estudo tem como área de aplicação a Petrografia Sedimentar, dentro do domínio da Geologia.

A Geologia é a ciência que estuda a Terra, sua composição, estrutura, propriedades físicas e os processos que lhe dão forma. Já a Petrografia Sedimentar é uma subárea de pesquisa da Geologia que estuda amostras de rochas extraídas de testemunhos de poços ou áreas de exploração de petróleo, com a finalidade de determinar, a partir da análise de uma pequena porção, as propriedades de um reservatório de petróleo (ABEL, 2001). Esse estudo tem como objetivo investigar o potencial econômico de reservatórios de petróleo, através da medição da qualidade desses reservatórios.

A qualidade de um reservatório de petróleo é determinada pelas características originais da rocha-reservatório (rocha porosa que retém o petróleo). É através da petrografia, uma atividade desenvolvida em laboratório, que são realizadas as diferentes análises sobre as rochas e suas características. A análise das amostras de rochas é realizada por meio da descrição detalhada das feições da rocha, gerada por um geólogo, contendo informações importantes sobre todos os aspectos da rocha. Entre outros dados mais sofisticados, são descritos, individualmente, os constituintes que formam a rocha, incluindo minerais, fósseis e poros. A classificação de uma rocha depende fortemente da classificação dos minerais que existem na rocha.

Devido a sua complexidade, a geologia do petróleo tem sido constante objeto de pesquisa em Inteligência Artificial nas últimas décadas. Uma das principais dificuldades encontradas neste domínio está no fato do conhecimento necessário para a solução de problemas ser restrito aos especialistas, devido às dificuldades inerentes aos problemas existentes. Um dos grandes desafios nesta área é, portanto, a captura, a padronização e a representação deste conhecimento, de forma processável por computador. A Engenharia de Conhecimento, através das ontologias, fornece o melhor meio para disponibilizar formalmente o modelo do conhecimento para este domínio.

A ontologia definida no domínio da petrografia de rochas sedimentares foi parcialmente representada no modelo do conhecimento do Sistema *PetroGrapher* para dar suporte a captura de dados e ao raciocínio. Esta ontologia representa o vocabulário aplicado por geólogos para realizar a descrição de amostras de rocha durante a análise para avaliação de reservatórios de petróleo, permitindo que a terminologia geológica seja padronizada e que esteja disponível para o usuário para todos os aspectos que podem ser descritos sobre uma amostra. Ela foi modelada a partir do conhecimento eliciado de um geólogo petrógrafo especialista e posteriormente desenvolvida e validada com a assistência do mesmo profissional. Para atender a aplicação deste trabalho e dar maior suporte na contextualização das consultas, a ontologia do Projeto *PetroGrapher* foi estendida e conceitos mais genéricos foram incluídos.

3.2 Processo de Desenvolvimento

O processo de desenvolvimento do sistema de integração W-KIS, o qual é baseado em Serviços Web Semânticos, segue as seguintes etapas (observadas na figura 3.2):

1. **Especificação de Requisitos.** Esta etapa compreende a coleta e formalização das informações essenciais sobre cada domínio de aplicação; a definição das funcionalidades esperadas; e a especificação do tipo de informação que deve ser integrada em cada domínio.
2. **Design dos Dados.** Tendo como base a análise realizada na especificação dos requisitos, o design dos dados é subdividido em duas etapas: a especificação da ontologia de domínio, utilizando a linguagem OWL; e a análise e seleção dos

portais Web a serem utilizados na aplicação. A especificação da ontologia pode consistir no desenvolvimento de uma nova ontologia, ou na reutilização de uma ontologia já existente (com as devidas modificações, se necessário). A seleção dos portais a serem integrados no sistema é realizada manualmente com base nos requisitos definidos na primeira etapa e na ontologia de domínio previamente especificada.

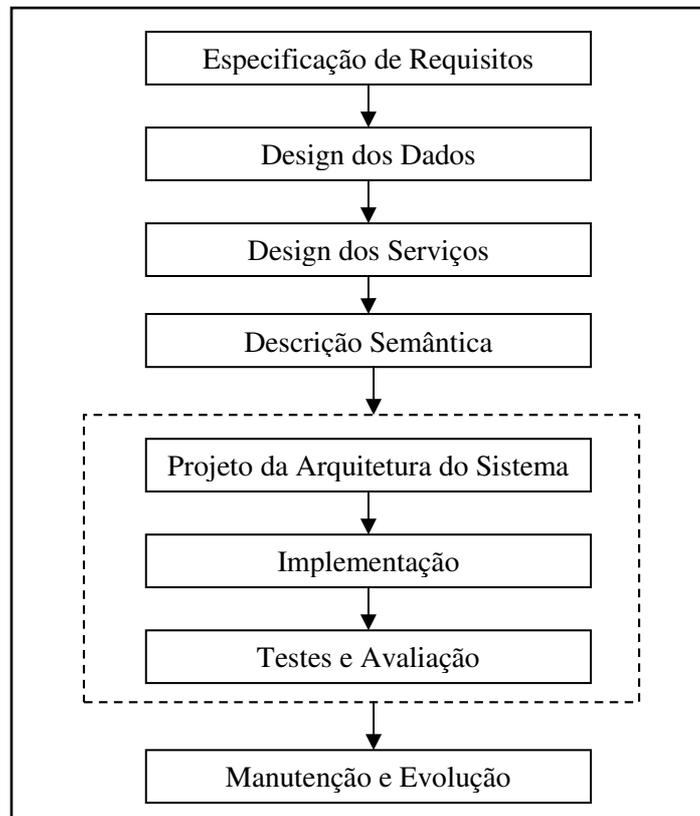


Figura 3.2: Processo de desenvolvimento da aplicação

3. **Design dos Serviços.** Nesta etapa, cada portal Web selecionado deve ser encapsulado por um Serviço Web, com o objetivo de expor as funcionalidades fornecidas pelas fontes de informação selecionadas.
4. **Descrição Semântica.** A descrição semântica consiste no enriquecimento de cada Serviço Web modelado com uma descrição formal rica de suas capacidades. Cada anotação semântica é realizada utilizando a abordagem OWL-S, descrita na seção 2.2.4.
5. **Projeto da Arquitetura do Sistema.** Esta etapa corresponde à especificação dos componentes que irão fazer parte do sistema e como eles serão interligados. Fazem parte desta etapa: a definição de um meio para acessar as descrições semânticas dos Serviços Web, através de um mecanismo de consulta; o desenvolvimento de um mecanismo para executar os serviços selecionados pelo mecanismo de consulta; e o design de uma interface para interação com o usuário final.
6. **Implementação.** Com a arquitetura do sistema definida, os componentes devem ser implementados e devidamente conectados, formando o sistema completo.

7. **Testes e Avaliação.** A validação do sistema deve ser feita com a realização de um estudo de caso em cada domínio de aplicação, comparando os resultados obtidos em cada um. Além disso, deve ser realizada uma comparação do comportamento do sistema com outras propostas de integração e mecanismos de busca na Web, para que a performance do sistema possa ser avaliada e possíveis falhas possam ser detectadas e corrigidas.
8. **Manutenção e Evolução.** O sistema deve permitir que novos portais Web sejam adicionados ao conjunto de fontes de informação, bem como a remoção de portais deste conjunto. Além disso, o sistema deve ser capaz de lidar com alterações nos conteúdos das fontes. Para isto, é necessário manter a manutenção do sistema.

As etapas 1-4 são específicas para cada domínio de aplicação utilizado, ou seja, elas devem ser realizadas toda vez que um novo domínio for selecionado para ser aplicado no sistema. A etapa 8 influencia apenas as tarefas 3 e 4, as quais devem ser realizadas cada vez que uma fonte for adicionada ao sistema, ou quando algum serviço oferecido por alguma fonte for alterado. Os componentes da arquitetura já implementados não precisam sofrer alteração alguma.

4 DESCRIÇÃO DA APLICAÇÃO

Este capítulo apresenta a descrição detalhada de cada componente do sistema de integração W-KIS. Tomando como base as etapas do processo de desenvolvimento do sistema descritas no capítulo anterior, primeiramente serão apresentados os modelos de domínio especificados como a primeira parte da tarefa de Design dos Dados. Em seguida, são descritas as fontes de informação, selecionadas como a segunda parte desta tarefa, juntamente com os serviços desenvolvidos (na etapa do Design dos Serviços) para encapsulá-las. Na seqüência, as descrições semânticas realizadas na tarefa da Descrição Semântica são introduzidas. Por fim, a arquitetura completa do sistema W-KIS, definida e implementada nas etapas de Projeto da Arquitetura do Sistema e de Implementação, é apresentada, detalhando seus componentes centrais.

4.1 Modelos de Domínio

Ontologias de domínio são utilizadas no sistema W-KIS para capturar o conhecimento de domínios particulares, fornecendo uma descrição detalhada dos conceitos em cada domínio e de seus relacionamentos. Com a utilização de ontologias, as fontes de informação no sistema podem compartilhar conceitos comuns ou realizar mapeamentos entre conceitos distintos. Além disso, através desta facilidade, os usuários formulam consultas orientados pela ontologia e o sistema tem a responsabilidade de gerenciar a heterogeneidade e a distribuição nas diferentes bases.

Para que uma ontologia seja definida é necessário primeiro capturar o conhecimento sobre os conceitos do domínio e sobre como eles se organizam. O engenheiro de conhecimento deve inicialmente compreender o domínio através de diversas fontes (agentes do domínio, especialistas, documentos, etc) e estruturar o conhecimento através de um modelo formal. Os modelos de domínio utilizados no presente trabalho representam os principais conceitos relacionados a cada domínio (Biomedicina e Geologia) representados através de uma hierarquia de classes e subclasses, seus atributos e a forma como os conceitos relacionam-se uns com os outros.

A ontologia utilizada no domínio da Biomedicina, denominada *Oncogen*, é uma ontologia de domínio para a oncologia genética: genes implicados no desenvolvimento de câncer. Ela é uma ontologia para terminologia biológica, a qual fornece um modelo de conceitos biológicos que podem ser utilizados para formar um *framework* semântico para tarefas tais como o armazenamento, recuperação e análise de dados no referente domínio. A ontologia *Oncogen* foi implementada em OWL, através do editor de ontologia Protégé. Seus principais conceitos podem ser visualizados na figura 4.1. A descrição da ontologia completa, listada textualmente para auxiliar a compreensão, encontra-se no Anexo A.

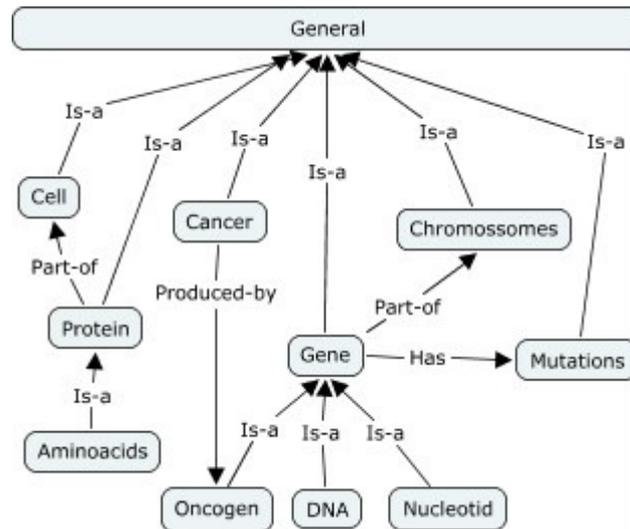


Figura 4.1: Principais conceitos da ontologia *Oncogen*

A segunda ontologia deste trabalho, aplicada no domínio da Geologia, é um refinamento de uma ontologia pré-existente, a qual faz parte do projeto *PetroGrapher* (ABEL, 2001 e 2004). Ela foi previamente modelada como uma partonomia de termos, seguindo a metodologia CommonKADS (SCHREIBER et al., 2000) e, no presente trabalho, ela foi traduzida para OWL, também através da utilização do editor de ontologia Protégé. Esta ontologia, denominada *PetroGrapher*, representa o vocabulário na área de Petrografia Sedimentar. Seus principais conceitos são demonstrados na figura 4.2. A descrição da ontologia completa, também lista textualmente para uma melhor visualização, encontra-se no Anexo B.

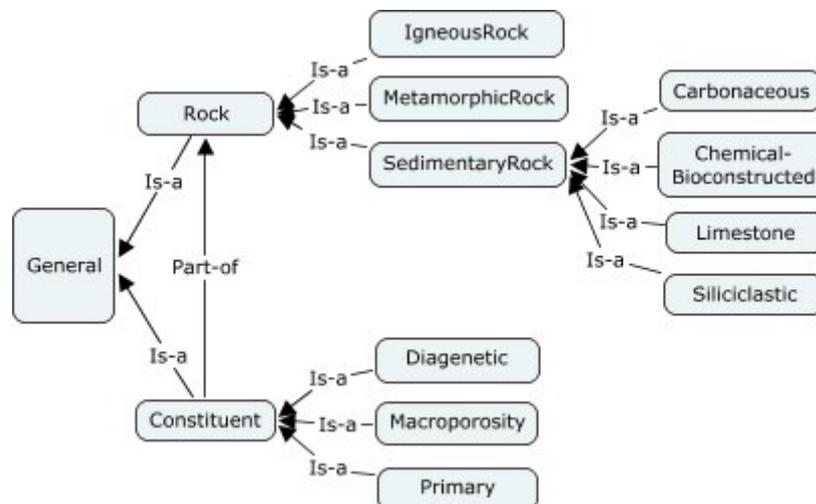


Figura 4.2: Principais conceitos da ontologia *PetroGrapher*

Cada parâmetro de entrada e de saída das operações oferecidas pelos portais Web utilizados como fontes de informação no sistema deve ser validado através da correspondência com um conceito da ontologia de domínio. Desta forma, o sistema não apenas reconhece os parâmetros como *strings*, *integers*, ou qualquer outro tipo básico de dados, mas também como conceitos relacionados com as classes da ontologia, tendo um significado associado, o qual pode ser interpretado de forma automática pelo sistema. Essa capacidade de tratar conteúdo semântico e de interpretar a intenção do usuário a

partir dos conhecimentos que ele utiliza, mostra-se como o diferencial do sistema W-KIS.

4.2 Fontes de Informação

Com base nos modelos de domínio previamente definidos, foram selecionadas diferentes fontes de informação, através do estudo da ontologia que descreve o domínio de aplicação e da análise dos distintos portais Web. Na análise de cada portal, foi determinada a viabilidade de acesso aos dados por meio de Serviços Web e, confirmada a viabilidade, foram definidas as funcionalidades oferecidas pelo portal a serem disponibilizadas no sistema.

A idéia foi selecionar manualmente alguns portais Web que fornecem informação sobre genes, proteínas e suas relações, no domínio da Biomedicina; e sobre minerais e rochas, no domínio da Geologia. Alguns portais web são simplesmente interfaces na Web para bases de dados, enquanto outros oferecem serviços mais sofisticados tais como motores de busca. Nove portais foram analisados no domínio da Biomedicina e doze no domínio da Geologia. Os portais primeiramente selecionados foram aqueles que disponibilizam mecanismos de busca ou glossários com definições e descrições de interesse para cada domínio. Em seguida, foram descartados os portais que não permitiram o acesso às suas funcionalidades por meio de Serviços Web. Dos portais Web analisados, quatro foram utilizados como fontes de informação no domínio da Biomedicina e cinco foram utilizados no domínio da Geologia. As fontes selecionadas são descritas a seguir.

Dentre as fontes ligadas à Biomedicina, está o projeto *Gene Ontology*² o qual provê um vocabulário controlado para descrever atributos de genes em qualquer organismo. O portal *Gene Ontology* oferece um motor de busca o qual permite a consulta de genes e proteínas através de seu nome ou de seu ID na base de dados. Outra fonte de informação é a base de dados *PubMed*³, disponível através do sistema de recuperação de informação NCBI Entrez, a qual fornece um motor de busca desenvolvido pela organização NCBI (*National Center for Biotechnology Information*), na biblioteca nacional de medicina (NLM – *National Library of Medicine*). O motor de busca deste portal Web permite a consulta de genes e proteínas, dentre outros conceitos. Já a terceira fonte, o portal *MPact*⁴, fornece um ponto de acesso comum às fontes de interação em MIPS (*The Munich Information Center for Protein Sequences*), disponibilizando recursos relacionados a informação sobre genoma. *MPact* fornece ao usuário formulários de consulta intuitivos para recuperar rapidamente tanto descrições de proteínas como as interações de interesse entre elas. O último portal, a base de conhecimento sobre proteína *Swiss-Prot*⁵, é uma base de dados de seqüências de proteínas, a qual pretende fornecer um alto nível de anotação, o mínimo de redundância e um alto nível de integração com outras bases de dados. Este portal apresenta um motor de busca para consultar descrições de proteínas.

² <http://www.geneontology.org/>

³ <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed>

⁴ <http://mips.gsf.de/genre/proj/mpact/>

⁵ <http://www.expasy.ch/cgi-bin/spot-search-ful>

As fontes de informação no domínio da Geologia incluem o portal *Schlumberger Oilfield Glossary*⁶, uma referência instantânea que oferece definições precisas para grande parte das atividades relacionadas à área do petróleo, revisadas por especialistas técnicos. Neste portal Web, existe um motor de busca e um glossário os quais permitem consultar definições de diversos termos da área da Petrologia. A segunda fonte é o portal *Geologynet*⁷, utilizado através de duas de suas bases de dados disponíveis na Web: uma base de dados sobre minerais e uma base de dados sobre rochas. Elas apresentam uma interface de busca, fornecendo acesso a uma coleção de bases de dados de rochas e minerais as quais podem ser utilizadas para educação e pesquisa. A terceira fonte de informação, denominada *Athena*⁸, é uma base de dados contendo uma grande quantidade de informações sobre minerais. Este portal disponibiliza um motor de busca que permite a consulta por minerais através de seu nome ou parte do nome. O quarto portal utilizado é o *Mindat*⁹, a maior base de dados sobre minerais e a maior página Web de referência mineralógica na Internet. Esta página contém dados mundiais sobre minerais, coleta de minerais, localização de minerais e outras informações mineralógicas. Este portal apresenta um índice em ordem alfabética de minerais, variedades, sinônimos e outras nomenclaturas, além de um motor de busca por descrições de minerais. A quinta e última fonte de informação utilizada, o portal *WebMineral*¹⁰, é outra base de dados sobre minerais contendo 4.442 descrições de espécies de minerais. Este portal Web apresenta tanto um motor de busca que permite consultar a base de dados mineralógica, como um glossário, listando, em ordem alfabética, as espécies de minerais.

O problema da seleção de fontes de informação na Web é análogo ao problema da descoberta de conteúdo na Web. A seleção dos portais utilizados como fontes de informação no sistema W-KIS foi realizada manualmente. No cenário da Web atual, ainda existem limitações para automatizar por completo este processo. Um pré-requisito para que esta seleção seja realizada de forma automatizada é o desenvolvimento e a evolução da Web Semântica. No cenário da Web Semântica, um modelo de domínio seria definido para especificar tanto o conteúdo desejado do portal como o tipo de funcionalidade desejada. Tendo o objetivo da seleção formalmente definido (através de uma ontologia) e o conteúdo dos portais na Web também formalmente definido (através da infra-estrutura proporcionada pela Web Semântica), a seleção das fontes poderia ser realizada de forma automática por agentes inteligentes. Entretanto, como o foco do trabalho foi atribuir contexto a buscas através do uso de ontologias, essa alternativa não foi aqui explorada.

Neste trabalho, um Serviço Web foi desenvolvido para encapsular cada portal Web selecionado, expondo suas funcionalidades e fornecendo acesso à sua informação e às suas operações. Se alguma operação disponibilizada pelo portal Web não é encapsulada através do Serviço Web, ela não estará disponível para o usuário do sistema de integração. Os serviços implementados são Serviços Web simples devido ao fato de eles não dependerem de outros Serviços Web, nem exigirem interação adicional com o

⁶ <http://www.glossary.oilfield.slb.com/default.cfm>

⁷ <http://geologynet.com/indexa.htm>

⁸ <http://un2sg4.unige.ch/athena/mineral/mineral.html>

⁹ <http://www.mindat.org/index.php>

¹⁰ <http://webmineral.com/>

usuário. Eles permitem o acesso tanto aos dados dos portais como aos processos disponibilizados por eles.

