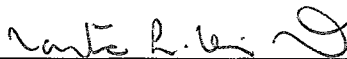


GOS: SERVIÇOS DE ONTOLOGIA NA INTEGRAÇÃO DE BASES DE DADOS

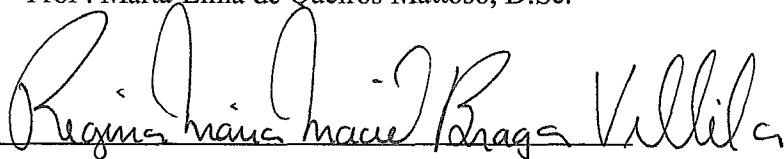
Alessandreia Marta de Oliveira

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

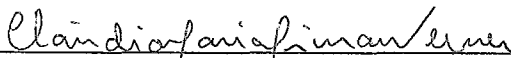
Aprovada por:



Prof.ª Marta Lima de Queirós Mattoso, D.Sc.



Prof.ª Regina Maria Maciel Braga Villela, D.Sc.



Prof.ª Cláudia Maria Lima Werner, D.Sc.



Prof.ª Neide Santos, D. Sc

RIO DE JANEIRO, RJ – BRASIL

MARÇO 2003

OLIVEIRA, ALESSANDREIA MARTA DE

GOS: Serviços de Ontologia na Integração
de Bases de Dados [Rio de Janeiro] 2003

XIII, 108 p. 29,7 cm (COPPE/UFRJ,
M.Sc., Engenharia de Sistemas e Computação,
2003)

Tese – Universidade Federal do Rio de
Janeiro, COPPE

1. Ontologia
2. Bases de Dados
3. Integração de Informação
4. GOA
5. Projeto OdysseyShare

I. COPPE/UFRJ II. Título (série)

Este trabalho quer ser uma homenagem à memória de meu pai Tarcísio e de meu tio Paulinho, que não puderam acompanhar minha caminhada de perto, mas que são presenças constantes em minha vida.

Por outro lado as nossas vidas são sempre marcadas por pessoas cuja presença nos incentiva e estimula na busca, na consecução e na realização de nossos sonhos, neste caso menciono especialmente a memória de minha querida avó, a minha grande e eterna paixão. Ela que nem sequer pode esperar o momento da minha aprovação ao curso de mestrado, tenho certeza de que se tornou a estrela, cuja luminosidade vai guiar e orientar minha existência na realização de todos os meus desejos; por causa dela eu aprendi que só quem ama pode ouvir estrelas.

Agradecimentos

A concretização deste trabalho deve-se à importante colaboração de várias pessoas, o que fez desse processo um aprendizado além da ciência e das letras. Eu agradeço a todos aqueles que, de forma direta ou indireta, colaboraram.

À minha orientadora, professora Marta Mattoso, por estimular o meu trabalho e por ser exigente e crítica para que eu sempre obtivesse o melhor resultado possível.

Num abraço afetuoso, minha gratidão à minha co-orientadora Professora Regina Braga. Seu zelo, sua paciência e sua disponibilidade me fizeram entender que o sucesso de um trabalho depende bastante das pessoas com quem o sonhamos juntos. Obrigada principalmente pela amizade!

À professora Cláudia Werner pela seriedade com que atua na pesquisa científica. Obrigada por estar participando de minha banca e pelo incentivo durante todo este tempo.

À professora Neide por estar participando de minha banca.

Aos professores do Programa de Engenharia de Sistemas e Computação (PESC) da COPPE/UFRJ, em particular ao prof. Jano Moreira de Souza.

À Patrícia Leal, secretária da linha de Banco de Dados do PESC e à Ana Paula secretária de Engenharia de Software pela simpatia.

Aos funcionários da Secretaria do PESC, pela atenção e eficiência no atendimento às minhas muitas demandas operacionais.

À minha mãe Maria Marta e à minha irmã Alessandra pela amizade e pelo amor que me levaram a perseguir a realização de meus sonhos. A elas, em tudo e por tudo, gratidão e afeto. A Leandro, meu cunhado, um abraço especial.

A Eduardo, o grande amor da minha vida, porque dizer sim vale a pena!

À minha sogra, Cilea, porque o sorriso estava sempre presente ainda que a vida a estivesse fazendo chorar.

A meus primos Marcelo, Daniella, Paula e Marcella porque sua presença e peraltices têm-me ajudado a descobrir o verdadeiro significado do amor.

Tia Irma, Tio Joaquim, Vitória, Cidinha e Hebert, obrigada pela paciência!

Ao pequeno Gabriel, que apesar da minha ausência sempre me recebeu com um abraço apertado dizendo que estava com saudade pois a madrinha tinha sumido.

A meu padrinho pela confiança e a certeza de que tudo terminaria bem.

Ao Professor Alberto de cuja sabedoria e humildade se está sempre aprendendo.

Ao grande amigo e colega de trabalho Marco Antônio, simplesmente por ser um profissional e amigo nota mil. Obrigada pelas oportunidades!

À Ana Paula pois alegrias, tristezas e tudo mais foram compartilhados nestes últimos 7 anos.

À Catarina, minha maior conquista no Mestrado. Pela amizade, pelo companheirismo e por tudo que fez por mim. Caty, eu me orgulho da nossa amizade!

Um beijinho especial para Pedro, Dró e Vó porque a distância não esquece uma grande amizade!

Aos amigos e companheiros do mestrado, em especial, André, Gabriela, Gustavo, Marcelo, Nicolaas e Robson pela torcida e pelos momentos de descontração.

Aos amigos do Projeto OdysseyShare pelas contribuições, amizade e motivação. Um abraço todo carinhoso e especial para Alexandre Dantas, Leonardo Murta, Márcio e Gustavo Veronese.

À equipe do projeto da Marinha, em especial Felipe Marun e Tenente Glauco. Ao Prof. Jano e Prof Blaschek., pela oportunidade.

Aos amigos da COPPE pelo incentivo e motivação. Um agradecimento especial para Sômulo, Marluce e Carla Valle.

A meus amigos e demais familiares, que também foram responsáveis por mais esta conquista. A meu avô, pelas orações.

Elevo agora meus pensamentos de terna gratidão a Deus que tem manifestado sua misericórdia e seu carinho em todos os momentos felizes e menos felizes de minha vida.

Ao CNPQ pelo apoio financeiro.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

GOS: SERVIÇOS DE ONTOLOGIA NA INTEGRAÇÃO DE BASES DE DADOS

Alessandrea Marta de Oliveira

Março/2003

Orientadoras: Marta Lima de Queirós Mattoso

Regina Maria Maciel Braga Villela

Programa: Engenharia de Sistemas e Computação

Existe uma necessidade de complementação das técnicas de recuperação de informação e navegação existentes, através de uma estratégia que tenha seu foco no conteúdo e na semântica da informação. No entanto, questões semânticas envolvem a existência de diferentes vocabulários descrevendo informações semelhantes em contextos similares. Uma solução para este problema é o uso de ontologias, incluindo relacionamentos semânticos entre termos de diferentes vocabulários. Nossa proposta associa ontologias e mediadores para reduzir a complexidade do problema de recuperação de informação em fontes distribuídas e heterogêneas para um problema menor que é conhecer os relacionamentos entre os termos através das ontologias. Nós apresentamos uma arquitetura e seu protótipo implementado com um exemplo no domínio do Poder Legislativo.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

GOS: ONTOLOGY SERVICES FOR DATABASE INTEGRATION

Alessandreia Marta de Oliveira

March/2003

Advisors: Marta Lima de Queirós Mattoso

Regina Maria Maciel Braga Villela

Department: Computer and System Engineering

Information retrieval and browsing techniques need to be complemented by strategies that focus in the content and semantics of information. However, semantic issues involve different vocabularies describing similar information in similar contexts. One solution to this problem is the adoption of ontologies, including semantic relationships among terms in different vocabularies. Our proposal associates ontologies and distributed mediators reducing the complexity of distributed heterogeneous information retrieval problem into a less complex problem which is the identification of ontologically related terms across ontologies. We present an architecture and its prototype implemented with an example in the Legislative Process domain.

Índice

| | | |
|-------------------|--|-----------|
| CAPÍTULO 1 | INTRODUÇÃO | 1 |
| 1.1. | MOTIVAÇÃO | 1 |
| 1.2. | OBJETIVOS | 4 |
| 1.3. | ORGANIZAÇÃO DO TEXTO | 5 |
| CAPÍTULO 2 | ONTOLOGIA | 7 |
| 2.1. | INTRODUÇÃO | 7 |
| 2.2. | CLASSIFICAÇÃO DE ONTOLOGIAS | 12 |
| 2.3. | USO DE ONTOLOGIAS | 14 |
| 2.4. | ENGENHARIA DE ONTOLOGIAS | 15 |
| 2.5. | MÉTODOS PARA A CONSTRUÇÃO DE ONTOLOGIAS | 16 |
| 2.6. | LINGUAGEM DE REPRESENTAÇÃO DE ONTOLOGIA | 18 |
| 2.6.1. | <i>KIF</i> | 19 |
| 2.6.2. | <i>Ontolingua</i> | 19 |
| 2.6.3. | <i>OKBC</i> | 20 |
| 2.6.4. | <i>OIL</i> | 21 |
| 2.6.5. | <i>DAML</i> | 22 |
| 2.6.6. | <i>DAML + OIL</i> | 22 |
| 2.6.7. | <i>OWL</i> | 23 |
| 2.7. | EDITORES DE ONTOLOGIA | 23 |
| 2.8. | CONCLUSÃO | 24 |
| CAPÍTULO 3 | ONTOLOGIAS NA INTEGRAÇÃO DE INFORMAÇÃO | 25 |
| 3.1. | INTRODUÇÃO | 25 |
| 3.2. | ODYSSEY-SEARCH | 25 |
| 3.2.1. | <i>Arquitetura</i> | 26 |
| 3.2.2. | <i>Ontologias no Sistema Odyssey-Search</i> | 27 |
| 3.2.3. | <i>Considerações quanto ao Odyssey</i> | 29 |
| 3.3. | OBSERVER | 29 |
| 3.3.1. | <i>Ontologias no Sistema OBSERVER</i> | 30 |
| 3.3.2. | <i>Descrição Global da Arquitetura do OBSERVER</i> | 30 |
| 3.3.3. | <i>Considerações sobre o sistema Observer</i> | 33 |
| 3.4. | INFOSLEUTH | 34 |
| 3.4.1. | <i>Arquitetura</i> | 35 |
| 3.4.2. | <i>Ontologias na Arquitetura InfoSleuth</i> | 37 |

| | | |
|--|---|-----------|
| 3.4.3. | <i>Considerações sobre o InfoSleuth</i> | 38 |
| 3.5. | OUTROS TRABALHOS RELACIONADOS | 39 |
| 3.5.1. | <i>Kess</i> | 39 |
| 3.5.1.1 | Considerações quanto ao Kess | 42 |
| 3.5.2. | <i>THETIS</i> | 43 |
| 3.5.2.1 | Integração de Recursos Científicos por um KBS - Ontologia para o Transporte de Lixo | 44 |
| 3.5.2.2 | A Base de Conhecimento | 45 |
| 3.5.2.3 | Geração de Caminhos de Produção de Dados | 45 |
| 3.5.2.4 | A Arquitetura para Repositórios Científicos Integrados | 45 |
| 3.5.2.5 | Considerações sobre o THETIS | 47 |
| 3.5.3. | <i>OntoSeek: Recuperação de Componentes Baseada em Ontologia</i> | 48 |
| 3.5.3.1 | Considerações sobre o OntoSeek | 50 |
| 3.6. | CONCLUSÃO | 51 |
| CAPÍTULO 4 GOS: SERVIÇOS DE ONTOLOGIA | | 53 |
| 4.1. | INTRODUÇÃO | 53 |
| 4.2. | O USO DO GOS NA ARQUITETURA ODYSSEYSHARE | 55 |
| 4.3. | O SISTEMA GOS | 57 |
| 4.4. | GOA | 58 |
| 4.5. | RELACIONAMENTOS INTRA E ÍTERONTOLÓGICOS | 59 |
| 4.6. | SERVIÇOS | 60 |
| 4.7. | A BASE DE REGRAS | 62 |
| 4.7.1. | <i>Forward Chaining</i> | 62 |
| 4.7.2. | <i>Cadastro das Regras</i> | 63 |
| 4.7.3. | <i>As Regras</i> | 65 |
| 4.7.3.1 | Relacionamentos Sinônimos | 67 |
| 4.7.3.2 | Relacionamentos Hipônimos e Hiperônimos | 69 |
| 4.7.4. | <i>Processamento das Regras</i> | 72 |
| 4.7.5. | <i>Exemplo de Utilização da Base de Regras</i> | 72 |
| 4.8. | CONCLUSÃO | 73 |
| CAPÍTULO 5 USANDO OS SERVIÇOS DE ONTOLOGIA NO DOMÍNIO LEGISLATIVO MUNICIPAL | | 75 |
| 5.1. | INTRODUÇÃO | 75 |
| 5.2. | CADASTRO E REMOÇÃO DE ONTOLOGIAS E TERMOS ONTOLÓGICOS | 76 |

| | | |
|-------------------|--|------------|
| 5.3. | UTILIZANDO A ARQUITETURA GOS | 81 |
| 5.3.1. | <i>Relacionamentos Intraontológicos</i> | 82 |
| 5.3.2. | <i>Relacionamentos Interontológicos</i> | 86 |
| 5.4. | CONCLUSÃO | 90 |
| CAPÍTULO 6 | CONCLUSÕES E TRABALHOS FUTUROS | 91 |
| | REFERÊNCIAS BIBLIOGRÁFICAS | 94 |
| | ANEXO I – PROJETOS RELACIONADOS: QUADRO COMPARATIVO | 107 |

Lista de Figuras

| | |
|---|----|
| FIGURA 3.1: ARQUITETURA ODYSSEY-SEARCH (BRAGA, 2000) | 27 |
| FIGURA 3.2: ELEMENTOS DA CAMADA DE MEDIAÇÃO (BRAGA, 2000) | 28 |
| FIGURA 3.3: PROCESSAMENTO DE CONSULTA (NIETO, 1998) | 32 |
| FIGURA 3.4: ARQUITETURA INFOSLEUTH (BAYARDO ET AL., 1997) | 36 |
| FIGURA 3.5: UMA ONTOLOGIA RESUMIDA (ANDRADE E SALTZ, 1999) | 40 |
| FIGURA 3.6: ONTOLOGIAS, OBJETOS E RELACIONAMENTOS (ANDRADE E SALTZ, 1999) | 41 |
| FIGURA 3.7: ARQUITETURA PARA REPOSITÓRIOS INTEGRADOS (CHRISTOPHIDES ET AL., 1999) | 46 |
| FIGURA 3.8. GRAFO REPRESENTANDO UM COMPONENTE DE SOFTWARE (BORGO ET AL., 1997) | 49 |
| FIGURA 3.9 ARQUITETURA ONTOSEEK (BORGO ET AL., 1997) | 50 |
| FIGURA 4.1: ARQUITETURA COMPUBLISH (OLIVEIRA ET AL., 2002) | 55 |
| FIGURA 4.2: ARQUITETURA GOS | 57 |
| FIGURA 4.3: RELACIONAMENTO SINÔNIMO: A TEM COMO SINÔNIMO B | 59 |
| FIGURA 4.4: RELACIONAMENTO HIPÔNIMO: A TEM COMO HIPÔNIMO B | 60 |
| FIGURA 4.5: RELACIONAMENTO HIPERÔNIMO: A TEM COMO HIPERÔNIMO B | 60 |
| FIGURA 4.6: AS REGRAS | 62 |
| FIGURA 4.7: CADASTRO DE VARIÁVEIS E CONDIÇÕES | 63 |
| FIGURA 4.8: CADASTRO DE CLÁUSULAS | 64 |
| FIGURA 4.9: REGRAS E CLÁUSULAS EXISTENTES | 64 |
| FIGURA 4.10: CADASTRO DE REGRAS | 65 |
| FIGURA 4.11: ONTOLOGIAS WORDNET E STANFORD I | 66 |
| FIGURA 4.12: RELACIONAMENTOS INTERONTOLÓGICOS: ONTOLOGIAS WORDNET E STANFORD I | 72 |
| FIGURA 5.1: ESPECIFICAÇÃO DE ONTOLOGIAS | 76 |
| FIGURA 5.2: ESPECIFICAÇÃO DE TERMOS: NOME, DESCRIÇÃO E DENOMINAÇÃO | 77 |
| FIGURA 5.3: ESPECIFICAÇÃO DE TERMOS DE UMA ONTOLOGIA: LISTA DE HIPÔNIMOS | 78 |
| FIGURA 5.4: INFORMAÇÕES RELACIONADAS AO TERMO PROJETO | 79 |
| FIGURA 5.5: O GOS E OS SERVIÇOS ONTOLÓGICOS | 81 |
| FIGURA 5.6: LEGISLATIVO MUNICIPAL E ONTOLOGIAS RELACIONADAS | 82 |
| FIGURA 5.7: RELACIONAMENTOS INTRAONTOLÓGICOS ANTES DO USO DO GOS | 83 |
| FIGURA 5.8: GOS: DESCOBERTA DE NOVOS RELACIONAMENTOS INTRAONTOLÓGICOS | 84 |
| FIGURA 5.9: RELACIONAMENTOS INTERONTOLÓGICOS ANTES DO USO DO GOS | 87 |
| FIGURA 5.10: DESCOBERTA DE NOVOS RELACIONAMENTOS INTERONTOLÓGICOS | 88 |

Lista de Tabelas

| | |
|---|----|
| TABELA 2.1 – FASES DO PROCESSO DE CONSTRUÇÃO DE ONTOLOGIA | 17 |
| TABELA 3.1: QUADRO RESUMO DOS TRABALHOS RELACIONADOS | 52 |
| TABELA 4.1: ONTOLOGIAS STANFORD I E WORDNET: RELACIONAMENTOS INTRAONTOLÓGICOS | 66 |
| TABELA 4.2: ONTOLOGIAS STANFORD I E WORDNET: RELACIONAMENTOS INTERONTOLÓGICOS | 66 |
| TABELA 4.3: ONTOLOGIAS STANFORD I E WORDNET: LISTA DE SINÔNIMOS | 67 |
| TABELA 4.4 – TERMOS ONTOLÓGICOS (II) | 68 |
| TABELA 4.5 – CLÁUSULAS ANTECEDENTES E CONSEQÜENTES (II) | 68 |
| TABELA 4.6 – TERMOS ONTOLÓGICOS (III) | 69 |
| TABELA 4.7 – CLÁUSULAS ANTECEDENTES E CONSEQÜENTES (III) | 69 |
| TABELA 4.8 – TERMOS ONTOLÓGICOS (IV) | 70 |
| TABELA 4.9 – CLÁUSULAS ANTECEDENTES E CONSEQÜENTES (IV) | 70 |
| TABELA 4.10 – TERMOS ONTOLÓGICOS (V) | 70 |
| TABELA 4.11 – CLÁUSULAS ANTECEDENTES E CONSEQÜENTES (V) | 70 |
| TABELA 4.12 – TERMOS ONTOLÓGICOS (VI) | 71 |
| TABELA 4.13 – CLÁUSULAS ANTECEDENTES E CONSEQÜENTES (VI) | 71 |
| TABELA 4.14 – TERMOS ONTOLÓGICOS (I) | 73 |
| TABELA 4.15 – CLÁUSULAS ANTECEDENTES E CONSEQÜENTES (I) | 73 |
| TABELA 5.1: ONTOLOGIA DO LEGISLATIVO MUNICIPAL | 80 |
| TABELA 5.2: ONTOLOGIA DO LEGISLATIVO ESTADUAL | 80 |
| TABELA 5.3: LEGISLATIVO MUNICIPAL: NOVOS RELACIONAMENTOS INTRAONTOLÓGICOS | 85 |
| TABELA 5.4: LEGISLATIVO ESTADUAL: NOVOS RELACIONAMENTOS INTRAONTOLÓGICOS | 86 |
| TABELA 5.5: LEGISLATIVO MUNICIPAL: NOVOS RELACIONAMENTOS INTERONTOLÓGICOS | 89 |
| TABELA 5.6: LEGISLATIVO ESTADUAL: NOVOS RELACIONAMENTOS INTERONTOLÓGICOS | 90 |

Capítulo 1 Introdução

O termo ontologia vem da filosofia e significa uma explicação sistemática da existência, ou seja, o que existe é o que pode ser representado (MAEDCHE, 2002).

Enquanto este termo foi mais limitado à área filosófica no passado, hoje está ganhando um papel específico em Inteligência Artificial, Linguística Computacional e Banco de Dados. Em particular, sua importância está sendo verificada em campos de pesquisa tais como, representação de conhecimento, projeto de banco de dados, orientação a objeto, recuperação de informação, gerência de conhecimento e projeto de sistemas baseados em agente, dentre outros (FENSEL, 2001). As áreas de aplicação atuais são as mais diversas como, por exemplo, medicina, engenharia mecânica, comércio eletrônico, sistemas de informação geográfica e sistemas de informação biológica (GUARINO, 1998).

No contexto da Web, a utilização de ontologias é crucial: permite que agentes de software compreendam a semântica embutida nas definições e vocabulário especificados com respeito a um domínio, sem ambigüidades, viabilizando o intercâmbio de informações através de consultas (MAEDCHE, 2002).

Este conceito adquiriu novas vertentes e a definição utilizada neste trabalho é a que se segue: Ontologia é um vocabulário de termos e a especificação do relacionamento entre estes termos. Uma definição deve ser criada para cada termo através de uma descrição informal, bem como a especificação dos relacionamentos entre os termos, formando então uma rede semântica.

1.1. Motivação

Diante da grande quantidade de informações disponíveis atualmente, fica difícil de se conhecer a localização, organização/estrutura, linguagem de consulta e semântica de dados disponíveis para consulta (FENSEL, 2001).

A WWW (World Wide Web), segundo MAEDCHE (2002), é hoje um exemplo disso, isto é, uma grande fonte de disseminação de informação nas principais áreas de conhecimento. O seu uso intensivo, aliado ao seu crescimento exponencial, têm mudado drasticamente o comportamento da sociedade, que dispõe atualmente de uma grande variedade de serviços (comércio, notícias, bibliotecas, etc.), simplificando enormemente muitas de suas tarefas diárias. Entretanto, esse sucesso e crescimento exponencial têm

dificultado muito a localização, acesso, apresentação e manutenção da informação para uso de um número tão grande de usuários. O compartilhamento de recursos distribuídos, autônomos, heterogêneos e disponibilizados sem a mínima padronização é um problema complexo. A insatisfação dos usuários com relação à realização de consultas na Web está relacionada sobretudo à demora na obtenção dos resultados e à recuperação de uma elevada quantidade de informações que, na maioria das vezes, não atendem às necessidades dos usuários.

Dentre os sistemas de busca e recuperação de informação existentes, uma das abordagens adotadas é a busca por palavras-chave, onde a consulta é um conjunto de palavras-chave (MAEDCHE, 2002). Neste tipo de abordagem, existe pouca semântica associada e o usuário que submete a consulta é o responsável pela filtragem das informações e correlações. Este tipo de busca exige ainda, em alguns casos, que o usuário esteja atento aos possíveis lugares onde poderia achar informações pertinentes. A busca por palavras-chave possui alto poder de processamento envolvido, mas nem sempre é eficaz pois muitas das informações encontradas não estão relacionadas com a que o usuário está procurando, conforme mencionado anteriormente, e a razão para isto está ligada à pouca semântica.

Existe, portanto, uma necessidade de complementação das técnicas de recuperação de informação e navegação existentes, através de uma estratégia que tenha seu foco no conteúdo e na semântica da informação. Um problema crítico em caracterizar o conteúdo da informação é o uso de vocabulários diferentes descrevendo informação semelhante, ou seja, diferentes termos são utilizados para caracterizar uma mesma informação.

No contexto da Web, o conteúdo encontrado hoje ainda é projetado para usuários (humanos) consultarem e não para os programas computacionais manipularem, o que explica a quantidade de informação inútil encontrada quando se está pesquisando algum assunto; e isso acontece porque o computador não conhece o significado (semântica) do que o usuário procura (BERNERS-LEE et al., 2001). O conteúdo da informação é disponibilizado principalmente em linguagem natural, havendo portanto uma grande lacuna entre a informação que é disponibilizada para a execução dos serviços de recuperação por ferramentas, e a que é hoje disponibilizada para a leitura humana. Com base nessas deficiências, a Web se torna um desafio para a comunidade científica que busca sucesso para integração, intercâmbio e entendimento semântico sobre essas informações (MAEDCHE e STAAB, 2001).

O problema de localização de informações também é crítico nas arquiteturas de integração de banco de dados. Uma solução bem sucedida é o uso de mediadores

(WIEDERHOLD, 1999) que apresenta um esquema de integração comum. Entretanto, o vocabulário desse esquema nem sempre é familiar ao usuário. Uma maneira de se resolver o problema de vocabulários diferentes é a utilização de relacionamentos semânticos entre os termos através do uso de ontologias.

Ontologias estão mostrando ser a resposta adequada também para a Web Semântica (HENDLER et al., 2002), cuja proposta é fornecer estruturas e dar significado semântico ao conteúdo das páginas Web, criando um ambiente onde agentes de software e usuários possam trabalhar de forma cooperativa. Neste novo contexto, a Web será capaz de representar associações entre termos que, em princípio, poderiam não estar relacionadas. Para isso, computadores necessitam ter acesso a coleções estruturadas de informações (dados e metadados) e de um conjunto de regras de inferência que ajudem no processo de dedução automática. Estas regras são especificadas através de ontologias, que permitem representar explicitamente a semântica dos dados (BERNERS-LEE et al., 2001). Através de ontologias é possível elaborar uma rede enorme de conhecimento humano, complementando o processamento da máquina e melhorando qualitativamente o nível de serviços na Web (MAEDCHE, 2002).

A integração de informações no contexto do Desenvolvimento Baseado em Componentes (DBC) também pode ser beneficiada com o uso de ontologias. O DBC (JACOBSON et al., 1997), (SAMETINGER, 1997) tem como principal proposta a construção de aplicações pela composição de produtos de software preexistentes, através da identificação e reutilização de diferentes requisitos já construídos e utilizados em outros sistemas. A reutilização pode proporcionar, entre outras vantagens, a redução da complexidade, dos custos e do tempo de desenvolvimento, além de aumentar a confiabilidade dos sistemas produzidos através do uso de componentes já testados. No entanto, não podemos reutilizar componentes prontos se não soubermos como e onde encontrá-los. Dentro deste contexto, o repositório surge como uma solução para a armazenagem de componentes, tornando-os públicos e acessíveis a todos. Tal repositório deve ser capaz de gerenciar estruturas de dados dos mais variados tipos, tamanhos e complexidade (atributos típicos dos componentes). Atualmente, muitos repositórios podem ser encontrados na Web. Localizá-los, no entanto, não é trivial. Por isso, a construção de uma infra-estrutura capaz de integrar informações de componentes armazenados em repositórios distribuídos torna-se uma questão importante para um processo de reutilização. Além disso, tal infra-estrutura deve prover também recursos avançados para pesquisa de componentes. As estratégias de busca atualmente disponíveis na Web, baseadas

em sua maioria no uso de palavras-chave, como dito anteriormente, permitem a execução de pesquisas de forma limitada, originando respostas muito volumosas e genéricas. Aliado a isto, os problemas associados a questões semânticas que envolvem a existência de diferentes vocabulários descrevendo informações semelhantes dentro de um mesmo contexto, diminuem ainda mais a precisão e qualidade dos mecanismos de busca existentes.

1.2. Objetivos

O objetivo dessa dissertação está em reduzir o problema da necessidade de conhecer a estrutura e semântica de dados espalhados nos vários repositórios de componentes para um problema menor que é conhecer os relacionamentos (termos sinônimos, por exemplo) entre os termos dentro e através das ontologias. Para alcançar estes objetivos, este trabalho apresenta o sistema GOS (GOA Ontology Services) com serviços para o gerenciamento de várias ontologias com diferentes vocabulários. Esse sistema foi desenvolvido sobre o GOA, Gerente de Objetos Armazenados desenvolvido na COPPE/UFRJ (GOA, 2002). Para efetuar o mapeamento de uma consulta sobre uma ontologia em termos de outras ontologias, o GOS utiliza alguns dos relacionamentos apresentados em Nieto et al. (2000). O GOS leva em consideração inicialmente, três tipos de relacionamentos: sinônimos (termos com a mesma semântica), hiperônimos (termos que são mais genéricos que um outro) e hipônimos (termos mais específicos que um outro). As consultas dos usuários são reescritas com a mesma semântica, substituindo os termos por termos sinônimos de diferentes ontologias. Os relacionamentos hiperônimos e hipônimos são usados de acordo com a necessidade do usuário.

Este trabalho está baseado em pesquisas na área de engenharia de domínio¹ (ARANGO, 1988), combinada com a tecnologia de mediadores² (WIEDERHOLD, 1999) provendo serviços de ontologia. O uso de ontologias vem sendo bem sucedido em diversas propostas (GUARINO et al., 1999), (BRAGA, 2000), (NIETO et al., 2000). O GOS usa a engenharia de domínio junto a mediadores e provê inferências para obter relacionamentos intra e interontológicos (termos pertencentes a uma mesma ontologia – ou a ontologias diferentes – e que possuem algum relacionamento entre si, como por exemplo um relacionamento de sinonímia) de modo transparente ao usuário. O sistema GOS foi

¹ Segundo ARANGO (1988), a engenharia de domínio representa um enfoque mais sistematizado para a análise de domínio, sob uma perspectiva mais voltada para a construção de componentes, criando-se um processo completo para a especificação de componentes reutilizáveis (i.e., análise, projeto e implementação).

² Segundo WIEDERHOLD (1999), mediadores apresentam um esquema de representação comum.

desenvolvido visando um contexto de mediadores para integração de informações. A engenharia de domínio é utilizada para a caracterização das informações associadas a cada mediador. Da mesma forma, ontologias são criadas associadas aos mediadores.

O sistema GOS faz parte da arquitetura do ambiente de desenvolvimento baseado em componentes OdysseyShare (WERNER et al., 2002) e interage através da ComPublish (SOUZA, 2002) com um sistema de mediadores. Além disso, o GOS utiliza um modelo de ontologias aliado a um mecanismo de inferência para lidar as informações sobre os componentes e com o relacionamento destas diferentes informações em uma ontologia. Cada mediador provê uma visão lógica das informações sobre componentes agrupadas segundo um determinado domínio de aplicação. O uso de ontologias facilita a localização de informações sobre os componentes baseado no vocabulário do domínio de aplicação. Além disso, relacionamentos intra e interontológicos permitem o mapeamento de relacionamentos entre diferentes componentes de um dado mediador ou de diferentes mediadores, aumentando o potencial da busca por informações.

A proposta desta dissertação é ainda motivada por um projeto desenvolvido pela equipe da COPPE/UFRJ no domínio do Poder Legislativo (WERNER et al., 2000a). O objetivo deste projeto foi disponibilizar uma infra-estrutura de apoio ao desenvolvimento de aplicações no domínio de processamento legislativo, baseado na reutilização de componentes previamente construídos. Usuários podem se beneficiar de informações deste domínio e de outros relacionados, como o domínio da justiça, entre outros. É importante que informações públicas pertinentes a todos os domínios relacionados, sejam apresentadas ao usuário, particularmente quando eles não estão cientes de sua existência. Foi implementado o sistema GOS capaz de identificar informações relevantes do domínio Poder Legislativo, bem como de outros domínios relacionados.

1.3. Organização do Texto

Os estudos e análises realizados no desenvolvimento desta tese foram organizados em cinco capítulos adicionais, como se segue.

O capítulo 2 apresenta uma caracterização de ontologias, tais como diversas definições nas diferentes áreas em que o termo ontologia está sendo utilizado, vantagens e desvantagens do uso de ontologias, classificação de uma ontologia de acordo com o nível de detalhe e nível de dependência da ontologia em relação a uma tarefa específica. Foram apresentados critérios específicos que devem ser observados durante todo o processo de construção de ontologias e

metodologias para a construção de ontologias, de acordo com algumas propostas existentes na literatura. Para finalizar, algumas linguagens de representação de ontologias são mostradas, juntamente com alguns editores de ontologia.

O capítulo 3 apresenta uma revisão bibliográfica onde são examinados alguns dos trabalhos relacionados à proposta GOS e que de alguma forma contribuíram para este trabalho. Neste capítulo são mostradas características gerais dos projetos relacionados, bem como a arquitetura destes sistemas. São citados também aspectos conceituais como, por exemplo, características das ontologias, modelo utilizado, representação lógica, representação física e manipulação. Por fim, é apresentado um quadro resumo comparativo entre GOS e os sistemas avaliados, tais como OBSERVER e Odyssey-Search. .

Os principais componentes e serviços do GOS são descritos no capítulo 4. O GOS tem como um de seus objetivos o estabelecimento de ligações entre termos pertencentes a uma mesma ontologia ou de ontologias diferentes. São apresentados os relacionamentos implementados: sinônimos, hipônimos e hiperônimos. Para que estas ligações sejam efetuadas, o algoritmo *Forward Chaining* foi utilizado; para derivar conhecimento foi usada uma máquina de inferência e para a representação de conhecimento, uma base de regras.

O capítulo 5 mostra os benefícios da utilização do GOS e de seus serviços de estabelecimentos de relações entre termos inter e intraontológicos no domínio Legislativo Estadual e outros domínios relacionados. Apresenta ainda, o funcionamento dos outros serviços oferecidos pelo sistema GOS, como inclusão de ontologias, inclusão de termos em uma ontologia e visualização das informações armazenadas.

No capítulo 6 são reportadas as conclusões gerais e as contribuições desta tese, bem como algumas indicações para trabalhos futuros.

Capítulo 2 Ontologia

2.1. Introdução

Computadores atualmente estão mudando de dispositivos isolados para pontos de entrada em uma rede mundial de troca de informação e transações de negócio. Neste contexto, suporte a dados, informação, troca de conhecimento, troca de componentes estão se tornando o assunto chave na tecnologia atual (FENSEL et al., 2000). Neste contexto, ontologias provêem um entendimento comum e compartilhado de um domínio que pode ser comunicado através de pessoas e sistemas de aplicação e estão representando um papel importante no suporte ao processo de troca de informação em várias áreas (FENSEL, 2001), (MAEDCHE, 2002).

Uma ontologia pode ser definida como um vocabulário específico e relacionamentos usados para descrever certos aspectos de realidade, e um conjunto de suposições relativas ao significado das palavras. Entre vários outros esquemas e estruturas de classificação, incluindo *thesaurus*³, técnicas baseadas em palavras chave e taxonomia⁴, ontologia é vista como sendo um esquema que fornece modelos de domínio precisos e mais completos (OUKSEL e SHETH, 1999).

O termo ontologia teve sua origem na filosofia e hoje apresenta diversas definições nas diferentes áreas em que está sendo utilizado (MAEDCHE, 2002). Várias comunidades científicas utilizam o termo ontologia, mas não se tem uma definição genérica para o termo. Vale a pena ressaltar que o interesse deste trabalho está na área de banco de dados. A seguir são apresentadas algumas definições, de acordo com a área de aplicação, para que se tenha uma visão geral:

Em Filosofia

- “Ontologia é uma disciplina da filosofia e significa uma explicação sistemática da existência” GUARINO E GIARRETA (1995): O termo ontologia foi adotado da Filosofia, onde tinha o intuito de desvendar o significado das coisas do mundo, procurando descrever a natureza das coisas. Aristóteles definiu ontologia como a ciência da existência, diferente das ciências especiais, investiga uma categoria da

³ *Thesaurus*: Esquema de classificação que se preocupa com a estruturação de conceitos relacionados entre si, através de vários significados (CAMPOS, 1994).

⁴ Taxonomia: Estrutura de conceitos sistematizados dentro de um domínio de conhecimento (CAMPOS, 1994).

existência e suas determinações. Em outras palavras, é a ciência que estuda o ser enquanto ser. Considera que ontologia diz respeito a todas as espécies de existência e aos atributos relacionados a esta existência. Já há muito tempo, filósofos têm usado ontologias para tentar descrever domínios naturais e a existência dos seres e coisas em si GUARINO E GIARRETA (1995).

Uma observação apresentada em GUARINO (1998) é que Ontologia (inicial maiúscula sem artigo indefinido) diz respeito a uma disciplina filosófica, que é uma ramificação da filosofia que lida com a natureza e organização da realidade enquanto que uma ontologia (o termo acompanhado de artigo indefinido e com a inicial minúscula) refere-se a um determinado objeto particular, onde o termo de algum modo, independente da linguagem, é relatado para especificar bases de conhecimentos projetadas com o propósito de expressar conhecimento compartilhado. No contexto do trabalho proposto nesta dissertação será utilizado o termo ontologia, pois o objetivo aqui é expressar conhecimento em uma determinada área de aplicação.

Em Ciência da Informação

- “Ontologia é a relação entre conceitos de objetos relacionados por contiguidade no espaço ou no tempo, ou por conexão de causa-efeito” (CAMPOS, 1994).

Em Ciência da informação, ontologia é a relação entre conceitos. Conceito é definido como uma unidade de conhecimento, constituído por características que refletem as propriedades significativas de um objeto ou de uma classe de objetos.

O conceito de Ontologia utilizado no contexto desta dissertação assemelha-se ao conceito de Classificação em Ciência da Informação: “Classificação” é “qualquer método de reconhecimento de relações, genéricas ou outras, entre itens de informação, não importa o grau de hierarquia usada, nem se aqueles métodos são aplicados em conexão com sistemas tradicionais ou computadorizados de informação” (CAMPOS, 1994).

Em Inteligência Artificial

Na Inteligência Artificial (IA), ontologias têm um caráter um pouco distinto do apresentado pela filosofia, sendo usadas para descrever domínios já consagrados, como Medicina, Engenharia e Direito, onde é possível saber o significado projetado das coisas (FENSEL, 2001). Assim, o que se busca com uma ontologia em IA é firmar um acordo sobre

o vocabulário do domínio de interesse, a ser partilhado por agentes que conversam sobre ele (FALBO, 1998). Seguem algumas definições de ontologia de acordo com a comunidade de IA:

- Ontologia é uma especificação de uma conceitualização.
 - Em GRUBER (1999) o autor define ontologia como uma especificação explícita de uma conceitualização. Ontologia é uma descrição de conceitos e relacionamentos que podem existir para uma determinada comunidade. Em outras palavras, uma ontologia consiste na especificação de um vocabulário representacional com definições precisas dos termos desse vocabulário seguidos de um conjunto formal de axiomas que restringe a interpretação e o uso desses termos e que inicialmente é escrito em linguagem natural; e depois formalizados na linguagem de formalização específica;
 - “Ontologia é um registro parcial e explícito de uma conceitualização de uma linguagem utilizada para uma base de conhecimento, conceitualização esta que depende do propósito desejado para a ontologia” (GUARINO e GIARRETA, 1995): Esta definição pode ser considerada como um refinamento da definição apresentada por GRUBER (1999);
 - “Ontologia é uma especificação parcial e explícita de uma conceitualização que é expressa como um ponto de vista meta-nível sobre um conjunto de possíveis teorias de domínio com o propósito de projeto modular, reprojeto e reuso de componentes de sistema” (SCHREIBER et al., 1995): Uma ontologia é descrição meta-nível de uma representação de conhecimento. O problema é que uma ontologia pode ou não ser meta-nível, dependendo da natureza de seu domínio. Esta definição pode ser considerada como compatível com as definições apresentadas por GRUBER e GUARINO;

- Ontologia é a especificação da teoria lógica.
 - “Ontologia é a especificação da teoria lógica” (GUARINO, 1995): Uma ontologia especifica os componentes arquiteturais usados em um domínio particular. É, portanto, uma representação de componentes e suas interações com o objetivo de fornecer uma estrutura explícita para o resto do sistema;
 - “Ontologia é uma teoria lógica que restringe os modelos pretendidos de uma linguagem lógica” (GUARINO, 1997a): O autor afirma que ontologia se refere

a um conjunto de símbolos não lógicos (funções e predicados) de uma linguagem lógica, usados como primitivas com um objetivo de representação específico. Uma ontologia provê os axiomas que restringem o significado dos predicados;

- Em GUARINO (1998), o autor reafirma que ontologia é uma teoria lógica que explica o significado pretendido de um vocabulário formal, isto é, seu compromisso ontológico com uma conceitualização específica do mundo. Os modelos pretendidos de uma linguagem lógica usando tal vocabulário são restritos por seu compromisso ontológico. Uma ontologia indiretamente reflete este compromisso por aproximar estes modelos pretendidos. Vale ressaltar que ontologia é dependente da linguagem enquanto a conceitualização é independente (GUARINO, 1998). Segundo GUARINO (1998), em *Inteligência Artificial*, no termo ontologia, estes dois conceitos se chocam, mas a separação é essencial para estudar o compartilhamento, fusão e tradução de ontologias, pois envolve múltiplos vocabulários e conceitualizações;

- Ontologia é um conjunto de termos e dos relacionamentos entre estes termos.
 - “Ontologia é um conjunto de termos de interesse em um domínio de informação particular e de relacionamentos entre eles” (ANDRADE e SALTZ, 1999): Ontologia é uma questão importante para a gerência de conhecimento. Funciona como um meio termo entre formalizar conhecimento e permitir informações compartilhadas baseadas em um vocabulário comum;
 - “Ontologia é um conjunto de definições de primitivas de conhecimento específico, tais como relações, funções e objetos, ou de outra forma, é a especificação de um vocabulário para um domínio de discurso compartilhado - definição de classe, relações, funções e outros objetos” (GRUBER, 1993b).
 - “Ontologias são coleções de informação” (BERNERS-LEE et al., 2001). No contexto da Web Semântica, ontologia é definida como um documento ou arquivo que formalmente define as relações entre os termos. Uma ontologia possui uma taxonomia e um conjunto de regras de inferência. Uma taxonomia define classes de objetos e as relações entre eles. As regras de inferência oferecem um poder adicional pois novas informações podem ser deduzidas.

Em Engenharia de Software

Em Engenharia de Software ontologia vem sendo definida como um vocabulário de termos e o relacionamento entre estes termos; definição esta que também vem sendo utilizada no contexto de Inteligência Artificial e da Web Semântica, como foi apresentado anteriormente:

- “Ontologia pode ser definida como sendo um vocabulário de termos e a especificação do relacionamento entre estes termos” (BRAGA, 1999). No processo de construção da ontologia, uma definição deve ser criada para cada termo através de uma descrição informal bem como a especificação dos relacionamentos entre os termos, formando então uma rede semântica;
- “Ontologia é uma teoria sobre um domínio que especifica um vocabulário de entidades, classes, propriedades, predicados e funções; e um conjunto de relações que amarra esse vocabulário. Engloba todos os conceitos definidos pela ontologia e é, pois, mais completa. Esses conceitos são explicitamente definidos e representados de forma que possam ser compartilhados, e possuem restrições expressas por axiomas que restringem o significado dos termos” (FIKES e FARQUHAR, 1999). Para o contexto do trabalho apresentado em OLIVEIRA (1999) foi utilizada esta definição onde consideraram os aspectos de que essa ontologia é uma descrição parcial projetada para ser compartilhada dentro da comunidade que concorda com a sua definição, e para o fim específico de desenvolvimento de software em um dado domínio.

Em Banco de Dados

- “Ontologia explica o conteúdo, as propriedades essenciais e os relacionamentos entre termos na base de conhecimento” (MITRA et al., 1999).
 - Uma ontologia é definida em WIEDERHOLD e JANNINK (1998) como uma estrutura que possibilita compartilhamento e reuso de conhecimento pela especificação dos termos e relacionamentos entre eles. Ontologia relata fontes de conhecimento bem como coleciona e organiza os termos de referência. Os termos referem-se a objetos abstratos e do mundo real, os objetos do mundo real fornecem uma instrução para as definições e um modo de validar o significado dos termos empregados. O mapeamento dos termos para objetos pode diferir em

vários domínios autônomos. O relacionamento tem semântica e também um significado estrutural.

- Para MELLO e HEUSER (2000) uma ontologia pode ser considerada uma especificação parcial de um universo de discurso que descreve basicamente conceitos, relações entre conceitos e regras de integridade. Porém, uma ontologia não é estritamente um esquema conceitual de dados. Um esquema conceitual descreve, dentre outras coisas, a estrutura de um banco de dados em um alto nível de abstração. Já uma ontologia não representa a estrutura das fontes de dados associadas a ela, apenas propõe uma estrutura de consenso para grupos de usuários, sendo essa estrutura instanciada através de mecanismos de integração de dados. Assim, uma ontologia é um mecanismo de interpretação parcial ou total do universo de dados de uma ou mais fontes, não existindo obrigatoriamente uma correspondência direta entre possíveis estruturas implícitas ou explícitas dessas fontes e a estrutura da ontologia.

A definição que melhor se adequa inicialmente, à proposta dessa dissertação, é a mesma utilizada em FIKES e FARQUHAR (1999) onde ontologia é considerada como sendo a especificação de um vocabulário de termos e das relações entre estes termos, existindo também a preocupação de que essa descrição seja compartilhada dentro da comunidade que concorda com sua definição. Esta definição posteriormente será aprimorada para atender a outros requisitos que farão parte deste trabalho.

2.2. Classificação de Ontologias

Uma ontologia pode ser classificada de acordo com duas dimensões (GUARINO, 1997b), (GUARINO, 1998):

- Granularidade: Nível de detalhe da ontologia
 - Uma ontologia baseada em um conjunto mínimo de conceitos pode ser desenvolvida de acordo com características particulares, ou seja, através de um conjunto mínimo de axiomas escritos em uma linguagem de expressividade mínima, que dê suporte a apenas um conjunto limitado de serviços, de forma a ser compartilhada entre usuários que concordem com o vocabulário utilizado;
 - Uma ontologia com elevado grau de detalhamento consegue especificar o significado pretendido de um vocabulário com o objetivo de estabelecer

consenso sobre o compartilhamento desse vocabulário ou de uma base de conhecimento que usa esse vocabulário. Uma das desvantagens da ontologia detalhada é que ela pode ser difícil de ser desenvolvida e até mesmo de ser entendida, diante do número de axiomas disponível e da expressividade da linguagem adotada.

- Dependência: Nível de dependência da ontologia em relação a uma tarefa específica ou ponto de vista
 - Ontologias genéricas descrevem conceitos gerais, como, por exemplo, espaço, tempo, matéria, ação, que são independentes de um problema ou domínio particular. Procuram construir teorias básicas do mundo, de caráter bastante abstrato, aplicáveis a qualquer domínio (MAEDCHE, 2002);
 - Ontologias de domínio referem-se à descrição do vocabulário relacionado a um domínio genérico. Ontologias de domínio são construídas para serem utilizadas em um micro-mundo e são tipos comumente desenvolvidos, sendo que diversos trabalhos são encontrados na literatura, enfocando áreas como química e medicina (FALBO, 1998), (MAEDCHE, 2002);
 - Ontologias de aplicação são em geral uma especialização das ontologias de domínio e tarefas e referem-se aos papéis desempenhados por entidades do domínio quando estão realizando alguma atividade;
 - Ontologias de representação explicam conceitualizações que fundamentam os formalismos de representação de conhecimento. Conceitos e relacionamentos definidos em outros tipos de módulos de ontologias são considerados instâncias e conceitos nas ontologias de representação;
 - Ontologias de tarefa descrevem uma tarefa ou atividade genérica, ou seja, expressam conceitualizações sobre a resolução de problemas, independente do domínio em que ocorrem, tal como vendas. O estudo de ontologias de tarefas, segundo FALBO (1998) é a vertente mais recente do estudo de ontologias e sua principal motivação é facilitar a integração dos conhecimentos de tarefa e domínio em uma abordagem mais uniforme e consistente, tendo por base o uso de ontologias.

2.3. Uso de Ontologias

Ontologias têm sido utilizadas com diferentes propósitos e desenvolvidas em diferentes domínios. De uma maneira geral, a utilização de ontologias se classifica em uma das seguintes categorias:

- Desenvolvimento de aplicações:
 - Ontologias podem ser utilizadas com a finalidade de promover o entendimento compartilhado de um problema ou de uma tarefa, o que facilita o processo de identificação dos requisitos do sistema e o entendimento das relações entre os componentes do sistema (OLIVEIRA, 1999). No contexto do desenvolvimento de aplicações, as ontologias permitem que o desenvolvedor pratique a reutilização em altos níveis, do que é o usual em Engenharia de Software (BRAGA e WERNER, 1999), (BRAGA et al., 2001);
 - Permite o compartilhamento de um vocabulário comum entre plataformas de software heterogêneas, possibilitando ainda que o desenvolvedor se concentre na estrutura do domínio e nas tarefas essenciais do domínio, sem se preocupar com detalhes implementacionais (BRAGA e WERNER, 1999), (BRAGA et al., 2001);
- Interoperabilidade entre sistemas: Indica a necessidade de diferentes usuários trocarem dados ou utilizarem diferentes ferramentas. Ontologias podem ser usadas para apoiar a tradução entre diferentes linguagens e representações (OLIVEIRA, 1999);
- Comunicação entre as pessoas e organização: Ontologias capacitam o entendimento compartilhado e a comunicação entre pessoas com diferentes necessidades e pontos de vista particulares sobre seu trabalho (OLIVEIRA, 1999);
- Compartilhamento e validação da base de conhecimento (ANDRADE e SALTZ, 1999):
 - Compartilhamento de conhecimento: Pode-se admitir que as bases de conhecimento são capazes de resolver problemas pela troca de mensagens de acordo com a consulta. A ontologia é essencial para compartilhar conhecimento entre sistemas diferentes;

- Permite a validação da base de conhecimento: A validação dos dados ocorre através de testes e se o procedimento está correto, assume-se que a base de conhecimento está correta, como nos sistemas tradicionais.

2.4. Engenharia de Ontologias

A engenharia de ontologias é o processo de desenvolvimento de ontologias que envolve disciplina e organização, sendo considerada mais do que uma ciência, pois não existe nenhuma definição e padronização de ciclos de vida, métodos e técnicas que orientem o desenvolvimento de ontologias. A engenharia de ontologias é uma tarefa que deve ser cuidadosa e bem conduzida através de princípios adequados que permitam sua melhor avaliação (GÓMEZ-PÉREZ et al., 1996).

Semelhante ao desenvolvimento de sistemas, a definição de ontologias deve considerar critérios específicos, durante todo o processo de construção, bem como um processo de avaliação que envolva verificação e validação, tais como (GRÜNINGER e FOX, 1995), (GRUBER, 1993b), (GÓMEZ-PÉREZ et al., 1995):

- **Clareza:** Uma ontologia deve comunicar efetivamente o significado planejado dos termos definidos, as definições devem ser objetivas, documentadas em linguagem natural e independentes do contexto social ou computacional, e onde for possível, uma definição completa é preferida em relação a uma parcial. Uma ontologia deve ser sempre clara e objetiva, pois qualquer teoria formal é sobre um domínio;
- **Consistência:** Uma ontologia deve garantir consistência com sua definição, tanto no que se refere à especificação dos axiomas lógicos quanto aos conceitos informais, como os comentários em linguagem natural;
- **Extensibilidade:** Uma ontologia deve ser projetada para antecipar usos do vocabulário compartilhado e, portanto, sua representação deve poder ser estendida e especializada. Deve ser possível, portanto, definir novos termos para usos especiais, com base no vocabulário existente, sem haver necessidade de rever definições existentes;
- **Generalidade:** Uma ontologia deve ser capaz de ser compartilhada entre atividades distintas tais como projeto e análise;
- **Suporte:** Uma ontologia deve apoiar o desenvolvimento de grandes aplicações;

- **Completude:** A completude de uma definição em uma ontologia depende do nível de granularidade da ontologia. Garantir a completude de uma definição implica em garantir a completude tanto da definição formal quanto da informal;
- **Concisão:** Uma ontologia deve capturar apenas a informação necessária e suas definições devem ser concisas (redundâncias formais e informais devem ser evitadas, quando possível);
- **Compromissos de Codificação Mínimos:** A conceitualização deve ser especificada sem depender de uma tecnologia particular de representação de conhecimento, ou seja, não escolhendo uma representação somente por causa de uma determinada implementação ou codificação;
- **Compromissos Ontológicos Mínimos:** Uma ontologia deve ser suficiente para apoiar as atividades pretendidas de compartilhamento de conhecimento. Isto é obtido através de uma especificação de uma teoria, definição somente de termos que são essenciais para comunicação de conhecimento consistente com esta teoria, e definição do mínimo de restrições possíveis sobre o mundo que está sendo modelado.

O objetivo da avaliação de uma ontologia, segundo OLIVEIRA (1999) é detectar a ausência de propriedades nas definições e deve incluir a verificação da ontologia, ou seja, a correção da ontologia, do seu ambiente de software associado e de sua documentação. Os passos dessa verificação incluem: a verificação da correta sintaxe das definições, do conteúdo das definições e a verificação da estrutura ou arquitetura da ontologia, cujo objetivo é investigar se as definições estão sendo construídas seguindo os critérios de projeto do ambiente na qual estão incluídas. Enquanto isso, a validação da ontologia é responsável por garantir que a ontologia, seu ambiente de software e documentação correspondam ao que se propõem, isto é, que as definições apresentadas definam realmente o mundo real para o qual a ontologia foi projetada e se estas definições são necessárias e suficientes para representar as tarefas e suas soluções por usuários diferentes.

2.5. Métodos para a Construção de Ontologias

Até o momento, uma grande quantidade de ontologias foi desenvolvida, sob diferentes abordagens, e usando diferentes métodos e técnicas, entre elas, Stanford e WordNet (NIETO et al., 2000). Entretanto, poucos trabalhos foram publicados sobre como proceder, mostrando

as práticas, critérios de projeto, atividades, métodos e ferramentas usadas para construir ontologias, resultando assim na ausência de atividades padronizadas, ciclos de vida e métodos sistemáticos, assim como de um conjunto de critérios de qualidade, técnicas e ferramentas, que conduzam o desenvolvimento (FALBO, 1998). Na tabela 2.1 são apresentadas algumas propostas existentes na literatura com fases de desenvolvimento distintas, porém com objetivos comuns.

Tabela 2.1 – Fases do Processo de Construção de Ontologia

| Proposta | Fases | Objetivo |
|----------------------------|---|---|
| USCHOLD e GRÜNINGER (1996) | Identificação do Propósito | Especificar propósito, requisitos, usuários |
| | Construção da Ontologia | Efetuar a captura, codificação e integração com ontologias existentes |
| | Avaliação | Avaliar questões de competência, especificação de requisitos |
| | Documentação | Facilitar compartilhamento de conhecimento |
| FERNANDEZ, et al., 1997 | Especificação | Identificar propósito, requisitos, usuários, formalização |
| | Conceituação | Estruturação do conhecimento |
| | Integração | Integração com ontologias existentes |
| | Implementação | Codificar a ontologia em linguagem formal |
| | Avaliação | Efetuar julgamento técnico da ontologia |
| | Documentação | Documentar todas as fases do desenvolvimento da ontologia |
| GRÜNINGER e FOX (1995) | Cenários de motivação | Estabelecer conjunto de problemas a ser resolvido e as possíveis soluções |
| | Definição de questões informais de competência | Avaliar os compromissos da ontologia |
| | Especificação da terminologia em lógica de primeira ordem | Especificar a terminologia em lógica de primeira ordem |
| | Definição formal das questões de competência | Definir questões de competência |
| | Especificação de axiomas em lógica de primeira ordem | Especificar as definições dos termos e restrições dos objetos |
| | Avaliação da ontologia | Avaliação da ontologia baseada nas questões formais de competência |
| FALBO (1998) | Identificação do propósito e especificação de requisitos | Especificar propósito, requisitos, usuários |
| | Captura da ontologia | Identificar conceitos e relações importantes |
| | Formalização da Ontologia | Fixar a terminologia da ontologia e a semântica |
| | Integração com ontologias existentes | Aproveitar conceituações previamente estabelecidas |
| | Avaliação | Verificar requisitos de especificação e critério de qualidade |
| | Documentação | Documentar desenvolvimento da ontologia |

As propostas apresentadas concordam que a documentação e a avaliação da ontologia devem ser feitas durante todas as fases de desenvolvimento da ontologia.

A proposta de USCHOLD e GRÜNINGER (1996) tem como característica marcante a preocupação com a correta definição do propósito da ontologia, enquanto FERNANDEZ et al. (1997) enfatizam a necessidade de documentação de todas as fases do desenvolvimento e de um ciclo de vida evolutivo para o desenvolvimento de uma ontologia.

Em GRÜNINGER e FOX (1995) existe a preocupação com uma formalização mais rigorosa com a finalidade de conduzir melhor o projeto e a avaliação da ontologia.

Para FALBO (1998), o processo de desenvolvimento de uma ontologia deve ser visto como um processo fortemente iterativo. A etapa de captura pode apontar novos requisitos ainda não identificados. Na avaliação, pode-se perceber que os termos descritos são insuficientes para o propósito planejado, o que pode provocar um retorno à etapa de captura. Na etapa de formalização incoerências podem ser detectadas, impondo uma revisão das especificações e dos termos definidos na ontologia. Finalmente, quando for necessário integrar uma ontologia com outras existentes, este processo pode ter substancial impacto na definição e formalização dos termos. Uma vez obtida a ontologia formal, muitas vezes é desejável torná-la operacional. Para tal, duas outras atividades devem ser realizadas: projeto e codificação. No projeto, os conceitos e relações da ontologia formal devem ser colocados em um formato compatível com a linguagem de implementação. Na codificação, a ontologia é codificada na linguagem escolhida. Pode ser destacado ainda o fato de que em FALBO (1998) é proposta uma solução mais completa que incorporou características dos outros métodos procurando executar com maior desempenho o processo de construção de uma ontologia.

2.6. Linguagem de Representação de Ontologia

Uma linguagem de representação pode ser usada para expressar ontologias. Uma semântica simples e bem definida é de grande importância para uma linguagem de troca de ontologia porque é usada para transferir conhecimento. Existem muitas linguagens candidatas à representação de ontologias, entre elas KIF (GENESERETH e FIKES, 1992), Ontolingua (GRUBER, 1992), (FARQUHAR et al., 1997), OKBC (CHAUDHRI et al., 1998), DAML (ANKOLEKAR et al., 2001), OIL (FENSEL et al., 2000), DAML+OIL (GIL e RATNAKAR, 2002), OWL (SMITH et al., 2002).

Algumas destas linguagens são baseadas na Web e têm como base padrões W3C: RDF/RDFS (RDF, 2003) e XML/XMLS (XML, 2000), (XML, 2003), como por exemplo a

OIL (FENSEL et al., 2000). A linguagem XML (XML, 2003) tem sido empregada com o intuito de fornecer interoperabilidade sintática aos recursos na Web. A XML possibilita separar o conteúdo do layout, o que não ocorre com a HTML, proporcionando a aplicação de um conjunto extenso de tags que podem ser utilizadas para capturar a estrutura de um documento. A arquitetura RDF e o Esquema RDF (RDF Schema) são tecnologias que compõem um poderoso mecanismo para prover facilidade de descrição de forma genérica (RDF, 2002). Desta forma, triplas (sujeito, predicado e objeto) denotam relações entre pares de objetos. A tecnologia RDF utiliza a linguagem XML, ambas iniciativas do consórcio W3C, no intercâmbio e processamento de modelos RDF. Proporcionam também o compartilhamento de informações compreensíveis pela máquina através da utilização de namespaces (FENSEL et al., 2000).

2.6.1. KIF

A linguagem KIF (Knowledge Interchange Format) é uma versão estendida do cálculo de predicado de primeira ordem que cuida de trocar conhecimento entre bases construídas em diferentes linguagens (GENESERETH e FIKES, 1992), (GRUBER, 1992).

KIF não deve ser considerada como uma representação interna de conhecimento. Na realidade, o que acontece é que um programa ao ler uma base de conhecimento no formato KIF, executa a tradução dos dados em um formato interno próprio e trabalha com este formato. Quando existe a necessidade de comunicação com um outro programa, ocorre o mapeamento das estruturas internas para o formato KIF.

As definições de KIF possuem grande detalhamento, sendo que alguns destes detalhes são essenciais, como o fato de possuir semântica declarativa, o que quer dizer que, o significado de expressões na representação pode ser compreendido sem necessidade de um interpretador para manipular aquelas expressões (ao contrário do que ocorre, por exemplo, com PROLOG) e também prover a representação de conhecimento sobre conhecimento.

Por outro lado, KIF não é uma linguagem apropriada para iteração direta com os usuários e, por ser uma linguagem de especificação, não inclui comandos para consultas ou manipulação de bases de conhecimento.

2.6.2. Ontolingua

Ontolingua é um mecanismo para escrever ontologias em um formato canônico que é compatível com algumas linguagens de representação (GRUBER, 1993a), (FARQUHAR et

al., 1997). Isto permite manter a ontologia em uma forma única quando usada em sistemas com sintaxes diferentes, ou seja, ontologias escritas em Ontolingua podem ser compartilhadas por múltiplos usuários e grupos de pesquisa que usam os seus próprios sistemas de representação (GRUBER, 1992).

Ontolingua é uma ferramenta de domínio público com um mecanismo para definir ontologias: lê as definições, executa a tradução delas para uma forma apropriada (de acordo com um sistema de representação implementado) e reconhece definições que não podem ser traduzidas e avisa o usuário.

Ontolingua provê formas para definir classes, relações, funções, objetos e conjuntos de axiomas (FARQUHAR et al., 1996). A sintaxe e semântica das definições estão baseadas na notação e semântica do KIF. É uma extensão da linguagem KIF com primitivas padrões para definir classes, relações e organizar conhecimento em uma hierarquia centrada em objetos com herança.