O desenvolvimento de cada Serviço Web foi realizado através de um processo semi-automatizado. Primeiramente, o portal Web foi modelado como uma classe comum em uma linguagem de programação orientada a objetos, onde cada método corresponde a uma operação do portal. Uma classe base genérica foi desenvolvida na qual o desenvolvedor apenas necessita especificar, para operação de cada portal, três parâmetros, os quais denotam as informações específicas para acessar cada portal (visto que os portais Web estão disponibilizados na Internet, através de páginas HTML): o endereço da página; o método de acesso, o qual pode ser GET ou POST; e a localização da resposta desejada na página. Posteriormente, o arquivo *.aar*, o qual denota o Serviço Web, juntamente com o arquivo WSDL são gerados automaticamente a partir da classe desenvolvida, através da biblioteca *Apache Axis2* (a ferramenta mais difundida atualmente para Serviços Web (MARTIN et al., 2005)), permitindo a especificação da sintaxe das mensagens de entrada e saída, assim como outros detalhes necessários para a execução do serviço.

O serviço que fornece acesso ao motor de busca do portal *Mindat* (figura 4.3), por exemplo, consiste em um método (ou operação): *getMineral*, o qual encapsula a ferramenta de busca “*Mineral Search*” disponibilizada pelo portal, retornando a descrição de um mineral, dado o seu nome. Ou seja, o parâmetro de entrada é o nome de algum mineral (*string*), e o parâmetro de saída é a descrição do devido mineral (outra *string*). Uma parte do arquivo WSDL que descreve as capacidades do serviço *Mindat* é demonstrada na figura 4.4. O arquivo WSDL inteiro está disponível no Apêndice A.

The screenshot shows the Mindat.org website interface. At the top, there is a navigation bar with links for Home Page, Log In, Message Board, Register, Search Pages, and Chat Now! (2 online). Below this, the main content area features a 'Mineral Search' section with a text input field labeled 'Enter name, or part of name:' and a 'Search' button. A red box highlights the search input field and button, with a red arrow pointing to the search button. To the right of the search bar, the date 'January 21, 2008' and a link 'Free Mineral Books! - Click Here' are visible. The main content area is divided into several sections: 'Welcome to mindat.org' with a description of the site, 'New discussions' with a list of recent forum posts, and 'Mindat News' with a list of recent news items. On the left side, there are links for 'Search Pages' and 'Other Pages'.

Figura 4.3: Portal Web *Mindat*

```

- <wsdl:definitions targetNamespace="http:// www.inf.ufrgs.br/~mtmello ">
  <wsdl:documentation>Mindat</wsdl:documentation>
+ <wsdl:types></wsdl:types>
+ <wsdl:message name="getMineralMessage"></wsdl:message>
+ <wsdl:message name="getMineralResponseMessage"></wsdl:message>
+ <wsdl:portType name="MindatPortType"></wsdl:portType>
- <wsdl:binding name="MindatSOAP11Binding" type="ns:MindatPortType">
  <soap:binding transport="http://schemas.xmlsoap.org/soap/http" style="document"/>
  + <wsdl:operation name="getMineral"></wsdl:operation>
  </wsdl:binding>
+ <wsdl:binding name="MindatSOAP12Binding" type="ns:MindatPortType"></wsdl:binding>
+ <wsdl:binding name="MindatHttpBinding" type="ns:MindatPortType"></wsdl:binding>
+ <wsdl:service name="Mindat"></wsdl:service>
</wsdl:definitions>

```

Figura 4.4: Parte do arquivo WSDL do serviço *Mindat*

Apesar da descrição WSDL apresentar diversas informações importantes sobre o Serviço Web, ela sozinha não fornece informação suficiente para dar suporte à decisão de quais Serviços Web são relevantes para atender a necessidades do cliente, ou de quais operações dentro de um Serviço Web devem ser executadas. WSDL fornece apenas informações sintáticas sobre como acessar um Serviço Web. Nenhum sistema de software é capaz de ler e utilizar a interface WSDL sem a assistência humana. Isto ocorre porque a linguagem de especificação WSDL, por si só, não fornece meios de incluir representações da semântica das operações definidas ou dos elementos de mensagem associados (MARTIN et al., 2005).

Por exemplo, diferentemente do serviço *Mindat* que oferece apenas uma operação, o serviço *Geologynet* oferece duas operações: *getMineral* e *getRock*, as quais são apresentadas na figura 4.5. Um ser humano é capaz de distinguir que cada operação possui um objetivo completamente diferente da outra, visto que a primeira retorna a descrição de um mineral e a segunda retorna a descrição de uma rocha. Porém, qualquer aplicação reconhece as duas operações como iguais sintaticamente, pois ambas possuem parâmetros e resultados especificados com o mesmo tipo. Portanto, a descrição semântica dos Serviços Web disponibilizados para a aplicação presente se faz necessária justamente para que o sistema seja capaz de reconhecer operações que tenham objetivos diferentes, mesmo que estas sejam iguais sintaticamente.

<p>Operação 1 String getMineral(String mineralName)</p> <p>Operação 2 String getRock(String rockName)</p>

Figura 4.5: Operações oferecidas pelo serviço *Geologynet*

4.3 Descrições Semânticas

Como já mencionado anteriormente, a falta de uma descrição semântica associada ao significado dos parâmetros de entrada e saída na linguagem WSDL torna impossível o desenvolvimento de sistemas que, sem a assistência humana, possam encontrar dinamicamente e executar com sucesso um Serviço Web. Portanto, depois de desenvolver os Serviços Web, suas descrições semânticas devem ser disponibilizadas para tornar possível o acesso a eles de forma automatizada.

Uma vez que os Serviços Web tenham sido implementados e sua descrição WSDL tenha sido definida, sua funcionalidade é descrita semanticamente através da especificação OWL-S, de forma que eles possam ser acessados automaticamente. Tal descrição permite que a escolha do serviço que melhor atende a necessidade do usuário seja feita de forma automática. É em cima destas descrições semânticas que as consultas são realizadas.

No arquivo WSDL estão definidas todas as estruturas de dados que serão trocadas como partes de mensagens entre aplicações. A estrutura de dados suportada pela linguagem WSDL permite tipos básicos, *built-in* e tipos de dados mais complexos definidos pelo usuário (como estruturas, por exemplo). Porém, essa estrutura de dados não permite que os dados tenham qualquer tipo de conhecimento associado, ou seja, não há significado algum associado aos parâmetros.

A anotação semântica foi realizada de uma forma semi-automática, através da linguagem OWL-S, com base na ontologia de domínio escrita em OWL, utilizando o editor *OWL-S Editor*, como um *plug-in* da ferramenta Protégé. Este editor oferece um ambiente de desenvolvimento no qual as ontologias de domínio são facilmente integradas com as descrições dos serviços. O significado (a semântica) de cada parâmetro é formalmente especificado, de forma a ser processável por computador. Cada termo, juntamente com cada operação de cada serviço é rotulado com um conceito da ontologia de domínio. As figuras 4.6 e 4.7 demonstram um exemplo de como o parâmetro de entrada da operação *getMineral* do serviço *Mindat*, que deve ser um mineral, é mapeado para o conceito '*Constituent*' da ontologia de domínio *PetroGrapher* na descrição semântica, utilizando o editor *OWL-S Editor*.

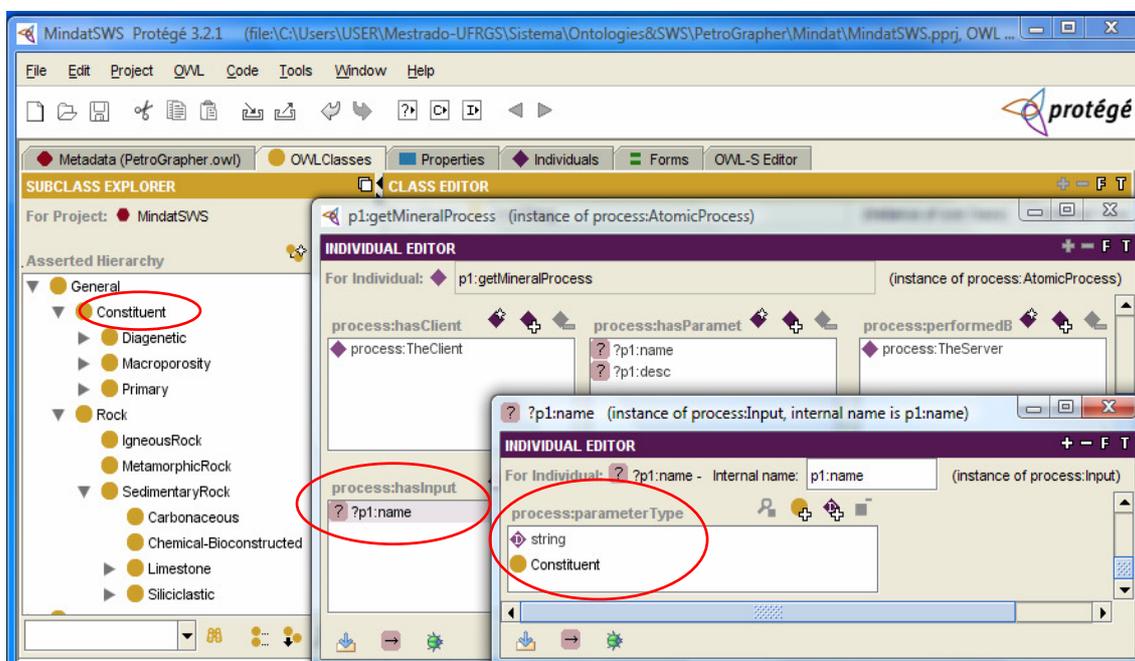


Figura 4.6: Anotação semântica através do *OWL-S Editor*

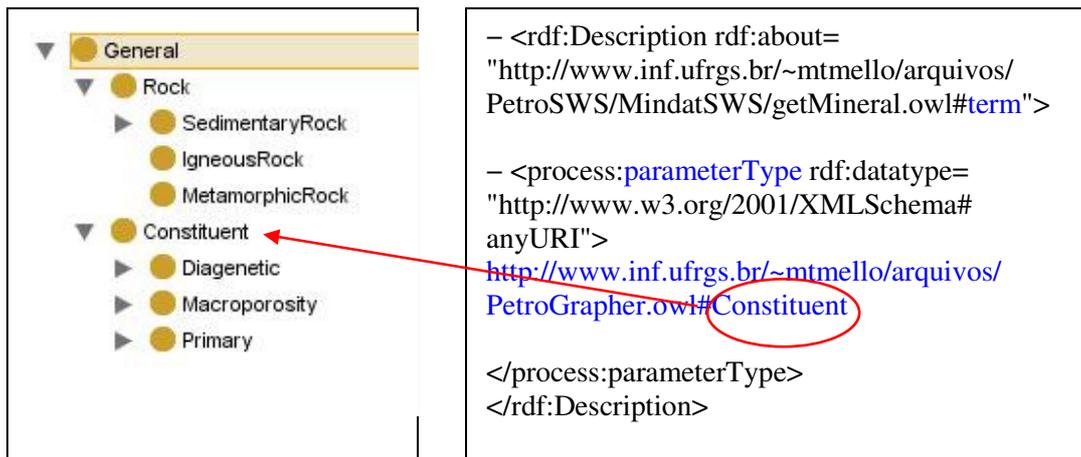


Figura 4.7: Mapeamento do parâmetro de entrada

O *OWL-S Editor* permite a geração automática de uma descrição OWL-S “esqueleto”, baseada no arquivo WSDL pré-existente. Através desta facilidade, o editor toma como entrada um arquivo WSDL e, a partir dele, gera partes da descrição OWL-S automaticamente, com base nas entradas e saídas definidas no arquivo WSDL. Posteriormente, esta descrição simples deve ser enriquecida manualmente com a parte semântica. Esta tarefa envolve a associação dos parâmetros de cada operação com as classes da ontologia de domínio, juntamente com a definição das pré-condições e efeitos. Além disso, a descrição semântica permite a especificação de todos os detalhes necessários para acessar o serviço, através da base OWL-S/WSDL na descrição OWL-S. Desta forma, o sistema é capaz de selecionar e executar os serviços automaticamente.

A descrição semântica OWL-S do serviço *Mindat* é apresentada na figura 4.8, demonstrando sua operação e seu parâmetro de entrada. Como já demonstrado anteriormente, este serviço apresenta uma operação, *getMineral*, a qual admite o nome de um mineral como único parâmetro de entrada. Porém, na descrição semântica, este parâmetro tem seu tipo definido não mais como uma string, como está no arquivo WSDL, e sim como a classe ‘Constituent’ da ontologia *PetroGrapher* (como demonstrado na figura 4.7). Parte da descrição OWL-S da operação *getMineral* e seus parâmetros é demonstrada na figura 4.9. O arquivo OWL-S inteiro está disponível no Apêndice B.

Dispondo das descrições semânticas de todos os serviços que encapsulam os portais Web conectados ao sistema, W-KIS é capaz de reconhecer automaticamente quais portais apresentam conteúdo relevante para responder a consulta do usuário. Desta forma, o sistema seleciona e executa, de forma automatizada, apenas os serviços capazes de responder a busca do usuário, dentro do contexto especificado por ele.

```

- <rdf:Description rdf:about=
  "http:// www.inf.ufrgs.br/~mtmello /ontology/PetroSWS/MindatSWS/getMineral.owl#name">
- <process:parameterType rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
  http:// www.inf.ufrgs.br/~mtmello /ontology /PetroGrapher.owl#Constituent
</process:parameterType>
</rdf:Description>

```

Figura 4.8: Parte do arquivo OWL-S do serviço *Mindat*

```

- <j.1:Service rdf:ID="getMineralService">
  <j.1:supports rdf:resource="#getMineralGrounding"/>
  - <j.1:presents>
    - <j.0:Profile rdf:ID="getMineralProfile">
      <j.0:serviceName
        rdf:datatype="http://www.w3.org/2001/XMLSchema#string">getMineral</j.0:serviceName>
      <j.0:textDescription rdf:datatype="http://www.w3.org/2001/XMLSchema#string">gets a mineral description
        from a given mineral name</j.0:textDescription>
    - <j.0:hasOutput>
      + <j.2:Output rdf:ID="description"> </j.2:Output>
    </j.0:hasOutput>
    - <j.0:hasInput>
      + <j.2:Input rdf:ID="name "> </j.2:Input>
    </j.0:hasInput>
    <j.1:presentedBy rdf:resource="#getMineralService"/>
  </j.0:Profile>
</j.1:presents>
- <j.1:describedBy>
+ <j.2:AtomicProcess rdf:ID="getMineralProcess"> </j.2:AtomicProcess>
</j.1:describedBy>
</j.1:Service>

```

Figura 4.9: Parte do arquivo OWL-S demonstrando a operação *getMineral* e seus parâmetros

4.4 Arquitetura do sistema W-KIS

Uma arquitetura foi projetada para o sistema completo, fazendo uso de todos os conceitos e técnicas especificados nas seções 2.1 e 2.2. A aplicação adota a abordagem de integração virtual na qual o usuário submete uma consulta e o sistema determina quais fontes de informação são capazes de responder a consulta. Assim que a resposta é dada pelas fontes, o sistema de integração é responsável pela organização das mesmas, excluindo as respostas nulas, e pelo retorno das respostas efetivas ao usuário final.

O sistema de integração W-KIS compreende todos os componentes ilustrados na figura 4.10: uma interface de usuário, um componente de consulta, um componente de execução e um componente *core*; além do modelo de domínio (a ontologia de domínio), os Serviços Web que disponibilizam as fontes de informação para o sistema e as descrições semânticas dos serviços. O modelo de processo do sistema é representado na figura 4.11. Esta modelagem representa as tarefas realizadas por cada componente do sistema toda vez que uma consulta é realizada, bem como o fluxo de informação entre eles.

Os componentes representados no interior da área tracejada na figura 4.10 formam o conjunto de componentes centrais da arquitetura do sistema de integração. Eles foram projetados para serem utilizados de forma independente do domínio de aplicação. Ou seja, depois de implementados, eles podem ser reutilizados em mais de um domínio sem a necessidade de alterações ou especificações. Estes componentes serão apresentados em maiores detalhes nas seções subseqüentes.

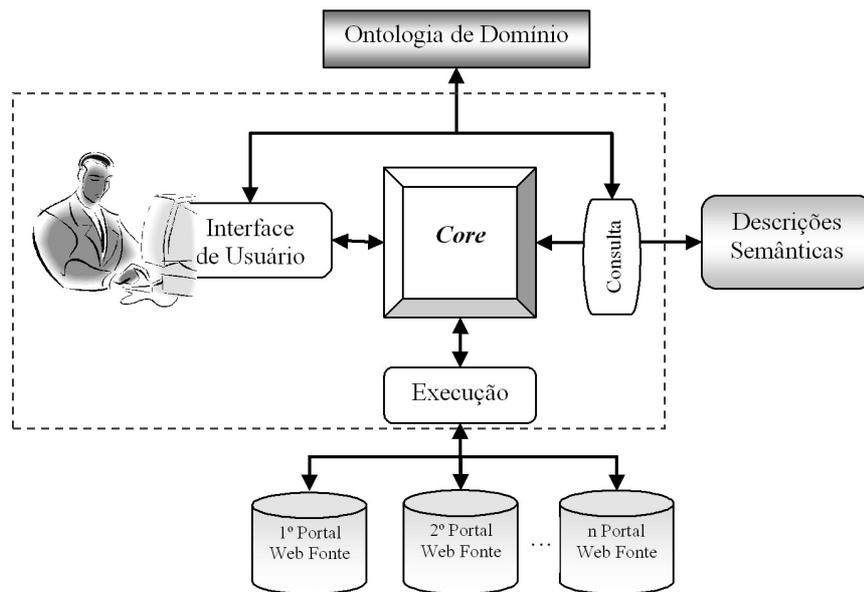


Figura 4.10: Arquitetura do sistema W-KIS

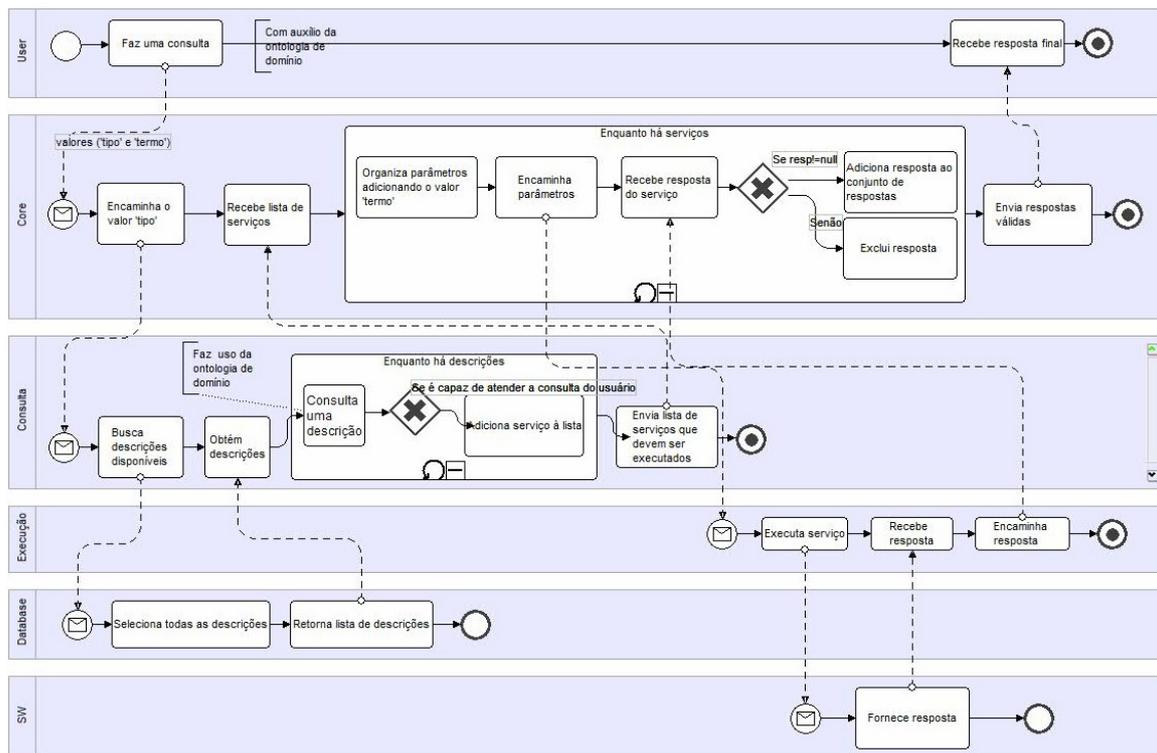


Figura 4.11: Modelo de processo do sistema W-KIS

4.4.1 Interface de Usuário

A interface de usuário atua como um mecanismo de meta-busca, no qual a consulta do usuário é respondida por uma ou mais fontes de informação que podem prover motores de busca individuais. A interface, demonstrada nas figuras 4.11 e 4.12, apresenta um campo onde o usuário deve digitar o “termo” que ele deseja consultar e um menu onde deve escolher o “tipo” que classifica o termo sendo buscado. Quando

nenhum valor é fornecido para o “termo”, W-KIS assume que o valor a ser buscado é o conceito selecionado no menu de “tipos”. Por outro lado, o valor do “tipo” deve sempre ser fornecido, portanto, a aplicação sempre inicia com algum valor pré-selecionado no menu. Caso o usuário não selecione algum conceito, o valor pré-selecionado será utilizado pelo sistema para executar a busca.

A interface em W-KIS permite que o usuário expresse sua consulta utilizando uma ferramenta baseada em um modelo de conhecimento, com a qual o usuário especifica a tradução dos dados como parte de sua consulta. Os principais conceitos da ontologia de domínio são utilizados para formar o menu que permite a escolha do “tipo”, através da API Protégé-OWL. Logo, quando o usuário seleciona um item, ele está selecionando o conceito (classe) da ontologia de domínio. Desta forma, o sistema não necessita realizar nenhuma tradução posterior da consulta fornecida pelo usuário. Os dados fornecidos pelo usuário são repassados diretamente para os devidos componentes do sistema.

Figura 4.12: Interface para o domínio da Biomedicina

Figura 4.13: Interface para o domínio da Geologia

Quando o usuário submete uma consulta, os valores do “termo” e do “tipo” são repassados para os componentes de execução e de consulta, respectivamente. O componente de consulta faz uso do “tipo” para buscar os serviços capazes de atender a consulta do usuário. Ou seja, o componente de consulta seleciona apenas os serviços capazes de disponibilizar informação sobre o conceito definido em “tipo”. Já o componente de execução utiliza o “termo” como parâmetro de entrada para os serviços que serão executados, isto é, este componente executa cada serviço selecionado pelo componente de consulta, buscando pelo valor “termo”. Os componentes de consulta e de execução serão descritos mais detalhadamente a seguir.

4.4.2 Componente de Consulta

Através do componente de consulta, o sistema busca e seleciona os Serviços Web apropriados através do acesso às suas descrições semânticas. Os serviços apropriados são aqueles que possuem meios para responder a consulta, de acordo com o valor especificado pelo “tipo” escolhido pelo usuário na interface. O sistema deve buscar e selecionar apenas as fontes relevantes para a consulta realizada (ex: quais fontes possuem informação sobre o conceito do tipo “mineral” no domínio da Geologia?).

As descrições semânticas dos serviços foram disponibilizadas na Internet através de URLs. A identificação de cada descrição, juntamente com a sua URL, foi armazenada em um banco de dados relacional. O componente de consulta acessa esse banco onde encontra a listagem de todas as descrições disponíveis para o sistema. Ele então carrega uma descrição por vez, a partir de sua URL encontrada no banco de dados, para realizar a consulta.

A consulta é formulada utilizando a ontologia de domínio como um esquema global de consulta. A tarefa da consulta é realizada através da correspondência do tipo do termo de entrada da consulta especificado pelo usuário com o tipo do parâmetro de entrada da operação oferecida pelo Serviço Web, especificado na descrição semântica das capacidades do serviço. A correspondência é feita com base no critério proposto por Paolucci et al. (2002), porém, no presente trabalho, somente as correspondências exatas são consideradas (na qual ambos os termos denotam exatamente o mesmo conceito, ou o termo da consulta é uma subclasse do conceito especificado no serviço que está sendo oferecido). Até o momento, a ferramenta fornece meios para explorar apenas os relacionamentos de hierarquia entre os conceitos nas ontologias de domínio. Entretanto, com o amadurecimento da tecnologia de Serviços Web Semânticos e com o desenvolvimento completo das estruturas para esta tecnologia, os demais relacionamentos poderão ser então explorados no futuro.

O componente de consulta lê a descrição OWL-S do serviço selecionado, através de sua URL, e consulta o tipo do parâmetro de entrada definido. Sabendo o tipo permitido para o parâmetro de entrada, o componente realiza a inferência, através da API *Jena*, para saber se o tipo escolhido pelo usuário corresponde a uma classe ou a uma subclasse da ontologia de domínio. Nos casos afirmativos, o componente obtém o endereço URL da operação a ser executada e, a partir deste endereço, obtém as informações necessárias para que o componente de execução possa executar os serviços. Por fim, o componente de consulta constrói uma lista contendo todos os serviços capazes de atender a necessidade do usuário e suas informações, e transmite esta lista para o componente de execução, através do componente *core*. Somente os serviços presentes nesta lista serão executados pelo sistema.

4.4.3 Componente de Execução

O componente de execução utiliza a base OWL-S/WSDL, a qual faz parte da descrição semântica dos Serviços Web, para obter as informações necessárias para executar cada serviço selecionado pelo componente de consulta. Ele recebe a lista de serviços formada pelo componente de consulta e executa um a um, armazenando cada resultado fornecido. Ele retorna o resultado da execução de cada serviço para o componente central do sistema, desconsiderando os resultados nulos.

Os artefatos de nível de mensagem são criados automaticamente pelo componente, assim como a conexão com o servidor. As demais informações essenciais utilizadas para a execução de cada Serviço Web são fornecidas pela base OWL-S/WSDL da descrição semântica de cada serviço. Este conjunto de informações inclui: a URL para acessar o serviço (ex: 'http://localhost:8080/axis2/services/Mindat'); a operação a ser executada (ex: 'getMineral'); o termo de entrada para a operação (ex: 'quartz'); a referência para os tipos básicos dos parâmetros (ex: 'http://www.w3.org/2001/XMLSchema#xsd'); e o protocolo de transporte a ser utilizado (ex: 'HTTP').

4.4.4 Componente *Core*

O componente *core* é o componente central do sistema. Ele atua como uma aplicação intermediária, estabelecendo a conexão entre todos os demais componentes, organizando o fluxo de dados entre eles. Depois que o usuário envia a consulta através da interface, o componente *core* é o responsável por encaminhar corretamente o valor

do “termo” para o componente de execução e o valor do “tipo” para o componente de consulta.

Ele também é responsável pela organização e transmissão dos resultados obtidos nas execuções dos serviços, pelo componente de execução, ao usuário final. Assim que os resultados são retornados pelo componente de execução, eles são integrados e repassados ao usuário pelo componente central.

5 AVALIAÇÃO DO SISTEMA W-KIS

A validação do sistema W-KIS foi realizada através de dois procedimentos diferentes. A primeira avaliação foi efetuada a partir da realização de um estudo de caso em cada domínio de aplicação, e da comparação dos resultados obtidos em cada estudo. Como segunda avaliação, a aplicação desenvolvida foi comparada com outras abordagens existentes tanto para a integração de informação na Web, como para a realização de busca também na Web. Além destas duas validações, este capítulo também apresenta o procedimento de manutenção e evolução do sistema W-KIS, o qual tem grande influência sobre a avaliação da aplicação.

5.1 Estudos de Caso

Como este trabalho pretende prover um sistema de integração que seja capaz de atuar em mais de um domínio de aplicação, dois estudos de caso foram realizados nesta pesquisa. O primeiro, referente ao domínio da Biomedicina, foi realizado na Universidad de Murcia, na Espanha, sobre a área da Oncologia para diagnóstico e tratamento de câncer de mama. O segundo, referente ao domínio da Geologia, foi realizado na Universidade Federal do Rio Grande do Sul sobre a área da Petrografia Sedimentar, onde são realizadas descrições de rochas sedimentares para qualificar reservatórios de petróleo e avaliar seu potencial produtor.

Foram realizadas diversas consultas em ambos os domínios, sendo as respostas obtidas com o sistema em cada consulta comparadas com a informação disponibilizada nos devidos portais Web.

5.1.1 Domínio da Biomedicina

No domínio da Biomedicina, existem onze conceitos provindos da ontologia de domínio *Oncogen* os quais são utilizados para formar o menu da interface de usuário com o objetivo de classificar o termo sendo consultado. Dentre esses onze conceitos, seis foram identificados como mais relevantes para as consultas, de acordo com as fontes de informação incorporadas ao sistema para este domínio. Foram realizadas quinze consultas para cada um destes conceitos, e cinco consultas para cada um dos cinco conceitos restantes. No total, 115 consultas foram efetuadas no domínio da Biomedicina, com o objetivo de validar a aplicação para o presente domínio. Um exemplo de consulta realizada, demonstrado na figura 5.1 e detalhado a seguir, é a busca pelo termo “A4”, tendo como tipo selecionado o conceito “Gene”. Outros exemplos de consultas são demonstrados no Apêndice C.

Quando o usuário envia a consulta, duas variáveis (“termo” e “tipo”) são utilizadas pelo sistema para realizar a busca. Elas são repassadas para o componente *core* e ele é

responsável por enviar a variável “tipo” para o componente de consulta e a variável “termo” para o componente de execução.

A primeira tarefa do componente de consulta é buscar, em uma tabela de banco de dados (exibida na tabela 5.1), todos os serviços disponíveis para o sistema no referente domínio. Em seguida, ele consulta um serviço por vez, a partir da URL de sua descrição semântica, para obter somente aqueles capazes de atender o objetivo do usuário. Neste exemplo, a partir do tipo “Gene” definido pelo usuário, o componente de consulta realiza a busca pelos serviços capazes de responder esta consulta.

Figura 5.1: Exemplo de consulta

Tabela 5.1: Tabela contendo os serviços no domínio da Biomedicina

id	serviceName	serviceDescURL
01	GeneOntologySWS	http://www.inf.ufrgs.br/~mtmello/ontology/GeneOntologySWS
02	MPactSWS	http://www.inf.ufrgs.br/~mtmello/ontology/MPactSWS
03	PubMedSWS	http://www.inf.ufrgs.br/~mtmello/ontology/PubMedSWS
04	SwissProtSWS	http://www.inf.ufrgs.br/~mtmello/ontology/SwissProtSWS

Dentre os quatro serviços disponibilizados para o domínio da Biomedicina, apenas dois admitem como parâmetro de entrada um termo especificado com o tipo “Gene”: o serviço *GeneOntology* e o serviço *PubMed*. Cada um destes serviços apresenta, em sua descrição semântica, o tipo do parâmetro de entrada especificado como o conceito “Gene” da ontologia de domínio, o que corresponde ao tipo definido pelo usuário, como demonstrado na figura 5.2. O componente de consulta seleciona então apenas estes dois serviços, formando uma lista contendo as informações necessárias para a execução de cada um ([“http://localhost:8080/axis2/services/GO_WebService”, “getGene”, <http://www.w3.org/2001/XMLSchema#xsd>, “TRANSPORT_HTTP”], [“http://localhost:8080/axis2/services/PubMed_WebService”, “getGene”, <http://www.w3.org/2001/XMLSchema#xsd>, “TRANSPORT_HTTP”]). Esta lista é então repassada para o componente *core*.

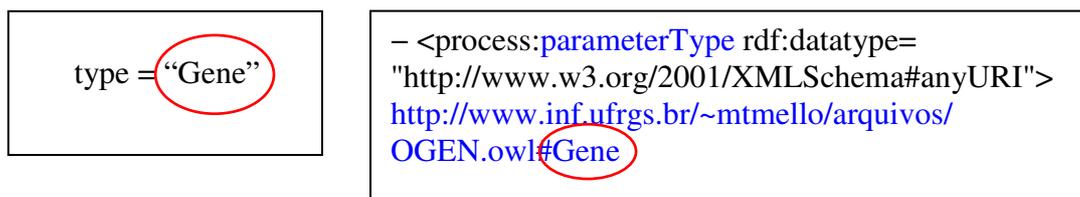


Figura 5.2: Correspondência do termo de entrada com o tipo definido como “Gene”

O componente *core* então envia um serviço por vez para o componente de execução, adicionando o “termo” digitado pelo usuário (“A4”) ao conjunto de informações do serviço. As respostas da execução dos serviços selecionados são repassadas para o usuário pelo componente *core* o qual é responsável por excluir as respostas nulas e transmitir ao usuário apenas as respostas válidas. O resultado retornado para a consulta do usuário neste estudo de caso, exibido na figura 5.3, apresenta a resposta fornecida pelo motor de busca por genes do portal *PubMed* contendo a descrição do gene A4. Neste exemplo, apenas uma resposta foi retornada, pois a execução do serviço do portal *GeneOntology*, o segundo portal capaz de responder a consulta, excedeu o tempo máximo de espera aceito pelo sistema W-KIS.

BIOMEDICINE

A4 description

from [PubMed Webservice](#) :

Official Symbol ATP6V0A4 and Name: ATPase, H+ transporting, lysosomal V0 subunit a4 [Homo sapiens] Other Aliases: tcag7.346, A4, ATP6N1B, ATP6N2, MGC130016, MGC130017, RDRTA2, RTA1C, RTADR, STV1, VPH1, VPP2Other Designations: ATPase, H+ transporting, lysosomal (vacuolar proton pump) non-catalytic accessory protein 1B; ATPase, H+ transporting, lysosomal (vacuolar proton pump) non-catalytic accessory protein 2 (38kD); H(+)-transporting two-sector ATPase, noncatalytic accessory protein 1B; vacuolar proton pump 116 kDa accessory subunit; vacuolar proton pump, subunit 2Chromosome: 7; Location: 7q33-q34 Annotation: Chromosome 7, NC_000007.12 (138041579..138133481, complement)MIM: 605239GeneID: 50617

Figura 5.3: Resultado da busca pelo termo “A4”

5.1.2 Domínio da Geologia

Já no domínio da Geologia, existem mais de duzentos conceitos definidos na ontologia de domínio *PetroGrapher* para serem utilizados no menu da interface de usuário para classificar o termo sendo consultado. Dentre esses conceitos, treze foram identificados como mais relevantes para as consultas, de acordo com as fontes de informação incorporadas ao sistema para este domínio. Foram realizadas dez consultas para cada um destes conceitos. Dentre os conceitos restantes, vinte e cinco foram selecionados aleatoriamente e duas consultas foram realizadas para cada um deles. No total, 180 consultas foram efetuadas no domínio da Geologia, com o objetivo de validar a aplicação para o presente domínio. Um exemplo de consulta realizada, demonstrado na figura 5.4, é a busca pelo termo “matrix”, tendo como tipo selecionado o conceito “Constituent”. Outros exemplos de consultas são demonstrados no Apêndice C.

Uma tabela de banco de dados, exibida na tabela 5.2, fornece uma listagem de todos os serviços disponíveis no referente domínio para o componente de consulta. A partir desta tabela, o componente obtém a URL da descrição semântica de cada serviço e,

através dela, realiza a busca pelo tipo do parâmetro de entrada definido no serviço, para conferir se admite o tipo “Constituent” escolhido pelo usuário.

The screenshot shows a web interface with a blue header containing the word "GEOLOGY". Below the header is a section titled "Query". There are two input fields: "term" and "type". The "term" field contains the text "matrix". Below the "term" field is a small blue link that says "enter the term you are searching for". The "type" field is a dropdown menu that is currently open, displaying a list of geological terms. The term "Constituent" is highlighted in blue. Other terms in the list include "Rock", "Goethite", "DiageneticIronOxide-HydroxideUndifferentiated", "Limonite", "Hematite", "ConglomerateRockFragment", "SiltstoneRockFragment", "SandstoneRockFragment", "PhosphaticRockFragment", "DolostoneRockFragment", "EvaporiticRockFragment", "ShaleRockFragment", "ChertRockFragment", "SedimentaryRockFragment", "MudrockFragmentUndifferentiated", "ClaystoneRockFragment", "SedimentaryRockFragmentUndifferentiated", "FerriferousRockFragment", and "MudrockFragment".

Figura 5.4: Exemplo de consulta

Tabela 5.2: Tabela contendo os serviços no domínio da Geologia

id	serviceName	serviceDescURL
01	AthenaSWS	http://www.inf.ufrgs.br/~mtmello/ontology/AthenaSWS
02	GeologynetSWS	http://www.inf.ufrgs.br/~mtmello/ontology/GeologynetSWS
03	MindatSWS	http://www.inf.ufrgs.br/~mtmello/ontology/MindatSWS
04	SchlumbergerSWS	http://www.inf.ufrgs.br/~mtmello/ontology/SchlumbergerSWS
05	WebMineralSWS	http://www.inf.ufrgs.br/~mtmello/ontology/WebMineralSWS

Os cinco serviços disponibilizados para o domínio da Geologia são capazes de responder a consulta tendo como entrada um termo definido com o tipo “Constituent”. O serviço *Schlumberger* admite como parâmetro de entrada um termo com o tipo definido como qualquer conceito da ontologia, visto que o tipo do parâmetro no serviço está especificado como o conceito “General” o qual abrange todos os demais conceitos como subclasses. Já os demais serviços admitem parâmetros de entrada do tipo “Constituent”. A correspondência do tipo definido pelo usuário com o tipo especificado nos casos citados é apresentada na figura 5.5.

O componente de consulta seleciona então todos os cinco serviços, inserindo-os em uma lista juntamente com as informações necessárias para a execução de cada um ([“<http://localhost:8080/axis2/services/Athena>”, “getMineral”, <http://www.w3.org/2001/XMLSchema#xsd>, “TRANSPORT_HTTP”], [“<http://localhost:8080/axis2/services/Geologynet>”, “getMineral”, <http://www.w3.org/2001/XMLSchema#xsd>, “TRANSPORT_HTTP”], [“<http://localhost:8080/axis2/services/Mindat>”, “getMineral”,

http://www.w3.org/2001/XMLSchema#xsd, “TRANSPORT_HTTP”],
 [“http://localhost:8080/axis2/services/Schlumberger”, “getDefinition”,
 http://www.w3.org/2001/XMLSchema#xsd, “TRANSPORT_HTTP”],
 [“http://localhost:8080/axis2/services/WebMineral”, “getMineral”,
 http://www.w3.org/2001/XMLSchema#xsd, “TRANSPORT_HTTP”]). Esta lista é
 então repassada para o componente *core*, o qual envia um serviço por vez para o
 componente de execução, adicionando o “termo” digitado pelo usuário (“matrix”) ao
 conjunto de informações do serviço. As respostas da execução dos serviços
 selecionados são repassadas para o usuário pelo componente *core* o qual é responsável
 por excluir as respostas nulas e transmitir ao usuário, através da interface, apenas as
 respostas válidas.