2.6.3. OKBC

OKBC⁵ (Ontology Knowledge Base Connectivity) é um padrão de troca de ontologia para sistemas baseados em frames. É uma API para acessar sistemas de representação de conhecimento baseados em frame (CHAUDHRI et al., 1998).

Está sendo desenvolvido sob a responsabilidade do programa de Base de Conhecimento de auto desempenho (HPKB) do DARPA, onde está sendo usado como um protocolo inicial para a integração de vários componentes tecnológicos, e é o sucessor do GFP⁶ (Generic Frame Protocol), que originalmente foi alvo de sistemas que podem ser vistos como sistemas de representação de frame.

OKBC provê um modelo uniforme de um Sistema de Representação de Conhecimento baseado em uma conceitualização de classes e herança. OKBC está definido em uma linguagem de programação independente e tem implementações em LISP, Java e C.

Consiste de um conjunto de operações que provêm uma interface genérica para um sistema de representação de conhecimento, interface esta que possibilita o desenvolvimento de ferramentas de inferência e de análise, entre outras, que operam em vários Sistemas de Representação de Conhecimento⁷.

⁵ <http://www.ai.sri.com/~okbc/>

⁶ <http://www.ai.sri.com/~gfp/>

⁷ What is a Knowledge Representation? (DAVIS et al., 1993)

2.6.4. OIL

OIL (Ontology Interchange Language) é uma representação baseada na Web e uma camada de inferência para ontologias que combina aspectos providos por diferentes comunidades: Lógica Descritiva (semântica formal e capacidade de raciocínio), Sistemas Baseados em Frames (primitivas de modelagem) e Linguagens Web que tem como base padrões W3C: RDF/RDFS e XML/XMLS (BECHHOFFER et al., 2000), (BROEKSTRA et al., 2000), (FENSEL et al., 2000), (KLEIN et al., 2000).

A lógica descritiva é uma linguagem de representação de conhecimento baseada em lógica e que descreve conhecimento em termo de conceitos e restrições que são usados para automaticamente derivar classificações.

OIL herda da Lógica Descritiva a semântica formal e a capacidade de raciocínio desenvolvida para esta linguagem e, por outro lado, apresenta como diferencial o fato de que classes (conceitos) podem ser definidas em termos de descrições que especificam as propriedades que objetos têm que satisfazer para pertencer ao conceito (FENSEL et al., 2000). Estas descrições são expressas usando uma linguagem que permite a construção de descrições compostas, incluindo restrições em relacionamentos binários (roles) conectando objetos.

Dos Sistemas Baseados em Frames, OIL incorpora as primitivas de modelagem essenciais em sua linguagem (FENSEL et al., 2000). O fato de OIL basear-se em padrões Web (XML x RDF) vem da necessidade de uma sintaxe de linguagem para a troca de ontologia baseada em padrões Web para a representação de informação. Como já provado com XOL (uma linguagem baseada em frame com sintaxe XML para a troca de ontologias de biologia molecular), XML pode ser usado como uma linguagem de definição sintática para uma linguagem de troca de ontologia. As definições ontológicas de XOL são projetadas tanto para codificar esquema (metadados) como também para codificar, por exemplo, definições de objeto. OIL pode ser visto como uma extensão de XOL, apesar da omissão de algumas primitivas originais.

Outros candidatos para sintaxe OIL baseados na Web são XMLS (XML Schema), que possui maior poder expressivo e é mais rico na definição de estruturas e gramáticas, além do fato de capturar semântica da ontologia e incorporar noções de herança; RDF, que provê meios para adicionar semântica a um documento sem se preocupar com a estrutura do documento e o fato de RDF ser uma infra-estrutura que habilita a codificação, troca e reuso de metadados estruturados, além de prover uma sintaxe padrão para escrever ontologias e um

conjunto padrão de primitivas; e RDFS (RDF schemes) que provê um esquema básico para RDF (FENSEL et al., 2000).

Uma ontologia OIL contém descrições para classes, relacionamentos (slots) e instâncias. Classes podem se relacionar com outras classes através de uma hierarquia (classes/subclasses) e através de relações binárias estabelecidas entre duas relações. Além disso, restrições de cardinalidade podem ser atribuídas aos relacionamentos. A definição de uma ontologia em OIL é constituída de dois tipos de componentes: o primeiro, que descreve as características da ontologia (ontology container) utilizando-se de descritores do padrão Dublin Core; e o segundo (ontology definitions) que define o vocabulário particular daquela ontologia (HORROCKS, 2000).

2.6.5. DAML

A linguagem DAML (DARPA Agent Markup Language) é uma iniciativa da agência DARPA que foi desenvolvida como uma extensão da XML e RDF (ANKOLEKAR et al., 2001). Esta iniciativa da DARPA tem como objetivo construir uma linguagem que possa dar suporte ao desenvolvimento de sistemas mais complexos de representação de conhecimento, envolvendo tipos de dados, expressões lógicas e outras facilidades que não podem ser projetados em RDF ou RDF Schema devido a sua simplicidade.

2.6.6. DAML + OIL

Uma outra iniciativa, oriunda da combinação de DAML e OIL é uma linguagem proposta para representação de ontologias e metadados pela W3C. Tal combinação, denominada DAML+OIL (GIL e RATNAKAR, 2002), foi desenvolvida a partir da OIL original, embora não se utilize do seu conceito original de frames. É constituído de uma coleção de classes, propriedades e objetos que são adicionados ao RDF e RDFS, resultando em uma nova linguagem de semântica bem definida. Assim, declarações (*statements*) em DAML+OIL também são declarações RDF.

Existe ainda uma especificação da linguagem DAML para a descrição de serviços na Web: DAML-S (ANKOLEKAR et al., 2002a), (ANKOLEKAR et al., 2002b). Em DAML+OIL, entidades, relacionamentos, eventos etc. são definidos em termos de classes e propriedades. DAML-S define um conjunto de classes e propriedades em DAML+OIL específicos para a descrição de serviços. A classe Service está no topo de uma ontologia DAML-S.

2.6.7. OWL

A OWL (Web Ontology Language) está sendo proposta pelo grupo WebOnt (W3C Web Ontology Working Group) como uma extensão de XML, RDF e RDF Schema e está se baseando na linguagem DAML+OIL, como pode ser visto em (PATEL-SCHNEIDER et al., 2002)

A OWL está sendo projetada para aplicações que precisam compreender o conteúdo da informação (semântica) e possibilita a representação do vocabulário dos termos e os relacionamentos entre estes termos (MCGUINNESS e HARMELEN, 2002).

A OWL é uma linguagem para definir ontologias da Web e as bases de conhecimentos associadas a estas ontologias. Em OWL, uma ontologia é um conjunto de definição de classes e suas propriedades, e pode incluir os relacionamentos entre as classes (SMITH et al., 2002).

2.7. Editores de Ontologia

A medida em que linguagens de representação de ontologias são utilizadas, tornam-se necessários mecanismos de edição e modelagem de ontologias que, a exemplo de ferramentas CASE (as quais possibilitam o uso de ferramentas distintas de modelagem), permitam a utilização de várias linguagens semânticas na Web (MOURA, 2002).

O sistema Protégé-2000 (NOY et al., 2000), (NOY et al., 2001) é uma ferramenta gráfica para a edição de ontologias e aquisição de conhecimento. Inclui um mecanismo de customização que permite a modelagem conceitual em várias linguagens semânticas Web a exemplo de RDF, OIL e DAML+OIL.

A Protégé-2000 é uma ferramenta utilizada para desenvolver sistemas baseados em conhecimento. As aplicações desenvolvidas com esta ferramenta podem ser utilizadas em solução de problemas e tomadas de decisão em um domínio particular. A Protégé-2000 funciona como se segue: primeiro, o usuário cria um esquema baseado em classes, subclasses e atributos para representar as informações do domínio de conhecimento que está sendo modelado (os atributos de uma classe podem ter restrições quanto ao tipo e cardinalidade dos dados); depois o usuário cria instâncias baseadas no domínio modelado. Essas instâncias são criadas a partir de um formulário de entrada de dados que é, automaticamente, definido de acordo com o esquema. A Protégé-2000 possui uma interface gráfica bastante intuitiva. Isso permite que especialistas sobre um determinado domínio (por exemplo, um pneumologista), modelem sua área de conhecimento (doenças respiratórias) de uma maneira fácil e natural. O

software ainda permite uma personalização do formulário de entrada de dados e possui uma ferramenta para realização de consultas à base de conhecimento criada (esquema + instâncias). O esquema do domínio pode ser salvo em conformidade com a sintaxe da RDF Schema, e as respectivas instâncias, em RDF.

Outras ferramentas, a exemplo de OntoEdit (ONTOEDIT, 2002a), (STAAB e MAEDCHE, 2000) e OntoBroker (ONTOBROKER, 2002), também caminham nessa direção.

OntoBroker, por exemplo, é um sistema orientado a objetos que provê compiladores em diversas linguagens para descrever ontologias, regras e fatos.

OntoEdit (ONTOEDIT, 2002b) é uma ferramenta de modelagem de ontologias em um nível conceitual, possibilitando uma independência de linguagens concretas de representação. As funções básicas do software e a interface gráfica são bastante semelhantes à da Protégé-2000. Entretanto, a OntoEdit possui um maior número de recursos, como a definição de axiomas, disjunções e meta-informações sobre a ontologia criada e seu desenvolvedor. Também é possível relacionar e descrever outras ontologias afins com a ontologia criada. As ontologias modeladas podem ser exportadas utilizando a sintaxe da RDF Schema ou DAML+OIL, por exemplo.

2.8. Conclusão

Neste capítulo foram discutidos vários aspectos de ontologias, tais como diversas definições nas diferentes áreas em que o termo ontologia está sendo utilizado, vantagens e desvantagens do uso de ontologias, classificação de uma ontologia de acordo com o nível de detalhe e nível de dependência da ontologia em relação a uma tarefa específica. Foram apresentados critérios específicos que devem ser observados durante todo o processo de construção de ontologias e metodologias para a construção de ontologias, de acordo com algumas propostas existentes na literatura. Para finalizar, algumas linguagens de representação de ontologias e alguns editores de ontologia foram mostrados.

Capítulo 3 Ontologias na Integração de Informação

3.1. Introdução

O objetivo deste capítulo é mostrar alguns dos trabalhos relacionados à proposta apresentada nesta dissertação e que de alguma forma contribuíram para este trabalho. Para tanto, foi criada uma estrutura, com a finalidade de manter uma linha de descrição e análise coerente, e que será seguida, sempre que possível: são apresentadas características gerais dos projetos relacionados, tais como: grupo de desenvolvimento, projeto, período de desenvolvimento (início e evoluções), objetivo principal do projeto e a razão pela qual este trabalho está relacionado à proposta descrita nesta dissertação. São mostrados também, aspectos conceituais como características das ontologias, modelo utilizado, representação lógica, representação física (armazenamento) e manipulação (operações típicas, linguagem de manipulação utilizada). A arquitetura do sistema será detalhada juntamente com diagramas, descrição dos serviços e componentes e a tecnologia predominante. A integração com outros sistemas é apresentada no tópico de considerações e para finalizar um quadro resumo comparativo entre os sistemas avaliados e o sistema GOS.

3.2. Odyssey-Search

O objetivo do projeto Odyssey (BRAGA et al., 1999a), (BRAGA et al., 1999b), desenvolvido na COPPE/UFRJ era a construção de uma infra-estrutura de reutilização baseada em modelos de domínio. A infra-estrutura Odyssey provê suporte ao Desenvolvimento Baseado em Componentes (DBC) em domínios de aplicação específicos, utilizando tecnologias como frameworks orientados a objetos (JOHNSON, 1992), arquiteturas de software (SHAW e GARLAN, 1996), agentes inteligentes (WOOLDRIDGE et al., 2000), mediadores (WIEDERHOLD, 1994), ontologias (GUARINO, 1998) e engenharia de domínio (ARANGO, 1988).

Atualmente um ambiente cooperativo para o desenvolvimento de componentes, denominado *Odyssey Share*, vem sendo construído por meio da combinação de técnicas do contexto de Desenvolvimento Baseado em Componentes, Trabalho Cooperativo e Banco de Dados (ODYSSEYSHARE, 2002).

Sobre o processo de Engenharia de Domínio do Odyssey, vale mencionar que o principal objetivo é a criação de componentes reutilizáveis em um dado domínio, para que

estes componentes venham a atender as necessidades dos desenvolvedores de aplicações naquele domínio, em todos os níveis de abstração requeridos. Um domínio é definido por um conjunto comum de problemas ou funções que aplicações neste domínio podem resolver. As figuras centrais desta fase de definição são os especialistas de domínio e os engenheiros de domínio que tem como objetivo especificar e evoluir os conceitos do domínio. (XAVIER, 2001).

A abordagem apresentada neste trabalho pode ser considerada como uma extensão da proposta da Odyssey-Search, especificada em BRAGA (2000) no que diz respeito ao processo de busca por componentes reutilizáveis em outros domínios de aplicação, onde um conjunto de relacionamentos básicos entre termos de diferentes ontologias foi implementado. A abordagem desta dissertação apresenta este mesmo conjunto de relacionamentos, bem como oferece um suporte semi-automático para a descoberta destes relacionamentos, uma vez que na Odyssey-Search este trabalho é efetuado manualmente.

3.2.1. Arquitetura

Odyssey-Search é um sistema multi-agente que realiza o armazenamento e a busca por informações de um dado domínio relevantes para a aplicação sendo especificada pelo engenheiro de software e a busca por informações em domínios correlatos ao domínio em questão, pelo engenheiro de domínio.

O papel dos principais módulos da arquitetura deste sistema (Figura 3.1) é descrito a seguir:

- **Modelo dos Usuários:** Especifica e disponibiliza o perfil do usuário (informações que indicam as preferências do usuário) para os agentes de interface e de filtragem;
- **Agente de Interface (AI):** A adaptação da interface é realizada de acordo com o perfil do usuário. A partir das necessidades iniciais do usuário em relação à aplicação a ser desenvolvida e às escolhas e informações que ele disponibilize, o agente é capaz de apresentar a informação ao usuário de acordo com suas necessidades. O agente determina os requisitos do usuário e provê uma visão virtual das bases de informações do domínio. A forma de apresentação desta base de informações é alterada dinamicamente pelo agente, a fim de acomodar as necessidades e requisitos do usuário;
- **Agente de Filtragem de Informação (AF):** Baseado nas necessidades iniciais do usuário, no modelo do usuário e na ontologia do domínio, o AF filtra para o agente

de interface apenas as informações mais adequadas para o usuário, decidindo também a respeito de como a informação será apresentada. Inicialmente, é utilizada a busca de informações baseadas no perfil do usuário ou nas necessidades da aplicação a ser desenvolvida pelo engenheiro de aplicação e, posteriormente, a busca baseia-se no conteúdo de uma fonte de informação e a partir daí deriva novas informações;

- Agente de Armazenamento e Recuperação de informação (AR): Possui as funções de armazenar componentes do domínio em repositórios adequados, de acordo com o domínio ao qual o componente está relacionado e permitir a recuperação destes componentes de maneira adequada.

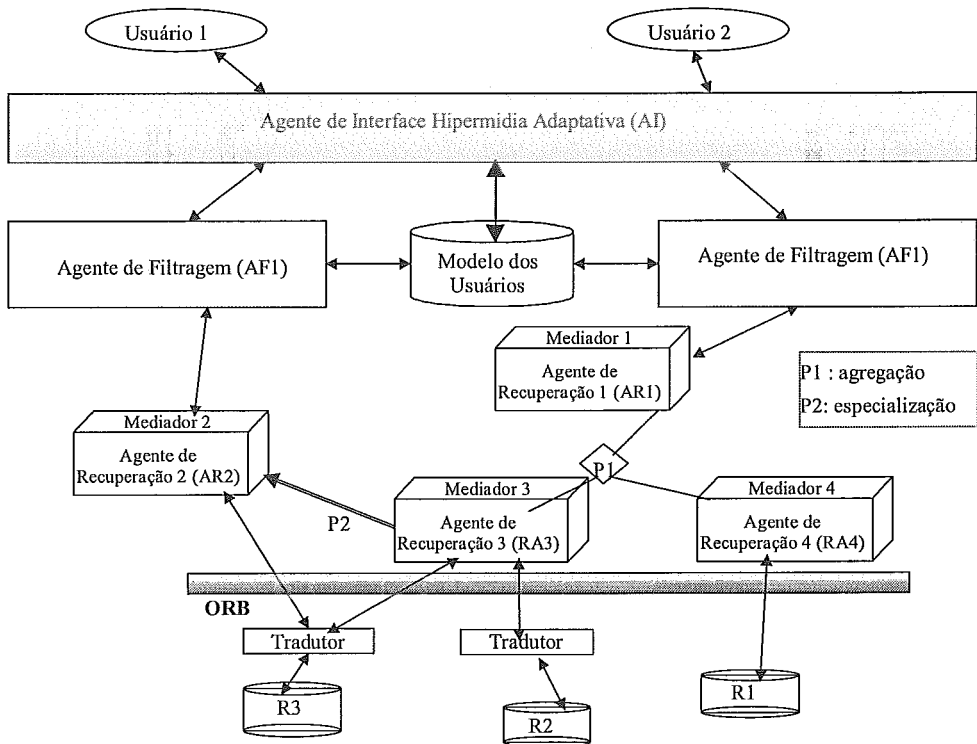


Figura 3.1: Arquitetura Odyssey-Search (BRAGA, 2000)

3.2.2. Ontologias no Sistema Odyssey-Search

Para Braga (2000), conforme apresentado no capítulo 2 deste trabalho, ontologia pode ser definida como sendo um vocabulário de termos e a especificação do relacionamento entre estes termos. No processo de construção da ontologia, uma definição deve ser criada para

cada termo através de uma descrição informal, bem como a especificação dos relacionamentos entre os termos, formando então uma rede semântica.

Odyssey-Search utiliza ontologia de domínio para a busca por informações do domínio, no contexto de um dado domínio ou entre domínios correlatos. O componente responsável por realizar as ligações interdomínios é o Gerente de Serviços (Figura 3.2). Este componente efetua todas as traduções necessárias entre termos de diferentes domínios e sempre tem informação atualizada a respeito do domínio. Esta informação é transmitida através do barramento CORBA para os mediadores requisitados, sendo que o Gerente de Serviços (GS) efetua a tradução da requisição para as respectivas ontologias dos mediadores, de acordo com os relacionamentos interdomínios capturados, usando os termos ontológicos corretos para cada domínio, representados pelos seus mediadores. Vale lembrar que a associação entre os termos é um processo manual. Estes relacionamentos entre termos ontológicos de diferentes domínios envolvem ligações semânticas tais como relações de sinônimos, hiperônimos e hipônimos, que são detalhadas no Capítulo 4.

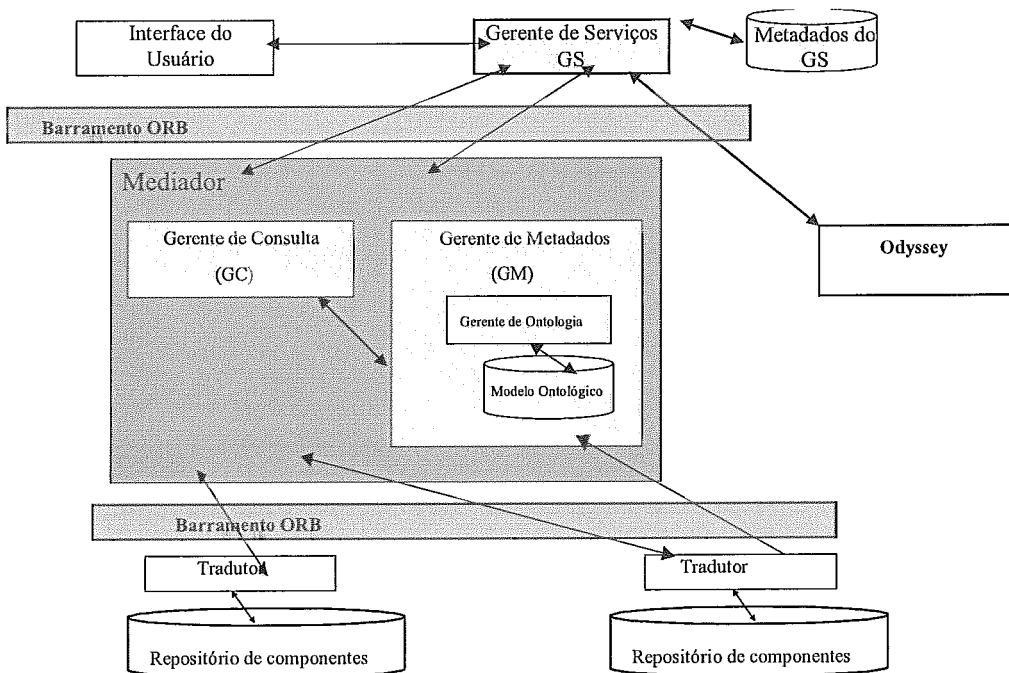


Figura 3.2: Elementos da Camada de Mediação (BRAGA, 2000)

3.2.3. Considerações quanto ao Odyssey

A aplicação mostrada neste trabalho pode ser considerada como uma extensão da proposta descrita em BRAGA (2000) no que diz respeito ao processo de busca por componentes reutilizáveis em outros domínios de aplicação, onde um conjunto de relacionamentos básicos entre termos de diferentes ontologias foi implementado. Como uma das contribuições da abordagem GOS, podem ser citados o suporte semi-automático desenvolvido para auxiliar no estabelecimento de novos relacionamentos. Suporte este que utiliza regras e um mecanismo de inferência para deduzir as novas combinações.

Sobre a Odyssey-Search, vale mencionar que o processo de recuperação de informação leva em conta as preferências do usuário e inclusive faz sugestões de informações desconhecidas pelo mesmo. Uma outra característica importante, que também é uma contribuição do projeto GOS, é a capacidade de execução de consulta envolvendo múltiplas ontologias: se uma consulta não for satisfeita a partir do uso de uma dada ontologia, esta é especificada utilizando os termos de outra ontologia, através dos relacionamentos ontológicos (OLIVEIRA et al., 2002).

3.3. OBSERVER

O objetivo do sistema OBSERVER (NIETO et al., 2000) é permitir que os usuários formulem consultas preocupando-se apenas com a semântica dos dados que lhes interessam. Aspectos relacionados à sintaxe, localização, estrutura e organização dos dados são manipulados pelo sistema. O sistema é baseado na idéia de usar ontologias como metadados para descrever um repositório de dados distribuídos e heterogêneos. Além disso, ontologias são consultadas pelos usuários e por isso apresentam um mecanismo que permite obter as respostas acessando os repositórios de dados subjacentes.

O trabalho apresentado em NIETO et al. (2000) é uma continuação de trabalhos anteriores desenvolvidos pelo grupo de pesquisa do qual o autor faz parte (Departamento de Informática e Engenharia de Sistemas, Universidade de Zaragoza), que lidavam com banco de dados federados.

O GOS está bastante relacionado ao projeto OBSERVER e vale mencionar como características em comum, que são descritas posteriormente, o uso de relacionamentos semânticos entre ontologias para efetuar o mapeamento de uma consulta sobre uma ontologia em termos de outras ontologias e também o gerenciamento de várias ontologias com

diferentes vocabulários, ou seja, o usuário escolhe a ontologia mais apropriada para sua necessidade e o sistema gerencia a navegação sobre outras ontologias.

3.3.1. Ontologias no Sistema OBSERVER

É importante notar a heterogeneidade existente entre as ontologias utilizadas pelo OBSERVER pelo fato de terem sido desenvolvidas de forma independente por diferentes organizações que correspondem às áreas de lingüística e representação de conhecimento. As ontologias foram criadas usando linguagens de representação de conhecimento diferentes, mas foram descritas usando Lógica Descritiva (LD), tentando manter a mesma semântica fornecida originalmente por seus criadores. Estas ontologias podem ser consideradas como metadados semanticamente ricos que capturam o conteúdo da informação dos repositórios de dados subjacentes.

As ontologias deste protótipo foram ligadas a repositórios de dados que armazenam dados relacionados aos termos destas ontologias. Vale observar que todas elas descrevem diferentes repositórios de dados contendo referências bibliográficas.

3.3.2. Descrição Global da Arquitetura do OBSERVER

A filosofia usada no projeto do sistema, segundo NIETO (1998), foi criar um sistema altamente independente do número de repositório de dados/ontologias como também ser capaz de lidar com heterogeneidade nos níveis estruturais, funcionais ou semânticos.

Cada nó no sistema de informação inclui capacidades do processamento de consultas e repositório de dados acessíveis para o resto dos nós através das ontologias que descrevem-nas semanticamente. Para resolver o problema do vocabulário, um repositório compartilhado contendo os relacionamentos entre ontologias pode também ser acessado pelos processadores de consulta.

Os principais módulos envolvidos na tarefa de processamento de consultas são o Servidor de Ontologia, o Gerente de Relacionamento entre Ontologias e o Processador de Consulta que são apresentados a seguir.

O Servidor de Ontologia é o módulo que provê informação sobre ontologias residentes em seus nós e nos repositórios de dados subjacentes. Existe apenas um servidor de ontologia por nó contendo termos ontológicos. Sua principal utilização é acessar dados de ontologias subordinadas. O servidor de ontologia mantém um catálogo de suas ontologias, com informações vinculadas sobre os repositórios de dados subjacentes de cada ontologia. O

objetivo do Servidor de Ontologia é dividido em duas partes: encapsular toda interação direta com ontologias e sistemas de Lógica Descritiva (LD) e encapsular todo acesso ao repositório de dados.

Os serviços providos pelo Servidor de Ontologia são classificados de acordo com a informação solicitada:

- Serviços que provêm informação intencional de uma ontologia: O Servidor de Ontologia pode ser invocado por outros módulos do sistema para obter informação sobre a estrutura de qualquer ontologia em seu nó. Como exemplo destes serviços podem ser citados: enumeração de termos, recuperação da definição de termos e representação gráfica da ontologia;
- Responder consultas formuladas sobre uma ontologia: Dada um consulta em LD e o nome da ontologia, o servidor retorna o dado correspondente armazenado nos repositórios subjacentes.

O Gerente de Relacionamentos entre Ontologias (Interontology Relationships Manager - IRM) é o módulo chave que permite ao OBSERVER lidar com diferentes ontologias criadas de diferentes pontos de vista e semânticas.

Qualquer tipo de propriedade semântica que pode ser definida através de componentes de ontologia do sistema é armazenado e gerenciado pelo IRM em um repositório independente. Isto resolve o problema do vocabulário, isto é, ontologias podem ter sido definidas sob semânticas diferentes e os relacionamentos armazenados no IRM ajudam o sistema a traduzir a consulta de termos em uma ontologia em termos em outra ontologia diferente, preservando a semântica.

O estabelecimento de propriedades semânticas através de abstrações criadas por diferentes pessoas, com diferentes pontos de vista, não é uma tarefa fácil, uma vez que pode ser relacionado de várias maneiras e o uso de vocabulários diferentes (sinônimos, hipônimos, entre outros) torna ainda mais complexo. Então é um processo crítico que precisa de um conhecimento semântico perfeito sobre os dois termos suspeitos de estarem relacionados de alguma forma. Isto explica porque o processo de criar uma nova ontologia deve incluir descrições textuais e exemplos para cada termo na ontologia. Afirmam que o administrador de ontologia é a melhor pessoa para definir esta informação e que se esta tarefa for excluída, o futuro processo de definir relações entre ontologias poderá ser mais difícil ainda, uma vez que relações existentes poderão ser omitidas e, no pior caso, poderão ser estabelecidas relações falsas.

OBSERVER utiliza o seguinte conjunto de tipos de relacionamentos entre ontologias, para ajudar a traduzir consultas da linguagem de uma ontologia para uma linguagem de outra: sinônimo, hipônimo, hiperônimo, sobreposição, disjunção, cobertura.

Com a ajuda do Processador de Consulta, apresentado detalhadamente na Figura 3.3, o usuário constrói uma consulta expressa em LD usando termos de uma ontologia escolhida. Depois de acessar os dados relevantes de acordo com a ontologia do usuário, caso o usuário queira mais dados, o processador de consultas navega por outras ontologias do sistema de informação e traduz termos na consulta do usuário, preservando a semântica de sua consulta, o que determina a importância dos repositórios de dados para os componentes de ontologia.

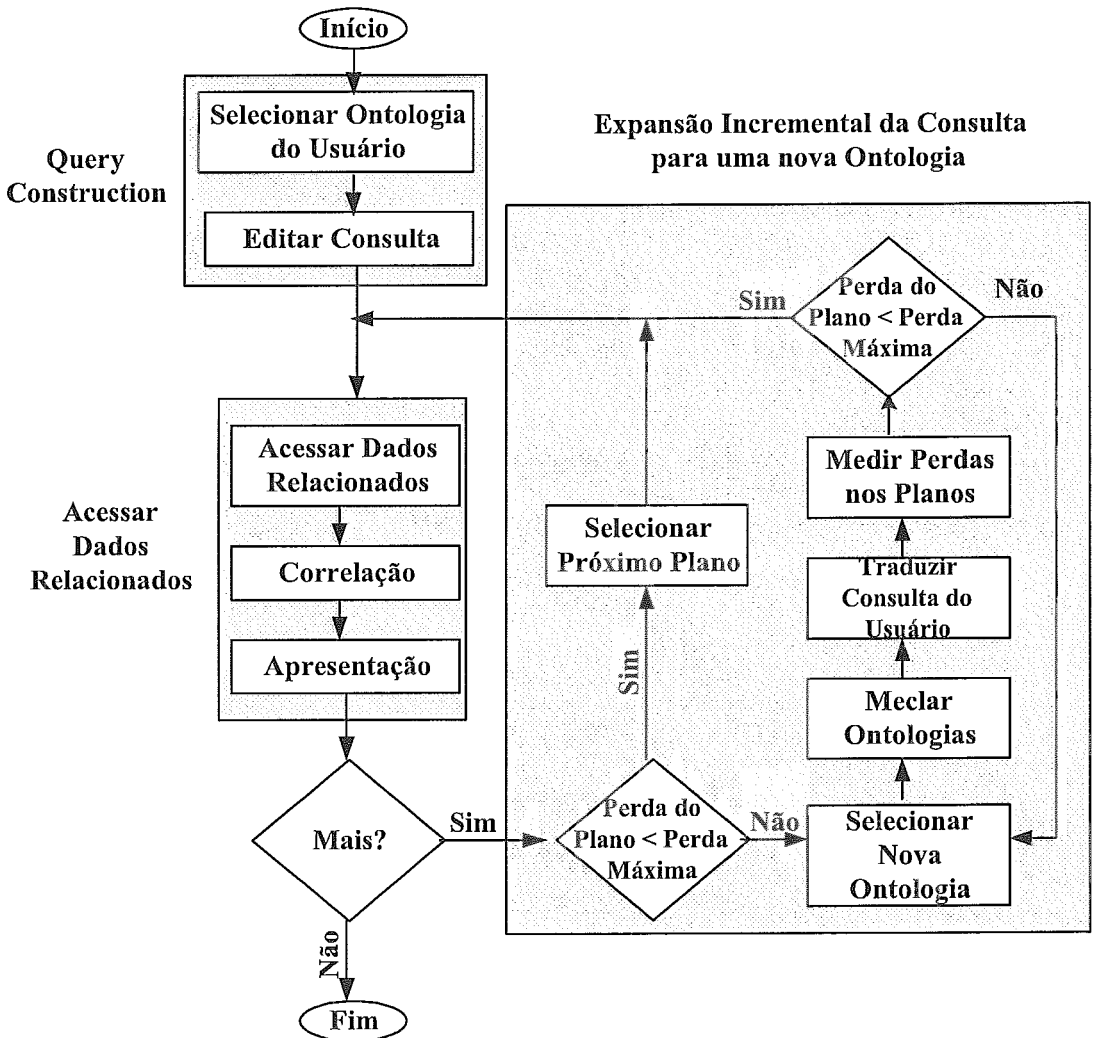


Figura 3.3: Processamento de Consulta (NIETO, 1998)

3.3.3. Considerações sobre o sistema Observer

Assim como o sistema OBSERVER, a proposta apresentada neste trabalho lida com o gerenciamento de várias ontologias com diferentes vocabulários, ou seja, o usuário escolhe a ontologia mais apropriada para sua necessidade e o sistema gerencia a navegação sobre outras ontologias. Diante da possível dificuldade por parte do usuário em efetuar a busca de um conjunto de informações, quando estas estão distribuídas por diversas bases de dados e com semânticas diferentes, esta dissertação sugere o uso de ontologias e para isso é proposta uma ferramenta que tem como objetivo facilitar a navegação sobre ontologias, uma vez que esta também não é uma tarefa simples.

No caso de múltiplas ontologias, uma consulta usa termos de uma ontologia selecionada pelo usuário e precisa ser traduzida em termos de outras ontologias que possuem informações relevantes. Um obstáculo neste contexto, de acordo com NIETO (1998), é o vocabulário compartilhado, especialmente quando diferentes ontologias de domínio são usadas para descrever informações similares através de domínios. OBSERVER está preocupado com isso e propõe o uso de múltiplas ontologias pré-existentes para acessar repositórios de dados heterogêneos, distribuídos e construídos de modo independente. O conteúdo de cada repositório de dados é descrito por uma ou mais ontologias usando um sistema baseado em lógica descritiva. Cada repositório de dados é visto no nível de conceitos semânticos relevantes. A informação requisitada pelo OBSERVER é especificada através de uma expressão em lógica descritiva baseada em conceitos em uma ontologia de domínio. O projeto GOS também está preocupado com a correta tradução de uma consulta baseada em uma ontologia em termos de outras e por isso segue a mesma linha de OBSERVER.

Um outro objetivo em comum com o sistema OBSERVER é permitir que os usuários formulem consultas, preocupando-se principalmente com a semântica dos dados que lhes interessam.

Outra característica em comum é o uso de relacionamentos semânticos entre ontologias para efetuar o mapeamento de uma consulta sobre uma ontologia em termos de outras ontologias. O diferencial do GOS está no fato de que serão utilizados um conjunto de regras e um mecanismo de inferência para deduzir novas combinações válidas, inclusive a partir das combinações já existentes. Este mecanismo automatiza a inferência de forma integrada com mediadores e banco de dados e permite obter as respostas através do acesso aos repositórios de dados subjacentes.

Um último aspecto a ser ressaltado é que o gerenciamento das respostas, levando-se em conta a perda de informação associada, de acordo com um limite determinado pelo usuário, e o controle desta perda de informação foram aspectos observados no OBSERVER e que também influenciaram esta proposta descrita nesta dissertação. Uma vez que ambas as propostas lidam com relacionamentos semânticos (hipônimos e hiperônimos, por exemplo), a mudança semântica e conseqüentemente à perda de informação citada acima acontece quando um termo em uma consulta é substituído por uma expressão de outra ontologia (na tentativa de obter uma resposta completa para a consulta do usuário).

3.4. InfoSleuth

InfoSleuth tem sua raiz no projeto Carnot do MCC (MicroElectronics e Computer Technology Corporation - Austin, Texas), com especialidade em integração de bases de informações heterogêneas e tem como objetivo explorar novas tecnologias em sistemas que recuperam e processam informações em fontes que estão em constante alteração (HWANG, 1999), (BAYARDO et al., 1997), (FOWLER et al., 1999).

De acordo com BAYARDO et al (1997), tecnologias emergentes, como a Internet, têm expandido significativamente os tipos e o volume de dados disponíveis em um sistema de informação. Além disso, nestes novos ambientes, não há controle formal sobre as novas fontes de informação, sendo assim, aplicações são desenvolvidas sem conhecimento completo das fontes que estarão disponíveis na sua execução. Banco de dados federados como Carnot, que fazem integração de dados estáticos, não lidam bem com estes ambientes variáveis. Por outro lado, tecnologias Web baseadas em busca por palavras chave, são escaláveis mas incapazes de acessar informações baseadas em conceitos.

O projeto InfoSleuth integra novas tecnologias como agentes, ontologias de domínio, barramento e Internet, em um ambiente aberto e dinâmico e mostra o uso de agentes de informação e ontologias de domínio como elemento chave para escalabilidade (BAYARDO et al., 1997).

O projeto Carnot, de acordo com BAYARDO et al. (1997), era usado para integrar pesquisas de informações heterogêneas, possibilitando a integração de pesquisas de informações estáticas e foi pioneiro no uso de agentes. O projeto InfoSleuth estende as capacidades da tecnologia Carnot em ambientes de mudanças dinâmica. Ao efetuar uma consulta, o usuário não se preocupa com a localização da informação desejada e nem com sua

existência. A consulta é especificada genericamente sem existir também a preocupação com a estrutura.

Para conseguir meios mais flexíveis para localizar uma informação e lidar com resultados parciais, InfoSleuth integra o seguinte desenvolvimento tecnológico (BAYARDO et al., 1997):

- **Tecnologia de agentes:** Agentes especializados que representam os usuários, as fontes de informação e o próprio sistema cooperam para enviar os requisitos de processamento de informação, permitindo a reconfiguração fácil e rápida das capacidades do sistema;
- **Ontologias de domínio:** Ontologias oferecem uma descrição declarativa, concisa e uniforme da informação semântica, independente da representação sintática subjacente ou de modelos conceituais de bases de informação;
- **Agentes de Informação:** Agentes procuram conciliar informações necessárias (especificadas em termos de alguma ontologia) com os recursos atualmente disponíveis, que recuperam e atualizam requisições que podem ser dirigidas apenas para recursos relevantes;
- **Internet:** Java e Applets Java (JAVA, 2002) são usados para fornecer aos usuários e administradores, interfaces independentes do sistema e também para promover a independência do agente em relação a sua localização ou plataforma.

3.4.1. Arquitetura

InfoSleuth, como descrito em BAYARDO et al. (1997), compreende uma rede de agentes que se comunicam por meio de uma linguagem de consulta de agentes de alto nível denominada KQML (FININ et al., 1994). Usuários estipulam requisições e consultas sobre ontologias especificadas via interface baseada em applet. A linguagem de consulta de banco de dados (SQL) é usada internamente para representar consultas sobre ontologias especificadas. As consultas são dirigidas de agentes mediadores para agentes especializados em recuperação de dados de recursos distribuídos e integração e análise de resultados. Usuários interagem com esta rede de agentes via applets rodando em um browser Web que se comunica com um agente inteligente personalizado.

Agentes anunciam seus serviços e processos requisitando outros através de inferência baseada em conhecimento local, ou de uma requisição para um agente mais apropriado ou ainda pela decomposição de uma requisição em uma coleção de sub requisições e posterior

direcionamento dessas requisições para o agente apropriado que cuida da integração dos resultados. Decisões sobre direção de requisições são baseadas na ontologia InfoSleuth, um corpo de metadados que descreve o conhecimento de agentes e seus relacionamentos com um outro. Decisões sobre decomposição de consultas são baseadas em ontologias de domínio, escolhidas pelo usuário que descreve o conhecimento sobre os relacionamentos dos dados armazenados pelos recursos que descrevem a ontologia (BAYARDO et al., 1997).

A construção de ontologias para uso pelo InfoSleuth é acompanhada mais facilmente pelo uso do IMTS (Integrated Management Tool Suite) que provê um conjunto de interface gráfica para este propósito.

A figura 3.4 apresenta a arquitetura do InfoSleuth e a funcionalidade de cada agente é descrita a seguir (FOWLER et al., 1999), (BAYARDO et al., 1997):

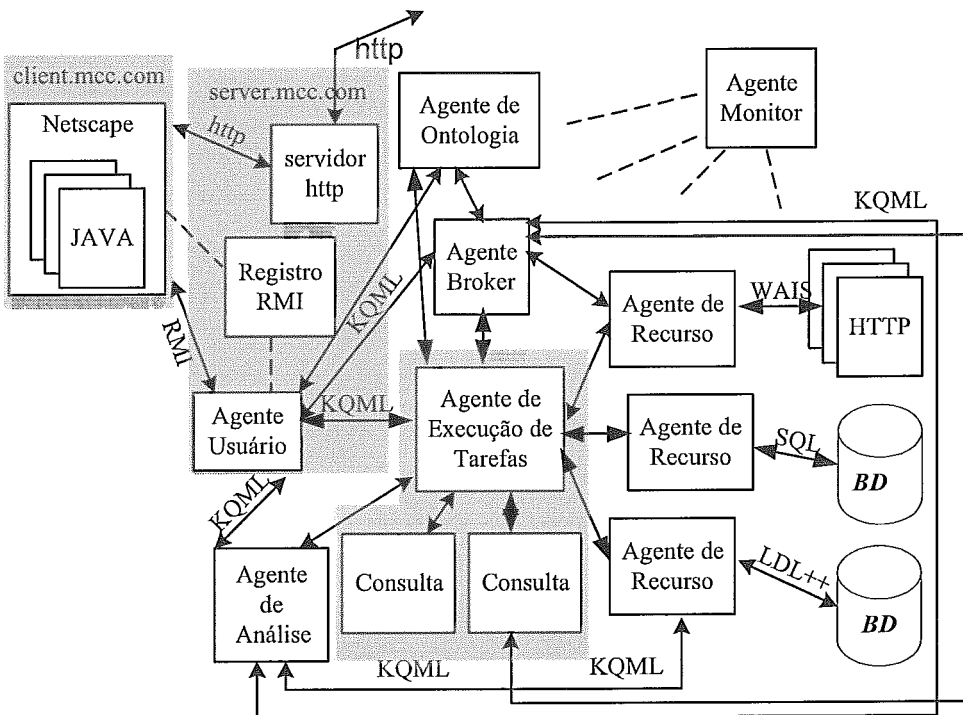


Figura 3.4: Arquitetura InfoSleuth (BAYARDO et al., 1997)

- **Agente Usuário:** constitui a passagem inteligente do usuário no InfoSleuth. Usa conhecimento das ontologias de domínio para assistir o usuário na formulação de consultas e na exibição de seus resultados;
- **Agente de Ontologia:** provê um conhecimento global de ontologias e responde consultas sobre elas;

- **Agente de Informação:** recebe e armazena publicações de todos os agentes InfoSleuth em suas capacidades. Baseado nisso, responde as consultas dos agentes sobre para onde dirigir suas requisições específicas;
- **Agente de Recursos:** provê um mapeamento da ontologia comum para o esquema de banco de dados e linguagem para seu recurso, executa os pedidos específicos para aquele recurso, incluindo consultas contínuas e notificações e também anuncia as capacidades de recursos;
- **Agente de Análise de Dados:** corresponde ao agente de recursos especializado em métodos de análise e mineração de dados;
- **Agente de Execução de Tarefas:** coordena a execução de sub tarefas de informação de alto nível necessárias a execução das consultas. Utiliza informação provida pelo Agente de Informação para identificar os recursos que tem as informações solicitadas e dirigir os pedidos ao agente de recursos correto e reagrupar os resultados;
- **Agente Monitor:** localiza a interação dos agentes e dos passos de execução de tarefas. Provê uma interface visual para exibir a execução.

3.4.2. Ontologias na Arquitetura InfoSleuth

A arquitetura InfoSleuth é baseada na comunicação entre agentes, cooperando para ajudar o usuário a encontrar e recuperar a informação necessária. Uma questão crítica na comunicação entre os agentes é a de compromissos ontológicos, isto é, entendimento entre os vários agentes dos termos para especificar o contexto do agente e o contexto da informação manipulada pelos agentes (BAYARDO et al., 1997).

Uma ontologia pode ser definida, segundo BAYARDO et al (1997) como uma especificação de um vocabulário para um domínio de discurso compartilhado que pode incluir definições de classes, relacionamentos, funções e outros objetos. Ontologias na arquitetura InfoSleuth são usadas para capturar esquema de banco de dados (relacional, orientado a objetos, hierárquico), modelos conceituais e aspectos da arquitetura de agentes InfoSleuth (configurações de agente e especificação de workflow). As motivações para o uso de ontologias são (BAYARDO et al., 1997):

- Capturar e argumentar sobre conteúdo da informação: Em um ambiente dinâmico, o volume de dados disponíveis é um problema que afeta a escalabilidade do sistema;

- Determinar a importância de uma informação sem acessar os dados subjacentes. Ontologias especificadas em uma linguagem de representação de conhecimento ou programação lógica podem ser usadas para argumentar sobre o conteúdo da informação e conseqüentemente capacitar a determinação de sua importância;
- Capturar novas e diferentes visões em ambientes abertos como modelos de domínio;
- Especificação da infra-estrutura do agente: Ontologias são usadas para especificar o contexto em que os vários agentes operam, ou seja, a informação manipulada pelos vários agentes e o relacionamento entre eles.

3.4.3. Considerações sobre o InfoSleuth

InfoSleuth provê uma infra-estrutura baseada em agente, cooperando para ajudar o usuário a encontrar e recuperar a informação necessária, e um processamento de consulta baseada em conceitos, por isso ontologias são usadas para prover especificação declarativa e precisa da informação semântica. São também usadas para descrever conteúdo da informação em repositórios de dados independentes de representação sintática subjacente dos dados. Um usuário expressa sua requisição de informação através de conceitos ontológicos que são compreendidos pelos agentes apropriados que comunicam com os demais para retornar a informação apropriada para o usuário.

Antes de escolher um formato universal de ontologia, InfoSleuth permite múltiplos formatos e representações, descrevendo cada formato de ontologia com um meta-modelo, o que facilita a integração entre diferentes tipos de ontologia. InfoSleuth integra, por exemplo, tecnologias de agente, que se comunicam através de KQML, Internet e ontologias de domínio. InfoSleuth não usa lógica descritiva para descrever ontologias, como acontece em OBSERVER. KIF é usado para representar conhecimento e SQL é usado internamente para representar consultas sobre ontologias específicas. No projeto descrito nesta dissertação, como foi dito anteriormente, serão usadas regras para cuidar da representação das ontologias e o algoritmo de *Forward Chaining* para gerar novos fatos a partir de um conjunto inicial de dados.

Embora InfoSleuth suporte múltiplas ontologias, não há comunicação entre ontologias de domínio, como propõe esta dissertação. O sistema escolhe a ontologia mais apropriada para a consulta e o texto do processamento é desempenhado por tal ontologia e seus repositórios subjacentes. Diferente do projeto desta dissertação, este protótipo lida apenas

com bases de dados e não se aprofunda no problema de definir o mapeamento de informações, para vários repositórios de dados heterogêneos, repositórios estes que, em sua maioria, possuem capacidades de consultas limitadas. Como consequência disso, InfoSleuth não considera vocabulário compartilhado através de ontologias de domínio, ou seja, não são usadas diferentes ontologias de domínio para descrever informações similares através de domínios e com isso não lida com os problemas quando ontologias são projetadas sob diferentes pontos de vista.

3.5. Outros Trabalhos Relacionados

3.5.1. Kess

Em ANDRADE e SALTZ (2000), é apresentada uma aplicação para a implementação de técnicas de otimização de consulta no *Knowledge Enhanced SQL Server* (Kess). Kess é um servidor SQL com um modelo de dados semântico baseado em ontologia que permite a execução de inferências limitadas. Foi projetado para aplicações que requerem a utilização de ontologias ou cadeias semânticas para responder consultas que necessitam de algum conhecimento extra, como por exemplo, exploração de dados.

O servidor Kess ainda está em desenvolvimento e sua arquitetura pode ser considerada como uma extensão de um servidor SQL convencional. A principal característica inovadora deste servidor, quando comparado aos servidores SQL convencionais, é a capacidade de armazenar conhecimento e usar esse conhecimento embutido. Estas técnicas estão divididas em duas categorias diferentes, a primeira baseada em semântica e a segunda em acesso a dados por caminhos relacionados (ANDRADE e SALTZ, 2000).

Em geral, a implementação de capacidades dedutivas em um KBMS (*Knowledge Database Management System*) é realizada usando predicados de consulta que normalmente requerem recursão. Para verificar a submissão de uma tupla a um predicado, é necessário atravessar o grafo semântico/ontológico, seguir as arestas, que denotam o relacionamento e verificar se os vértices são compatíveis com os predicados. Portanto, é importante para Kess, tanto quanto para banco de dados dedutivos em geral, transformar o predicado de consulta em tempo de compilação, de tal forma que a recursão seja menor e mais barata para a avaliação (ANDRADE e SALTZ, 2000).

Como tarefas de exploração de dados, ANDRADE e SALTZ (2000) consideram, por exemplo, cenários como o domínio de pesquisa médica. Banco de dados da área médica tem atributos com nomes extensos, bem como alguns atributos que possuem dependência

semântica (ontológica) com cada um dos outros atributos e onde a interpretação de uma instância pode depender de uma outra. A vantagem apresentada pelo Kess é que os usuários, neste caso, especialistas médicos, são capazes de expressar consultas mais complexas sem a responsabilidade de se tornarem especialistas no modelo de dados, isto é, os usuários não têm que estar completamente familiarizados com o conteúdo de banco de dados, quando estão procurando fatos interessantes, correlações.

Outra tarefa de exploração de dados, considerada no projeto, é a mineração de dados. A maioria das aplicações de Mineração de Dados gera regras baseadas somente no conteúdo da base de dados. A utilização de algum conhecimento pode ser acrescentado ao processo de recuperação e geração de regras com significados semânticos baseados, por exemplo, em agregações sobre uma ontologia.

Um exemplo clássico no cenário de supermercado é que ao invés de gerar regras como COCA → RUFFLES, isto é, um cliente que compra Coca Cola também compra Ruffles e PEPSI → DORITOS, significando que um cliente que compra Pepsi também compra Doritos, podem ser geradas regras semanticamente mais complexas e mais abstratas, que abrangem um universo maior de dados, como por exemplo, REFRIGERANTE → SALGADINHO (Figura 3.5), significando que um cliente que compra refrigerante também compra salgadinhos, se for utilizado algum conhecimento (ANDRADE e SALTZ, 1999).

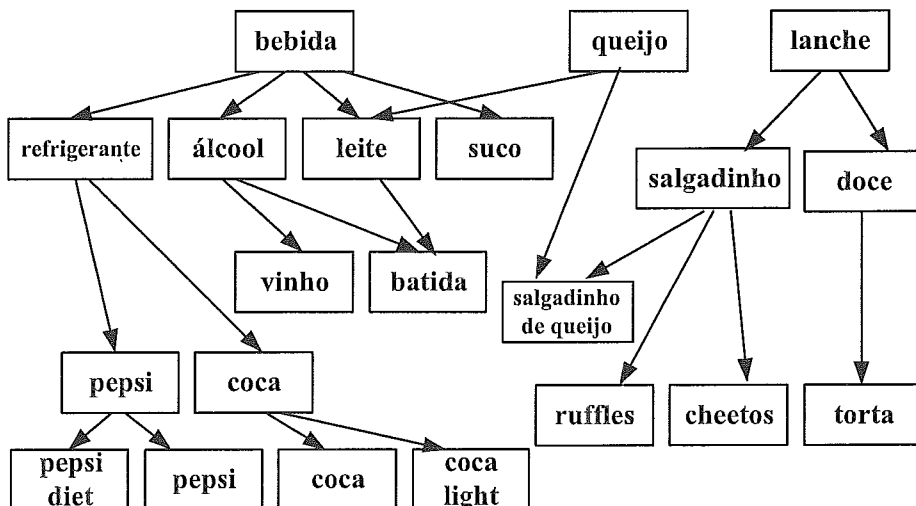


Figura 3.5: Uma Ontologia Resumida (ANDRADE e SALTZ, 1999)

Kess define um formalismo para incluir o conhecimento do domínio, isto é, objetos descrevendo por exemplo, conhecimento real, propriedades, relações, classes e subclasses, estados e processos. Depois são definidos os operadores sobre o conhecimento do domínio que fornecerão ao DBMS (Database Management System) o poder de inferência sobre os dados armazenados em suas tabelas.

Uma ontologia, no contexto do projeto, é definida como um tipo abstrato de dados com três estruturas de dados:

- Um conjunto de objetos O: A figura 3.6 mostra duas ontologias A e B, um conjunto A, B, C, e, F, G, H de objetos que pertencem à ontologia A e um conjunto D e E de objetos que pertencem à ontologia B;
- Um conjunto de relacionamentos (funções) que ligam um objeto a outro. Uma função é definida da seguinte forma $f: O \rightarrow \text{relationship_name } O$. Na figura 3.7 existe, por exemplo, um relacionamento ligando o objeto A ao objeto C da mesma ontologia e também um relacionamento conectando os objetos C e D de diferentes ontologias;
- Um grafo de ontologia $G(V, E)$, onde os vértices V são objetos e as arestas E são relacionamentos.

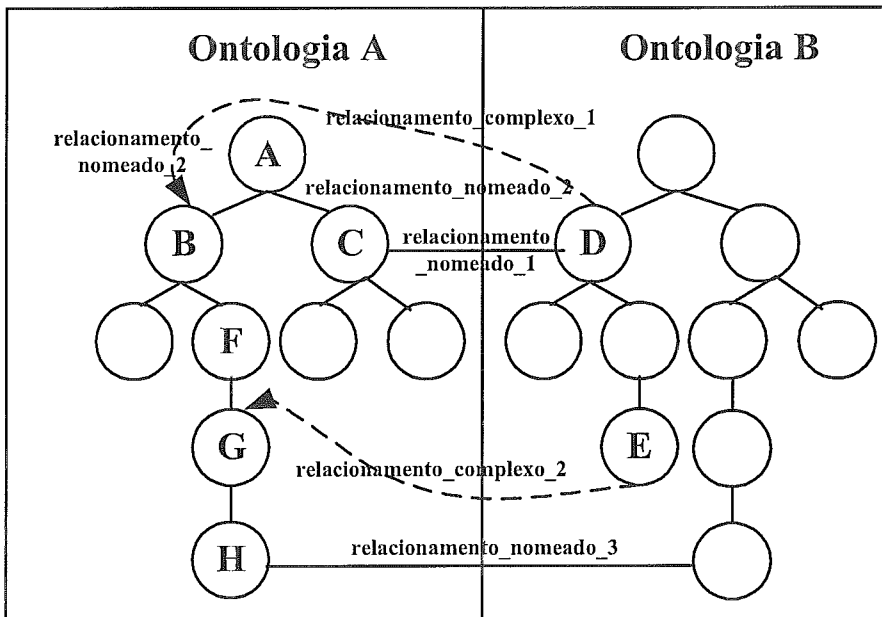


Figura 3.6: Ontologias, Objetos e Relacionamentos (ANDRADE e SALTZ, 1999)

Neste projeto foram implementadas várias extensões aos comandos SQL convencionais para apoiar o modelo de dados semântico. Estas extensões consistem basicamente de primitivas para criar uma ontologia, seus objetos, os relacionamentos conectando os objetos e os caminhos de inferência. Para exemplificar, é apresentada a sintaxe para alguns dos comandos que implementam estas características:

- Create ontology <ontology-name>, onde ontology-name é um identificador nomeado pelo usuário. Aloca espaço para um estrutura de dados que armazena os objetos e relacionamentos entre estes objetos;
- Insert object into <ontology-name> values (<object-name>), onde object-name é o nome de um objeto pertencente a uma ontologia;
- Insert relationshipname into <ontology-name> values (<relationship-name>), onde relationship-name é o nome de um relacionamento que pode conectar objetos da ontologia;
- Insert inferencepathname into <ontology-name> values (<inference-path-name>), onde inference-path-name é o nome de um caminho de inferência conectando objetos de uma mesma ontologia ou de ontologias diferentes;
- Selectwhere inference-path-name (object-name-1, object-name-2); onde este comando (select ou join) tem sua semântica padrão, mas as tuplas são mostradas apenas quando um caminho de inferência inference-path-name é encontrado e é conectado a dois objetos especificados.

3.5.1.1 Considerações quanto ao Kess

A aplicação apresentada em ANDRADE e SALTZ (2000) cuida da implementação de técnicas de otimização de consulta no servidor Kess, que foi projetado para aplicações que requerem a utilização de ontologias ou cadeias semânticas para responder consultas que necessitam de algum conhecimento extra. A proposta do GOS é usar ontologias como metadados para descrever repositórios de dados distribuídos e heterogêneos e prover o mecanismo que, utilizando-se de conhecimento e de um processo de geração e recuperação de regras baseadas em agregações sobre ontologias, permite obter as respostas acessando os repositórios de dados subjacentes.

Uma característica comum à esta proposta é o fato de prover liberdade aos usuários no que diz respeito à construção de consultas mais complexas sem a responsabilidade de estarem completamente familiarizados com o conteúdo de banco de dados.

Kess define um formalismo para incluir o conhecimento do domínio, isto é, objetos descrevendo por exemplo, conhecimento real, propriedades, relações, classes e subclasses, estados e processos. Depois são definidos os operadores sobre o conhecimento do domínio que fornecerão ao DBMS o poder de inferência sobre os dados armazenados em suas tabelas. Neste projeto foram implementadas várias extensões aos comandos SQL convencionais para apoiar o modelo de dados semântico, que consistem basicamente de primitivas para criar uma ontologia, seus objetos, os relacionamentos conectando os objetos e os caminhos de inferência. No caso do trabalho mostrado nesta dissertação, para a representação da ontologia e posteriormente para a inferência de novas combinações de dados válidas, são utilizadas algumas técnicas de Inteligência Artificial, tais como regras para representar conhecimento e mecanismo de *forward chaining* que utiliza um conjunto de regras para gerar novos fatos a partir de um conjunto inicial de dados. Outro ponto importante é que enquanto no Kess são utilizadas extensões SQL, como já foi dito, o GOS utilizará a OQL (*Object Query Language*).

3.5.2. THETIS

Em CHRISTOPHIDES et al. (1999), é tratado o problema de produzir combinações significativas de conjuntos de dados e programas científicos disponíveis para gerar dados precisos para uma tarefa particular (por exemplo, planejamento ambiental). Propõem um Sistema Baseado em Conhecimento (KBS – *Knowledge Based System*) que captura a relação semântica entre dados e programas apoiados em uma ontologia específica da aplicação. Baseado nesta informação, o KBS dinamicamente deduz caminhos de produção de dados válidos sem a intervenção do usuário, o que é importante para aplicações multidisciplinares. Este trabalho teve como suporte o Projeto de pesquisa THETIS, que é uma coleção de repositórios distribuídos com foco no Gerenciamento da Zona Costeira da Região Mediterrânea na Europa e que teve seu início em 1997.

Para validar este trabalho, estão desenvolvendo um protótipo para a Administração da Zona Costeira. Os cientistas ambientais e de instituições públicas que trabalham na Administração da Zona Costeira, segundo CHRISTOPHIDES et al.(1999), freqüentemente precisam extrair e combinar dados de áreas científicas diferentes, como biologia marítima, oceanografia física e química, geologia e engenharia, armazenados em repositórios distribuídos. Por exemplo, considere o transporte de lixo em uma área costeira particular dada uma fonte de poluição. Autoridades locais poderiam exigir esta informação para determinar a melhor localização para instalar um canal de escoamento de lixo.

Defendem o fato de que a descoberta manual de conjuntos de dados e programas que podem conduzir à produção de dados desejados é tediosa. Requer uma investigação rigorosa e entendimento de descrições de metadados. Além disso, o usuário tem que manter o caminho de buscas relacionadas para descobrir possíveis modos de produzir dados para uma tarefa particular. Então, o suporte para derivar combinações válidas de dados e programas de um modo direto é inestimável, particularmente para grandes sistemas de informação. Para alcançar isto, usam um KBS que captura as habilidades dos cientistas considerando as relações semânticas entre recursos científicos. Isto permite a geração por demanda de caminhos de produção de dados com esforço reduzido no lado do usuário, segundo CHRISTOPHIDES et al. (1999).

3.5.2.1 Integração de Recursos Científicos por um KBS - Ontologia para o Transporte de Lixo

Em Ciências Ambientais, conceitos como ar, água, onda e lixo dão o significado intrínseco de dados e programas que residem nos repositórios. Associados a estes conceitos há várias noções físicas como localização ou tempo que são encontradas em vários padrões de metadados ambientais e geoespaciais. Uma fundamentação ontológica deste conhecimento provê um framework formal para deduzir combinações válidas de recursos científicos disponíveis relativo a problemas ambientais concretos, por exemplo, administração do lixo.

Para o escopo deste trabalho, Ontologia conceitua conjuntos abstratos de dados de informações ambientais. Relacionamentos entre os conceitos representam os programas empregados para produzir tais dados.

A semântica desta ontologia é determinada pela interpretação dos vários conceitos e relações que resumem os recursos científicos subordinados. Para este propósito, seguem uma definição intencional de noção de Ontologia que usa atributos como localização e resolução de *grid*, que descrevem as propriedades de recursos científicos e são usados para especificar os relacionamentos entre eles. Do ponto de vista lógico, as descrições formam um conjunto de regras axiomáticas, que são consideradas verdadeiras, independente do conteúdo atual dos repositórios subordinados.

3.5.2.2 A Base de Conhecimento

Para permitir argumentação na combinação de alternativas entre dados e programas, defendem uma definição de noções de Ontologia em um KBS que usa Cláusulas de Horn. Uma noção de Ontologia N é definida como uma cláusula $N(A_1, A_2, \dots, A_n)$, onde A_1, A_2, \dots, A_n são atributos. Relacionamentos entre conceitos são expressos como regras da forma: $N(A_1, A_2, \dots, A_n) : - N_1(A_1, A_2, \dots, A_n), \dots, N_n(A_1, A_2, \dots, A_n)$ onde ":-" denota implicação e ",", "conjunção".

Fatos extensionais indicam conjuntos de dados disponíveis ou intencionais que denotam conjunto de dados que podem ser gerados por programas. Não há nenhuma necessidade de armazenar explicitamente fatos no KBS. Fatos intencionais são dinamicamente deduzidos por regras. Fatos extensionais podem ser construídos "on-line" por uma máquina de busca de metadados que localiza os recursos correspondentes.