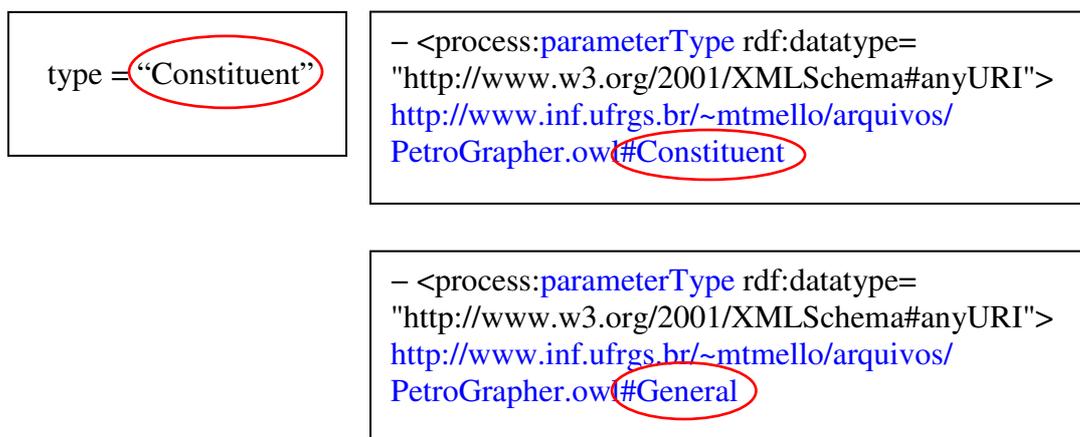


Figura 5.5: Correspondência do parâmetro de entrada

O resultado retornado para o usuário, demonstrado na figura 5.6, apresenta a resposta fornecida pelo portal *Schlumberger* contendo a definição do termo *matrix*. Neste exemplo, apenas uma resposta foi retornada devido ao fato dos demais portais não fornecerem qualquer informação sobre o termo buscado.

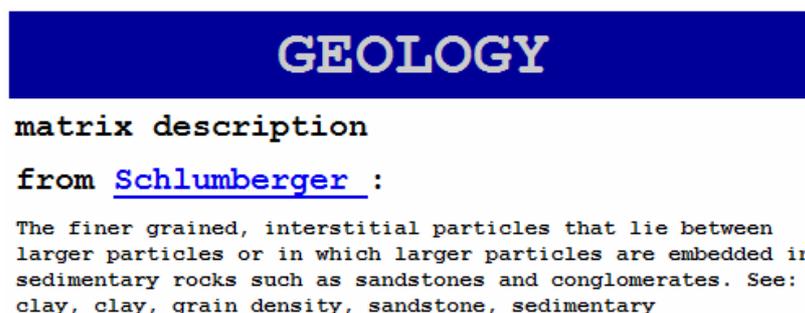


Figura 5.6: Resultado da busca pelo termo “matrix”

5.1.3 Avaliação nos Dois Domínios

No total do processo de testes com o sistema W-KIS foram realizadas 295 consultas, 115 no domínio da Biomedicina e 180 no domínio da Geologia. Tanto no domínio da Biomedicina como no domínio da Geologia, W-KIS apresentou um ótimo desempenho, retornando respostas válidas e relevantes para cada domínio. As respostas são consideradas corretas quando estão consistentes com a informação disponibilizada nos portais utilizados como fontes. A única situação em que W-KIS não retorna resposta alguma, mesmo a resposta existindo em algum portal Web incorporado ao sistema, é

quando o tempo de execução de portal fonte excede o tempo máximo de espera aceito pela aplicação. Esta situação pode ser observada no exemplo apresentado no estudo de caso no domínio da Biomedicina. Tanto o portal *GeneOntology* como o portal *PubMed* apresentam informação sobre o termo “A4” sendo buscado. Porém, apenas a resposta fornecida pelo portal *PubMed* é retornada para o usuário, pois o tempo de retorno do portal *GeneOntology* excedeu o tempo máximo permitido por W-KIS.

Nos demais casos, sempre que pelo menos um portal conectado ao sistema apresentar informação sobre o termo buscado pelo usuário, W-KIS é capaz de retornar esta resposta para o usuário final. No exemplo apresentado como estudo de caso no domínio da Geologia, todos os portais Web vinculados ao sistema para este domínio são selecionados e executados pelo sistema, porém apenas uma resposta, fornecida pelo portal *Schlumberger*, é retornada para o usuário. Isto ocorreu devido ao fato de que somente este portal apresentou informação sobre o termo buscado pelo usuário (“matrix”), apesar de todos serem capazes de responder consultas sobre o tipo selecionado (“Constituent”).

Através dos estudos de caso realizados, o sistema provou oferecer um mecanismo de consulta baseado na intenção do usuário, retornando apenas respostas contextualizadas ao domínio no qual a aplicação está rodando. Com os resultados obtidos com os estudos de caso em ambos os domínios, W-KIS também provou ser capaz de atuar em domínios de aplicação diferentes, apresentando uma arquitetura contendo componentes básicos centrais os quais são reutilizáveis para os diferentes domínios.

Além disso, o sistema é capaz de localizar e executar automaticamente os serviços que encapsulam os portais Web, fornecendo as respostas retornadas por cada portal diretamente para o usuário e não apenas endereços das páginas Web onde as respostas podem ser encontradas. Caso o usuário deseje obter informações adicionais, o endereço de cada portal que forneceu cada resposta é disponibilizado juntamente a resposta. O tempo de retorno da resposta final ao usuário é decorrente do tempo que cada portal leva para executar seu serviço. Quanto mais eficiente for o serviço disponibilizado pelo portal, mais rápida é a execução da consulta no W-KIS, não sendo percebido um atraso maior pela contextualização da busca atribuída pelas ontologias. Como W-KIS acessa as fontes de informação somente no instante da execução da consulta, ele sempre obtém a versão mais atualizada da fonte, refletindo automaticamente qualquer alteração nos dados estáticos dos portais.

Uma das principais vantagens do sistema proposto no presente trabalho é que a aplicação integra os resultados de diferentes motores de busca relacionados a um domínio de aplicação específico em uma única ferramenta. Esta característica facilita a tarefa de consulta do usuário, excluindo a tarefa de busca por diferentes fontes de informação e focando as respostas retornadas no domínio de interesse do usuário.

5.2 Comparação com Outras Abordagens

A primeira comparação realizada foi com as nove aplicações para integração na Web citadas na seção 2: *SHOE*, *OBSERVER*, *Manifold*, *SEMEDA*, *SEWISE*, *BioMeke*, *Integra*, *DIP* e *BIRD*; excluindo os motores de busca e meta-busca. As principais vantagens do sistema W-KIS são:

1. W-KIS foi aplicado e validado em dois domínios de aplicação: Biomedicina e Geologia, podendo ser facilmente readaptado para ser aplicado em outros

contextos através da utilização dos componentes centrais da arquitetura do sistema. Estes componentes não necessitam sofrer qualquer alteração no processo de readaptação para domínios distintos. Dentre as abordagens listadas na seção 2, somente as ferramentas *SHOE* e *DIP* podem ser aplicados em mais de um domínio de aplicação. Já as soluções *SEMEDA* e *SEWISE* podem ser estendidas para serem aplicadas em outros domínios. As demais ferramentas são específicas para algum determinado domínio de aplicação.

2. W-KIS propõe uma solução completamente automatizada para o processo de busca e execução dos serviços que encapsulam as funcionalidades oferecidas pelos portais Web, através da utilização de Serviços Web Semânticos. As únicas abordagens que também automatizam por completo esse processo são as ferramentas *DIP* e *BIRD*. Porém, *DIP* é um projeto voltado para a área de Comércio Eletrônico, apesar de poder ser aplicado em diferentes domínios de aplicação, enquanto *BIRD* é uma ferramenta específica para o domínio da Biomedicina.
3. W-KIS lida com a heterogeneidade semântica das fontes de informação utilizando ontologias de domínio através de Serviços Web Semânticos. A adoção de ontologias de domínio é a abordagem mais utilizada pelas ferramentas com o objetivo de lidar com a heterogeneidade de semântica entre as fontes a serem integradas. Dentre as ferramentas analisadas, apenas *Manifold* não lida com este tipo de heterogeneidade, enquanto *BioMeke* lida de forma limitada, não utilizando ontologias.
4. W-KIS integra não apenas dados estáticos, mas também processos, isto é, os serviços oferecidos pelos portais Web. Esta capacidade ainda não é explorada pela grande maioria das propostas para integração na Web. Dentre as abordagens revisadas neste trabalho, apenas as aplicações *DIP* e *BIRD* são capazes de integrar processos além de dados estáticos.
5. W-KIS permite que fontes sejam removidas e/ou que novas fontes sejam adicionadas ao repositório do sistema, sem a necessidade de qualquer alteração na arquitetura do sistema. Além disso, as fontes de informação podem ter seus dados atualizados sem interferir no funcionamento da aplicação. Do conjunto de abordagens analisadas na seção 2, as soluções *SHOE*, *Manifold*, *SEMEDA*, *SEWISE*, *BioMeke* e *Integra* lidam com as alterações nas fontes de informação de forma muito limitada.

As abordagens analisadas, citadas na seção 2 do presente trabalho, apresentam uma ou algumas das características consideradas essenciais para um sistema de integração para a Web. Porém, nenhuma delas é capaz de apresentar todas estas características em uma única ferramenta. Por outro lado, W-KIS reúne todas as características enumeradas acima como vantagens do sistema, disponibilizando uma ferramenta mais robusta para integrar informação na Web, com o objetivo de atender a consultas contextualizadas em domínios específicos.

Além da comparação com as principais abordagens atuais para integração na Web, a performance do sistema W-KIS ainda foi comparada com as ferramentas de busca convencionais mais utilizadas atualmente na Web: *Google*¹¹, *Yahoo*¹² e *AOL*¹³. Nestes

¹¹ <http://www.google.com/>

¹² <http://search.yahoo.com/>

motores de busca, o usuário fornece uma ou mais palavras-chave as quais descrevem a informação que ele está buscando. Então, uma lista de documentos é retornada, ordenada por relevância. Entretanto, na maioria das vezes, esta lista é bastante longa, contendo muitos documentos irrelevantes para a pesquisa do usuário. Isto acontece visto que quando o usuário fornece as palavras-chave, ele possui um tópico relevante em seu pensamento. Portanto, a busca que ele realiza está relacionada a este tópico, e os resultados que ele espera obter também devem estar dentro deste contexto. Porém, nestes portais de busca, não há meios para o usuário especificar o contexto no qual ele deseja realizar a sua consulta.

As figuras 5.7 e 5.8 demonstram os primeiros resultados retornados pelo portal *Google* em duas consultas realizadas com os mesmos termos utilizados nos estudos de caso com o sistema W-KIS no domínio da Biomedicina e no domínio da Geologia respectivamente: “A4” e “matrix”. Estes termos foram escolhidos justamente por serem considerados ‘problemáticos’, isto é, eles podem ter mais de uma interpretação dependendo do contexto em que estão inseridos.

The image shows a screenshot of a Google search results page for the query "A4". At the top, the Google logo is on the left, followed by a search box containing "A4", a "Search" button, and links for "Advanced Search" and "Preferences". Below the search bar, a "Web" section lists several search results:

- A4 paper format / International standard paper sizes**: Explains standard paper sizes like ISO A4 that are today widely used all over the world. www.ci.cam.ac.uk/~mgk25/iso-paper.html - 59k - [Cached](#) - [Similar pages](#) - [Note this](#)
- a4**: www.a4.com/ - [Similar pages](#) - [Note this](#)
- ISO 216 - Wikipedia, the free encyclopedia**: It is the standard which defines the commonly available A4 paper size. ... For example, if an A4 page is folded in half so that it is A5 in size, ... en.wikipedia.org/wiki/ISO_216 - 31k - [Cached](#) - [Similar pages](#) - [Note this](#)
- Audi A4 - Wikipedia, the free encyclopedia**: The Audi A4 is a compact executive car produced by the German automaker Audi since 1994. It is the successor of the Audi 80. ... en.wikipedia.org/wiki/Audi_A4 - 98k - [Cached](#) - [Similar pages](#) - [Note this](#)
- Welcome to A4Tech website**: Global, USA, China, Estonia, Germany, Czech, Poland, Japan, Latvia, Lithuania, Belarus, Russia, Romania, Taiwan, Ukraine, Ukraine, Netherlands ... www.a4tech.com/ - 10k - [Cached](#) - [Similar pages](#) - [Note this](#)
- Audi UK > New Cars > A4**: Owning a car from the A4 range sets you apart from the rest. Each model's looks, performance and safety features have been fine-tuned to meet the ... www.audi.co.uk/audi/uk/en2/new_cars/a4.html - 78k - [Cached](#) - [Similar pages](#) - [Note this](#)
- Audi of America > See All Models > Audi A4**: Driving pleasure and superior technology, exciting styling and a level of equipment that is worthy of the luxury class: these are the qualities that make ... www.audiusa.com/audi/us/en2/new_cars/Audi_A4.html - 94k - [Cached](#) - [Similar pages](#) - [Note this](#)
- A4 MAGAZINE**: fashion, art, culture, design monthly. a4mag.com/ - 2k - [Cached](#) - [Similar pages](#) - [Note this](#)
- Welcome to A4**: Autism & Aspergers Advocacy Australia was formed in response to the growing demand from people with autism spectrum condition and their families who wish to ... www.a4.org.au/ - 17k - [Cached](#) - [Similar pages](#) - [Note this](#)

Figura 5.7: Primeiros resultados da busca pelo termo “A4” utilizando *Google*

Foram analisados os primeiros 100 resultados fornecidos em ambas as consultas e todos deles retornaram respostas fora de contexto. Ou seja, de acordo com a intenção do usuário, a primeira consulta deveria retornar respostas dentro do contexto da Biomedicina, já a segunda deveria retornar respostas dentro do contexto da Geologia.

Um segundo teste realizado no portal *Google* foi a busca pelos mesmos termos utilizados anteriormente, porém com a adição de uma palavra para classificar cada termo sendo consultado, na tentativa de atribuir o contexto no qual a consulta deve ser efetuada: “A4+gene” e “matrix+constituent”. Para contextualizar a busca no domínio da Biomedicina foi utilizado o mesmo conceito utilizado no exemplo do estudo de caso para este domínio: “gene”. Os resultados fornecidos para a consulta “A4+gene” estão contextualizados na área da Biomedicina, entretanto, os primeiros 50 resultados representam referências para trabalhos de pesquisa ou artigos acadêmicos. Nenhum dos 50 primeiros resultados retornados fornece uma definição ou descrição do termo buscado. As primeiras respostas são apresentadas na figura 5.9.

The image shows a screenshot of a Google search results page for the query "matrix". At the top, the Google logo is visible next to a search bar containing the word "matrix". To the right of the search bar are buttons for "Search", "Advanced Search", and "Preferences". Below the search bar, the results are categorized under "Web". The first result is "The Matrix (1999)", which includes a brief description of the film directed by Andy Wachowski and Larry Wachowski, listing Keanu Reeves and Laurence Fishburne. The second result is "The Matrix Reloaded (2003)", also directed by the Wachowski brothers, listing Ray Anthony, Christine Anu, and Andy Amess. The third result is "THE MATRIX TRILOGY", an official site for the series. The fourth result is "The Matrix - Wikipedia, the free encyclopedia", providing a general overview of the 1999 film. The fifth result is "Matrix (mathematics) - Wikipedia, the free encyclopedia", defining the term in a mathematical context. The sixth result is "Matrix - Hairstyles", the official website for Matrix hair products. The seventh result is "The matrix Standard", a quality framework for information delivery. The eighth result is "Matrix | Interpreting the Real Estate Economy", a website related to real estate.

Figura 5.8: Primeiros resultados da busca pelo termo “matrix” utilizando *Google*

Google [Advanced Search](#) [Preferences](#)

Web

[Brazilian Journal of Medical and Biological Research - APOA1/C3/A4 ...](#)
APOA1/C3/A4 **gene** cluster variability and lipid levels in Brazilian children ... Several polymorphic sites have been identified in the APOA1-C3-A4 **gene** ...
www.scielo.br/scielo.php?pid=S0100-879X2005000400006&script=sci_arttext - 41k - [Cached](#) - [Similar pages](#)

[Structure, Evolution, and Polymorphisms of the Human ...](#)
Structure, Evolution, and Polymorphisms of the Human Apolipoprotein A4 **Gene** (APOA4). Sotirios K. Karathanasis, Peter Oettgen, Issam A. Haddad, and Stylianos ...
www.pnas.org/cgi/content/abstract/83/22/8457 - [Similar pages](#)

[Structure, evolution, and polymorphisms of the human ...](#)
Structure, evolution, and polymorphisms of the human apolipoprotein A4 **gene** (APOA4). S K Karathanasis, P Oettgen, I A Haddad, and S E Antonarakis ...
www.pubmedcentral.nih.gov/articlerender.fcgi?artid=386949 - [Similar pages](#)

[IngentaConnect Cloning and genomic organization of the human ...](#)
The present work reports the isolation and the characterization of a human glutathione transferase (GST) A4 subunit **gene**. This **gene** was localised on ...
www.ingentaconnect.com/.../1998/00000095/9000001/art80188;jsessionid=jg2o4lgmofg.alexandra?format=print - [Similar pages](#)

[A common deletion polymorphism in the apolipoprotein A4 gene and ...](#)
A common deletion polymorphism in the apolipoprotein A4 **gene** and its significance in lipid metabolism. MI Kamboh, JS Friedlaender, YI Ahn and RE Ferrell ...
atvb.ahajournals.org/cgi/content/abstract/14/5/656 - [Similar pages](#)

[Frequencies of Apolipoprotein A4 Gene Polymorphisms and ...](#)
Frequencies of Apolipoprotein A4 **Gene** Polymorphisms and Association with Serum Lipid

Figura 5.9: Primeiros resultados da busca por “A4+gene” utilizando *Google*

Já a figura 5.10 demonstra os primeiros resultados retornados para a consulta “matrix+constituent”, onde o conceito “constituent” é utilizado na tentativa de contextualizar a busca no domínio da Geologia. Para esta busca, as 50 primeiras respostas fornecidas não estão contextualizadas na área da Geologia.

Google [Advanced Search](#) [Preferences](#)

Web

[Matrix Biology : Differential cell adhesion to vocal fold ...](#)
Differential cell adhesion to vocal fold extracellular **matrix constituents**. Tannin J. Fujaa, Corresponding Author Contact Information ...
linkinghub.elsevier.com/retrieve/pii/S0945053X06000126 - [Similar pages](#)

[Materials Characterization : Metallographic Characterization of ...](#)
9, illustrates the **matrix constituents** in Specimen III which consists of martensite with a relative proportion of retained austenite. ...
linkinghub.elsevier.com/retrieve/pii/S1044580399000029 - [Similar pages](#)
[More results from linkinghub.elsevier.com >](#)

[book micromechanics of composites : composite properties of fibre ...](#)
Micromechanics of composites : composite properties of fibre and **matrix constituents** (with 3.5 diskette). Author(s) : STELLBRINK Publication date : 01-1996 ...
www.lavoisier.fr/notice/gb045370.html - 18k - [Cached](#) - [Similar pages](#)

[Dental cell interaction with extracellular-matrix constituents ...](#)
by another **constituent** of the extracellular **matrix** or by cell-surface-associated proteoglycans. Introduction. The involvement of epitheliomesenchymal ...
www.blackwell-synergy.com/doi/pdf/10.1111/j.1432-0436.1985.tb00312.x - [Similar pages](#)

[MGI - Gene Ontology Browser](#)
extracellular **matrix constituent** conferring elasticity [GO:0030023] (2 ... extracellular **matrix** structural **constituent** conferring compression resistance + ...
lists.informatics.jax.org/searches/GO.cgi?id=GO:0030023 - 14k - [Cached](#) - [Similar pages](#)

[Extracellular Matrix in Disc Degeneration -- Feng et al. 88 \(1 ...](#)
This interaction leads to the production of **matrix constituents** but not to cell migration or division. Chondroadherin may thus have a role in maintaining ...

Figura 5.10: Primeiros resultados da busca por “matrix+constituent” utilizando *Google*

Por outro lado, se realizarmos a busca utilizando o sistema W-KIS, apenas fornecendo os termos “A4” no domínio da Biomedicina e “matrix” no domínio da Geologia, sem especificar um tipo para cada termo, ou seja, deixando o tipo “General” selecionado no menu da interface, W-KIS ainda assim retorna as respostas corretas. Isto ocorre devido ao fato de todos os portais incorporados ao sistema serem executados quando o tipo “General” é selecionado.

Os portais de busca *Yahoo* e *AOL* retornaram resultados similares para as mesmas consultas. Portanto, nenhuma das três ferramentas de busca foi capaz de atender as necessidades do usuário buscando informações nos domínios específicos da Biomedicina e da Geologia.

Além dos portais de busca convencionais, também existem os motores de meta-busca disponibilizados na Web. Como a ferramenta de meta-busca *WebFusion* citada no capítulo 2 não está disponível atualmente, a ferramenta *MetaCrawler*¹⁴ será utilizada no processo de avaliação. Este portal propõe um acesso integrado a diversos motores de busca tradicionais através de um mecanismo de meta-busca. Como já explicado anteriormente, as ferramentas de meta-busca funcionam como uma interface no topo de diversos motores de busca, fornecendo um acesso uniforme a cada mecanismo de busca local. Nesse tipo de busca, o usuário submete palavras-chave na caixa de consulta e o sistema transmite o termo a ser pesquisado simultaneamente para vários motores de busca individuais e suas bases de dados de páginas Web. Em poucos segundos, o usuário recebe os resultados fornecidos por todos os motores de busca pesquisados. Os mecanismos de meta-busca não possuem uma base de dados própria de páginas Web; eles enviam os termos de busca para as bases mantidas pelos simples motores de busca individuais.

Os sistemas de busca do tipo meta-busca enviam a consulta para todos os motores de busca da sua base de conhecimento, o que pode facilitar bastante a tarefa de busca do usuário, o qual fica dispensado da tarefa de consultar cada portal de busca individualmente. Porém, os resultados fornecidos por esta ferramenta ainda não atendem os objetivos do usuário de forma satisfatória.

Duas buscas foram realizadas no portal *MetaCrawler*, repetindo as mesmas consultas feitas anteriormente nos estudos de caso e nas ferramentas de busca convencionais. Na primeira consulta, buscando pelo termo “A4”, foram retornados 86 resultados dentre os quais os 40 primeiros fornecem respostas fora do contexto da Biomedicina, como pode ser visto na figura 5.11, a qual demonstra os primeiros 10 resultados retornados. Apenas o quadragésimo primeiro documento retornado fornece uma resposta contextualizada. A segunda consulta realizada, buscando pelo termo “matrix”, retornou 77 resultados, todos fora do contexto da Geologia. A figura 5.12 ilustra os 10 primeiros resultados fornecidos pela ferramenta *MetaCrawler* na consulta pelo termo “matrix”.

¹⁴ <http://www.metacrawler.com>

The screenshot shows the MetaCrawler search interface. At the top, there are navigation links for Web, Images, Audio, Video, News, Yellow Pages, and White Pages. A search bar contains the text 'A4' and a red 'SEARCH' button. Below the search bar, there are links for 'Exact Phrase', 'Advanced', and 'Preferences'. The main heading is 'Web Search Results for "A4"'. Below this, there are logos for various search engines: All Search Engines, Google, Yahoo!, msn, and others. The search results are listed as follows:

- A4 paper format / International standard paper sizes**
Explains standard paper sizes like ISO A4 that are today widely used all over the world.
www.dl.cam.ac.uk/~mgk25/iso-paper.html [Found on Google, Windows Live, Yahoo! Search, Ask.com]
- Carros da Audi**
Aqui você encontra variedade de marcas e modelos no Meu Citroën.
Sponsored by: www.MeuCitroen.com.br/Audi/ [Found on Ads by Google]
- Audi of America > See All Models > Audi A4**
Driving pleasure and superior technology, exciting styling and a level of equipment that is worthy of the luxury class: these are the qualities that make ...
www.audiusa.com/audi/us/en2/new_cars/Audi_A4.html [Found on Google, Windows Live, Yahoo! Search]
- Audi A4 - Wikipedia, the free encyclopedia**
The Audi A4 is a compact executive car produced by the German automaker Audi since 1994. It is the successor of the Audi 80. ...
en.wikipedia.org/wiki/Audi_A4 [Found on Google, Yahoo! Search]
- Carro Audi**
Pesquise e Encontre Carro Audi. Carros Audi no BuscaPé. Confira!
Sponsored by: www.BuscaPe.com.br/Audi/ [Found on Ads by Google]
- ISO 216 - Wikipedia, the free encyclopedia**
It is the standard which defines the commonly available A4 paper size. ... For example, if an A4 page is folded in half so that it is A5 in size, ...
en.wikipedia.org/wiki/ISO_216 [Found on Google, Ask.com]
- A4 - Wikipedia, the free encyclopedia**
A4 may refer to: A4 paper size , an international standard paper size (210x297 mm) defined by ISO 216; A440 , the A above middle C; Audi A4 , a brand of car; HMS A4 , an A-class ...
en.wikipedia.org/wiki/A4 [Found on Windows Live, Yahoo! Search, Ask.com]
- Audi of America > Home**
Welcome to Audi of America Home ... Welcome to Audi of America. You need a current version of the Flash plugin to view this content. ...
www.audiusa.com/model_home/0,,bodyStyleId-1,00.htm... [Found on Yahoo! Search]
- Welcome to AudiWorld.com - The Audi Enthusiast Website**
Audi A4 Cabriolet: A Touch of Springtime for Open-Air Fans. 14-Dec-07: ... Audi A4 Sedan: S line Sport Package Plus for Even More Dynamism. 10-Dec-07: ...
www.audiworld.com/ [Found on Google]
- European Standard A4 Papers for Inkjets and Lasers**
European standard A4 papers for inkjet, laser, and color laser printers is available in bright white bond, matte, and gloss and others. A4 paper, A3 and A5 ...
www.graytex.com/a4-papers.htm [Found on Google, Windows Live]

Figura 5.11: Primeiros resultados da busca pelo termo “A4” utilizando *MetaCrawler*

A última comparação foi realizada com o portal *About.com*¹⁵. *About.com* é um portal Web de informações gerais o qual tem o objetivo de prover soluções práticas para diversos problemas do cotidiano. Este portal disponibiliza seções contendo conteúdos específicos a determinados domínios. Cada seção, que tem o suporte de um guia responsável pela sua construção e manutenção, oferece conselhos, tutoriais e diversos outros materiais, além de um motor de busca no domínio específico. Para realizar a comparação com a aplicação presente, foram acessados os motores de busca das seções da *Biomedicina*¹⁶ e da *Geologia*¹⁷ deste portal.

¹⁵ <http://www.about.com/>

¹⁶ <http://biotech.about.com/>

¹⁷ <http://geology.about.com/>

The screenshot shows the MetaCrawler search interface. At the top, there are navigation links for 'Web', 'Images', 'Audio', 'Video', 'News', 'Yellow Pages', and 'White Pages'. A search bar contains the text 'matrix' and a red 'SEARCH' button. Below the search bar, there are options for 'Exact Phrase', 'Advanced', and 'Preferences'. The main heading is 'Web Search Results for "matrix"'. Below this, there are links for 'All Search Engines', 'Google', 'Yahoo! Search', 'msn Search', and 'More Engines | View Demo'. The search results are displayed as a list of 10 items, each with a title, a brief description, and a source link. The first item is a sponsored result for 'Matrix Determinant' from NextTag.com. The second item is 'The Matrix (1999)' from IMDb.com. The third item is 'Matrix Cars' from Calibex.com. The fourth item is 'The Matrix - Wikipedia, the free encyclopedia'. The fifth item is 'Matrix Car' from Dealtime.com. The sixth item is 'THE MATRIX TRILOGY' from whatisthematrix.warnerbros.com. The seventh item is 'Matrix (mathematics) - Wikipedia, the free encyclopedia'. The eighth item is 'THE MATRIX TRILOGY' from whatisthematrix.com. The ninth item is 'Matrix -- from Wolfram MathWorld'. The tenth item is 'Matrix Games - Whats Your Strategy?' from a forum.

metacrawler
SEARCH THE SEARCH ENGINES!

Web | Images | Audio | Video | News | Yellow Pages | White Pages

matrix **SEARCH**

Exact Phrase [Advanced](#) | [Preferences](#)

Web Search Results for "matrix"

View Results From: [All Search Engines](#) | [Google](#) | [Yahoo! Search](#) | [msn Search](#) | [More Engines](#) | [View Demo](#)

All Search Engines 1 - 20 of 77 ([About Results](#)) 1 | 2 | 3 | 4 | [Next >](#)

WEBSITE MATCH <http://www.matrix.com>

- Matrix Determinant**
Compare multiple free quotes on a new car from local dealers.
Sponsored by: www.NextTag.com/cars [Found on Ads by Yahoo!]
- The Matrix (1999)**
Directed by Andy Wachowski, Larry Wachowski. With Keanu Reeves, Laurence Fishburne, Carrie-Anne Moss. A computer hacker learns from mysterious rebels about ...
www.imdb.com/title/tt0133093/ [Found on Google, Windows Live, Yahoo! Search]
- Matrix Cars**
Cheap Prices and Huge Selection. **Matrix** Cars on Sale.
Sponsored by: www.Calibex.com [Found on Ads by Yahoo!]
- The Matrix - Wikipedia, the free encyclopedia**
The **Matrix** is a 1999 science fiction action film written and directed by Larry and Andy Wachowski and starring Keanu Reeves, Laurence Fishburne, Carrie-Anne ...
en.wikipedia.org/wiki/The_Matrix [Found on Google, Windows Live, Yahoo! Search]
- Matrix Car**
Find great deals and save. Compare products, prices & stores.
Sponsored by: www.Dealtime.com/carmonitors [Found on Ads by Yahoo!]
- THE MATRIX TRILOGY**
Official site for the series. Cast and crew biographies, interviews, photographs, production notes, visual effects, posters, comics, press, chat, games, ...
whatisthematrix.warnerbros.com/ [Found on Google, Ask.com]
- Matrix (mathematics) - Wikipedia, the free encyclopedia**
In mathematics, a **matrix** (plural **matrices**) is a rectangular table of elements (or entries), which may be numbers or, more generally, any abstract quantities ...
[en.wikipedia.org/wiki/Matrix_\(mathematics\)](http://en.wikipedia.org/wiki/Matrix_(mathematics)) [Found on Google, Yahoo! Search]
- THE MATRIX TRILOGY**
... site for the movie series, featuring trailers, DVD info, games, screen savers, and behind-the-scenes information for all three films: The Matrix, The **Matrix** ...
www.whatisthematrix.com/ [Found on Yahoo! Search]
- Matrix -- from Wolfram MathWorld**
A **matrix** is a concise and useful way of uniquely representing and working with linear transformations. In particular, for every linear transformation, ...
mathworld.wolfram.com/Matrix.html [Found on Google, Windows Live, Yahoo! Search]
- Matrix Games - Whats Your Strategy?**
Official site includes news, company information, game lists, support, shopping, and a forum.

Figura 5.12: Primeiros resultados da busca pelo termo “matrix” utilizando *MetaCrawler*

As consultas realizadas anteriormente foram repetidas neste portal com o objetivo de avaliar as respostas retornadas em cada domínio de aplicação. A busca pelo termo “A4” foi realizada na seção correspondente ao domínio da Biomedicina. Mesmo na seção específica, os 100 primeiros resultados analisados fornecem respostas fora do contexto neste domínio. Os oito primeiros resultados retornados nesta consulta estão ilustrados na figura 5.13. Já a busca pelo termo “matrix” foi realizada na seção correspondente ao domínio da Geologia. Nesta seção, a maioria dos resultados retornados está relacionada ao domínio da Geologia. Porém, dentre os 100 primeiros resultados retornados e analisados, nenhum forneceu a definição ou descrição do conceito buscado como resposta. Os nove primeiros resultados retornados nesta consulta podem ser visualizados na figura 5.14.

Os experimentos realizados com os motores de busca convencionais foram avaliados com base na medida de precisão comumente utilizada na avaliação de ferramentas de busca. Esta medida indica o número de respostas retornadas analisadas para a avaliação dos resultados. Esta avaliação é descrita na tabela 5.3.

Tabela 5.3: Avaliação das consultas realizadas com motores de busca

Portal	Consulta	Retorno	Precisão
Google	A4	Respostas fora de contexto	P@100
Google	matrix	Respostas fora de contexto	P@100
Google	A4+gene	Respostas sem definição ou descrição	P@50
Google	matrix+constituent	Respostas fora de contexto	P@50
MetaCrawler	A4	Respostas fora de contexto	P@40
MetaCrawler	matrix	Respostas fora de contexto	P@77
About.com	A4	Respostas fora de contexto	P@100
About.com	matrix	Respostas sem definição ou descrição	P@100

About.com : Biotech / Biomedical

Business & Finance

↳ **Biotech / Biomedical**

Essentials

- What is Biotechnology?
- What Venture Investors are Looking For
- Nanomedicine and Disease
- Pros and Cons of Stem Cell Research
- Essentials for Cloning Genes

Topics

- Intro to Biotechnology
- Investing in Biotech
- The Biotech Business
- Intellectual Property
- Technical Theory
- Application
- Case Studies
- Careers
- FAQ
- Glossary

Tools

- Find a Job
- Get a Quote
- Yellow Pages

Forums

- Most Popular Articles
- Latest Articles
- Help

Sponsored Links

[Audi A3](#)
Pesquise carros e compare preços. Encontre Audi A3 aqui no BuscaPê
www.BuscaPe.com.br/Audi-A3

[A6 em Oferta](#)
Compre e Venda A6. Negocie em Tempo Real. Confira!
www.TodaOferta.com.br

[Carros da Audi](#)
Aqui você encontra variedade de marcas e modelos no Meu Citroën.
www.MeuCitroen.com.br/Audi

In Biotech / Biomedical: "A4"
Displaying results 1 to 10 of 1,840

[2009 Audi A4 photo gallery -- Pictures of the 2009 Audi A4](#)
2009 Audi A4 photo gallery -- Pictures of the 2009 Audi A4.
<http://cars.about.com/od/audi/fr/2009-Audi-A4-gallery/>

[A4 Paper -- A4 Paper Size](#)
Definition: A4 paper is the standard in Europe and other parts of the world. The dimensions of A4 paper are 8.27 X 11.69 inches. Most, if not all, ...
<http://wordprocessing.about.com/cs/glossary/g/A4.htm>

[Measuring Paper - Measurement Systems for Desktop Publishing](#)
Side-by-side comparison of letter size to A4 size paper. ... TIP: When photocopying A4 pages to Letter size paper, reduce the pages 94% to fit the shorter ...
http://desktoppub.about.com/cs/intermediate/a/measure_paper.htm

[Weber Summit Gold A4 \(Discontinued\) - Gas Grill Review](#)
Weber Summit Gold A4 Gas Grill Review. This is possibly the best of the "mass market" grills on the market. To get more quality and appearance in a grill ...
<http://bbq.about.com/od/gasgrillreviews/gr/aapr051205a.htm>

[2007 Audi S4 - Test drive and new car review - 2007 Audi S4](#)
The 2007 Audi S4 is the hot-rod version of Audi's A4 and the smallest of its sporty S-line cars (which also includes the bigger S6 and S8).
http://cars.about.com/od/audi/fr/07_audi_s4.htm

[New Job Descriptions](#)
A4.017 PERFORM ANTENNA MAINTENANCE AND INSPECTIONS. A4.018 COMPLY WITH TEMPEST REQUIREMENTS ... A4.023 TROUBLESHOOT ERROR CONDITIONS ON PERIPHERAL ...
<http://usmilitary.about.com/library/milinfo/navyjobs/blrm.htm>

Figura 5.13: Primeiros resultados da busca pelo termo "A4" utilizando *About.com*

Com base nos resultados obtidos com as ferramentas de busca e meta-busca e nos resultados fornecidos pela aplicação proposta neste trabalho, o sistema W-KIS pode ser avaliado como a melhor opção para consultas em contextos específicos, tais como os domínios da Biomedicina e Geologia. A ferramenta W-KIS apresenta as seguintes vantagens:

1. Realiza buscas em domínios específicos – separadamente. O sistema, através do conhecimento agregado, é personalizado aos usuários de acordo com sua área de interesse.

- Retorna somente respostas relevantes, contextualizadas em um domínio específico, mesmo quando utilizados termos comuns a outros domínios como os exemplificados nos testes aqui realizados. As respostas retornadas pelo sistema respondem adequadamente a consulta do usuário na grande maioria das buscas.
- Fornecer resultados efetivos proporcionados por outros portais de busca diretamente e não apenas o endereço para a página Web que fornece a resposta. Os resultados de várias buscas realizadas separadamente são retornados de forma unificada para o usuário.

The screenshot shows the About.com website interface for a search on "matrix" within the Geology section. On the left, there is a navigation menu with categories like Education, Essentials, Topics, Buyer's Guide, and Tools. The main content area displays search results for "matrix", showing 1 to 10 of 23,700 results. The first result is "Minerals with Metallic Luster - Chalcopyrite in Matrix", followed by "Rock Identification Tables", "Aphanitic Extrusive Rock Types", "Conglomerate", "Minerals with Metallic Luster", "Rock Picture Gallery", "Copper", "Sandstone", and "Mineral Picture Gallery". Each result includes a brief description and a URL.

Figura 5.14: Primeiros resultados da busca pelo termo “matrix” utilizando *About.com*

5.3 Manutenção e Evolução

Faz parte da tarefa de manutenção e evolução do sistema a atualização do repositório de Serviços Web e de suas descrições semânticas, permitindo a adição e remoção de serviços, além da atualização do próprio serviço quando necessário.

Os dados nas bases de dados locais de cada portal Web podem ser alterados sem qualquer interferência no funcionamento do sistema W-KIS. Os serviços vinculados ao sistema são dependentes dos portais Web, ou seja, o sistema acessa os devidos portais, através dos serviços, no momento da execução da consulta. Portanto, a aplicação

sempre obtém a versão mais atualizada de cada portal Web. Entretanto, quando ocorre alguma alteração nos processos, isto é, nos serviços oferecidos pelos portais, tanto os serviços Web como as suas descrições semânticas devem ser atualizados, caso contrário o sistema não será capaz de executar os devidos processos.

No caso de algum portal conectado a W-KIS não estar mais disponível na Web, o sistema remove automaticamente o serviço que encapsula tal portal, juntamente com sua descrição semântica. Toda vez que isso acontecer, o sistema deve reportar a remoção juntamente com o nome do portal não mais disponibilizado.

Já no caso de um novo portal ser disponibilizado na Web, o sistema ainda não é capaz de realizar algum tipo de rastreamento de forma automática. Atualmente, o desenvolvedor é responsável por verificar, manualmente, quando novos portais de interesse para um determinado domínio são disponibilizados na Web. Porém, a adição de um novo portal ao conjunto de fontes de informação do sistema, assim como a remoção de portais deste conjunto, pode ser realizada sem a necessidade de qualquer alteração nos componentes centrais da arquitetura da aplicação.

6 CONCLUSÃO

Nos últimos anos, a Web emergiu como uma tecnologia para compartilhamento de informação, tornando-se o principal meio de acesso a conteúdos nos mais diversos domínios de aplicação. Porém, buscar e integrar o conteúdo disponível na Web ainda é uma tarefa difícil de ser realizada. Um sistema de busca e integração de informação na Web deve ser capaz de responder consultas que podem exigir a extração e combinação de dados e/ou processos de múltiplos portais Web.

Vários trabalhos propõem diversas soluções, utilizando tecnologias diferentes, para o problema da integração na Web. Entretanto, a maioria das abordagens permite apenas a integração de dados estáticos, não dispondo de meios para lidar com os processos oferecidos pelos portais Web. Além disso, quase todas as soluções propostas não são completamente automatizadas. Um outro ponto negativo é o fato da maioria das aplicações serem específicas para algum determinado domínio de aplicação, não existindo nenhuma proposta para o domínio da Geologia.