3.5.2.3 Geração de Caminhos de Produção de Dados

Uma consulta é uma descrição do recurso desejado em termos de um conceito de ontologia e deve ser satisfeita por fatos extensionais ou intencionais, como consultas posteriores que requerem expansão adicional. Este processo iterativo leva em conta todas as possíveis combinações de regras e fatos extensionais. O resultado é um conjunto de árvores cujos nós são fatos intencionais e as folhas são extensionais, incorporando todos os caminhos de produção válidos através dos quais dados para a idéia examinada podem ser gerados.

Através de uma interface gráfica implementada no KBS, o usuário pode explorar produções de dados válidas de um modo direto. Esta interface é responsável por remeter consultas de usuário para o KBS, receber e interpretar as árvores resultantes, e exibir esta informação.

3.5.2.4 A Arquitetura para Repositórios Científicos Integrados

A arquitetura é constituída de três módulos principais: a Máquina de Busca de Metadados, o Sistema de Base de Conhecimento e o "Workflow" de tempo real que cooperam para proporcionar ao usuário a funcionalidade desejada. A arquitetura é mostrada na Figura 3.7. A funcionalidade de cada componente é descrita a seguir:

- A Máquina de Busca de Metadados é responsável por localizar recursos externos, conjunto de dados ou programas. Também pode recuperar informação complementar armazenada nos repositórios, por exemplo, documentação de usuário

nas fontes disponíveis. A Máquina de Busca aceita consultas de metadados nas propriedades fonte e retorna uma lista de descrições de metadados e referências. Referências apontam para o repositório de *wrappers* que provê acesso e chamada da interface para os sistemas legados subordinados onde os dados e programas residem;

- O Sistema de Base de Conhecimento aceita questões relativas à disponibilidade de conceitos de ontologia. Gera e retorna as produções de dados correspondentes baseados nos recursos disponíveis e restrições impostos pelas regras de ontologia. Estas produções provêm toda a informação que é necessária para construir especificações de workflow. O KBS comunica-se regularmente com a Máquina de Busca de Metadados para atualizar seu banco de dados extensional;
- O Sistema Workflow de tempo real monitora e coordena a execução de workflows. Executa cada passo intermediário da especificação de workflow, acessando dados e chamando programas através do repositório *wrappers*. Podem ser empregadas técnicas de recuperação e controle para aumentar tolerância a falhas.

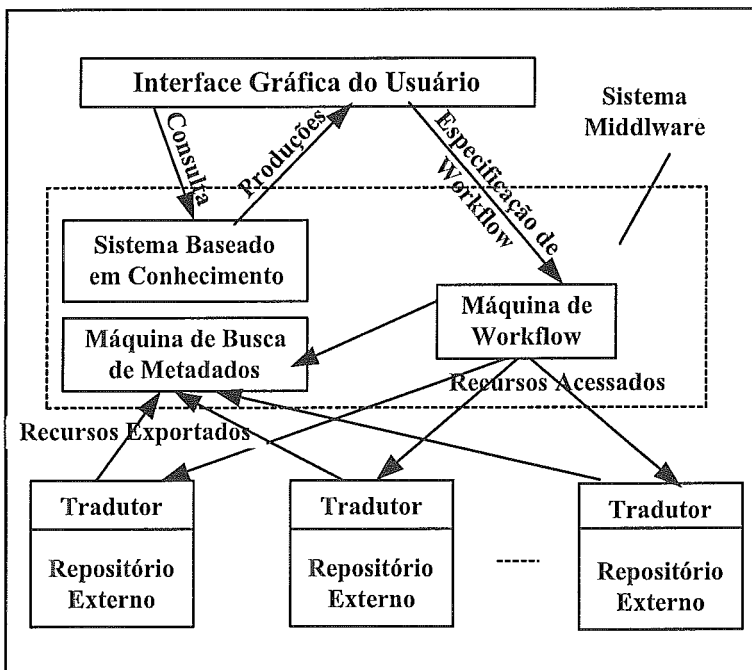


Figura 3.7: Arquitetura para Repositórios Integrados (CHRISTOPHIDES et al., 1999)

Esta arquitetura assegura a escalabilidade e extensibilidade requeridas em sistemas federados. Permite que organizações geograficamente dispersas e autônomas operacionalmente "exportem" os seus recursos. Registrar um novo recurso com o sistema requer somente prover *wrappers* apropriados e descrições de metadados.

Para aumentar o desempenho e tolerância a falha, a Máquina de Busca de Metadados pode ser distribuída em várias máquinas. Também, várias unidades de conhecimento que aderem a domínios diferentes podem ser ligadas ao KBS para apoiar uma variedade de aplicações e grupos de usuário.

3.5.2.5 Considerações sobre o THETIS

Este projeto, como dito anteriormente, cuida do problema de produzir combinações significativas de conjuntos de dados e programas científicos disponíveis para gerar dados precisos para uma tarefa particular. Propõem um KBS que captura a relação semântica entre dados e programas apoiados em uma ontologia específica da aplicação. Baseado nesta informação, o KBS dinamicamente deduz caminhos de produção de dados válidos sem a intervenção do usuário. A aplicação desenvolvida e mostrada nesta dissertação também leva em conta o fato de que a transparência para o usuário é um fator prioritário e se preocupa com a captura da relação semântica entre os termos da ontologia.

Usam um KBS que captura as habilidades dos cientistas considerando as relações semânticas entre recursos científicos, o que permite a geração por demanda de caminhos de produção de dados sem esforço significativo no lado do usuário. Este projeto, assim como a proposta deste trabalho, se preocupa com o fato de prover liberdade aos usuários na construção de consultas mais complexas.

Para permitir argumentação na combinação de alternativas entre dados e programas, defendem uma definição de noções de Ontologia em um KBS que usa Cláusulas de Horn, como já foi dito anteriormente. Esta dissertação propõe um conjunto de regras "if-then" para representar conhecimento e utiliza um mecanismo de inferência para deduzir novos dados.

No contexto do projeto, uma consulta é uma descrição do recurso desejado em termos de um conceito de ontologia e o resultado é um conjunto de árvores incorporando todos os caminhos de produção válidos através dos quais dados para a idéia examinada podem ser gerados. Através de uma interface gráfica implementada no KBS, o usuário pode explorar produções de dados válidas de um modo direto. Esta interface é responsável por remeter consultas de usuário para o KBS, receber e interpretar as árvores resultantes, e exibir esta

informação. O projeto GOS fornece ao usuário uma interface para cadastrar a ontologia e seus termos, bem como uma interface para o administrador cadastrar as regras.

3.5.3. **OntoSeek: Recuperação de Componentes Baseada em Ontologia**

OntoSeek é uma ferramenta orientada a objetos, baseada em ontologia, com alguma capacidade de inferência e um processo de recuperação baseado em palavras chave para a recuperação de componentes (BORGO et al., 1997), (GUARINO et al., 1999).

OntoSeek é resultado de uma cooperação entre Corinto (Consorzio di Ricerca Nazionale Tecnologia Oggetti - consórcio este que tem como um dos objetivos o estudo da tecnologia orientada a objeto), IBM, Apple Itália e o Ladseb-CNR (National Research Council Institute of Systems Science e Biomedical Engineering).

OntoSeek adota uma linguagem de expressividade limitada, priorizando a simplicidade dos requisitos mais importantes. Utiliza uma estrutura de grafos bem simples para representar as consultas e os componentes, e assume que o usuário tem familiaridade com o vocabulário utilizado para codificar componente, apoiando-se em uma grande ontologia lingüística para executar o “casamento” entre consultas e dados, a Ontologia Sensus (BORGO et al., 1997), (SWARTOUT et al., 1996). O componente é descrito por um grafo (Figura 3.8), onde arcos e nós são rotulados com palavras em inglês. Uma vez que relações binárias não são normalmente denotadas por nomes, uma semântica especial é utilizada para este grafo, chamado Grafo Léxico Semântico. Nomes em inglês no grafo são reconhecidos por uma interface léxica, baseada no Wordnet (thesaurus que indica os diversos significados associados a cada palavra). O analista escolhe o significado correto e o grafo é transformado em um grafo de sentidos baseado na Ontologia utilizada no sistema. O grafo de consultas é construído de modo similar, mas as palavras escolhidas e o sentido correspondente podem ser diferentes, assim como a estrutura do grafo. Conceitualmente, o processo de busca implementa um algoritmo de casamento de grafo, retornando os identificadores de todos os componentes com descrição incluída na consulta.

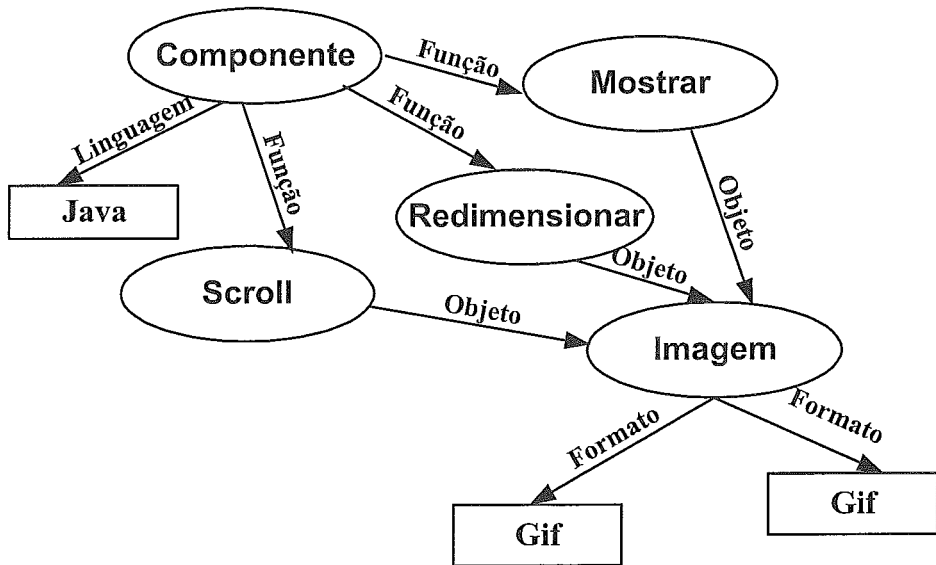


Figura 3.8. Grafo Representando um Componente de Software (BORGÓ et al., 1997)

OntoSeek foi projetado para codificar uma informação através de um grafo simples de conceitos e relações rotuladas por palavras em inglês; consultar uma base de dados previamente codificados e executar o casamento semântico baseado em ontologia. E para possibilitar a gerência de repositórios de informação baseada em ontologias compartilhadas, um número de requisitos mais prioritários tem sido assumidos, tais como: eficiência no volume de dados e portabilidade/escalabilidade do sistema.

O papel dos principais módulos da arquitetura OntoSeek (Figura 3.9) é descrito a seguir: A descrição funcional de um componente é representada pelo Gráfico Léxico Semântico. A principal característica deste grafo é o fato de que seus componentes são denotados por itens léxicos e depois refinados interativamente em um grafo de sentido por uma Interface de Codificação. Os possíveis significados das palavras são mostrados através da Interface Léxica. O grafo de sentido é a entrada da fase de classificação, onde o grafo descrevendo o componente é adicionado ao Banco de Dados, devidamente ligado a Ontologia para explorar algoritmos de busca eficiente. No processo de recuperação de dados, a consulta é representada como um grafo de palavra, que é refinada no grafo de sentido na Interface de Recuperação de Dados. A base de dados é então pesquisada para encontrar os componentes que atendem a consulta de acordo com as restrições taxonômicas presentes na ontologia.

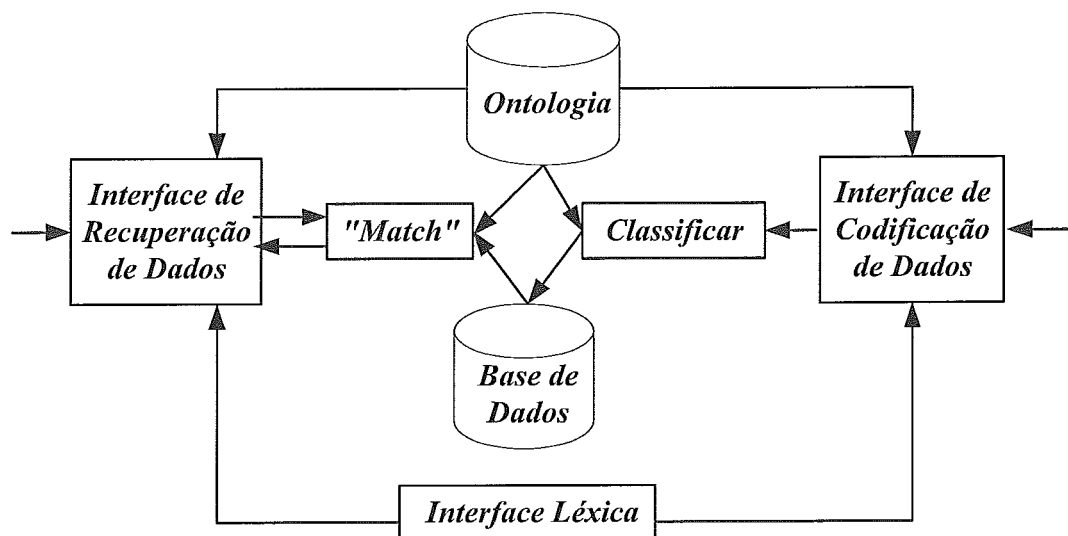


Figura 3.9 Arquitetura OntoSeek (BORGIO et al., 1997)

3.5.3.1 Considerações sobre o OntoSeek

OntoSeek é uma ferramenta para a recuperação de componentes baseada em ontologia, e que apresenta algumas características que são propostas também em GOS, como por exemplo o fato de utilizarem alguma capacidade de inferência.

Enquanto na proposta desta dissertação, são usadas ontologias de forma integrada e não se limita a uma representação, o projeto OntoSeek adota uma linguagem de expressividade limitada, priorizando a simplicidade dos requisitos mais importantes, utilizando uma estrutura de grafos bem simples para representar as consultas e os componentes.

Ao contrário do GOS, OntoSeek assume que o usuário tem familiaridade com o vocabulário utilizado para codificar componente, apoiando-se em uma grande ontologia lingüística para executar o “casamento” entre consultas e dados; conceitualmente, o processo de busca implementa um algoritmo de casamento de grafo, retornando os identificadores de todos os componentes com descrição incluídas na consulta.

OntoSeek foi projetado para codificar uma informação através de um grafo simples de conceitos e relações rotuladas por palavras em inglês; consultar uma base de dados previamente codificados e executar o casamento semântico baseado em ontologia. Como no trabalho proposto nesta dissertação, estão preocupados com a gerência de repositórios de informação baseados em ontologias compartilhadas.

3.6. Conclusão

Como foi visto, alguns sistemas de recuperação de informação usam, por exemplo, mecanismos de inferência sobre ontologias para derivar novas informações. Estes trabalhos incluem OntoSeek, Thetis e Kess. Uma diferença entre estes projetos e o nosso é que usamos um processo semi-automático para derivar novas informações e relacionamentos interontológicos. Estes projetos não fornecem uma solução para acessar informações de diferentes domínios, de modo transparente para o usuário como fazemos em nosso trabalho.

NIETO et al (2000) apresentam uma proposta que permite que os usuários formulem consultas preocupando-se apenas com a semântica dos dados que lhes interessam. Aspectos relacionados à sintaxe, localização, estrutura e organização dos dados são manipulados pelo sistema. O sistema é baseado na idéia de usar ontologias como metadados para descrever repositório de dados distribuídos e heterogêneos. Além disso, ontologias são consultadas pelos usuários e por isso apresentam um mecanismo que permite obter as respostas acessando os repositórios de dados subjacentes. Em NIETO et al (2000), relacionamentos semânticos entre ontologias são usados para efetuar o mapeamento de uma consulta sobre uma ontologia em termos de outras ontologias e também o gerenciamento de várias ontologias com diferentes vocabulários, de modo similar à nossa proposta. O diferencial de nossa abordagem é que o estabelecimento de relacionamentos interontológicos é efetuado de maneira semi-automática, utilizando uma base de regras juntamente com um mecanismo de inferência para derivar novos relacionamentos.

Outro trabalho relacionado à nossa proposta é o sistema InfoSleuth. Esse sistema usa ontologias para a recuperação de informação de bases de dados distribuídas e heterogêneas. InfoSleuth pode ser visto como um framework que pode ser modelado para um determinado propósito. Uma aplicação interessante é o projeto EDEN na área ambiental. Neste caso, algumas ferramentas de InfoSleuth foram personalizadas para este projeto.

O GOS pode ser considerado como uma extensão da Odyssey-Search, trabalho apresentado em BRAGA (2000), onde foi implementado um protótipo que realiza o armazenamento e a busca por informações relevantes de um dado domínio para o desenvolvimento de aplicações e a busca por informações em domínios correlatos ao domínio em questão. Em BRAGA (2000), também é utilizada a ontologia de domínio para a busca por informações do domínio, no contexto de um dado domínio ou entre domínios correlatos. Uma diferença é que na Odyssey-Search, a associação entre os termos é um processo manual, enquanto que em nossa proposta existe um suporte semi-automático para ajudar neste

processo. Em BRAGA (2000), o processo de recuperação de informação leva em conta as preferências do usuário e inclusive faz sugestões de informações desconhecidas pelo mesmo. No sistema GOS, serviços ontológicos são oferecidos ao usuário através de uma interface gráfica, tais como cadastro e remoção de ontologias, inclusão de novos termos, dentre outros, serviços estes que não estão disponíveis nas outras propostas apresentadas, inclusive BRAGA (2000).

O diferencial da nossa proposta em relação aos demais trabalhos está no fato de que utilizamos um conjunto de regras e um mecanismo de inferência para deduzir novas combinações válidas, inclusive a partir das combinações já existentes. Este mecanismo automatiza a inferência de forma integrada com mediadores e banco de dados e permite obter as respostas através do acesso aos repositórios de dados subjacentes.

Para finalizar, é apresentado um quadro resumo dos projetos relacionados à proposta GOS, tentando mostrar cada projeto e as suas principais características (mencionadas anteriormente e marcadas com o “x”). O sistema GOS aparece no quadro para que a combinação das características possa ser observada (GOS e os outros sistemas) e será apresentado em detalhes no capítulo 4.

Tabela 3.1: Quadro Resumo dos Trabalhos Relacionados

| <i>Projeto</i> | <i>GOS</i> | <i>Odyssey-Search</i> | <i>OBSERVER</i> | <i>OntoSeek</i> | <i>InfoSleuth</i> | <i>Kess</i> | <i>THETIS</i> |
|-----------------------------------|------------|-----------------------|-----------------|-----------------|-------------------|-------------|---------------|
| Características | | | | | | | |
| Ontologias de domínio | x | x | x | x | x | x | |
| Ontologias de aplicação | | | | | | | x |
| Tecnologia de Agentes | x | x | | | x | | |
| Metadados | x | x | x | | | | x |
| Mediadores | x | x | | | | | |
| Inferência | x | x | | x | | | |
| Regras | x | | | | | | x |
| Uso de SGBD | x | x | | | x | x | |
| KBS | | | | | | x | x |
| Relacionamentos Semânticos | x | x | x | | | | |
| Serviços Ontológicos | x | | | | | | |

É apresentado ainda um quadro comparativo mais geral (Anexo I), onde além de algumas características são apresentadas contribuições de cada projeto, definição para ontologia, tecnologias utilizadas entre outros tópicos.

Capítulo 4 GOS: Serviços de Ontologia

4.1. Introdução

Segundo NIETO et al. (2000), existe atualmente uma necessidade de complementação das técnicas de recuperação de informação e navegação existentes, através de uma estratégia que tenha seu foco no conteúdo e na semântica da informação. Um problema crítico em caracterizar o conteúdo da informação é o uso de vocabulários diferentes descrevendo informação semelhante, ou seja, diferentes termos são utilizados para caracterizar uma mesma informação.

BERNERS-LEE et al. (2001) completam afirmando que, no contexto da Web, o conteúdo encontrado hoje ainda é projetado para usuários (humanos) consultarem e não para os programas computacionais manipularem, o que explica a quantidade de informação inútil encontrada quando se está pesquisando algum assunto; e isso acontece porque o computador não conhece o significado (semântica) do que o usuário procura.

Esse problema também é crítico nas arquiteturas de integração de banco de dados. Uma solução bem sucedida é o uso de mediadores (WIEDERHOLD, 1994) que apresenta um esquema de representação comum. Entretanto, o vocabulário desse esquema nem sempre é familiar ao usuário. Uma solução para o problema de vocabulários diferentes é a utilização de relacionamentos semânticos entre os termos através do uso de ontologias (NIETO et al., 2000).

Ontologias estão mostrando ser a resposta adequada também para a Web Semântica (HENDLER et al., 2002). Neste novo contexto, a Web será capaz de representar associações entre coisas que, em princípio, poderiam não estar relacionadas. Para isso, computadores necessitam ter acesso a coleções estruturadas de informações (dados e metadados) e de um conjunto de regras de inferência que ajudem no processo de dedução automática. Estas regras são especificadas através de ontologias, que permitem representar explicitamente a semântica dos dados (BERNERS-LEE et al., 2001). Através de ontologias é possível elaborar uma rede enorme de conhecimento humano, complementando o processamento da máquina e melhorando qualitativamente o nível de serviços na Web (MAEDCHE, 2002).

A integração de informações no contexto do Desenvolvimento Baseado em Componentes (DBC), que tem como principal proposta a construção de aplicações a partir de diferentes quesitos já construídos e utilizados em outros sistemas também pode ser

beneficiada com o uso de ontologias. Não se pode reutilizar componentes prontos se não se sabe como e onde encontrá-los. Dentro deste contexto, o repositório surge como uma solução para o armazenamento de componentes, tornando-os públicos e acessíveis a todos. Por isso, a construção de uma infra-estrutura capaz de integrar informações de componentes armazenados em repositórios distribuídos torna-se uma questão importante para um processo de reutilização. Além disso, tal infra-estrutura deve prover também recursos avançados para pesquisa de componentes. As estratégias de busca atualmente disponíveis na Web, baseadas em sua maioria no uso de palavras-chave, permitem a execução de pesquisas de forma limitada, originando respostas muito volumosas e genéricas. Aliado a isto, a falta de informações descritivas (metadados) relacionadas aos componentes publicados, além dos problemas associados a questões semânticas que envolvem a existência de diferentes vocabulários descrevendo informações semelhantes dentro de um mesmo contexto, diminuem ainda mais a precisão e qualidade dos mecanismos de busca existentes.

O sistema GOS (GOA Ontology Services) proposto neste trabalho faz parte da arquitetura do ambiente de desenvolvimento baseado em componentes OdysseyShare (WERNER et al., 2002) e interage através da ComPublish (SOUZA, 2002) com um sistema de mediadores. Além disso, o GOS utiliza um modelo de ontologias aliado a um mecanismo de inferência para lidar com as informações sobre os componentes e com o relacionamento destas diferentes informações em uma ontologia. Cada mediador provê uma visão lógica das informações sobre componentes agrupadas segundo um determinado domínio de aplicação. O uso de ontologias facilita a localização de informações sobre os componentes baseado no vocabulário do domínio de aplicação. O GOS possui serviços para o gerenciamento de várias ontologias com diferentes vocabulários. Para efetuar o mapeamento de uma consulta sobre uma ontologia em termos de outras ontologias, o GOS leva em consideração três tipos de relacionamentos: sinônimos (termos com a mesma semântica), hiperônimos (termos mais genéricos) e hipônimos (termos mais específicos). As consultas dos usuários são reescritas com a mesma semântica, substituindo os termos por termos sinônimos de diferentes ontologias. Os relacionamentos hiperônimos e hipônimos são usados de acordo com a necessidade do usuário. O objetivo do GOS é, portanto, reduzir o problema da necessidade de conhecer a estrutura e semântica de dados espalhados nos vários repositórios para um problema menor que é conhecer os relacionamentos (termos sinônimos, por exemplo) entre os termos através das ontologias.

4.2. O uso do GOS na Arquitetura OdysseyShare

O projeto OdysseyShare tem como objetivo explorar os aspectos colaborativos no DBC, através da construção de mecanismos para o suporte à interação em grupo no ambiente Odyssey (WERNER et al, 2000b). O trabalho em equipe, no contexto do desenvolvimento cooperativo de componentes, pode envolver desenvolvedores em todo o mundo, produzindo e armazenando componentes em repositórios dos mais variados tipos e independentes uns dos outros (diferentes formatos de armazenamento, capacidades de consulta, etc). Por esse motivo, a localização e recuperação de componentes de software, em especial os disponíveis na Internet, é uma tarefa complexa (WERNER et al., 2002).

Neste sentido, foi elaborada uma arquitetura, denominada ComPublish (SOUZA, 2002) que visa auxiliar desenvolvedores de software a publicar variados artefatos de software na Internet, tal como modelos, diagramas, código fonte, programas ou qualquer outro tipo de artefato utilizado ou produzido nas diferentes etapas do processo de desenvolvimento de um software. Através de sua arquitetura baseada em mediadores, a ComPublish (Figura 4.1) provê uma visão lógica de vários componentes publicados local ou remotamente dentro de um mesmo domínio de aplicação (OLIVEIRA et al., 2002). Alguns de seus serviços são:

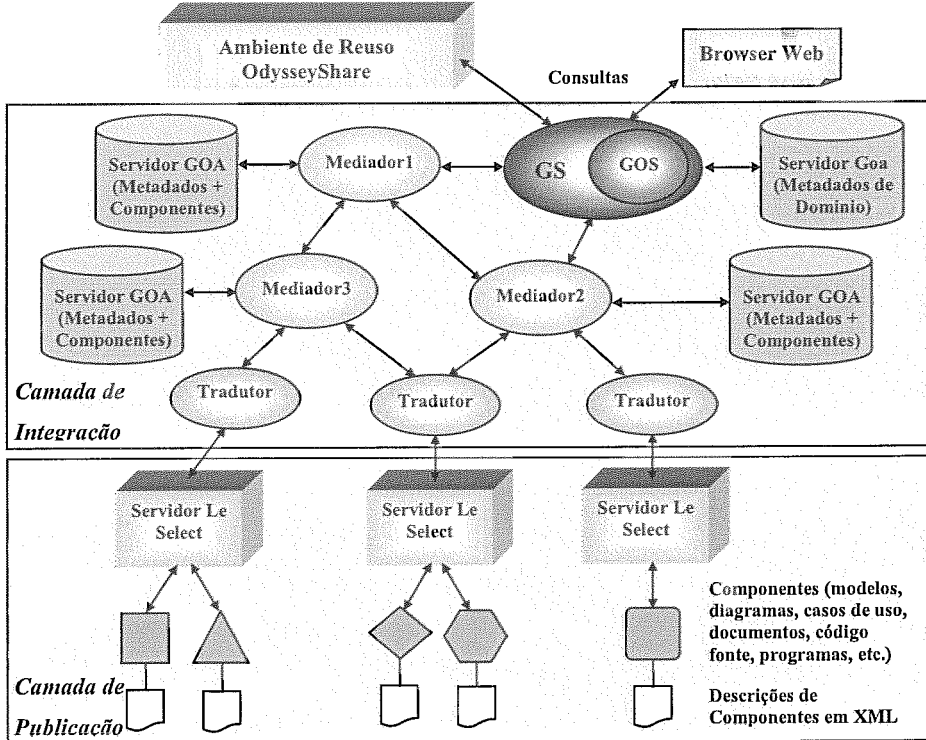


Figura 4.1: Arquitetura ComPublish (OLIVEIRA et al., 2002)

- **Serviços para Publicação de Componentes:** A publicação de componentes em repositórios remotos é feita pelo módulo representado pelo Servidor Le Select. Um repositório pode ser qualquer gerenciador de dados, desde um simples sistema de arquivos a um poderoso sistema de banco de dados. O Le Select fornece os mecanismos necessários para a abstração da heterogeneidade de diferentes repositórios de componentes a serem publicados na Internet. Através dos serviços de publicação de metadados do Le Select, componentes são descritos segundo o padrão XML. Um DTD XML pré-definido é fornecido para cada categoria de componentes, divididos em código, documentos e serviços. Para registrar um componente, o publicador deve fornecer informações como: domínio de aplicação, autor, linguagem, etc. Uma vez publicados, os componentes armazenados no repositório remoto passam a ser referenciados por um mediador específico dentro da camada de integração, através dos metadados exportados de cada componente pelo módulo publicador. Desta forma, cada componente se torna acessível às consultas submetidas por usuários na camada de aplicações. Uma vez descoberto e selecionado, o módulo publicador oferece serviços de ftp para transferir o componente remoto para a máquina local do usuário (OLIVEIRA et al., 2002);
- **Serviços de Mediação para Integração de Metadados:** A camada de integração da ComPublish (Figura 4.1) é formada por um conjunto de mediadores e tradutores que se comunicam entre si segundo um barramento CORBA. Os tradutores encapsulam a API cliente Le Select, que provê os serviços necessários para comunicação com os servidores Le Select remotos. Os mediadores, por sua vez, são responsáveis por organizar as informações dos componentes publicados segundo um determinado domínio de aplicação. Os serviços de armazenamento e consulta dos metadados (incluindo as informações de componentes, os modelos ontológicos e a base de regras) e dos componentes locais a cada mediador são realizados pelo servidor de objetos GOA. Todas as informações de componentes exportadas pelos servidores Le Select no padrão XML são convertidas e armazenadas dentro do GOA, segundo o modelo orientado a objetos. Desta forma, a camada de mediação consegue prover aos usuários serviços de pesquisas de maior eficiência e qualidade, através de consultas seguindo o padrão OO. Dentre os mediadores, existe ainda um mediador especial denominado Gerente de Serviços (GS), que gerencia as informações relacionadas aos demais mediadores desta camada. O GS é

responsável por direcionar as consultas submetidas pela camada de aplicações aos mediadores adequados e realizar as ligações ontológicas intra e inter-domínios através do sistema GOS (OLIVEIRA et al., 2002).

4.3. O Sistema GOS

O sistema GOS (Figura 4.1) é responsável por realizar as ligações ontológicas intra e interdomínios. Este sistema realiza os mapeamentos necessários entre termos do mesmo e de diferentes domínios e mantém as informações atualizadas a respeito do domínio. O GOS recebe do GS solicitações de serviços ontológicos, como por exemplo, criação de ontologias e inclusão de termos ontológicos, descoberta de informações sobre componentes relacionados ao termo em questão e estabelecimento de relacionamentos intra e interontológicos.

A integração de ontologias no contexto do GOS é realizada através do "casamento" das características constantes de cada modelo ontológico, através da identificação de relacionamentos entre estes termos ontológicos de diferentes domínios tais como relações de sinônimos, hipônimos e hiperônimos, descritos neste capítulo. Através destes mecanismos, utilizados em outras abordagens para a integração de esquemas heterogêneos (NIETO, 1998), ontologias de domínio similares podem ser integradas e/ou direcionar as buscas em um dado domínio para domínios correlatos, como também foi apresentado em (BRAGA, 2000). O estabelecimento de relações intra e interontológicas é feito de modo semi-automático e usa uma base de regras juntamente com um mecanismo de inferência para derivar novas relações. A figura 4.2 apresenta os principais módulos da arquitetura GOS descritos a seguir.

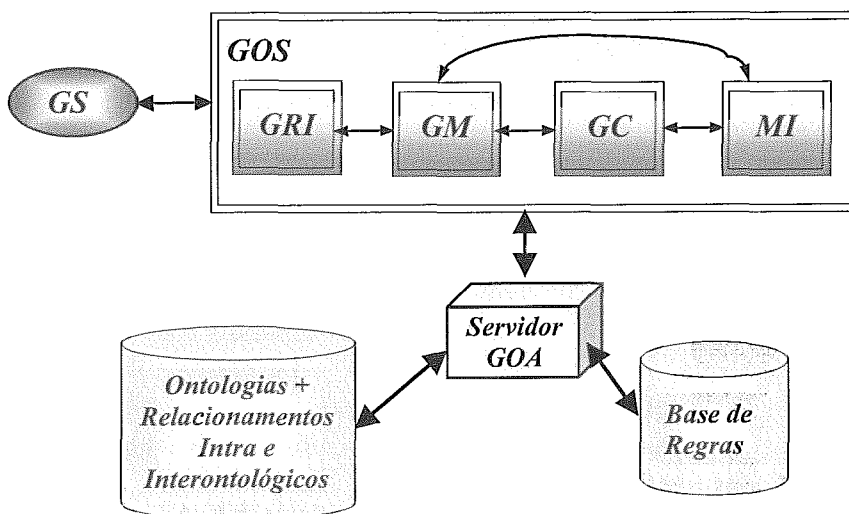


Figura 4.2: Arquitetura GOS

- **Gerente de Consultas (GC):** É encarregado de recuperar informações relacionadas a um determinado termo de uma ontologia específica. O GC navega através das ligações entre termos da ontologia, como sinônimos, hipônimos, hiperônimos e apresenta o resultado como um conjunto de termos relacionados. Caso haja solicitação de verificação em outros domínios (ontologias relacionadas), o GC submete uma consulta ao GRI, que é responsável por verificar termos de outras ontologias, apresentando assim um resultado mais abrangente;
- **Gerente de Relacionamentos Interontológicos (GRI):** Provê informações sobre os relacionamentos interontológicos. Para isso, o GRI gerencia as listas de sinônimos, hiperônimos e hipônimos, entre termos de diferentes ontologias. GRI é o módulo chave que permite à arquitetura GOS lidar com diferentes ontologias criadas de diferentes pontos de vista e semânticas. Os relacionamentos armazenados no GRI permitem que o GOS mapeie termos de uma ontologia sobre termos da mesma ou de outra ontologia diferente, desde que haja um relacionamento semântico entre eles;
- **Gerente de Metadados (GM):** Oferece serviços para visualização das ontologias e de seus termos, criação de ontologias e inclusão de termos em uma ontologia;
- **Máquina de Inferência (MI):** Gera novos relacionamentos intra e interontológicos, a partir de relacionamentos ontológicos já existentes através de um conjunto de regras pré-definido. Estas regras estão armazenadas no Servidor de Objetos GOA (GOA, 2002). Este processamento depende de uma solicitação explícita do administrador.

4.4. GOA

Pela descrição da arquitetura GOS e de seus módulos principais, observa-se que são utilizados os serviços de um gerente de objetos para o armazenamento e consulta aos metadados armazenados. Este gerente de objetos é o GOA (MAURO, 1998).

GOA possui uma arquitetura cliente/servidor de objetos. Em um servidor de objetos, a unidade de transferência entre o servidor e o cliente é o objeto. O servidor conhece o conceito de objeto e é capaz de executar consultas sobre os mesmos. Além disso, quando solicitado, o servidor fornece ao cliente uma cópia do objeto e registra o tipo de bloqueio desejado. Maiores informações sobre o GOA podem ser obtidas em (GOA, 2002).

Uma das funções do GOA é ser utilizado como repositório dos metadados dos mediadores, tradutores e fontes de dados distribuídas com integração na arquitetura de mediação, atuando também como um repositório dos modelos ontológicos e da base de regras utilizada. Além disso, a linguagem de consulta do GOA é utilizada para a consulta aos metadados armazenados. Todos os serviços providos pelo GOS são realizados através de consultas OQL (OMG, 2002) a bases GOA.

4.5. Relacionamentos Intra e Interontológicos

Relacionamentos intra e interontológicos são relacionamentos semânticos entre termos pertencentes a uma mesma ontologia e entre ontologias diferentes, respectivamente. Para realizar o mapeamento ontológico, são utilizados alguns relacionamentos intra e interontológicos. Na literatura são descritos diversos outros relacionamentos (MOKRANE, 2001), (NIETO et al., 2000), no entanto, no contexto da proposta deste trabalho, são considerados os relacionamentos de sinônimos, hiperônimos e hipônimos, uma vez que estes satisfazem de modo adequado a necessidade inicial. Vale mencionar que os construtores básicos para esses três relacionamentos podem ser especializados para a implementação de outros. A descrição destes relacionamentos é apresentada a seguir.

- **Sinônimo:** Um termo ontológico A tem como sinônimo um termo B se possuírem a mesma semântica (Figura 4.3). Para o relacionamento sinônimo, a representação utilizada é *Synonym (A, B)* (lê-se: A tem como sinônimo B). Um relacionamento de sinônimo associa termos ontológicos que são sinônimos entre si;

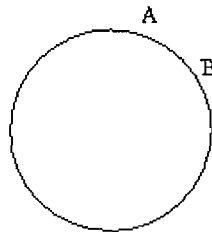


Figura 4.3: Relacionamento Sinônimo: A tem como sinônimo B

- **Hipônimo:** Quando um termo ontológico B é mais específico que um outro termo A, isto é o termo B tem sua abstração incluída pela abstração representada pelo termo A (Figura 4.4). Para o relacionamento hipônimo, utiliza-se a seguinte representação *Hyponym (A, B)* (lê-se: A tem como hipônimo B). Relacionamentos

de hipônimos representam ligações entre termos ontológicos que são mais específicos que o termo ontológico considerado;

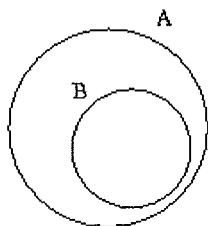


Figura 4.4: Relacionamento Hipônimo: A tem como hipônimo B

- **Hiperônimo:** Quando um termo ontológico B é mais geral que um outro termo A, isto significa que o termo B inclui a abstração representada pelo termo A (Figura 4.5). Para o relacionamento hiperônimo, a representação utilizada é Hypernym (A, B) (lê-se: A tem como hiperônimo B). Relacionamentos de hiperônimos representam ligações entre termos ontológicos que são mais genéricos que o termo ontológico considerado.

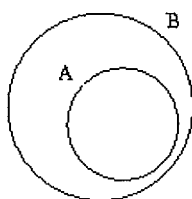


Figura 4.5: Relacionamento Hiperônimo: A tem como hiperônimo B

Desta forma, com estes relacionamentos, é possível associar termos ontológicos de um mesmo domínio ou de diferentes domínios, permitindo a busca integrada por informações reutilizáveis. Através da estrutura de mediação, usuários podem então procurar informações de um modo transparente e uniforme. Isto facilita o trabalho do usuário, uma vez que a busca por informações do domínio leva em consideração que o usuário muitas vezes não sabe o que está procurando ou mesmo onde está localizada a informação de que necessita e às vezes utiliza termos não técnicos, ou não muito utilizados no domínio, para iniciar a busca.

4.6. Serviços

O GOS recebe solicitações de serviços ontológicos, como por exemplo, criação de ontologias e inclusão de termos ontológicos, descoberta de informações sobre componentes

relacionados ao termo em questão e estabelecimento de relacionamentos intra e interontológicos. Os serviços oferecidos pelo GOS são descritos a seguir.

- **Adicionar Ontologia:** Uma ontologia é registrada a partir da informação de seu nome. Vale lembrar que uma ontologia está associada a um domínio bem específico, como por exemplo, a ontologia do Legislativo Municipal que está associado ao Domínio Legislativo. A ontologia é cadastrada com um nome único que a identifica;
- **Incluir Termos em uma Ontologia:** A qualquer momento, termos podem ser incluídos em uma ontologia. Para adicionar um termo, é necessário fornecer um nome, uma denominação (outras designações que o termo possui) e uma descrição (definição deste termo a ser inserido), bem como associá-lo a uma ontologia. Para que os termos de uma ontologia estejam acessíveis para outras ontologias, os relacionamentos intra e interontológicos são estabelecidos. Neste caso, assim que o termo for inserido, são estabelecidos inicialmente, os relacionamentos deste novo termo com os existentes naquela ontologia. Esse processo, semi-automático, é feito através da interface da aplicação. Terminada essa fase, passa-se para o relacionamento deste termo com termos de qualquer outra ontologia, processo este semelhante ao anterior;
- **Relacionar termos em uma mesma ontologia:** Como mencionado anteriormente, após a inclusão de um termo em uma ontologia, passa-se para a fase de estabelecimento de relacionamentos intraontológicos (sinônimos, hiperônimos e hipônimos) através de inferência;
- **Relacionar termos em ontologias diferentes:** Um termo de uma ontologia pode se relacionar com outros termos da mesma ontologia ou até mesmo de ontologias diferentes. Nesta fase são estabelecidos, através de inferência, os relacionamentos dos termos incluídos em uma determinada ontologia com todos os outros termos de todas as outras ontologias;
- **Fornecer Definição:** Dado um termo e uma ontologia é possível visualizar uma descrição (definição) do termo em questão;
- **Fornecer Denominação:** Dado um termo e uma ontologia é possível visualizar as outras denominações que este termo possui;
- **Fornecer Ontologia:** Retorna o nome de todas as ontologias cadastradas;

- **Fornecer Termos Ontológicos:** Dada uma ontologia, retorna o nome de todos os termos cadastrados;
- **Fornecer Relacionamentos Intra e Interontológicos:** Dado um termo e uma ontologia é possível visualizar todos os termos que se relacionam com o termo em questão, através dos relacionamentos sinônimo, hiperônimo e hipônimo, sejam estes termos pertencentes ou não a mesma ontologia do termo selecionado.

4.7. A Base de Regras

O GOS utiliza algumas regras (Figura 4.6) para derivar novos relacionamentos entre os termos. Por atender a necessidade desta proposta, adotou-se o formato de regras SE-ENTÃO para a representação de fatos e o algoritmo *Forward Chaining* para o estabelecimento destes novos relacionamentos intra e interontológicos.

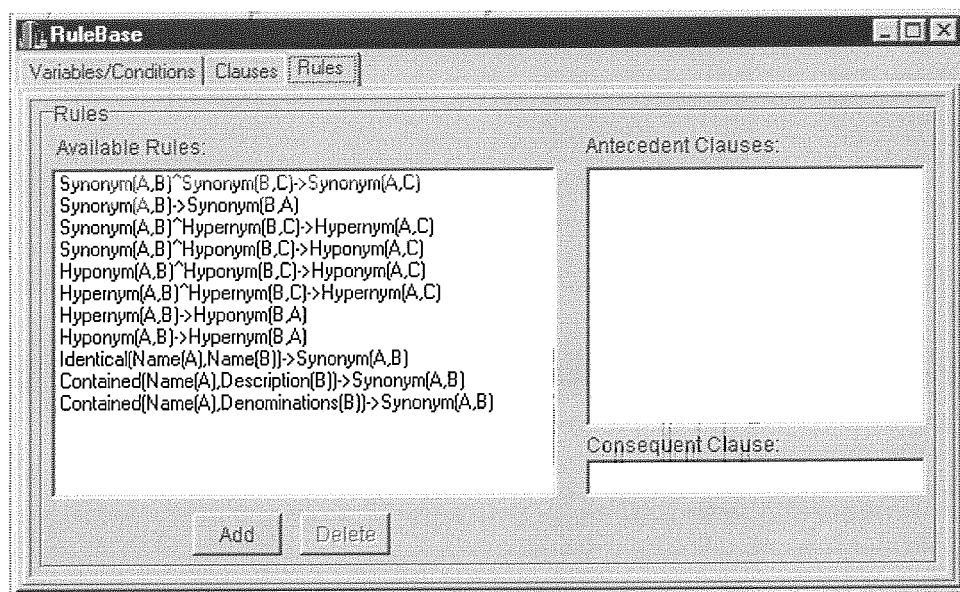


Figura 4.6: As regras

4.7.1. Forward Chaining

Uma maneira simples de representar conhecimento em aplicações de Inteligência Artificial é através da construção de regras SE-ENTÃO. Estas regras são de fácil entendimento, sendo consideradas como parte individual ou uma unidade de informação em uma base de conhecimento. Sendo assim, facilita a adição de novas regras, aumentando o

conhecimento da base. O conhecimento atual pode ser mudado, criando ou modificando as regras individuais (RUSSEL e NORVIG, 1995).

O GOS usa o algoritmo *Forward Chaining* que é um processo de inferência onde um conjunto de regras é utilizado para gerar novos fatos a partir de um conjunto inicial de dados. O *Forward Chaining* gera novos dados pela aplicação simples e direta ou execução das regras. Uma regra trata o relacionamento entre cláusulas (fatos) e, dependendo da situação, pode ser usada para gerar nova informação ou provar a verdade de uma asserção (RUSSEL e NORVIG, 1995).

4.7.2. Cadastro das Regras

Como mencionado, o GOS utiliza algumas regras para gerar novos fatos a partir de um conjunto inicial de dados. O cadastro destas regras é apresentado a seguir. Uma regra é formada a partir de cláusulas antecedentes e conseqüentes. Cada cláusula por sua vez é formada por variáveis e condições. A partir da figura 4.7 novas variáveis e condições podem ser criadas de acordo com a necessidade.

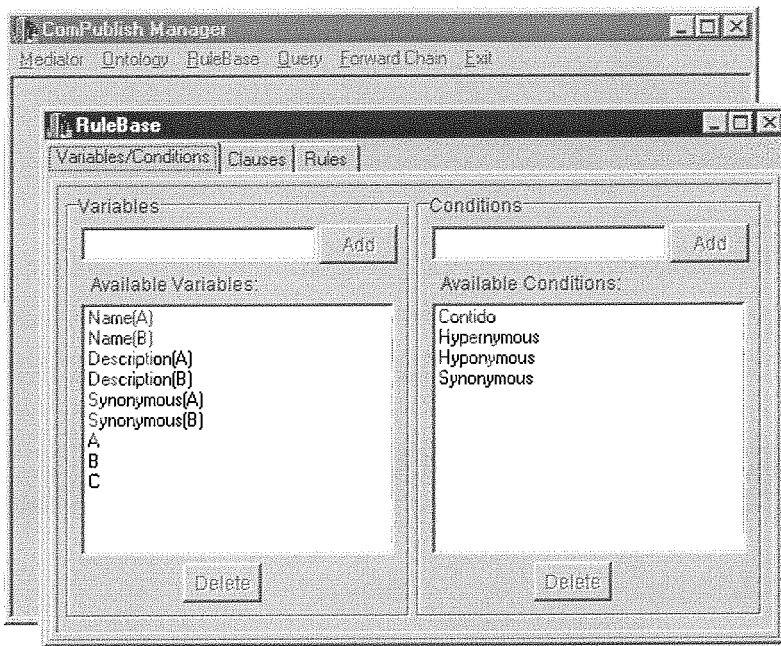


Figura 4.7: Cadastro de Variáveis e Condições

Uma nova cláusula quando adicionada deve ser baseada nas variáveis e condições preexistentes. Uma cláusula possui duas variáveis e uma condição, como mostrado na Fig.4.8.

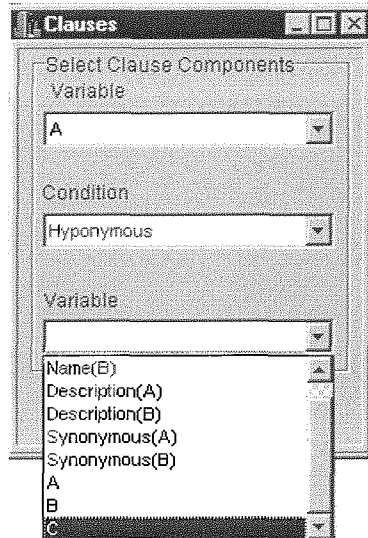


Figura 4.8: Cadastro de Cláusulas

Uma regra é formada por pelo menos uma cláusula antecedente e uma única cláusula conseqüente. Novas regras podem ser definidas e isto deve ser feito a partir das cláusulas cadastradas. Ao cadastrar uma regra, como pode ser visto na figura 4.9, um teste é feito para verificar quais são as regras (não podem existir duas regras iguais) e cláusulas (que podem ser escolhidas para cláusulas conseqüentes e antecedentes) existentes.

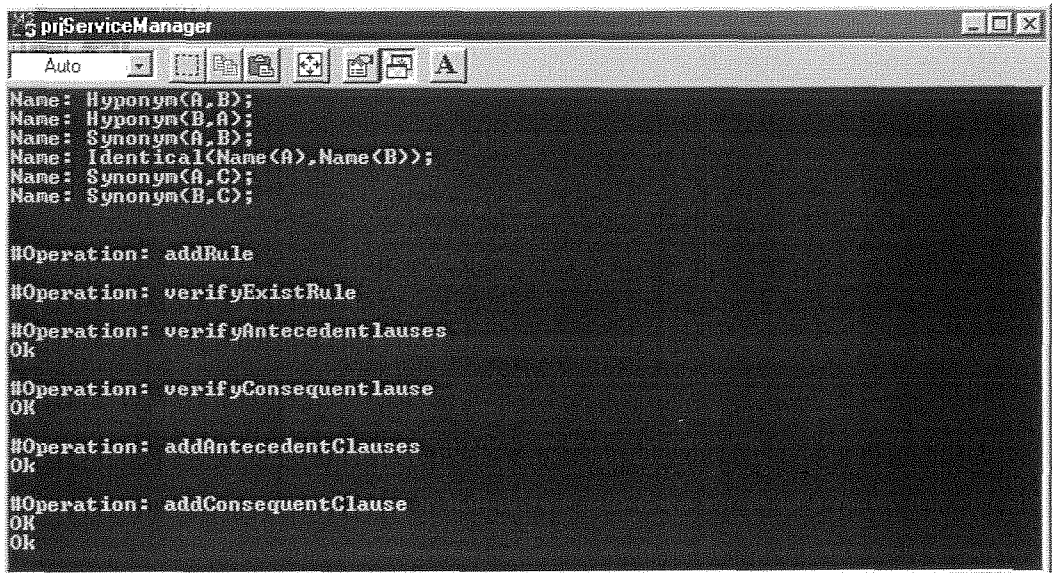


Figura 4.9: Regras e Cláusulas Existentes

Para exemplificar, a figura 4.10 apresenta o cadastramento de uma das regras definidas na seção 4.7.3. Esta regra possui duas cláusulas antecedentes (Synonymous (A, B), Hypernymous (B, C)) e uma cláusula conseqüente que está sendo selecionada (Hypernymous (A, C)) e significa: Se um termo A tem como sinônimo um termo B e este termo B, por sua vez, tem como hipônimo o termo C, então o termo A tem como hipônimo o termo C (como será detalhado na seção 4.7.3.2).

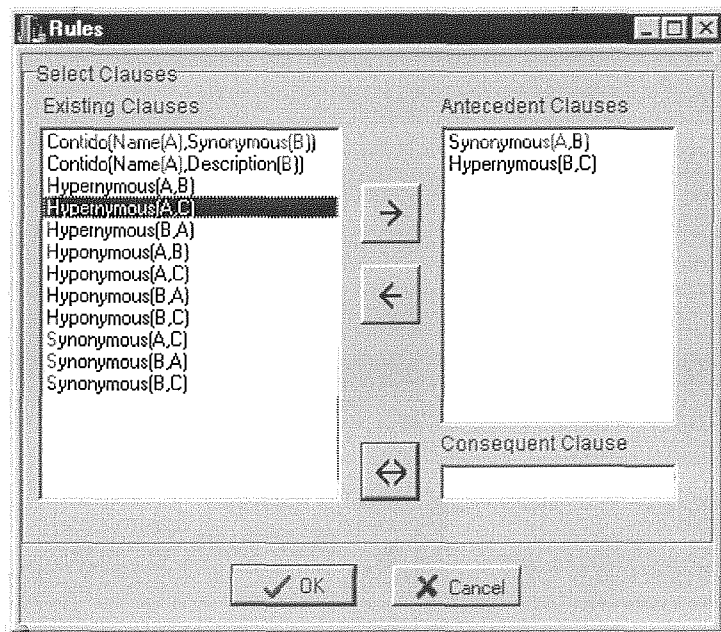


Figura 4.10: Cadastro de Regras

4.7.3. As Regras

A seguir são apresentadas as regras usadas nesta aplicação. Para exemplificar o uso destas regras no estabelecimento de novos relacionamentos entre os termos, foi utilizada uma adaptação do subconjunto das ontologias WordNet e Stanford I (NIETO, 1998). A Figura 4.11 apresenta estas ontologias com alguns relacionamentos preexistentes.

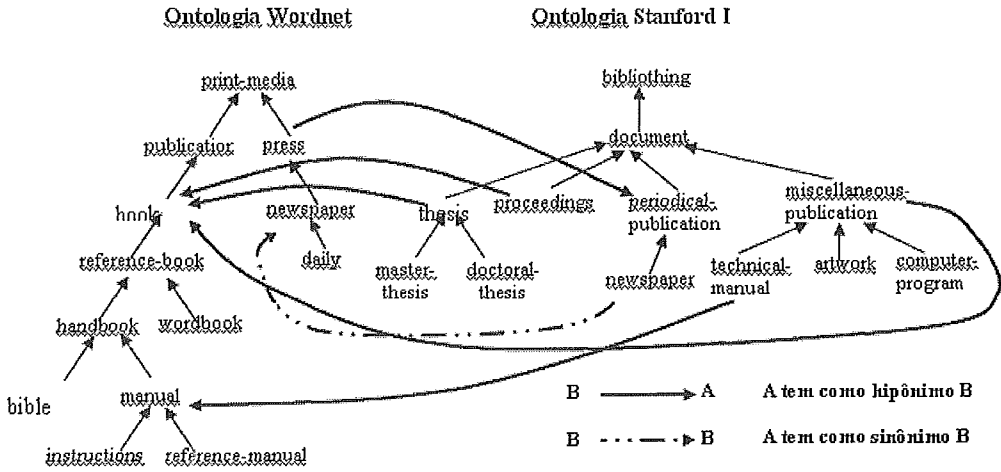


Figura 4.11: Ontologias WordNet e Stanford I

As Tabelas 4.1, 4.2 e 4.3 também apresentam as ontologias WordNet e Stanford I bem como os relacionamentos entre os termos, permitindo-nos assim uma melhor visualização das informações.

Tabela 4.1: Ontologias Stanford I e WordNet: Relacionamentos Intraontológicos

| Ontologia WordNet - Hipônimos | | Ontologia Stanford I - Hipônimos | |
|-------------------------------|------------------|----------------------------------|---------------------------|
| Termo A | Termo B | Termo A | Termo B |
| handbook | bible | bibliothing | document |
| publication | book | document | periodical-publication |
| newspaper | daily | document | proceedings |
| reference-book | hand-book | document | thesis |
| manual | instructions | periodical-publication | newspaper |
| hand-book | manual | thesis | master-thesis |
| press | newspaper | thesis | doctoral-thesis |
| print-media | press | document | miscellaneous-publication |
| print-media | publication | miscellaneous-publication | technical-manual |
| book | reference-book | miscellaneous-publication | artwork |
| manual | reference-manual | miscellaneous-publication | computer-program |
| reference-book | Wordbook | | |

Tabela 4.2: Ontologias Stanford I e WordNet: Relacionamentos Interontológicos

| Ontologias Stanford I e WordNet - Lista de Hipônimos | | | |
|--|------------|-----------------|------------|
| Termo A | Ontologia | Termo B | Ontologia |
| technical-manual | Stanford I | manual | WordNet |
| proceedings | Stanford I | book | WordNet |
| thesis | Stanford I | book | WordNet |
| misc-publication | Stanford I | book | WordNet |
| press | WordNet | periodical-publ | Stanford I |

Tabela 4.3: Ontologias Stanford I e WordNet: Lista de Sinônimos

| Lista de Sinônimos | |
|--------------------|-----------|
| Stanford I | WordNet |
| newspaper | newspaper |

Com o objetivo de facilitar a compreensão, as regras foram divididas de acordo com os tipos de relacionamentos estabelecidos a partir de tais regras (novos relacionamentos de sinônimos, novos relacionamentos hiperônimos e hipônimos).

4.7.3.1 Relacionamentos Sinônimos

No contexto deste trabalho um termo em uma ontologia possui nome, descrição, termos sinônimos entre outras informações. Vale ressaltar que dois termos em uma mesma ontologia não podem possuir o mesmo nome, mas em ontologias diferentes isso pode acontecer. O “campo” sinônimo é responsável por “armazenar” todos os outros termos de uma mesma (ou de outra) ontologia que são sinônimos do termo em questão, ou seja, é uma lista contendo o nome de todos os outros termos da mesma (ou de outra) ontologia que são sinônimos deste primeiro. A descrição é responsável por fornecer uma informação mais detalhada do termo ontológico. O termo ontológico pode possuir outras designações, que não necessariamente são termos ontológicos e por isso o campo denominações foi criado.

Na primeira fase em que se utilizam regras para derivar novos relacionamentos, estas são responsáveis pelo estabelecimento de relacionamentos sinônimos entre os termos ontológicos.

As propriedades definidas são apresentadas a seguir:

- **Identical (Name (A), Name (B)) -> Synonym (A, B):** Se um termo A da ontologia 1 possui o mesmo nome que um termo B da ontologia 2, então o termo A tem como

sinônimo o termo B. Como mencionado anteriormente, termos pertencentes a uma mesma ontologia devem possuir nomes diferentes. Esta regra, portanto, somente será aplicada para termos pertencentes a ontologias diferentes;

- **Contained (Name (A), Description(B)) -> Synonym (A, B):** Se um termo A está contido na descrição de um termo B então o termo A pode ter como sinônimo o termo B. Para tanto, é solicitada a confirmação ou não desta possibilidade por parte do usuário. Em caso positivo, o termo B é tido como sinônimo do termo A.
- **Contained (Name (A), Denomination(B)) -> Synonym (A, B):** Se um termo A está contido no campo denominações de um termo B então o termo A pode ter como sinônimo o termo B. Para tanto, também é solicitada a confirmação ou não desta possibilidade por parte do usuário. Em caso positivo, o termo B é tido como sinônimo do termo A.

As propriedades foram mencionadas em ordem de aplicação, ou seja, em um primeiro momento, verifica-se se um termo é sinônimo de outro termo através do nome. Posteriormente, verifica-se a possibilidade de dois termos serem sinônimos pela observação dos campos descrição e denominação.

Ainda em busca de estabelecer novos termos sinônimos são definidas as seguintes propriedades apresentadas a seguir:

- **Synonym (A, B) -> Synonym (B, A):** Se um termo A tem como sinônimo um termo B então o termo B tem como sinônimo A. A aplicação da regra é exemplificada pelas tabelas 4.4 e 4.5:

Tabela 4.4 – Termos Ontológicos (II)

| Termos | Ontologia |
|-----------|------------|
| newspaper | Stanford I |
| newspaper | Wordnet |

Tabela 4.5 – Cláusulas Antecedentes e Conseqüentes (II)

| Antecedente | Conseqüente |
|--------------------------------|--------------------------------|
| Synonym (newspaper, newspaper) | Synonym (newspaper, newspaper) |

- **Synonym (A, B) ^ Synonym (B, C) -> Synonym (A, C):** Se um termo A tem como sinônimo um termo B e B tem como sinônimo um termo C então o termo A tem como sinônimo o termo C (Synonym (A, C));

4.7.3.2 Relacionamentos Hipônimos e Hiperônimos

Buscando o estabelecimento de novas ligações entre os termos utilizando os relacionamentos hiperônimos e hipônimos são definidas as propriedades a seguir:

- **Hipônimo:** Para o relacionamento hipônimo, é válida a seguinte propriedade, exemplificada pelas tabelas 4.6 e 4.7:
 - **Hyponym (A, B) ^ Hyponym (B, C) -> Hyponym (A, C):** Se um termo A tem como hipônimo um termo B e B tem como hipônimo um termo C então o termo A tem como hipônimo o termo C;

Tabela 4.6 – Termos Ontológicos (III)

| Termos | Ontologia |
|---------------|------------|
| master-thesis | Stanford I |
| thesis | Stanford I |
| document | Stanford I |
| book | Wordnet |

Tabela 4.7 – Cláusulas Antecedentes e Conseqüentes (III)

| Antecedente | Conseqüente |
|--|-----------------------------------|
| Hyponym (document, thesis) ^ Hyponym (thesis, master-thesis) | Hyponym (document, master-thesis) |
| Hyponym (book, thesis) ^ Hyponym (thesis, master-thesis) | Hyponym (book, master-thesis) |

- **Hiperônimo:** Para o relacionamento hiperônimo, é válida a seguinte propriedade, exemplificada pelas tabelas 4.8 e 4.9:
 - **Hypernym (A, B) ^ Hypernym (B, C) -> Hypernym (A, C):** Se um termo A tem como hiperônimo um termo B e B tem como hiperônimo C então o termo A tem como hiperônimo o termo C;

Tabela 4.8 – Termos Ontológicos (IV)

| Termos | Ontologia |
|----------------|------------|
| reference-book | Wordnet |
| hand-book | Wordnet |
| Book | Wordnet |
| master-thesis | Stanford I |
| thesis | Stanford I |

Tabela 4.9 – Cláusulas Antecedentes e Conseqüentes (IV)

| Antecedente | Conseqüente |
|---|--------------------------------|
| Hypernym (hand-book, reference-book) ^ Hypernym (reference-book, book) | Hypernym (hand-book, book) |
| Hypernym (master-thesis, thesis) ^ Hypernym (thesis, book) | Hypernym (master-thesis, book) |

- **Hiperônimo e hipônimo:** Para os relacionamentos hiperônimo e hipônimo são válidas as seguintes propriedades, exemplificadas pelas tabelas 4.10 e 4.11:
 - **Hypernym (A, B) -> Hyponym (B, A):** Se um termo A tem como hiperônimo um termo B então B tem como hipônimo A;
 - **Hyponym (A, B) -> Hypernym (B, A):** Se um termo A tem como hipônimo um termo B então B tem como hiperônimo A.