O presente trabalho propôs um sistema, denominado W-KIS (*Web Knowledge Integration System*), o qual faz uso de Serviços Web Semânticos para prover a busca e a integração de dados e processos oferecidos por diferentes portais disponibilizados na Web. O Serviço Web Semântico é um paradigma de pesquisa que, através da utilização de ontologias de domínio, é capaz de lidar com a heterogeneidade semântica entre as fontes de informação. Além disso, ele estabelece uma plataforma comum para a integração de aplicações distribuídas, fornecendo meios para automatizar os processos de descoberta, composição, execução e monitoramento dos serviços, no ambiente da Web.

Através da utilização de Serviços Web Semânticos, este trabalho propõe uma extensão da linguagem de descrição de Serviços Web capaz de tratar conceitos de conhecimento e seus relacionamentos. Com o desenvolvimento do sistema W-KIS, foi demonstrado que a utilização de conceitos e dos relacionamentos hierárquicos entre eles permite contextualizar as consultas do usuário, ampliando a precisão da busca.

O processo de desenvolvimento da aplicação W-KIS envolveu a utilização de ontologias de domínio existentes; a análise detalhada de cada portal Web; a modelagem de um Serviço Web para cada fonte de informação; a descrição semântica de cada serviço; o design e a implementação da arquitetura do sistema; provendo, por fim, o acesso integrado as fontes de informação. Depois de desenvolvido, o sistema foi validado através de dois estudos de caso: o primeiro no domínio da Biomedicina e o segundo no domínio da Geologia, para garantir o desenvolvimento de uma solução de

consulta que fosse eficaz independentemente do domínio de aplicação. W-KIS foi capaz de retornar respostas válidas e contextualizadas ao domínio no qual a aplicação estava atuando, provando sua capacidade de atuar em mais de um domínio de aplicação. W-KIS ainda foi avaliado em comparações com outras abordagens para integração na Web e com mecanismos de busca também na Web.

A principal contribuição do sistema desenvolvido no presente trabalho é a ampliação da capacidade descritiva do Serviço Web, disponibilizando um mecanismo de consulta baseado na intenção do usuário. Através do conhecimento agregado pela ontologia de domínio, o sistema atua em áreas de conhecimento específicas, respondendo a consulta do usuário de acordo com o contexto de seu interesse. Através dos Serviços Web Semânticos, W-KIS localiza e executa os serviços de forma automática, retornando os resultados fornecidos por cada portal Web diretamente para o usuário. Além disso, a aplicação desenvolvida provê uma arquitetura para um sistema de integração para a Web contendo quatro componentes principais básicos (interface de usuário, componente de consulta, componente de execução e componente *core*), os quais podem ser utilizados para mais de um domínio de aplicação. Os componentes centrais do sistema realizam a integração de acordo com o domínio especificado e vinculado ao sistema através da ontologia de domínio e das descrições semânticas dos serviços.

Apesar de todas as vantagens citadas, o sistema W-KIS ainda apresenta deficiências. A primeira limitação é o fato da seleção dos portais Web a serem utilizados como fontes de informação no sistema ser realizada manualmente. Além disso, alterações nas operações oferecidas pelos portais Web causam desatualização nos serviços vinculados ao sistema e na descrição semântica dos mesmos. Ou seja, se as operações oferecidas por um portal encapsulado por um Serviço Web Semântico são modificadas de alguma forma, tanto o serviço como a sua descrição semântica também devem ser modificados, isto é, um novo serviço ser gerado e uma nova descrição semântica deve ser realizada, caso contrário o sistema torna-se incapaz de acessar tal portal. É importante ressaltar que quando alterações são realizadas apenas sobre os dados que alimentam algum portal, nenhuma alteração é necessária no sistema o qual acessa sempre a versão mais atualizada do portal Web.

6.1 Sugestões para Trabalhos Futuros

Como mencionado anteriormente, o sistema W-KIS ainda pode ser melhorado. Como primeira sugestão para trabalho futuro, sugere-se a implementação de um componente adicional ao sistema o qual seja responsável pelo monitoramento da Web com o objetivo de detectar automaticamente quando novos portais Web relacionados ao domínio de interesse são disponibilizados na rede. Quando detectados, estes portais devem então ser adicionados ao conjunto de fontes de informação pelo desenvolvedor.

A segunda sugestão é a realização da busca e seleção de mais portais Web tanto no domínio da Biomedicina como no da Geologia, com o objetivo de melhorar o conjunto de testes para a validação do sistema W-KIS. Uma outra sugestão é a extensão do estudo de caso, através da adição de outros domínios de aplicação, com o objetivo de tornar o sistema uma solução independente de domínio para o problema da integração das funcionalidades de portais Web através da utilização de Serviços Web Semânticos.

Além disso, com o amadurecimento da tecnologia de Serviços Web Semânticos e com o desenvolvimento completo das ferramentas para esta tecnologia, será possível explorar a semântica de outros relacionamentos além dos hierárquicos de forma a

capturar a semântica descrita nas ontologias e aplicar esta melhoria no aperfeiçoamento da busca na Web.

REFERÊNCIAS

ABEL, M. **Estudo da Perícia em Petrografia Sedimentar e sua Importância para a Engenharia de Conhecimento**. 2001. 239p. Tese (Doutorado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre.

ABEL, M.; SILVA, L. A. L.; DE ROS, L. F.; MASTELLA, L. S.; CAMPBELL, J. A.; NOVELLO, T. PetroGrapher: Managing Petrographic Data and Knowledge Using an Intelligent Database Application. **Expert Systems with Applications**, New York, v.26, n.1, p. 9-18, Jan. 2004.

ACUÑA, C. J. et al. Toward Web Portals Integration through Semantic Web Services. In: INTERNATIONAL CONFERENCE ON NEXT GENERATION WEB SERVICES PRACTICES, NWESP, 2005. **Proceedings...** Washington, DC: IEEE Computer Society, 2005. p. 22-26.

ALESSO, H. P.; SMITH, C. F. **Developing Semantic Web Services**. Canada: AK Peters, 2005.

ALONSO, G. et al. **Web Services: Concepts, Architectures and Applications**. Berlin: Springer, 2004.

ANTONIOU, G.; VAN HARMELEN, F. **A Semantic Web Primer**. London: The MIT Press, 2004.

BELIAN, R. B.; SALGADO, A. C. Aspectos Semânticos em um Sistema de Integração de Informações na Web. In: BRAZILIAN WORKSHOP ON SEMANTIC WEB, WWS, 1., 2004. **Proceedings...** Brasília: CEUR-WS, 2004.

BERGAMASCHI, S. et al. An Intelligent Approach to Information Integration. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGIES IN INFORMATION SYSTEMS, FOIS, 1., 1998. **Proceedings...** [S.l.]: IOS Press, 1998. p. 253-267.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, New York, v. 284, n.5, p. 34-43, May 2001.

BERTINO, E.; FERRARI, E. XML and Data Integration. **IEEE Internet Computing**, Piscataway, NJ, v.5, n.6, p. 75-76, Nov. 2001.

BORST, W.N. **Construction of engineering ontologies**. 1997. PhD Thesis. University of Twente, Enschede.

CABRAL, L. et al. Approaches to Semantic Web Services: an Overview and Comparisons. In: EUROPEAN SEMANTIC WEB SYMPOSIUM, ESWS, 1., 2004. **The Semantic Web: Research and Applications: proceedings**. Berlin: Springer-Verlag, 2004. p. 225-239. (Lecture Notes in Computer Science, v. 3053).

CABRAL, L. et al. IRS-III : A Broker for Semantic Web Services based Applications. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, ISWC, 5., 2006. **Proceedings...** [S.l. : s.n.], 2006.

CHINNICI, R. et al. **Web Services Description Language (WSDL)**. Version 2.0, Part 1: Core Language. W3C Recommendation. 2007. Disponível em : <<http://www.w3.org/TR/wsdl20/>>. Acesso em: nov. 2006.

CRUZ, I. F.; XIAO, H. The Role of Ontologies in Data Integration. **International Journal of Engineering Intelligent Systems**, [S.l.], v.13, n.4, Dec. 2005.

DE MELLO, M. T.; ABEL, M.; GARCÍA-SÁNCHEZ, F. Using Semantic Web Services to Integrate Data and Processes from Different Web Portals. In: INTERNATIONAL WORKSHOP ON INTELLIGENT WEB BASED TOOLS, IWBT; IEEE ICTAI, 19., 2007. **Proceedings...** [S.l. : s.n.], 2007. p. 9-16.

DEAN, M.; SCHREIBER, G. **OWL Web Ontology Language Reference**: W3C Recommendation. 2004. Disponível em: <<http://www.w3.org/TR/owl-ref/>>. Acesso em: mar. 2007.

DIP Integrated Project. **DIP Annual Project Report**. Dec. 2006. Disponível em: <<http://dip.semanticweb.org/>>. Acesso em: ago. 2007.

ELENIUS, D. et al. The OWL-S Editor – A Development Tool for Semantic Web Services. In: EUROPEAN SEMANTIC WEB CONFERENCE, ESWC, 2., 2005. Berlin: Springer-Verlag, 2005. p. 78-92. (Lecture Notes in Computer Science, v. 3532).

EYAL, A.; MILO, T. Integrating and Customizing Heterogeneous E-commerce Applications. **The International Journal on Very Large Data Bases**, New York, v.10, n.1, p. 16-38, Aug. 2001.

FARRELL, J.; LAUSEN, H. **Semantic Annotations for WSDL and XML Schema**. W3C Recommendation. 2007. Disponível em: <<http://www.w3.org/TR/sawSDL/>>. Acesso em: set. 2007.

GARDARIN, G. et al. SEWISE: An Ontology-based Web Information Search Engine. In: INTERNATIONAL CONFERENCE ON APPLICATIONS OF NATURAL LANGUAGE TO INFORMATION SYSTEMS, NLDB, 2003. **Proceedings...** [S.l. : s.n.], 2003. p. 106-119.

GOMEZ, J. M.; RICO, M.; GARCÍA-SÁNCHEZ, F.; LIU, Y.; DE MELLO, M. T. BIRD: Biomedical Information Integration and Discovery with Semantic Web Services. In: INTERNATIONAL WORK CONFERENCE ON THE INTERPLAY BETWEEN NATURAL AND ARTIFICIAL COMPUTATION, IWINAC, 2., 2007, La Manga del Mar Menor, Spain. **Nature Inspired Problem – Solving Methods in Knowledge Engineering: proceedings**. Berlin: Springer-Verlag, 2007. p. 561-570. (Lecture Notes in Computer Science, v. 4528).

GÓMEZ-PÉREZ, A. Ontological engineering: a state of the art. **Expert Update**, [S.l.], v.2, n.3, p. 33-43, 1999. Disponível em: <<http://citeseer.ist.psu.edu/444416.html>>. Acesso em: out. 2006.

GÓMEZ-PÉREZ, A.; FERNANDEZ-LÓPEZ, M.; CORCHO, O. **Ontological Engineering**. London: Springer-Verlag, 2004.

GOOGLE. **Por que usar o Google**. Tudo sobre o Google, 2008. Disponível em: <http://www.google.com.br/intl/pt-BR/why_use.html>. Acesso em: out. 2007.

GRAU, B.C. et al. Next Steps for OWL. In: OWL EXPERIENCES AND DIRECTIONS WORKSHOP, OWLED, 2., 2006. **Proceedings...** [S.l.: s.n.], 2006.

HEFLIN, J.; HENDLER, J. Searching the Web with SHOE. In: ARTIFICIAL INTELLIGENCE FOR WEB SEARCH, 2000. **Proceedings...** Menlo Park, CA: AAAI Press, 2000. p. 35-40.

HANSEN, M.; MADNICK, S.; SIEGEL, M. Data Integration Using Web Services. In: VLDB WORKSHOP EEXIT AND CAISE WORKSHOP DTWEB ON EFFICIENCY AND EFFECTIVENESS OF XML TOOLS AND TECHNIQUES AND DATA INTEGRATION OVER THE WEB, 2002. **Proceedings...** Berlin: Springer-Verlag, 2002. p. 165-182. (Lecture Notes in Computer Science, v. 2590).

HSU, C. et al. Reconfigurable Web Wrapper Agents for Web Information Integration. In: IJCAI WORKSHOP ON INFORMATION INTGRATION ON THE WEB, IWeb, 2003. **Proceedings...** [S.l.], 2003. p.15-20. Disponível em: <<http://www.isi.edu/info-agents/workshops/ijcai03/papers/Hsu2-wndl20032.pdf>>. Acesso em: dez. 2006.

HUSEK, D. et al. Meta-Search Utilizing Evolutionary Recommendation: A Web Search Architecture Proposal. In: INTERNATIONAL WORKSHOP ON INTELLIGENT WEB BASED TOOLS, IWBT; IEEE ICTAI, 19., 2007. **Proceedings...** [S.l.: s.n.], 2007. p. 27-33.

IVES, Z.G. et al. An Adaptative Query Execution System for Data Integration. **ACM SIGMOD Record**. New York, USA, v. 28, n. 2, p. 299-310, June 1999.

KAMBHAMPATI, S.; KNOBLOCK, C. Guest Editor's Introduction: Information Integration on the Web. **IEEE Intelligent Systems**, [S.l.], v.18, n. 5, p. 14-15, Sept./Oct. 2003.

KÖHLER, J.; SCHULZE-KREMER, S. The Semantic Metadatabase (SEMEDA): Ontology based integration of federated molecular biological data sources. **Silico Biology**, [S.l.], v.2, n. 0021, 2002. Disponível em: <<http://www.bioinfo.de/isb/2002/02/0021/main.html>>. Acesso em: out. 2006.

LEVY, A.Y. Logic-Based Techniques in Data Integration. In: MINKER, J. (Ed.) **Logic-Based Artificial intelligence**. Norwell, MA: Kluwer Academic Publishers, 2000. p. 575-595. (Kluwer International Series in Engineering and Computer Science, v. 597).

MARQUET, G.; GOLBREICH, C.; BURGUN, A. From an Ontology-based Search Engine Towards a More Flexible Integration for Medical and Biological Information. In: SEMANTIC INTEGRATION WORKSHOP, 2003, Sanibel, Island. **Proceedings...** Florida: CEUR-WS, 2003.

MARTIN, D. et al. Bringing Semantics to Web Services: The OWL-S Approach. In: INTERNATIONAL WORKSHOP ON SEMANTIC WEB SERVICES AND WEB PROCESS COMPOSITION, SWSWPC, 1., 2004, San Diego, CA. **Proceedings...** Berlin: Springer-Verlag, 2005. p. 26-42.

MCGUINNESS, D. L.; VAN HARMELEN, F. **OWL Web Ontology Language Overview**. W3C Recommendation. 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em: jun. 2006.

MCILRAITH, S.; SON, T. C.; ZENG, H. Semantic Web Services. **IEEE Intelligent Systems**, [S.l.], v.16, n.2, p. 46-53, 2001.

MENA, E. et al. OBERVER: An Approach for Query Processing in Global Information Systems Based on Interoperation Across Pre-Existing Ontologies. **Distributed and Parallel Databases**, Hingham, MA, v.8, n.2, p.223-271, 2000.

PAOLUCCI, M. et al. Semantic Matching of Web Services Capabilities. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, ISWC, 1., 2002. **Proceedings...** Berlin: Springer-Verlag, 2002. p. 333-347. (Lecture Notes in Computer Science, v. 2342).

ROMAN, D. et al. Web Service Modeling Ontology, **Applied Ontology**, [S.l.], v.1, n.1, p. 77-106, 2005.

SCHREIBER, G. et al. **Knowledge engineering and management: the CommonKADS methodology**. Cambridge: The MIT Press, 2000.

TOSIC, V.; PATEL, K.; PAGUREK, B. WSOL – Web Service Offerings Language. In: INTERNATIONAL WORKSHOP ON WEB SERVICES, E-BUSINESS, AND THE SEMANTIC WEB, 2002. **Proceedings...** Berlin: Springer-Verlag, 2002. p. 57-67. (Lecture Notes in Computer Science, v. 2512).

USCHOLD, M.; GRUNINGER, M. Ontologies and Semantics for Seamless Connectivity. **ACM SIGMOD Record**, [S.l.], v.33, n.4, p. 58-64, Dec. 2004.

VASILIU, L.; HARAND, S.; CIMPIAN, E. The DIP Project: Enabling Systems & Solutions for Processing Digital Content with Semantic Web Services. In: EUROPEAN WORKSHOP ON THE INTEGRATION OF KNOWLEDGE, SEMANTICS AND DIGITAL MEDIA TECHNOLOGY, EWIMT, 2004. **Proceedings...** [S.l. : s.n.], 2004.

VINOSKI, S. Integration with Web Services. **IEEE Internet Computing**, Piscataway, NJ, v.7, n.6, p. 75-77, 2003.

W3C. **Semantic Web Activity**. 2003. Disponível em: <<http://www.w3.org/2001/sw/>>. Acesso em: maio 2006.

ZHU, F. et al. Dynamic Data Integration Using Web Services. In: IEEE INTERNATIONAL CONFERENCE ON WEB SERVICES, ICWS, 2004, San Diego, CA. **Proceedings...** Washington, DC: IEEE Computer Society, 2004. p. 262-269.

ANEXO A ONTOLOGIA *ONCOGEN*

Concept: General	
Is-a	Object
Name	string
DataOrigin	date
Synonymous	string
XReference	string

Concept: Cancer	
Is-a	General
CancerName	string
CancerComplications	string
CancerNutrition	string
CancerSymptoms	string
CancerCauses	string
RiskFactors	string
Tests	string
ClinicalTrials	string
IsProducedBy	list-of [instance-of < Oncogen >]
ActivateTo	one-of [instance-of < Gene >]

Concept: Cancer Types	
Is-a	Cancer

Concept: Cancer by Body Location	
Is-a	Cancer

Concept: Carcinogenesis	
Is-a	General
Is-aProcessOf	one-of [instance-of < Cancer >]

Concept: Cell	
Is-a	General
CellName	string
CellParts	string
HasPhase	list-of [instance-of < Cell Cycle >]
HasProtein	one-of [instance-of < Protein >]

Concept: Cell Cycle	
Is-a	General
Is-aCellPhase	one-of [instance-of < Cell >]

Concept: Chromosomes	
Is-a	General
ChromosomeName	string
CodingRegion	string
CPGIsland	string
Exon	string
ImmunoglobulinSwitch	string
Intron	string
PolyAAdditionSite	string
RegulatoryElements	string
TATABox	string
UntranslatedRegion3	string
UntranslatedRegion5	string

Concept: Gene	
Is-a	General
Part-of	Chromosomes
GeneName	string
GeneCode	string
GeneLocation	string
HasMutation	one-of [instance-of <Mutations >]

Concept: DNA	
Is-a	Gene
Part-of	Gene
DNACode	string
DNAFunction	string
NucleotidCode	char, one-of [A, C, T, G]
NucleotidName	string
TranscriptTo	one-of [instance-of < mRNA >]

Concept: mRNA	
Is-a	Gene
mRNACode	string
mRNAFunction	string
Synthesized	one-of [instance-of <Polypeptide >]

Concept: Nucleotid	
Is-a	Gene
Part-of	DNA
Sequence	string

Concept: Nitrogen Base	
Is-a	Nucleotid
CodeBaseNitrogen	char, one-of [A, C, T, G, U]
NameBaseNitrogen	string, one-of [adenine, cytosine, thymine, guanine, uracil]

Concept: Pentosa

Is-a	Nucleotid
------	-----------

Concept: Phosphoric Acid

Is-a	Nucleotid
------	-----------

Concept: Oncogen

Is-a	Gene
------	------

HasMutations	list-of [instance-of < Mutations >]
--------------	-------------------------------------

Concept: DNA Repair

Is-a	Oncogen
------	---------

Concept: Promotors Oncogen

Is-a	Oncogen
------	---------

Concept: Tumor Suppressor Gene

Is-a	Oncogen
------	---------

Concept: Mutations

Is-a	General
------	---------

MutationName	string
--------------	--------

MutationCode	string
--------------	--------

MutationSynonymous	String
--------------------	--------

Concept: Amplification

Is-a	Mutations
------	-----------

Produces	one-of [instance-of < Protein Kinases >]
----------	------------------------------------------

Concept: Deletion

Is-a	Mutations
------	-----------

Concept: Insertions	
Is-a	Mutations

Concept: Inversion	
Is-a	Mutations

Concept: Translocation	
Is-a	Mutations
ActivateTo	one-of [instance-of < Gene >]

Concept: Protein	
Is-a	General
Part-of	Cell
ProteinName	string
ProteinClassification	string
ProteinCode	string
ProteinStructure	string
ProteinFunction	string

Concept: Aminoacids	
Is-a	Protein
AminoacidName	string
AminoacidCode	string
Is-aChainAminoacids	one-of [instance-of < Polypeptide >]

Concept: Polypeptide	
Is-a	Aminoacids
PolypeptideName	string
PolypeptideCode	string
SynthesizedBy	one-of [instance-of <mRNA >]

Concept: Cyclin Protein	
Is-a	Protein

Concept: Dependent Cyclin Protein Kinases

Is-a

Protein

Concept: Dependent Cyclin Proteins Inhibit Kinases

Is-a

Protein

Concept: Protein GT Pasa Activity

Is-a

Protein

Concept: Protein Kinases

Is-a

Protein

ANEXO B ONTOLOGIA *PETROGRAPHER*

Concept: General	
Is-a	Object

Concept: Rock	
Is-a	General
RockName	string
IsMadeOf	list-of [instance-of < Constituent >]

Concept: Igneous Rock	
Is-a	Rock

Concept: Metamorphic Rock	
Is-a	Rock

Concept: Sedimentary Rock	
Is-a	Rock

Concept: Carbonaceous Sedimentary Rock	
Is-a	Sedimentary Rock

Concept: Chemical/Bioconstructed Sedimentary Rock	
Is-a	Sedimentary Rock

Concept: Limestone Sedimentary Rock	
Is-a	Sedimentary Rock

Concept: Limestone Carbonate	
ls-a	Limestone Sedimentary Rock

Concept: Siliciclastic Sedimentary Rock	
ls-a	Sedimentary Rock
ConcaveConvexContacts	string(20), one-of [abundant, common, rare, trace]
GrainSizes	string(30), one-of [boulder, cobble, pebble, granule, gravel, very coarse sand, coarse sand, medium sand, fine sand, very fine sand, silt, clay]
LongContacts	string (20), one-of [abundant, common, rare, trace]
ModalGrainSizes	string (30), one-of [gravel, very coarse sand, coarse sand, medium sand, fine sand, very fine sand, silt, clay]
Orientation	string (20), one-of [homogeneous, parallel, oriented, imbricated, heterogeneous, chaotic]
Packing	string (20), one-of [loose, normal, tight, heterogeneous]
PointContacts	string (20), one-of [abundant, common, rare, trace]
Roundness	string (20), one-of [well rounded, rounded, sub-rounded, sub-angular, angular]
RoundnessModifiers	string (30), one-of [soft intraclasts, intraclasts, pressure dissolution, corrosion / deformation, replacement, overgrowths]
Shape	string (20), one-of [equant, oblate, prolate, spheroidal, rod, blade, disc]
Sorting	string (30), one-of [very well sorted, well sorted, sorted, moderately sorted, poorly sorted]
Sphericity	string (20), one-of [high, medium, low]
Structures	string (30), one-of [parallel bedding, normal-graded bedding, reverse-graded bedding, massive bedding, parallel lamination, climbing lamination, liner lamination, flaser lamination, massive, bioturbation, fluidized, load, folded, fractured, stylolites, nodules, crystalline, crusts, roots, spotted, vesicular, amygdaloidal, flow, bioconstruction, bioaccumulation]
Support	string (30), one-of [grain-supported, grain to matrix-supported, grain to cement-supported, matrix-supported, matrix to cement-supported, cement-supported]
SuturedContacts	string (20), one-of [abundant, common, rare, trace]

Concept: Limestone Dolomite	
Is-a	Limestone Sedimentary Rock

Concept: Siliciclastic Conglomerate	
Is-a	Siliciclastic Sedimentary Rock

Concept: Siliciclastic Sandstone	
Is-a	Siliciclastic Sedimentary Rock

Concept: Siliciclastic Shale	
Is-a	Siliciclastic Sedimentary Rock

Concept: Constituent	
Is-a	General
Part-of	Rock
ConstituentName	string

Concept: Diagenetic	
Is-a	Constituent
Habit	string (30), one-of [blocky, booklet, botryoid, bridge, coarsely-crystalline, coarse mosaic, coating, discrete crystal, fibro-radiated, fibrous, fine mosaic, framboid, ingrowth, internal sediment, lamella, large rhomb, massive, meniscus, microcrystalline, ooid, outgrowth, parallel-prismatic, pelletoid, peloid, pigment, poikilotopic, prismatic, prismatic-radiated, radiated, rim, rosette, sheaf, small rhomb, spherulite, vermicule]
Location	string(30), one-of [intergranular continuous pore-lining, intergranular discontinuous pore-lining, intergranular pore-filing, intergranular discrete, intergranular displacive, intragranular replacive, intragranular pore-lining, intragranular pore-filling, intragranular discrete crystals, intragranular displacive, moldic pore-lining, moldic pore-filling, oversized pore-lining, oversized pore-filling, grain fracture-filling, grain fracture-lining, rock fracture-filling, rock fracture-lining, concretions/nodules, massive beds/lenses]
NominalAmount	string (20), one-of [abundant, common, rare, trace]
Modifier	string (20), one-of [dissolved, zoned, fractures, recrystallized]

Concept: Macroporosity	
Is-a	Constituent
Name	string (30), one-of [intergranular porosity, intragranular porosity, intracrystalline porosity, intercrystalline porosity, moldic porosity, grain fracture porosity, rock fracture porosity, shrinkage porosity, oversized porosity, lamellar porosity]
Location	string(30), one-of [conceptwork, interstitial, conceptwork and interstitial]
NominalAmount	string (20), one-of [abundant, common, rare, trace]
Modifier	string (30), one-of [reduced by compaction, reduced by cementation, enlarged by dissolution]

Concept: Primary Constituent	
Is-a	Constituent
Location	string(30), one-of [in metamorphic rock fragment, in plutonic rock fragment, in sedimentary rock fragment, in volcanic rock fragment, in intrabasinal fragment, as monomineralic grain]
NominalAmount	string (20), one-of [abundant, common, rare, trace]

Constituents' subclasses:

[Primary Constituents]
1. Detrital quartz, Detrital quartz, 1
2. Detrital quartz monocrystalline, Detrital quartz, 3
3. Detrital quartz monocrystalline common plutonic, Detrital quartz, 5
4. Detrital quartz monocrystalline volcanic, Detrital quartz, 6
5. Detrital quartz monocrystalline hydrothermal, Detrital quartz, 7
6. Detrital quartz monocrystalline with abrupt extinction, Detrital quartz, 8
7. Detrital quartz monocrystalline with slightly undulose extinction, Detrital quartz, 9
8. Detrital quartz monocrystalline with strongly undulose extinction, Detrital quartz, 10
9. Detrital quartz polycrystalline, Detrital quartz, 11
10. Detrital quartz polycrystalline granoblastic, Detrital quartz, 13
11. Detrital quartz polycrystalline hydrothermal, Detrital quartz, 14
12. Detrital quartz polycrystalline planar, Detrital quartz, 15
13. Detrital quartz polycrystalline stretched, Detrital quartz, 16
14. Detrital feldspar, Detrital feldspar, 50
15. Detrital K-feldspar, Detrital feldspar, 52
16. Detrital microcline, Detrital feldspar, 54
17. Detrital orthoclase, Detrital feldspar, 55
18. Detrital perthite, Detrital feldspar, 56
19. Detrital sanidine, Detrital feldspar, 57

20. Detrital plagioclase, Detrital feldspar, 58
21. Detrital plagioclase andesine, Detrital feldspar, 60
22. Detrital plagioclase anorthite, Detrital feldspar, 61
23. Detrital plagioclase albite, Detrital feldspar, 62
24. Detrital plagioclase bytownite, Detrital feldspar, 63
25. Detrital plagioclase oligoclase, Detrital feldspar, 64
26. Plutonic rock fragment, Plutonic rock fragment, 150
27. Granitic/gneissic plutonic rock fragment, Plutonic rock fragment, 152
28. Basic/ultrabasic plutonic rock fragment, Plutonic rock fragment, 153
29. Volcanic rock fragment, Volcanic rock fragment, 200
30. Volcanic rock fragment with felsitic texture, Volcanic rock fragment, 202
31. Volcanic rock fragment with lathwork/orthitic texture, Volcanic rock fragment, 203
32. Volcanic rock fragment with microlitic texture, Volcanic rock fragment, 204
33. Volcanic shard, Volcanic rock fragment, 205
34. Volcanic pumice fragment, Volcanic rock fragment, 206
35. Volcanic scoria fragment, Volcanic rock fragment, 207
36. Volcanic sideromelane fragment, Volcanic rock fragment, 208
37. Volcanic perlitic fragment, Volcanic rock fragment, 209
38. Volcanic vitroclastic fragment, Volcanic rock fragment, 210
39. Sedimentary rock fragment, Sedimentary rock fragment, 250
40. Shale rock fragment, Sedimentary rock fragment, 254
41. Claystone rock fragment, Sedimentary rock fragment, 255
42. Mudrock fragment, Sedimentary rock fragment, 252
43. Siltstone rock fragment, Sedimentary rock fragment, 257
44. Sandstone rock fragment, Sedimentary rock fragment, 258
45. Conglomerate rock fragment, Sedimentary rock fragment, 259
46. Limestone rock fragment, Sedimentary rock fragment, 260
47. Dolostone rock fragment, Sedimentary rock fragment, 261
48. Chert rock fragment, Sedimentary rock fragment, 262
49. Ferriferous rock fragment, Sedimentary rock fragment, 264
50. Evaporitic rock fragment, Sedimentary rock fragment, 265
51. Metamorphic rock fragment, Metamorphic rock fragment, 300
52. Slate rock fragment, Metamorphic rock fragment, 302
53. Meta-siltite rock fragment, Metamorphic rock fragment, 303
54. Meta-sandstone rock fragment, Metamorphic rock fragment, 304
55. Phyllite rock fragment, Metamorphic rock fragment, 305
56. Schist rock fragment, Metamorphic rock fragment, 306
57. Meta-carbonate rock fragment, Metamorphic rock fragment, 307
58. Marble rock fragment, Metamorphic rock fragment, 308
59. Metavolcanic rock fragment, Metamorphic rock fragment, 309
60. Amphibolite rock fragment, Metamorphic rock fragment, 309
61. Hornfelse rock fragment, Metamorphic rock fragment, 310
62. Serpentinite rock fragment, Metamorphic rock fragment, 311
63. Itabirite rock fragment, Metamorphic rock fragment, 312
64. Micas/chlorite, Micas/chlorite, 400
65. Biotite, Micas/chlorite, 402

66. Muscovite, Micas/chlorite, 403
67. Chlorite, Micas/chlorite, 404
68. Heavy minerals, Heavy minerals, 450
69. Amphibole, Heavy minerals, 452
70. Garnet, Heavy minerals, 453
71. Pyroxene, Heavy minerals, 454
72. Tourmaline, Heavy minerals, 455
73. Rutile, Heavy minerals, 456
74. Olivine, Heavy minerals, 457
75. Epidote, Heavy minerals, 458
76. Zircon, Heavy minerals, 459
77. Monazite, Heavy minerals, 460
78. Sphene, Heavy minerals, 461
79. Calcite monocrystalline detrital, Heavy minerals, 462
80. Dolomite monocrystalline detrital, Heavy minerals, 463
81. Opaques, Heavy minerals, 464
82. Intrabasinal grains, Intrabasinal grains, 500
83. Mud intraclast, Mud intraclast, 550
84. Glauconite grain, Glauconite grain, 850
85. Glauconite peloid, Glauconite grain, 852
86. Glauconite ooid, Glauconite grain, 853
87. Glauconite intraclast, Glauconite grain, 854
88. Glauconitized mica, Glauconite grain, 855
89. Talc/stevensite ooid, Clay ooid/peloid, 805
90. Calcrete intraclast, Carbonate intraclast, 952
91. Carbonate bioclast, Carbonate bioclast, 900
92. Red Algae bioclast, Carbonate bioclast, 904
93. Green Algae bioclast, Carbonate bioclast, 905
94. Coccolith bioclast, Carbonate bioclast, 906
95. Calcisphere bioclast, Carbonate bioclast, 907
96. Cyanobacteria bioclast, Carbonate bioclast, 908
97. Benthonic foraminifer bioclast, Carbonate bioclast, 910
98. Planktonic foraminifer bioclast, Carbonate bioclast, 911
99. Macroforaminifer bioclast, Carbonate bioclast, 912
100. Coral bioclast, Carbonate bioclast, 926
101. Bryozoan bioclast, Carbonate bioclast, 927
102. Brachiopod bioclast, Carbonate bioclast, 928
103. Pelecypod bioclast, Carbonate bioclast, 914
104. Gastropod bioclast, Carbonate bioclast, 915
105. Cephalopod bioclast, Carbonate bioclast, 916
106. Echinoid bioclast, Carbonate bioclast, 918
107. Crinoid bioclast, Carbonate bioclast, 919
108. Ostracode bioclast, Carbonate bioclast, 922
109. Trilobite bioclast, Carbonate bioclast, 923
110. Crustacean bioclast, Carbonate bioclast, 924
111. Annelid (worm) bioclast, Carbonate bioclast, 925
112. Carbonate intraclast, Carbonate intraclast, 950
113. Calcite carbonate intraclast, Carbonate intraclast, 954

114.Dolomite carbonate intraclast,Carbonate intraclast,955
115.Siderite carbonate intraclast,Carbonate intraclast,956
116.Microcrystalline carbonate intraclast,Carbonate intraclast,957
117.Coarsely crystalline carbonate intraclast,Carbonate intraclast,958
118.Carbonate peloid,Carbonate peloid/pellet,1051
119.Limeclast,Other detrital constituents,1101
120.Siliceous intraclast,Siliceous grain,601
121.Silcrete intraclast,Siliceous grain,602
122.Silicified wood,Siliceous grain,603
123.Phosphate intraclast,Phosphate grain,655
124.Conodont bioclast,Phosphate grain,652
125.Fish bioclast,Phosphate grain,653
126.Phosphatic Brachiopod bioclast,Phosphate grain,654
127.Phosphate peloid,Phosphate grain,656
128.Phosphate ooid,Phosphate grain,657
129.Ferriferous intraclast,Ferriferous grain,701
130.Ferriferous ooid,Ferriferous grain,702
131.Evaporitic intraclast,Evaporitic grain,751
132.Gypsum monocrystalline detrital,Evaporitic grain,752
133.Carbonaceous fragment,Other detrital constituents,1102
134.Alterite fragment,Other detrital constituents,1103
135.Undetermined grains,Other detrital constituents,1104
136.Detrital matrix,Detrital matrix,1150
137.Siliciclastic mud matrix syndepositional,Detrital matrix,1153
138.Siliciclastic mud matrix from bioturbation,Detrital matrix,1154
139.Siliciclastic mud matrix from fluidization,Detrital matrix,1155
140.Micritic matrix syndepositional,Detrital matrix,1159
141.Micritic matrix from bioturbation,Detrital matrix,1160
142.Micritic matrix from fluidization,Detrital matrix,1161
143.Marly matrix syndepositional,Detrital matrix,1163
144.Marly matrix from bioturbation,Detrital matrix,1164
145.Marly matrix from fluidization,Detrital matrix,1165
146.Detrital quartz undifferentiated,Detrital quartz,2
147.Detrital quartz monocrystalline undifferentiated,Detrital quartz,4
148.Detrital quartz polycrystalline undifferentiated,Detrital quartz,12
149.Detrital feldspar undifferentiated,Detrital feldspar,51
150.Detrital K-feldspar undifferentiated,Detrital feldspar,53
151.Detrital plagioclase undifferentiated,Detrital feldspar,59
152.Plutonic rock fragment undifferentiated,Plutonic rock fragment,151
153.Volcanic rock fragment undifferentiated,Volcanic rock fragment,201
154.Sedimentary rock fragment undifferentiated,Sedimentary rock fragment,251
155.Mudrock fragment undifferentiated,Sedimentary rock fragment,253
156.Metamorphic rock fragment undifferentiated,Metamorphic rock fragment,301
157.Mica/chlorite undifferentiated,Micas/chlorite,401
158.Heavy minerals undifferentiated,Heavy minerals,451
159.Intrabasinal grain undifferentiated,Intrabasinal grains,501

160.Phosphate bioclast undifferentiated,Phosphate grain,651
161.Glaucinite grain undifferentiated,Glaucinite grain,851
162.Carbonate pellet,Carbonate peloid/pellet,1050
163.Carbonate bioclast undifferentiated,Carbonate bioclast,901
164.Algae bioclast undifferentiated,Carbonate bioclast,902
165.Foraminifer bioclast undifferentiated,Carbonate bioclast,909
166.Mollusk bioclast undifferentiated,Carbonate bioclast,913
167.Echinoderm bioclast undifferentiated,Carbonate bioclast,917
168.Arthropod bioclast undifferentiated,Carbonate bioclast,921
169.Carbonate intraclast undifferentiated,Carbonate intraclast,951
170.Dolocrete intraclast,Carbonate intraclast,953
171.Other detrital constituents,Other detrital constituents,1100
172.Detrital matrix undifferentiated,Detrital matrix,1151
173.Siliciclastic mud matrix undifferentiated,Detrital matrix,1152
174.Carbonate matrix undifferentiated,Detrital matrix,1156
175.Carbonate pelletoidal matrix,Detrital matrix,1157
176.Micritic matrix undifferentiated,Detrital matrix,1158
177.Marly matrix undifferentiated,Detrital matrix,1162
178.Mud intraclast undifferentiated,Mud intraclast,551
179.Argillaceous mud intraclast,Mud intraclast,552
180.Siltic mud intraclast,Mud intraclast,553
181.Soil intraclast,Mud intraclast,554
182.Siliceous grain,Siliceous grain,600
183.Siliceous bioclast,Siliceous grain,604
184.Phosphate grain,Phosphate grain,650
185.Ferriferous grain,Ferriferous grain,700
186.Lateritic grain,Ferriferous grain,704
187.Evaporitic grain,Evaporitic grain,750
188.Clay ooid/peloid,Clay ooid/peloid,800
189.Clay ooid/peloid undifferentiated,Clay ooid/peloid,801
190.Chamosite/berthierine ooid,Clay ooid/peloid,802
191.Chamosite/berthierine peloid,Clay ooid/peloid,803
192.Carbonate oolith,Carbonate ooid,1002
193.Rock fragment,Rock fragment,100
194.Rock fragment undifferentiated,Rock fragment,101
195.Mudstone rock fragment,Sedimentary rock fragment,256
196.Detrital Accessories,Detrital Accessories,350
197.Detrital accessories undifferentiated,Detrital Accessories,351
198.Carbonate oncolith,Carbonate ooid,1003
199.Carbonate ooid undifferentiated,Carbonate ooid,1001
200.Carbonate ooid,Carbonate ooid,1000
201.Carbonate pisolith,Carbonate ooid,1004
202.Phosphatic rock fragment,Sedimentary rock fragment,263
203.Detrital undifferentiated,Detrital undifferentiated,0
[Diagenetic Constituents]
1. feldspar undifferentiated,Diagenetic Feldspar,51