Tabela 4.10 – Termos Ontológicos (V)

| Termos | Ontologia |
|------------------|------------|
| reference-book | Wordnet |
| hand-book | Wordnet |
| manual | Wordnet |
| technical-manual | Stanford I |

Tabela 4.11 – Cláusulas Antecedentes e Conseqüentes (V)

| Antecedente | Conseqüente |
|--------------------------------------|-------------------------------------|
| Hypernym (hand-book, reference-book) | Hyponym (reference-book, hand-book) |
| Hyponym (manual, technical-manual) | Hypernym (technical-manual, manual) |

Nesta fase de aplicação das regras a seguir, o interesse está em descobrir novos relacionamentos de hiperônimos e hipônimos, baseados em relacionamentos sinônimos, hiperônimos e hipônimos já estabelecidos. As propriedades definidas são apresentadas a seguir:

- **Sinônimo e hiperônimo:** Para os relacionamentos sinônimo e hiperônimo é válida a seguinte propriedade, exemplificada pelas tabelas 4.14 e 4.15:
 - **Synonym (A, B) ^ Hypernym (B, C): -> Hypernym (A, C):** Se um termo A tem como sinônimo um termo B e B tem como hiperônimo um termo C então A tem como hiperônimo C;
- **Sinônimo e hipônimo:** Para os relacionamentos sinônimo e hipônimo é válida a seguinte propriedade, também exemplificada pelas tabelas 4.12 e 4.13:
 - **Synonym (A, B) ^ Hyponym (B, C): -> Hyponym (A, C):** Se um termo A tem como sinônimo um termo B e B tem como hipônimo um termo C então A tem como hipônimo C;

Tabela 4.12 – Termos Ontológicos (VI)

| Termos | Ontologia |
|------------------------|------------|
| newspaper | Stanford I |
| newspaper | Wordnet |
| press | Wordnet |
| periodical-publication | Stanford I |

Tabela 4.13 – Cláusulas Antecedentes e Conseqüentes (VI)

| Antecedente | Conseqüente |
|---|--|
| Synonym (newspaper, newspaper) ^ Hypernym (periodical-publication, newspaper) | Hypernym (periodical-publication, newspaper) |
| Synonym (newspaper, newspaper) ^ Hyponym (press, newspaper) | Hyponym (press, newspaper) |

Apresentadas as regras utilizadas nesta proposta, a figura 4.12 mostra alguns dos novos relacionamentos criados acima.

Como exemplo, são apresentados os detalhes da utilização de raciocínio através de uma regra usada no estabelecimento de relações entre os termos ontológicos de diferentes domínios (Tabela 4.14): Se um termo A tem como sinônimo um termo B (**Synonym (A, B)**) e B tem como hiperônimo C (**Hypernym (B, C)**) então A tem como hiperônimo C (**Hypernym (A, C)**). A regra (**Synonym (A, B) ^ Hypernym (B, C) -> Hypernym (A, C)**) apresenta duas cláusulas antecedentes e uma cláusula conseqüente. Neste exemplo, as cláusulas antecedentes mostram a existência de um relacionamento de sinônimo e outro de hiperônimo, estabelecidos previamente, entre dois termos; e a partir daí conclui-se que B é hiperônimo de C (Tabela 4.15).

Este relacionamento apresentado na tabela 4.15, bem como outros, não foram contemplados inicialmente pelo especialista ao cadastrar os termos e suas ligações. A partir do uso das regras propostas no GOS, porém, foram estabelecidos, resultando assim em uma resposta mais completa para o usuário.

Tabela 4.14 – Termos Ontológicos (I)

| Termos | Ontologia |
|---------------|------------------|
| newspaper | Stanford I |
| newspaper | Wordnet |
| Press | Wordnet |

Tabela 4.15 – Cláusulas Antecedentes e Conseqüentes (I)

| Antecedente | Conseqüente |
|---|-----------------------------|
| Synonym (newspaper, newspaper) ^ Hypernym (newspaper, press) | Hypernym (newspaper, press) |

4.8. Conclusão

Neste capítulo, a arquitetura GOS (Goa Ontology Services) e seus principais componentes foram descritos. O GOS tem como um de seus objetivos o estabelecimento de ligações entre termos de uma mesma ontologia ou de ontologias diferentes. Para tanto, são apresentados e caracterizados os relacionamentos que são importantes dentro do contexto do sistema GOS: sinônimos, hipônimos e hiperônimos. Para que estas ligações sejam efetuadas utilizou-se o algoritmo Forward Chaining para derivar conhecimento, uma máquina de inferência e uma base de regras para a representação de conhecimento (as regras estão no formato se-então). Alguns serviços, como incluir ontologia e incluir termos em uma ontologia

são oferecidos para facilitar o trabalho do usuário. Para melhor compreensão, uma adaptação das ontologias WordNet e Stanford I foram usadas para exemplificar o estabelecimento de ligações entre os termos e a utilização das regras.

Capítulo 5 Usando os Serviços de Ontologia no Domínio Legislativo Municipal

5.1. Introdução

Neste capítulo são mostrados os benefícios da utilização do GOS e de seus serviços inter e intraontológicos no domínio Legislativo Municipal. O GOS foi desenvolvido para ajudar na busca e identificação da informação apropriada. O uso do GOS promove a integração de informações através de mecanismos para traduzir a requisição de informação com o uso de ontologias. Para efetuar o mapeamento de uma consulta sobre uma ontologia em termos de outras ontologias, o GOS utiliza os relacionamentos definidos no seu contexto. Este capítulo mostra como o estabelecimento semi-automático de novos relacionamentos intra e interontológicos é efetuado e apresenta também o uso da base de regras juntamente com o mecanismo de inferência proposto no capítulo 4.

A proposta desta dissertação é motivada por um projeto que foi conduzido no domínio do Poder Legislativo WERNER et al (2000a). Usuários podem se beneficiar de informações deste domínio e de outros relacionados, como o domínio da justiça, entre outros. É importante que informações pertinentes a todos os domínios relacionados, que possam ser utilizadas, sejam apresentadas ao usuário, particularmente quando eles não estão cientes de sua existência. O objetivo do trabalho em WERNER et al (2000a) era disponibilizar uma infraestrutura de apoio ao desenvolvimento de aplicações no domínio de processamento legislativo, baseado na reutilização de componentes de software previamente construídos. Tal iniciativa partiu de uma parceria entre o grupo de reutilização de software da COPPE/UFRJ responsáveis pelo projeto Odyssey (BRAGA et al., 2001) e a equipe de informática da Câmara Municipal do Rio de Janeiro e visava oferecer a possibilidade de melhoria na produção do software legislativo, com custo reduzido, nas diversas Casas Legislativas, na medida em que as soluções encontradas podiam ser disponibilizadas a nível nacional.

Como o processamento legislativo é comum a todas as Casas Legislativas, resguardadas as particularidades de cada uma, este foi o domínio escolhido para apresentação dos serviços de ontologia propostos nesta dissertação. Estendendo o exemplo apresentado em BRAGA (2000), onde foram cadastradas as ontologias do Domínio Legislativo do Rio de Janeiro, as ontologias escolhidas e cadastradas na seção 5.2 através do sistema GOS são as

ontologias da Câmara Municipal de Belo Horizonte (CMBH, 1996) e da Assembléia Legislativa de Minas Gerais (ALMG, 2000).

O uso do sistema GOS no domínio do Poder Legislativo é particularmente interessante, pois no Brasil e em outros países existem várias casas legislativas que estão em níveis diferentes de tecnologia; enfrentando assim dificuldades no desenvolvimento de software, uma vez que tais casas não contam com recursos para tal desenvolvimento. O uso do GOS permite a disponibilização de informações relacionadas aos componentes, que podem ser usadas no domínio do legislativo para a construção de aplicações. Estas informações são relacionadas com a ontologia do domínio Legislativo, o que facilita sua busca, uma vez que os usuários destas casas legislativas não conhecem estas informações, mas sim a semântica do domínio.

5.2. Cadastro e Remoção de Ontologias e Termos Ontológicos

Como já foi dito anteriormente, o processo de especificação das ontologias é feito por especialistas do domínio. Para o cadastro de uma nova ontologia (Figura 5.1), devem ser seguidos os seguintes passos:

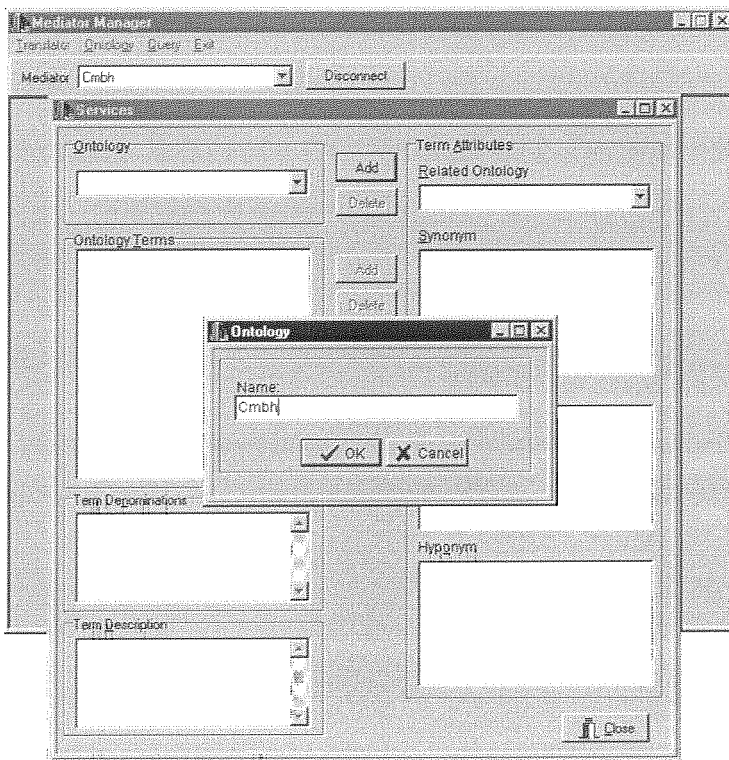


Figura 5.1: Especificação de Ontologias

Uma ontologia ao ser criada deve ser associada a um mediador e deve possuir um nome. Na Figura 5.1 está sendo criada a Ontologia do *Legislativo Municipal* que está sendo chamada de *CMBH* (provê informações do domínio legislativo da cidade de Belo Horizonte - **Câmara Municipal de Belo Horizonte**) associada ao mediador do *Legislativo Municipal* (CMBH, 1996).

Uma ontologia é composta de termos e dos relacionamentos entre estes termos. Para que um termo seja incluído, primeiro deve-se selecionar a ontologia a que ele pertence. Um termo possui nome, descrição, denominação e uma lista de sinônimos, hipônimos e hiperônimos. Vale lembrar que um termo pode possuir termos sinônimos (hipônimos e hiperônimos) não somente na própria ontologia a que ele pertence mas em outras ontologias.

Na Figura 5.2 está sendo cadastrado o termo ontológico denominado *Proposição* pertencente à ontologia do *Legislativo Municipal*, cuja descrição é *para os fins deste regimento, considera-se proposição toda matéria sujeita a deliberação da câmara*.

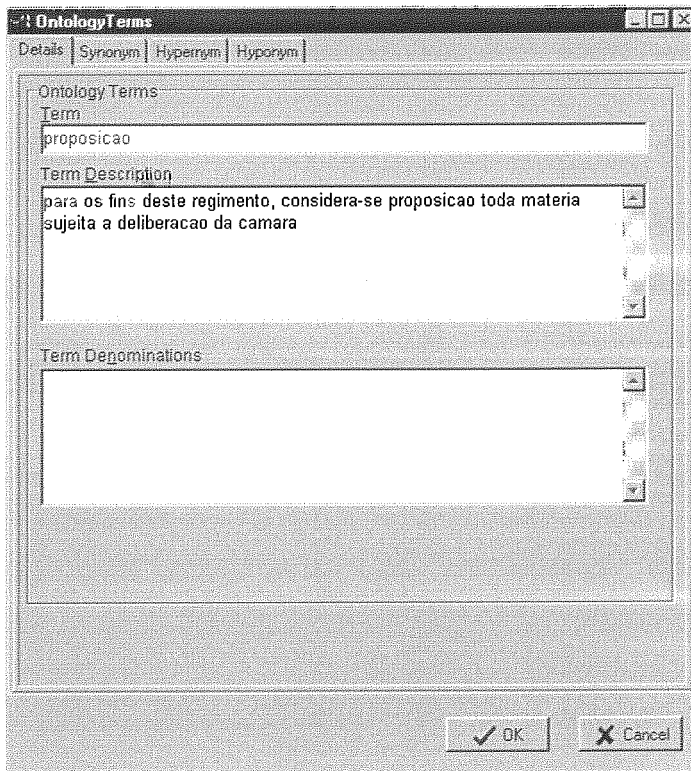


Figura 5.2: Especificação de Termos: Nome, Descrição e Denominação

A Figura 5.3 mostra o cadastramento da lista de hipônimos do termo *Proposição*. Todos estes termos pertencem a Ontologia *CMBH*. Da mesma forma como estão sendo

cadastrados os hipônimos, podem ser cadastrados os outros termos que se relacionam com **Proposição** através de relações de sinônimos ou hiperônimos, sejam eles da mesma ontologia ou não.



Figura 5.3: Especificação de Termos de uma Ontologia: Lista de Hipônimos

Antes de cadastrar um termo, por exemplo o termo **Projeto**, como hipônimo de **Proposição**, pode ser necessário confirmar alguma informação sobre o termo candidato, tal como a descrição do termo ou a sua própria lista de sinônimos, hipônimos ou hiperônimos. Para tanto, ao selecionar o termo candidato e clicar no botão **Show**, todas estas informações sobre o termo são apresentadas (Figura 5.4). De posse da informação necessária, o termo **Projeto** pode ou não ser incluído como hipônimo de **Proposição**.

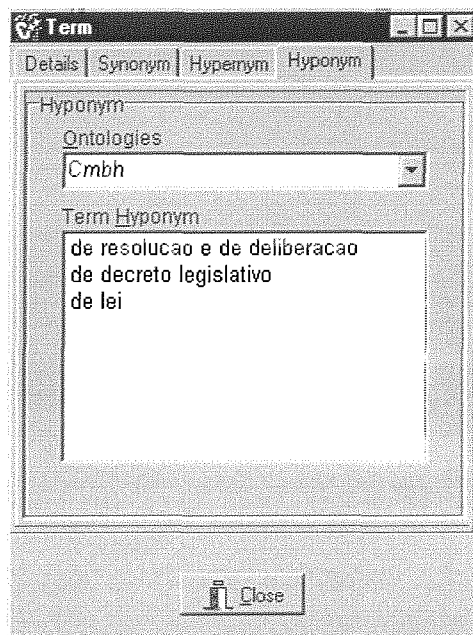


Figura 5.4: Informações Relacionadas ao Termo Projeto

Sobre a remoção de um termo, foi adotado o mesmo critério de remoção de ontologias, ou seja, um termo para ser retirado de uma ontologia, não pode estar relacionado com nenhum outro termo da mesma (ou de outra) ontologia.

Foi efetuado ainda, o cadastramento da ontologia do *Legislativo Estadual*, denominada **ALMG** e que provê informações do domínio legislativo do estado de Minas Gerais – **Assembléia Legislativa de Minas Gerais (ALMG, 2000)**. Como pode ser visto nas tabelas 5.1 e 5.2, foram também cadastrados outros 18 termos ontológicos associados às Ontologias do Legislativo Municipal e Legislativo Estadual. Estes termos serão utilizados na seção 5.3 para mostrar o uso do GOS na recuperação mais completa de informações a respeito de um dado termo. As tabelas 5.1 e 5.2 apresentam ainda os relacionamentos sinônimos, hipônimos e hiperônimos existentes entre os termos da mesma ontologia. Inicialmente, considere que estes são os termos cadastrados e que existe ainda uma relação de sinônimo entre os termos *Emenda* do *Legislativo Municipal* e *Emenda* do *Legislativo Estadual*. Este relacionamento foi estabelecido no momento em que o termo *Emenda* do Legislativo Estadual foi cadastrado e não está sendo contemplado nas tabelas a seguir.

Tabela 5.1: Ontologia do Legislativo Municipal

| Ontologia | Termo | Sinônimos | Hiperônimos | Hipônimos |
|-----------------------|-------------------------------|-----------|-------------|--|
| Legislativo Municipal | de decreto legislativo | | | |
| | de emenda a lei orgânica | | | |
| | de lei | | | de lei delegada; de lei complementar; de emenda a lei orgânica |
| | de lei complementar | | | |
| | de lei delegada | | | |
| | de resolução e de deliberação | | | |
| | emenda | | proposição | |
| | projeto | | proposição | de resolução e de deliberação; de decreto legislativo; de lei |
| | proposição | | | subemenda |
| | subemenda | | | |
| | substitutivo | | proposição | |

Tabela 5.2: Ontologia do Legislativo Estadual

| Ontologia | Termo | Sinônimos | Hiperônimos | Hipônimos |
|----------------------|-----------------------------------|-----------|-------------|---|
| Legislativo Estadual | emenda | | proposição | aditiva, modificativa, supressiva |
| | aditiva | | | |
| | modificativa | | | |
| | supressiva | | | |
| | proposição | | | |
| | sessão legislativa da assembléia | | | sessão legislativa ordinária, sessão legislativa extraordinária |
| | sessão legislativa ordinária | | | |
| | sessão legislativa extraordinária | | | |

A figura 5.5 apresenta uma visão geral dos serviços oferecidos por esta aplicação e que foram definidos no capítulo anterior (Item 4.3). Além do cadastramento das ontologias e termos, podem ser consultados os termos pertencentes a uma ontologia, bem como os relacionamentos de sinônimos, hiperônimos e hiponônimos existentes entre os termos.

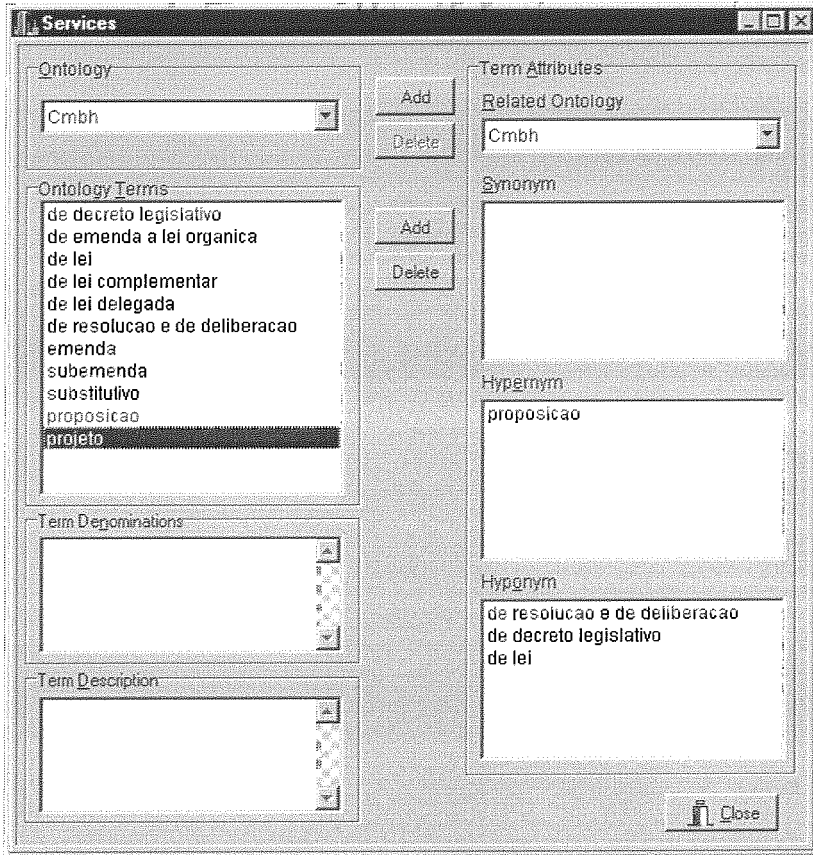


Figura 5.5: O GOS e os Serviços Ontológicos

5.3. Utilizando a Arquitetura GOS

Considere um usuário que está desenvolvendo uma aplicação no domínio do Poder Legislativo Municipal e quer saber informações mais gerais a respeito de determinado termo ontológico. Assim, ele pode usar a interface GOS para saber sobre a disponibilidade (ou não) deste tipo de informação, e recuperar a informação certa.

Para facilitar o entendimento, será analisado um exemplo mais detalhadamente. Considere que, quando o Mediador do Legislativo Municipal foi registrado, ele foi associado ao Mediador do Legislativo Estadual, como pode ser visto na figura 5.6. Assim, quando o

usuário acessa a interface para recuperar informações do *Legislativo Municipal* relacionadas, por exemplo, ao termo *Proposição*, ele pode querer acessar informações pertencente não somente ao próprio domínio mas também dos mediadores relacionados, isto é, informações neste caso, pertencentes ao *Legislativo Estadual*.

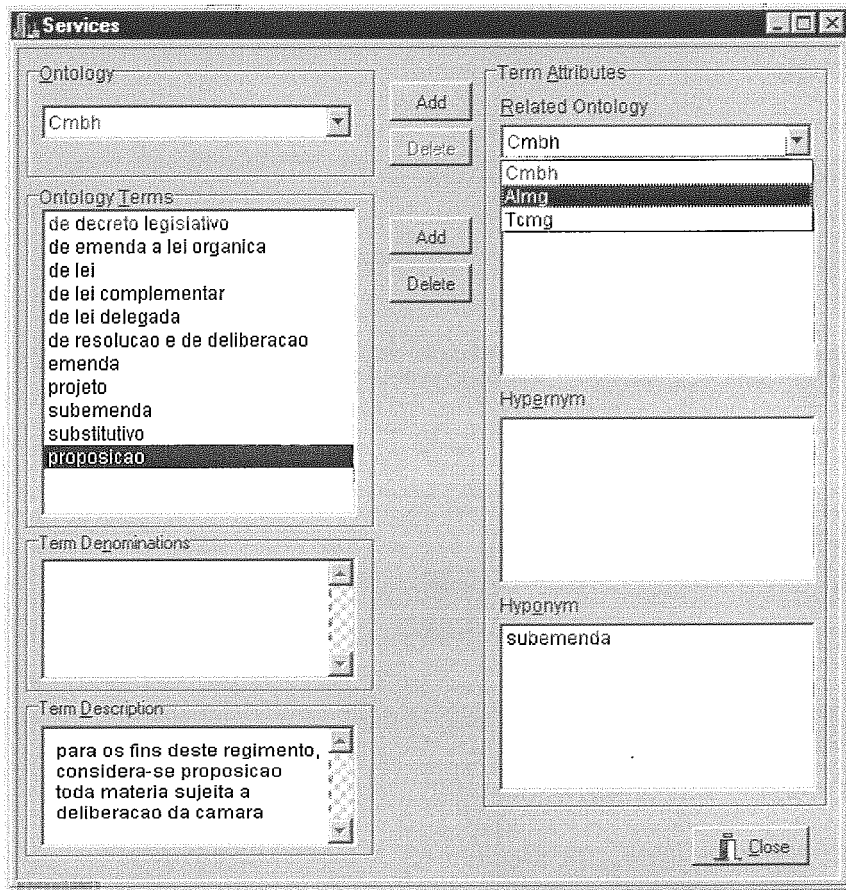


Figura 5.6: Legislativo Municipal e Ontologias Relacionadas

5.3.1. Relacionamentos Intraontológicos

Suponha que o usuário decidiu recuperar informações intraontológicas, relacionadas à *Proposição*, do *Legislativo Municipal*. Pela estrutura da arquitetura GOS, usuários podem então procurar informações de um modo transparente e uniforme. No exemplo acima, usuários não têm que saber onde as informações são armazenadas e nem saber como ter acesso às fontes de dados.

Num primeiro momento, o especialista que cadastrou a ontologia pode não ter efetuado o estabelecimento completo de relacionamentos de cada um dos termos desta ontologia, ou seja, alguns relacionamentos de termos na própria ontologia ou entre ontologias podem não ter sido descobertos. Com isso, quando o usuário resolver buscar informações relacionadas à *Proposição* do *Legislativo Municipal*, possivelmente um conjunto menor de informações será encontrado: *Proposição* tem como hipônimo, o termo *Subemenda* (Figura 5.7).

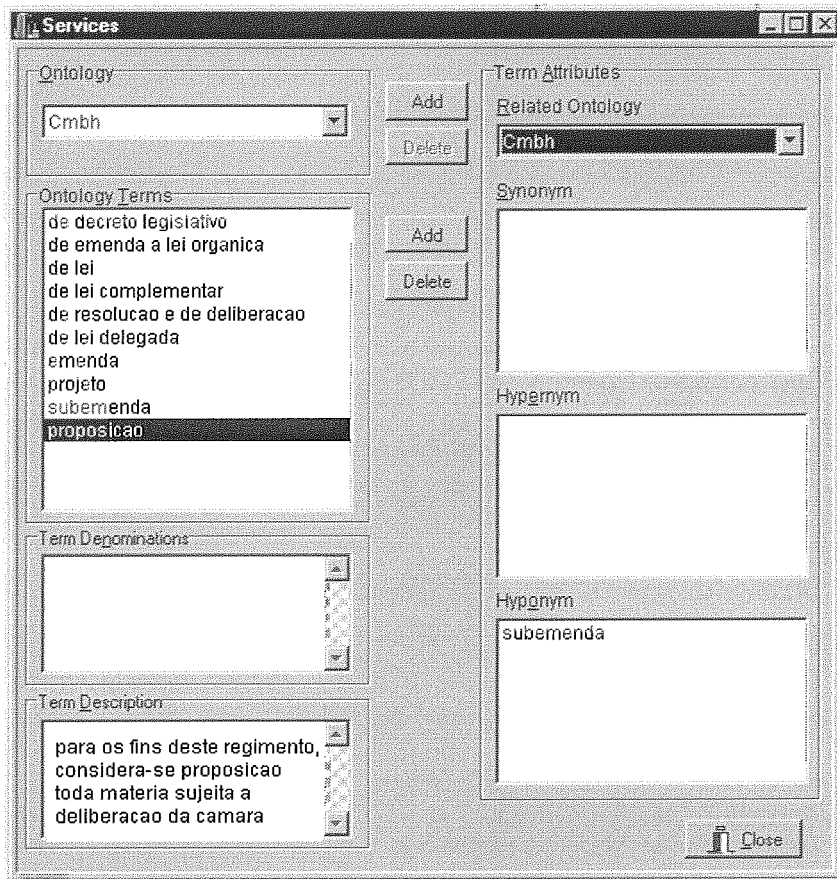


Figura 5.7: Relacionamentos Intraontológicos antes do Uso do GOS

Usando a interface GOS, após o administrador ter efetuado o estabelecimento de novos relacionamentos, através do mecanismo de inferência e da base de regras, pode-se obter um conjunto maior de informações, pois relacionamentos que num primeiro momento não foram contemplados pelo especialista, podem ser descobertos (Figura 5.8): *Proposição* tem como hipônimos os termos *Subemenda*, *Projeto*, *Emenda*, *Substitutivo*, *de Resolução* e *de*

Deliberação, de Decreto Legislativo, de Lei. Assim o GOS recupera a informação ontológica de cada domínio relacionado ao domínio pré-determinado de um modo mais completo, transparente e uniforme.

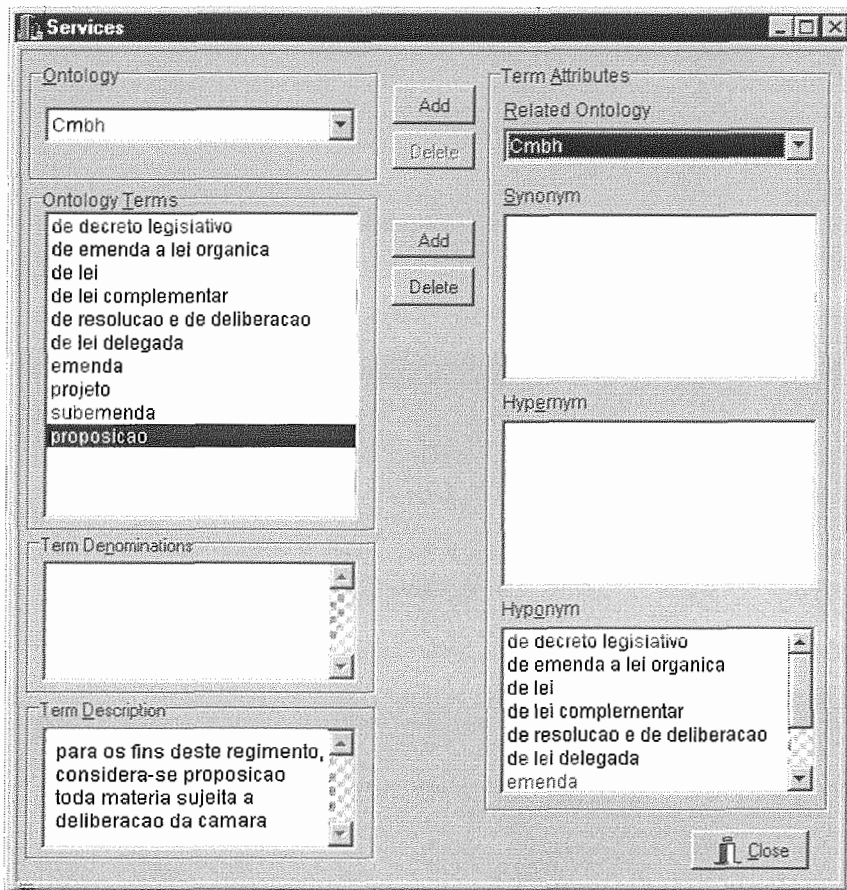


Figura 5.8: GOS: Descoberta de Novos Relacionamentos Intraontológicos

Voltando aos termos das ontologias do Legislativo Municipal e do Legislativo Estadual, cadastrados e apresentados nas tabelas 5.1 e 5.2, pode-se verificar através das tabelas 5.3 e 5.4 que novos relacionamentos foram gerados, não só relacionados ao termo *Proposição* mostrado anteriormente. Os novos relacionamentos entre os termos, gerados a partir do mecanismo de inferência e do uso das regras estão em **negrito** para facilitar a visualização.

Tabela 5.3: Legislativo Municipal: Novos Relacionamentos Intraontológicos

| Ontologia | Termo | Sinônimos | Hiperônimos | Hipônimos |
|--------------------------|----------------------------------|------------------|--|--|
| Legislativo Municipal | de decreto legislativo | | projeto, proposição | |
| | de emenda a lei orgânica | | de lei, projeto, proposição | |
| | de lei | | projeto, proposição | de lei delegada, de lei complementar, de emenda a lei orgânica |
| | de lei complementar | | de lei, projeto, proposição | |
| | de lei delegada | | de lei, projeto, proposição | |
| | de resolução e de deliberação | | projeto, proposição | |
| | emenda | | proposição | |
| | projeto | | proposição | de resolução e de deliberação, de decreto legislativo, de lei |
| | proposição | | | subemenda, projeto, emenda, substitutivo, de resolução e de deliberação, de decreto legislativo, de lei, de emenda a lei orgânica, de lei delegada, de lei complementar |
| | subemenda | | proposição | |
| | substitutivo | | proposição | |

Tabela 5.4: Legislativo Estadual: Novos Relacionamentos Intraontológicos

| Ontologia | Termo | Sinônimos | Hiperônimos | Hipônimos |
|----------------------|-----------------------------------|-----------|---|---|
| Legislativo Estadual | emenda | | proposição | Aditiva, modificativa, supressiva |
| | aditiva | | emenda, proposição | |
| | modificativa | | emenda, proposição | |
| | supressiva | | emenda, proposição | |
| | proposição | | | emenda, aditiva, modificativa, supressiva |
| | sessão legislativa da assembléia | | | sessão legislativa ordinária, sessão legislativa extraordinária |
| | sessão legislativa ordinária | | sessão legislativa da assembléia | |
| | sessão legislativa extraordinária | | sessão legislativa da assembléia | |

5.3.2. Relacionamentos Interontológicos

Suponha agora que o usuário decidiu recuperar informações interontológicas, relacionadas à *Emenda* (do Legislativo Municipal), ou seja, além das informações do termo contidas no próprio Mediador do Legislativo Municipal, o usuário gostaria de verificar a existência de informações relacionadas à *Emenda* no Mediador do Legislativo Estadual.