2. Infiltrated clay, Infiltrated clays, 100
3. Pseudomatrix clay, Pseudomatrix clays, 150
4. Zeolite undifferentiated, Zeolite, 251
5. carbonate undifferentiated, Diagenetic Carbonate, 301
6. Sulphate undifferentiated, Diagenetic Sulphate, 351
7. Sulfide undifferentiated, Diagenetic Sulfide, 401
8. Titanium mineral undifferentiated, Diagenetic Titanium minerals, 501
9. Diagenetic silica undifferentiated, Diagenetic Silica, 1
10. Diagenetic silica, Diagenetic Silica, 0
11. Quartz, Diagenetic Silica, 2
12. Chalcedony, Diagenetic Silica, 3
13. Opal, Diagenetic Silica, 4
14. Opal-A, Diagenetic Silica, 5
15. Opal-CT, Diagenetic Silica, 6
16. Diagenetic feldspar, Diagenetic Feldspar, 50
17. K-feldspar, Diagenetic Feldspar, 52
18. Albite, Diagenetic Feldspar, 53
19. Authigenic clay mineral, Authigenic clays, 200
20. Kaolin, Authigenic clays, 202
21. Kaolinite, Authigenic clays, 204
22. Dickite, Authigenic clays, 205
23. Smectite, Authigenic clays, 206
24. Illite/smectite, Authigenic clays, 207
25. Illite, Authigenic clays, 208
26. Chlorite, Authigenic clays, 209
27. Chlorite/smectite, Authigenic clays, 210
28. Berthierine/chamosite, Authigenic clays, 211
29. Glauconite, Authigenic clays, 212
30. Attapulgite/sepiolite, Authigenic clays, 213
31. Diagenetic zeolite, Zeolite, 250
32. Analcime, Zeolite, 252
33. Fibrous zeolites, Zeolite, 253
34. Chabasite, Zeolite, 254
35. Laumontite, Zeolite, 255
36. Heulandite/clinoptylolite, Zeolite, 256
37. Phillipsite, Zeolite, 257
38. Diagenetic carbonate, Diagenetic Carbonate, 300
39. Aragonite, Diagenetic Carbonate, 302
40. Calcite, Diagenetic Carbonate, 303
41. Mg-calcite, Diagenetic Carbonate, 304
42. Fe-calcite, Diagenetic Carbonate, 305
43. Dolomite, Diagenetic Carbonate, 306
44. Fe-dolomite/ankerite, Diagenetic Carbonate, 307
45. Siderite, Diagenetic Carbonate, 308
46. Magnesite, Diagenetic Carbonate, 309
47. Rhodocrosite, Diagenetic Carbonate, 310
48. Diagenetic sulphate, Diagenetic Sulphate, 350
49. Gypsum, Diagenetic Sulphate, 352

50. Anhydrite, Diagenetic Sulphate, 353
51. Baryte, Diagenetic Sulphate, 354
52. Diagenetic sulfide, Diagenetic Sulfide, 400
53. Pyrite, Diagenetic Sulfide, 402
54. Sphalerite, Diagenetic Sulfide, 403
55. Galena, Diagenetic Sulfide, 404
56. Chalcopyrite, Diagenetic Sulfide, 405
57. Pyrrhotite, Diagenetic Sulfide, 406
58. Marcassite, Diagenetic Sulfide, 409
59. Diagenetic iron oxide/hydroxide, Iron oxide/hydroxide, 450
60. Hematite, Iron oxide/hydroxide, 453
61. Goethite, Iron oxide/hydroxide, 454
62. Limonite, Iron oxide/hydroxide, 455
63. Diagenetic titanium mineral, Diagenetic Titanium minerals, 500
64. Diagenetic anatase, Diagenetic Titanium minerals, 502
65. Diagenetic rutile, Diagenetic Titanium minerals, 503
66. Diagenetic brookite, Diagenetic Titanium minerals, 504
67. Diagenetic sphene, Diagenetic Titanium minerals, 505
68. Diagenetic leuco, Diagenetic Titanium minerals, 506
69. Diagenetic phosphate, Other diagenetic constituents, 602
70. Apatite, Other diagenetic constituents, 604
71. Francolite, Other diagenetic constituents, 605
72. Collophane, Other diagenetic constituents, 606
73. Diagenetic halide, Other diagenetic constituents, 607
74. Halite, Other diagenetic constituents, 609
75. Silvite, Other diagenetic constituents, 610
76. Fluorite, Other diagenetic constituents, 611
77. Diagenetic uranium/vanadium mineral, Other diagenetic constituents, 612
78. Uraninite, Other diagenetic constituents, 614
79. Pitchblende, Other diagenetic constituents, 615
80. Carnotite, Other diagenetic constituents, 616
81. Vanadinite, Other diagenetic constituents, 617
82. Diagenetic tourmaline, Other diagenetic constituents, 618
83. Diagenetic amphibole, Other diagenetic constituents, 619
84. Diagenetic pyroxene, Other diagenetic constituents, 620
85. Diagenetic epidote, Other diagenetic constituents, 621
86. Hydrocarbon, Other diagenetic constituents, 622
87. Bitumen, Other diagenetic constituents, 624
88. Paraffinic hydrocarbon, Other diagenetic constituents, 625
89. Oil stain, Other diagenetic constituents, 626
90. Infiltrated clays undifferentiated, Infiltrated clays, 101
91. Pseudomatrix clays undifferentiated, Pseudomatrix clays, 151
92. Authigenic clay mineral undifferentiated, Authigenic clays, 201
93. Kaolin undifferentiated, Authigenic clays, 203
94. Other diagenetic constituents, Other diagenetic constituents, 600
95. Other diagenetic constituents undifferentiated, Other diagenetic constituents, 601

96. Diagenetic phosphate undifferentiated, Other diagenetic constituents, 603
97. Diagenetic halide undifferentiated, Other diagenetic constituents, 608
98. Diagenetic uranium/vanadium mineral undifferentiated, Other diagenetic constituents, 613
99. Hydrocarbon undifferentiated, Other diagenetic constituents, 623
100. Polymerized bitumen halo, Other diagenetic constituents, 629
101. Bornite, Diagenetic Sulfide, 408
102. Diagenetic iron oxide/hydroxide undifferentiated, Iron oxide/hydroxide, 452
103. Chalcocite, Diagenetic sulfide, 407
[Macroporosity Constituents]
1. Macroporosity, Macroporosity, 1
2. Macroporosity Intergranular, Macroporosity, 2
3. Macroporosity Intragranular, Macroporosity, 3
4. Macroporosity Intracrystalline, Macroporosity, 4
5. Macroporosity Moldic, Macroporosity, 5
6. Macroporosity Grain Fracture, Macroporosity, 6
7. Macroporosity Rock Fracture, Macroporosity, 7
8. Macroporosity Shrinkage, Macroporosity, 8
9. Macroporosity Oversized, Macroporosity, 9
10. Macroporosity Lamellar Decompaction, Macroporosity, 10

APÊNDICE A DESCRIÇÃO WSDL DO SERVIÇO *MINDAT*

```

- <wsdl:definitions targetNamespace="http://localhost">
  <wsdl:documentation>Mindat</wsdl:documentation>
  - <wsdl:types>
    - <xs:schema attributeFormDefault="qualified" elementFormDefault="qualified"
targetNamespace="http://www.w3.org/2001/XMLSchema#xsd">
      - <xs:element name="getMineral">
        - <xs:complexType>
          - <xs:sequence>
            <xs:element name="term" nillable="true" type="xs:string"/>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
      - <xs:element name="getMineralResponse">
        - <xs:complexType>
          - <xs:sequence>
            <xs:element name="return" nillable="true" type="xs:string"/>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
    </xs:schema>
  </wsdl:types>
  - <wsdl:message name="getMineralMessage">
    <wsdl:part name="part1" element="xsd:getMineral"/>
  </wsdl:message>
  - <wsdl:message name="getMineralResponseMessage">
    <wsdl:part name="part1" element="xsd:getMineralResponse"/>
  </wsdl:message>
  - <wsdl:portType name="MindatPortType">
    - <wsdl:operation name="getMineral">
      <wsdl:input message="ns:getMineralMessage" wsaw:Action=
        "urn:getMineral"/>
      <wsdl:output message="ns:getMineralResponseMessage" wsaw:Action=
        "urn:getMineral"/>
    </wsdl:operation>
  </wsdl:portType>
  - <wsdl:binding name="MindatSOAP11Binding" type="ns:MindatPortType">
    <soap:binding transport="http://schemas.xmlsoap.org/soap/http"
      style="document"/>
    - <wsdl:operation name="getMineral">
      <soap:operation soapAction="urn:getMineral" style="document"/>
      - <wsdl:input>
        <soap:body use="literal"/>

```

```

        </wsdl:input>
      - <wsdl:output>
        <soap:body use="literal"/>
      </wsdl:output>
    </wsdl:operation>
  </wsdl:binding>
- <wsdl:binding name="MindatSOAP12Binding" type="ns:MindatPortType">
  <soap12:binding transport="http://schemas.xmlsoap.org/soap/http"
    style="document"/>
  - <wsdl:operation name="getMineral">
    <soap12:operation soapAction="urn:getMineral" style="document"/>
    - <wsdl:input>
      <soap12:body use="literal"/>
    </wsdl:input>
    - <wsdl:output>
      <soap12:body use="literal"/>
    </wsdl:output>
  </wsdl:operation>
</wsdl:binding>
- <wsdl:binding name="MindatHttpBinding" type="ns:MindatPortType">
  <http:binding verb="POST"/>
  - <wsdl:operation name="getMineral">
    <http:operation location="getMineral"/>
    - <wsdl:input>
      <mime:content type="text/xml"/>
    </wsdl:input>
    - <wsdl:output>
      <mime:content type="text/xml"/>
    </wsdl:output>
  </wsdl:operation>
</wsdl:binding>
- <wsdl:service name="Mindat">
  - <wsdl:port name="MindatSOAP11port_http" binding="
    ns:MindatSOAP11Binding">
    <soap:address location="
      http://189.32.103.223:8080/axis2/services/Mindat"/>
  </wsdl:port>
  - <wsdl:port name="MindatSOAP12port_http" binding="
    ns:MindatSOAP12Binding">
    <soap12:address location="
      http://189.32.103.223:8080/axis2/services/Mindat"/>
  </wsdl:port>
  - <wsdl:port name="MindatHttpport" binding="ns:MindatHttpBinding">
    <http:address location="
      http://189.32.103.223:8080/axis2/services/Mindat"/>
  </wsdl:port>
</wsdl:service>
</wsdl:definitions>

```

APÊNDICE B DESCRIÇÃO OWL-S DA OPERAÇÃO GETMINERAL DO SERVIÇO MINDAT

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

  xmlns:p1="http://localhost:8080/ontology/MindatSWS/getMineral.owl#"
  xmlns:j.0="http://www.daml.org/services/owl-s/1.2/Service.owl#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:j.1="http://www.daml.org/services/owl-s/1.2/Process.owl#"
  xmlns:j.2="http://www.daml.org/services/owl-s/1.2/Grounding.owl#"
  xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
  xmlns:j.3="http://www.daml.org/services/owl-s/1.2/Profile.owl#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://localhost:8080/ontology/MindatSWS/getMineral.owl">
  <owl:Ontology rdf:about="">
    <owl:imports rdf:resource="http://www.daml.org/services/owl-
s/1.2/Grounding.owl"/>
    <owl:imports rdf:resource="http://www.daml.org/services/owl-
s/1.2/Service.owl"/>
    <owl:imports rdf:resource="http://www.daml.org/services/owl-
s/1.2/Profile.owl"/>
    <owl:imports rdf:resource="http://www.daml.org/services/owl-
s/1.2/Process.owl"/>
  </owl:Ontology>
  <j.2:WsdAtomicProcessGrounding
rdf:ID="getMineralAtomicProcessGrounding">
  <j.2:owlsProcess>
    <j.1:AtomicProcess rdf:ID="getMineralProcess">
      <j.0:describes>
        <j.0:Service rdf:ID="getMineralService">
          <j.0:presents>
            <j.3:Profile rdf:ID="getMineralProfile">
              <j.3:serviceName
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
              >getMineral</j.3:serviceName>
              <j.3:textDescription
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
              >gets a mineral description from a given mineral
name</j.3:textDescription>
              <j.3:hasOutput>
                <j.1:Output rdf:ID="desc">
                  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
                  >desc</rdfs:label>
                <j.1:parameterType
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI"

```

```

>http://www.w3.org/2001/XMLSchema#string</j.1:parameterType>
  </j.1:Output>
  </j.3:hasOutput>
  <j.3:hasInput>
    <j.1:Input rdf:ID="name">
      <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >name</rdfs:label>
      <j.1:parameterType
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI "
>http://www.w3.org/2001/XMLSchema#string</j.1:parameterType>
  </j.1:Input>
  </j.3:hasInput>
  <j.0:presentedBy rdf:resource="#getMineralService"/>
  </j.3:Profile>
  </j.0:presents>
  <j.0:describedBy rdf:resource="#getMineralProcess"/>
  <j.0:supports>
    <j.2:WsdLGrounding rdf:ID="getMineralGrounding">
      <j.0:supportedBy rdf:resource="#getMineralService"/>
      <j.2:hasAtomicProcessGrounding
rdf:resource="#getMineralAtomicProcessGrounding"/>
      </j.2:WsdLGrounding>
    </j.0:supports>
  </j.0:Service>
  </j.0:describes>
  <j.1:hasInput rdf:resource="#name"/>
  <j.1:hasOutput rdf:resource="#desc"/>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >getMineralProcess</rdfs:label>
  </j.1:AtomicProcess>
  </j.2:owlsProcess>
  <j.2:wsdlOperation>
    <j.2:WsdLOperationRef>
      <j.2:portType
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI "
>http://localhost:8080/axis2/services/Mindat?wsdl#MindatSOAP11port_htt
p</j.2:portType>
    <j.2:operation
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI "
>http://localhost:8080/axis2/services/Mindat?wsdl#getMineral</j.2:oper
ation>
    </j.2:WsdLOperationRef>
  </j.2:wsdlOperation>
  <j.2:wsdlDocument
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI "
>http://localhost:8080/axis2/services/Mindat?wsdl</j.2:wsdlDocument>
  <j.2:wsdlOutput>
    <j.2:WsdLOutputMessageMap>
      <j.2:wsdlMessagePart
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI "
>http://localhost:8080/axis2/services/Mindat?wsdl#return</j.2:wsdlMess
agePart>

```

```

        <j.2:owlsParameter rdf:resource="#desc"/>
    </j.2:WsdOutputMessageMap>
</j.2:wsdlOutput>
<j.2:wsdlOutputMessage
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI"

>http://localhost#getMineralResponseMessage</j.2:wsdlOutputMessage>
    <j.2:wsdlInput>
        <j.2:WsdInputMessageMap>
            <j.2:wsdlMessagePart
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI"

>http://localhost:8080/axis2/services/Mindat?wsdl#term</j.2:wsdlMessag
ePart>
        <j.2:owlsParameter rdf:resource="#name"/>
    </j.2:WsdInputMessageMap>
</j.2:wsdlInput>
<j.2:wsdlInputMessage
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI"
    >http://localhost#getMineralMessage</j.2:wsdlInputMessage>
    </j.2:WsdAtomicProcessGrounding>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.2.1, Build 365)
http://protege.stanford.edu -->

```


APÊNDICE C EXEMPLOS DE CONSULTAS REALIZADAS COM O SISTEMA W-KIS

Exemplo 1: Busca pelo termo “fungi”, selecionando como tipo o conceito “Protein”.

O resultado retornado para esta consulta, exibido na figura A, apresenta duas descrições da proteína “fungi”, uma fornecida pelo motor de busca por proteínas do portal *PubMed* e outra pelo motor de busca do portal *SwissProt*.

BIOMEDICINE

fungi description

from [PubMed Webservice](#) :

Chain A, Conformational Lability Of Lipases Observed In The Absence Of An Oil-Water Interface: Crystallographic Studies Of Enzymes From The Fungi Humicola Lanuginosa And Rhizopus Delemargi|157833953|pdb|1TIB|A[157833953]

from [SwissProt Webservice](#) :

ID: Putative fungistatic metabolite ORFNames: CHGG_05463 Description: Taxonomy Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetidae; Sordariales; Chaetomiaceae; Chaetomium.

Figura A: Resultado da busca pelo termo “fungi” no sistema W-KIS

Exemplo 2: Busca pelo termo “PYK2”, selecionando como tipo o conceito “Protein”.

Assim como no exemplo anterior, o resultado retornado para esta consulta, exibido na figura B, apresenta duas descrições da proteína “PYK2”, uma fornecida pelo motor de busca por proteínas do portal *PubMed* e outra pelo motor de busca do portal *SwissProt*.

BIOMEDICINE

PYK2 description

from [PubMed Webservice](#) :

PYK2 [Saccharomyces cerevisiae]gi|1420756|emb|CAA99675.1|[1420756]

from [SwissProt Webservice](#) :

ID: Development and differentiation-enhancing factor 2 Name: DDEF2
Synonyms: KIAA0400 Description: Homo sapiens (Human) [TaxID: 9606]

Figura B: Resultado da busca pelo termo “PYK2” no sistema W-KIS

Exemplo 3: Busca pelo termo “calcite”, selecionando como tipo o conceito “Constituent”.

O resultado retornado para esta consulta, exibido na figura C, apresenta quatro descrições do mineral “calcite”, fornecidas pelos portais *Athena*, *Geologynet*, *Mindat* e *WebMineral*, e uma definição fornecida pelo portal *Schlumberger*. Neste exemplo, todos os cinco portais conectados ao sistema forneceram alguma resposta para o usuário final.

GEOLOGY

calcite description

from [Athena](#) :

Name: CALCITE Formula: CaCO3 CrystalSystem: Rhomboedral (Trigonal)

from [Geologynet](#) :

Name: CALCITE Formula: CaCO3 CrystalSystem: Trigonal

from [Mindat](#) :

Name: From Calx, the Latin for Lime. Formula: CaCO 3 CrystalSystem: Trigonal
Colour: White, Yellow, Red, Orange, Blue, Green, Brown, Grey etc.

from [Schlumberger](#) :

[CaCO3]The crystalline form of calcium carbonate and chief constituent of limestone and chalk. Calcite reacts readily with dilute hydrochloric acid [HCl], so the presence of calcite can be tested by simply placing a drop of acid on a rock specimen. See: caliche, caliche, chalk, limestone

from [WebMineral](#) :

Formula: CaCO3 Name: From the Latin, calx, meaning lime.

Figura C: Resultado da busca pelo termo “calcite” no sistema W-KIS

Exemplo 4: Busca pelo termo “biotite”, selecionando como tipo o conceito “Constituent”.

O resultado retornado para esta consulta, exibido na figura D, apresenta três descrições do mineral “biotite”, as quais foram fornecidas pelos portais *Athena*, *Geologynet* e *Mindat* e *WebMineral*.

GEOLOGY

biotite description

from [Athena](#) :

Name: BIOTITE (-1M,-2M,-5M1,-6A) = Series name Formula:
 $K(Mg,Fe^{2+})_3(Al,Fe^{3+})Si_3O_{10}(OH,F)_2$ (?) CrystalSystem: Monoclinic / Anorthic
 (Triclinic)

from [Geologynet](#) :

Name: BIOTITE Formula: $K(Mg,Fe^{++})_3(Al,Fe^{+++})Si_3O_{10}(OH,F)_2$ CrystalSystem:
 Monoclinic

from [Mindat](#) :

Name: Named after the French physicist, Jean Baptiste Biot (1774 - 1862), who studied the optical properties of the micas. Formula: Not provided
 CrystalSystem: Not provided Colour: Not provided

from [WebMineral](#) :

Formula: $K(Mg,Fe^{++})_3[AlSi_3O_{10}(OH,F)_2]$ Name: Named in 1847 after the French physicist, Jean Baptiste Biot (1774 - 1862), who studied the optical properties of the micas.

Figura D: Resultado da busca pelo termo “biotite” no sistema W-KIS