Neste caso, acontece o mesmo problema que no anterior. Inicialmente, um conjunto menos completo de informações relacionadas à *Emenda*, do *Mediador do Legislativo Municipal* será mostrado ao usuário. Aliás, neste caso inicialmente nenhuma informação relacionada existe. Isso acontece pelo mesmo motivo citado acima: no cadastramento dos termos, nem todos os relacionamentos são mapeados (Figura 5.9).

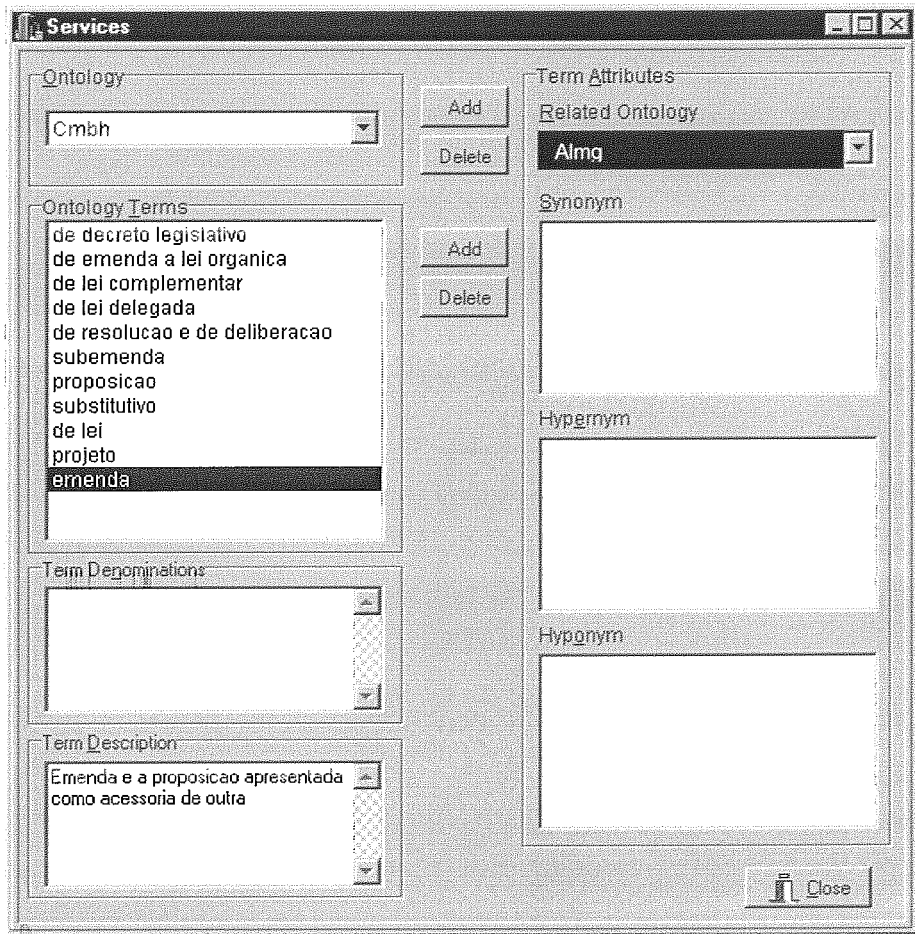


Figura 5.9: Relacionamentos Interontológicos antes do Uso do GOS

Assim, usando a interface GOS, para obter os relacionamentos interontológicos, pode-se obter um conjunto mais completo de informações, com este novo mapeamento. Pela Figura 5.10, pode-se observar a existência de termos pertencentes ao *Legislativo Estadual* relacionados à *Emenda do Mediador do Legislativo Municipal: Emenda* (Sinônimo), *Proposição* (Hiperônimo), *Aditiva*, *Modificativa* e *Supressiva* (Hipônimo). Isso acontece devido ao estabelecimento de novos relacionamentos a partir do conjunto de regras e do mecanismo de inferências mencionado no capítulo 4.

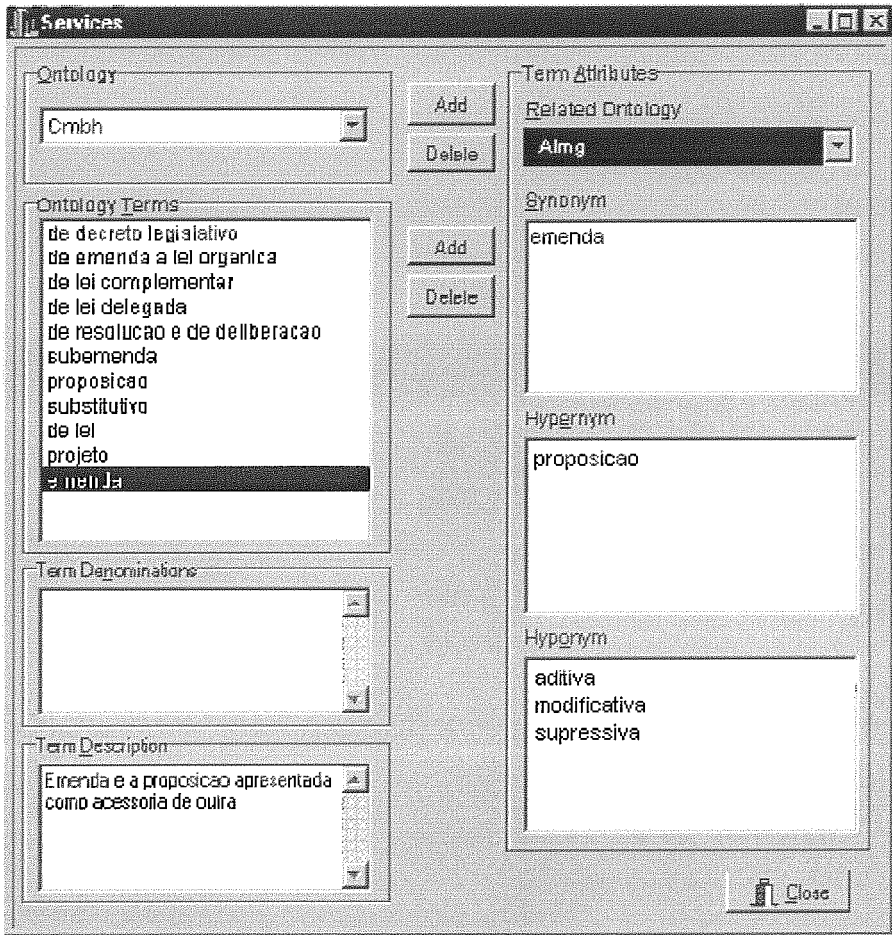


Figura 5.10: Descoberta de Novos Relacionamentos Interontológicos

Para finalizar, são apresentadas as tabelas 5.5 e 5.6 contendo também os novos relacionamentos interontológicos. Os relacionamentos interontológicos descobertos estão em itálico, para melhor entendimento, como por exemplo o termo *Emenda* do *Legislativo Municipal* tem como hipônimo o termo *Aditiva* (que está em itálico), pertencente ao *Legislativo Estadual*.

Tabela 5.5: Legislativo Municipal: Novos Relacionamentos Interontológicos

| Ontologia | Termo | Sinônimos | Hiperônimos | Hipônimos |
|--------------------------|-------------------------------|------------------|------------------------------------|--|
| Legislativo Municipal | de decreto legislativo | | projeto, proposição | |
| | de emenda a lei orgânica | | de lei, projeto, proposição | |
| | de lei | | projeto, proposição | de lei delegada, de lei complementar, de emenda a lei orgânica |
| | de lei complementar | | de lei, projeto, proposição | |
| | de lei delegada | | de lei, projeto, proposição | |
| | de resolução e de deliberação | | projeto, proposição | |
| | emenda | <i>emenda</i> | proposição, <i>proposição</i> | <i>aditiva, modificativa, supressiva</i> |
| | projeto | | proposição | de resolução e de deliberação, de decreto legislativo, de lei |
| | proposição | | | subemenda, projeto, emenda, substitutivo , de resolução e de deliberação, de decreto legislativo, de lei, de emenda a lei orgânica, de lei delegada, de lei complementar, <i>emenda</i> |
| | subemenda | | proposição | |
| substitutivo | | proposição | | |

Tabela 5.6: Legislativo Estadual: Novos Relacionamentos Interontológicos

| Ontologia | Termo | Sinônimos | Hiperônimos | Hipônimos |
|----------------------|-----------------------------------|------------------|---|--|
| Legislativo Estadual | emenda | <i>emenda</i> | proposição, <i>proposição</i> | aditiva, modificativa, supressiva |
| | aditiva | | emenda, emenda | |
| | modificativa | | emenda, emenda | |
| | supressiva | | emenda, emenda | |
| | proposição | | | <i>emenda, emenda,</i> aditiva, modificativa, supressiva |
| | sessão legislativa da assembleia | | | sessão legislativa ordinária, sessão legislativa extraordinária |
| | sessão legislativa ordinária | | sessão legislativa da assembleia | |
| | sessão legislativa extraordinária | | sessão legislativa da assembleia | |

5.4. Conclusão

O protótipo GOS, baseado em mediadores e ontologias, foi desenvolvido para ajudar na busca e identificação da informação apropriada. Para auxiliar na identificação de informações relacionadas e seus domínios apropriados, cada mediador inclui uma ontologia e provê o mapeamento para repositórios de informações relacionados ao domínio.

Neste capítulo, foi apresentado o uso do GOS no domínio do Poder Legislativo. Usuários podem se beneficiar de informações deste domínio e de outros relacionados, como o domínio da justiça, entre outros. Assim, informações de todos os domínios relacionados, que possam ser utilizadas, são apresentadas ao usuário.

Usando o GOS, o usuário pode obter um conjunto maior de informações, pois termos que num primeiro momento não foram contemplados pelo especialista, podem ser descobertos através do mecanismo de inferência e da base de regras. Assim, o GOS recupera a informação ontológica de cada domínio relacionado ao domínio pré-determinado de um modo mais completo, transparente e uniforme.

Capítulo 6 Conclusões e Trabalhos Futuros

Apresentamos neste trabalho uma proposta de um mecanismo para facilitar a busca por informações heterogêneas e distribuídas, mais especificamente, componentes de software distribuídos, através do uso de ontologias.

A integração de informações no contexto do Desenvolvimento Baseado em Componentes (DBC), que tem como principal objetivo a construção de aplicações a partir de diferentes artefatos já construídos e utilizados em outros sistemas, pode ser beneficiada com o uso de ontologias. Não se pode reutilizar componentes prontos se não se sabe como e onde encontrá-los. Por isso, a construção de uma infra-estrutura capaz de integrar informações de componentes armazenados em repositórios distribuídos torna-se uma questão importante para um processo de reutilização. Além disso, tal infra-estrutura deve prover também recursos avançados para pesquisa de componentes.

Diante disso, foi implementado, utilizando-se C++ Builder, o sistema GOS baseado em ontologias, para oferecer serviços ontológicos para serem utilizados na busca e identificação de informações ligadas a um determinado domínio de aplicação. No contexto do GOS, cada mediador provê o mapeamento para repositórios relacionados a um determinado domínio. Assim, uma solução para o problema de vocabulários diferentes relacionados a diferentes domínios e conseqüentemente para a integração de ontologias é a utilização de relacionamentos semânticos entre os termos (NIETO et al., 2000). Para realizar o mapeamento ontológico, utilizamos alguns relacionamentos interontológicos. Na literatura são descritos vários relacionamentos, no entanto, no contexto de nossa proposta, consideramos, os relacionamentos de sinônimos, hiperônimos e hipônimos, uma vez que estes satisfazem de modo adequado nossa necessidade inicial. Acreditamos que os construtores básicos para esses três relacionamentos podem e devem ser especializados para a implementação de outros.

Mediadores fornecem uma visão uniforme das informações disponíveis e ontologias, ligadas a estes mediadores, são usadas na busca destas informações através da representação de conceitos semanticamente relacionados. Esta solução pode ser utilizada junto ao OdysseyShare, além de poder também ser utilizada independente do mesmo.

O uso do GOS junto a uma arquitetura de mediadores promove a integração de informações através de mecanismos para traduzir a requisição de informação com o uso de ontologias. O estabelecimento de novos relacionamentos intra e interontológicos é efetuado

de modo semi-automático, utilizando uma base de regras juntamente com um mecanismo de inferência. O uso de serviços de mediação adiciona uma funcionalidade a mais na arquitetura, que é o acesso transparente a repositórios de informações. No entanto, sem o estabelecimento de relações interontológicas, a busca por informações poderia ficar restrito a um único mediador. No sistema GOS, consultas OQL, mapeadas de forma específica, utilizando os termos ontológicos apropriados a cada domínio, são enviadas aos diversos mediadores e processadas pelo servidor de objetos GOA.

Portanto, a principal contribuição deste trabalho é o desenvolvimento do sistema GOS que mostra o potencial do modelo de ontologias aliado a um mecanismo de inferência para lidar com a complexidade dos repositórios de informações e com o relacionamento de diferentes informações em uma ontologia.

Ontologia é um mecanismo formal capaz de ajudar na viabilização do processamento semântico da informação através de uma máquina. O uso de ontologias contribui para que o entendimento compartilhado de termos possa ser utilizado por homens e programas no intercâmbio de informações. No contexto da Web, a utilização de ontologias também é importante no que diz respeito à compreensão por agentes de software da semântica embutida nas definições e vocabulário especificados com respeito a um domínio, diminuindo as ambigüidades e viabilizando o intercâmbio de informações através de consultas. Bases de conhecimento podem também ser criadas especializando e instanciando aquela ontologia, através de uma aplicação específica. Diversas linguagens e mecanismos para a definição de ontologias foram criados nos últimos anos, a exemplo de: OIL (FENSEL et al., 2000), DAML (ANKOLEKAR et al., 2001), DAML+OIL (GIL e RATNAKAR, 2002), OWL (SMITH et al., 2002), dentre outros. A principal característica dessas linguagens está na capacidade de representar ontologias em XML ou RDF/RDFS, arquiteturas já consagradas pela W3C para interoperabilidade de informações na Web. Alguns editores de ontologia também estão sendo desenvolvidos, tais como OntoEdit (ONTOEDIT, 2002b) e Protégé (NOY et al., 2001). Como trabalho futuro, a integração do protótipo GOS com alguns destes mecanismos de definição de ontologia (como por exemplo, a OWL que deve se tornar um padrão da W3C) e edição de ontologia (OntoEdit) é viável. Para tanto, um estudo mais profundo desse assunto se faz necessário.

Devemos ressaltar também que a integração de ontologias torna-se hoje relevante para o desenvolvimento e consolidação de várias atividades na Web, tais como: prover mecanismos de busca mais precisos; oferecer maior suporte aos serviços de comércio

eletrônico; prover interoperabilidade entre serviços e aplicações, além de prover o entrelaçamento semântico entre os conteúdos das páginas, permitindo o uso eficiente desse ambiente por humanos e máquinas. Desta forma, uma possível utilização do sistema GOS nestes contextos seria viável, desde que adaptações necessárias fossem realizadas.

Ontologias estão também mostrando ser uma das respostas para a Web Semântica. Neste novo contexto, a Web será capaz de representar associações entre coisas que, em princípio, poderiam não estar relacionadas. Para isso, computadores necessitam ter acesso a coleções estruturadas de informações (dados e metadados) e de um conjunto de regras de inferência que ajudem no processo de dedução automática. Estas regras são especificadas através de ontologias, que permitem representar explicitamente a semântica dos dados (BERNERS-LEE et al., 2001). Através de ontologias é possível elaborar uma rede enorme de conhecimento humano, complementando o processamento da máquina e melhorando qualitativamente o nível de serviços na Web (MAEDCHE, 2002).

Referências Bibliográficas

- (ALMG, 2000) *Regimento Interno da Assembléia Legislativa de Minas Gerais*, Disponível em: <http://www.almg.gov.br/reg2/regimento.doc> (último acesso: 15/02/2003), 2000.
- (ANDRADE e SALTZ, 1999) Andrade, H. C. M., Saltz, J., *Towards a Knowledge Base Management System (KBMS): An Ontologie-Aware Database Management System (DBMS)*, University of Maryland, College Park, XIV SBBB, pag. 27-39, Florianópolis, SC, Brasil, Out, 1999.
- (ANDRADE e SALTZ, 2000) Andrade, H. C. M., Saltz, J., *Query Optimization in Kess - An Ontology-Based KBMS*, University of Maryland, College Park, XV SBBB, Disponível em: <http://citeseer.nj.nec.com/388401.html>, (último acesso em: 08/04/2003), João Pessoa, PB, Brasil, Out, 2000.
- (ANKOLEKAR et al., 2001) Ankolekar, A., Burstein, M., Hobbs, J., Lassila, O., Martin, D., McIlraith, S., Narayanan, S., Paolucci, M., Payne, T., Sycara, K., Zeng, H., DAMLS: Semantic Markup for Web Services, Proceedings of the First Semantic Web Working Symposium (SWWS'01), págs. 411–430, Stanford University, California, USA, 2001.
- (ANKOLEKAR et al., 2002a) Ankolekar, A., Burstein, M., Hobbs, J., Lassila, O., Martin, D., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Payne, T., Sycara, K., *DAML-S: Web Service Description for the Semantic Web*, In The First International Semantic Web Conference (ISWC), Lecture Notes in Computer Science 2342, pp.348-363, Springer Verlag, 2002.
- (ANKOLEKAR et al., 2002b) Ankolekar, A., Huch, F., Sycara, K., *Concurrent Execution Semantics for DAML-S with Subtypes*, In The First International Semantic Web Conference (ISWC), Disponível em: http://citeseer.nj.nec.com/ankolekar02_concurrent.html, (último acesso: 08/04/2003), Sardinia, Itália, Junho, 2002.

- (ARANGO, 1988) Arango, G., *Domain Engineering for Software Reuse*, Technical Report UCI-ICS 88-27, Universidade da Califórnia, 1988.
- (BAYARDO et al., 1997) Bayardo, R. J., Bohrer, W., Brice, R., Cichocki, A., Fowler, J., Helal, A., Kashyap, V., Ksiezzyk, T., Martin, G., Nodine, M., Rashid, M., Rusinkiewicz, M., Shea, R., Unnikrishnan, C., Unruh, A., Woelk, D., *InfoSleuth: Agent-Based Semantic Integration of Information in Open e Dynamic Environments*, Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data, págs. 195-206, Tucson, Arizona, United States, 1997.
- (BECHHOFFER et al., 2000) Bechhofer, S., Broekstra, J., Decker, S., Erdmann, M., Fensel, D., Goble, C., Harmelen, F., Horrocks, I., Klein, M., McGuinness, D., Motta, E., Schneider, P., Staab, S., Studer, R., *An Informal Description of Standard OIL e Instance OIL*, Disponível em: <http://www.ontoknowledge.org/oil/download/oil-whitepaper.pdf>, (último acesso: 08/04/2003), nov, 2000.
- (BERNERS-LEE et al., 2001) Berners-Lee, T., Hendler, J., Lassila, O., *The Semantic Web*, Scientific American, maio, 2001
- (BORGO et al., 1997) Borgo, S., Guarino, N., Masolo C., e Vetere, G., *Using a Large Linguistic Ontology for Internet-Retrieval of Object-Oriented Components*, Proc. of 9 th Int. Conf on Software Engineering e Knowledge Engineering (SEKE 97), págs. 528-534, Madrid, Spain, 1997
- (BRAGA, 1999) Braga, R. M. M., *Ferramenta de Navegação Inteligente Odyssey_Search*, COPPE/UFRJ, Relatório Técnico, Programa de Engenharia de Sistemas, Rio de Janeiro, 1999.
- (BRAGA, 2000) Braga, R. M. M., *Busca e Recuperação de Componentes em Ambientes de Reutilização de Software*, Tese de Doutorado, COPPE/UFRJ, Programa de Engenharia de Sistemas, Rio de Janeiro, Dez, 2000

- (BRAGA e WERNER, 1999) Braga, R. M. M., Werner, C. L., Processo de Engenharia de Domínio do Ambiente Odyssey, COPPE/UFRJ, Relatório Técnico, Programa de Engenharia de Sistemas, Rio de Janeiro, 1999.
- (BRAGA et al., 1999a) Braga, R. M. M., Werner, C. L., Mattoso, M., 1999b, *Odyssey: A Reuse Environment Based on Domain Models*, 2nd IEEE Symposium on Application-Specific System e Software Engineering Technology (ASSET'99), págs. 50-57, Richardson, Estados Unidos, 1999.
- (BRAGA et al., 1999b) Braga, R.M.M., Mattoso, M.L.Q., Werner, C.M.L., *The Use of Mediators for Component Retrieval in a Reuse Environment*, Workshop on Component-Based Software Engineering Process, Technology of Object-Oriented Languages e Systems Conference, págs. 542-546, Santa Barbara, California, Estados Unidos, agosto, 1999.
- (BRAGA et al., 2001) Braga, R.M.M., Mattoso, M.L.Q., Werner, C.M.L., *The Use of Mediation e Ontology Technologies for Software Component Information Retrieval*, Proceedings of ACM Symposium on Software Reusability (SSR'01), págs. 19-28, Toronto, Ontario, Canadá, 2001.
- (BROEKSTRA et al., 2000) Broekstra, J., Klein, M., Decker, S., Fensel, D., Horrocks, I., *Adding Formal Semantics to the Web: Building on top of RDF Schema*, in Proceedings of the Workshop ECDL Workshop on the Semantic Web, Disponível em: <http://www.ontoknowledge.org/oil/extending-rdfs.pdf>, (último acesso: 08/04/2003), Lisboa, Portugal, Setembro, 2000
- (CAMPOS, 1994) Campos, M. L. A., *Em Busca de Princípios Comuns na Área de Representação da Informação: Uma Comparação entre o Método de Classificação Facetada, o Método de Tesouro-Baseado-em-Conceito e a Teoria Geral da Terminologia*, Dissertação de Mestrado em Ciência da Informação, ECO/UFRJ e IBICT/DEP, 1994

- (CHAUDHRI et al., 1998) Chaudhri, V. K., Farquhar, A., Fikes, R., Karp, P. D., Rice, J. P., *Open Knowledge Base Connectivity 2.0.3*, Disponível em: <http://www-ksl-svc.stanford.edu:5915/doc/release/okbc/okbc-spec/okbc-2-0-3.pdf>, (último acesso: 08/04/2003), 1998.
- (CHRISTOPHIDES et al., 1999) Christophides, V., Houstis, C., Lalis, S., e Tsalpata, H., *Ontology-Driven Integration of Scientific Repositories*, 4th Workshop on Next Generation Information Technologies e Systems (NGITS'99), págs. 190-202, Zikhron-Yaakov, Israel, Julho, 1999.
- (CMBH, 1996) *Regimento Interno da Câmara Municipal de Belo Horizonte*, Disponível em: <http://www.pbh.gov.br/camaramunicipal/regimentointerno.htm> (último acesso: 15/02/2003), 1996.
- (DAVIS et al., 1993) Davis, R., Shrobe, H., Szolovits, P., *What is a Knowledge Representation?*, AI Magazine, págs. 19-33, Spring, 1993
- (FALBO, 1998) Falbo, R. A., *Integração de Conhecimento em um Ambiente de Desenvolvimento de Software*, Tese de Doutorado, COPPE/UFRJ, Programa de Engenharia de Sistemas, Rio de Janeiro, Dez, 1998.
- (FARQUHAR et al., 1996) Farquhar, A., Fikes, R., Rice, J., *The Ontolingua Server: A Tool for Collaborative Ontology Construction*, Relatório Técnico, KSL-96-26, Knowledge Systems Laboratory, Stanford, CA, 1996
- (FARQUHAR et al., 1997) Farquhar, A., Fikes, R., Rice J. *Tools for Assembling Modular Ontologies in Ontolingua*, In: Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI '97), págs. 436-441, Providence, Rhode Island; Julho, 1997.
- (FENSEL et al., 2000) Fensel, D., Horrocks, I., Van Harmelen, F., Decker, S., Erdmann, M., Klein, M., *OIL in a Nutshell*, Proceedings of the Workshop on Applications of Ontologies e Problem-solving Methods, 14th, European Conference on Artificial

Intelligence ECAI'00, Disponível em: <http://citeseer.nj.nec.com/horrocks00oil.html>, (último acesso: 08/04/2003), Berlin, Alemanha, Agosto, 2000

(FENSEL, 2001) Fensel, D., *Ontologies: A Silver Bullet for Knowledge Management e Electronic Commerce*, ISBN: 3540416021, 2001.

(FERNANDEZ et al., 1997) Fernandez, M., Gomez-Perez, A., Juristo, N., *Methontology: From Ontological Art Towards Ontological Engineering*, Laboratório de Inteligência Artificial, Faculdade de Informática, Universidade Politécnica de Madrid, AAAI Spring Symposium, págs. 33-40, 1997.

(FIKES e FARQUHAR, 1999) Fikes, R., Farquhar, A., *Distributed Repositories of Highly Expressive Reusable Ontologies*, IEEE Intelligent Systems & their applications, v. 14, n. 2 (Mar/Apr), págs. 73-79, 1999.

(FININ et al., 1994) Finin, T., Fritzson, R., McKay, D., McEntire, R., *KQML as an Agent Communication Language*, Proceedings of the 3rd International Conference on Information e Knowledge Management (CIKM'94), ACM Press, págs. 291-316, Gaithersburg, Maryland, Novembro, 1994.

(FOWLER et al., 1999) Fowler, J., Perry, B., Nodine, M., e Bargmeyer, B., *Agent-Based Semantic Interoperability in InfoSleuth*, SIGMOD Record 28, págs. 60-67. 1999.

(GENESERETH e FIKES, 1992) Genesereth, M. R., Fikes, R.E., *Knowledge Interchange Format, version 3.0, Reference Manual, Relatório Técnico-92-1*, Computer Science Dept., Stanford University, 1992.

(GIL e RATNAKAR, 2002) Gil, Y., Ratnakar, V., *A Comparison of (Semantic) Markup Languages*, Proceedings of the 15th International FLAIRS Conference, Special Track on Semantic Web, Disponível em: trellis.semanticweb.org/expect/web/semanticweb/flairs02_comparison.pdf, (último acesso: 08/04/2003), Pensacola, FL, Maio, 2002.

- (GOA, 2002) Disponível em: <http://www.cos.ufrj.br/~goa> (último acesso: 15/02/2003), 2002
- (GÓMEZ-PÉREZ et al., 1995) Gómez-Perez, A., Juristo, N. , Pazos, J., *Evaluation e Assessment of the Knowledge Sharing Technology*, In: Proceedings of the Second International Conference on Building e Sharing Very Large-Scale Knowledge Bases, Editorial IOS Press, Amsterdam, págs. 289-296, 1995.
- (GÓMEZ-PÉREZ et al., 1996) Gómez-Pérez, A., Fernandez, M., Vicente, A. J., *Towards a Method to Conceptualize Domain Ontologies*, in Proceedings of Workshop on Ontological Engineering/ECAI96, págs. 41-52, Budapest, Hungary, 1996
- (GRUBER, 1992) Gruber, T. R., *Ontolingua: A Mechanism to Support Portable Ontologies*, Relatório Técnico KSL 91-66, Stanford University, Knowledge Systems Laboratory, 1992.
- (GRUBER, 1993a) Gruber, T. R., *A Translation Approach to Portable Ontology Specifications*, Knowledge Acquisition, 5 (2): págs. 199 - 220, 1993.
- (GRUBER, 1993b) Gruber, T. R., *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*, In: International Journal of Human-Computer Studies, special issue on Formal Ontology in Conceptual Analysis e Knowledge Representation, págs. 907-928, 1993
- (GRUBER, 1999) Gruber, T. R., *What is an Ontology*, Disponível em: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> (último acesso: 15/02/2003), 1999.
- (GRÜNINGER e FOX, 1995) Grüninger, M., Fox, M. S., *Methodology for the Design e Evaluation of Ontologies*, Relatório Técnico, University of Toronto, 1995.
- (GUARINO, 1995) Guarino, N., *Formal Ontology, Conceptual Analysis e Knowledge, Representation*. International Journal of Human e Computer Studies, págs. 625-640, 1995.

- (GUARINO e GIARRETA, 1995) Guarino, N., Giarreta, P., *Ontologies e Knowledge Bases - Towards a Terminological Clarification*, In: Mars, N. J. I. (ed.), *Towards Very Large Knowledge Bases: Knowledge Building e Knowledge Sharing*, IOS Press, pags. 25-32, Amsterdam, 1995.
- (GUARINO, 1997a) Guarino, N., *Understanding, Building, e Using Ontologies*, A Commentary to "Using Explicit Ontologies in KBS Development", by van Heijst, Schreiber, e Wielinga, *International Journal of Human e Computer Studies*, vol 46, n. 2/3, págs. 293 - 310, 1997.
- (GUARINO, 1997b) Guarino, N., *Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, e Integration*, In: M. T. Pazienza (ed) *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, Springer Verlag: págs.139-170, 1997.
- (GUARINO, 1998) Guarino, N., *Formal Ontology e Information Systems*, In: Guarino, N. (ed.), *Proc. Of the 1st International Conference*, págs. 3-15, Trento, Italy, IOS Press, 1998.
- (GUARINO et al., 1999) Guarino, N., Masolo C., e Vetere, G., *OntoSeek: Content-Based Access to the Web*, *IEEE Intelligent Systems* 14 (3), págs. 70-80, May/June 1999.
- (HENDLER et al., 2002) Hendler, J., Berners-Lee, T., Miller, E., *Integrating Applications on the Semantic Web*, *Journal of the Institute of Electrical Engineers of Japan*, Vol 122(10), págs 676-680, Outubro, 2002.
- (HORROCKS et al., 2000) Horrocks, I., Fensel, D., Broekstra, J., Decker, S., Erdmann, M., Goble, C., Harmelen, F., Klein, M., Staab, S., Studer, R., *OIL: The Ontology Inference Layer*, Relatório Técnico, University of Manchester, Vrije Universiteit Amsterdam, 2000.
- (HWANG, 1999) Hwang, C. H., *Incompletely e Imprecisely Speaking: Using Dynamic Ontologies for Representing e Retrieving Information*, InfoSleuth Group, MCC,

Texas, In. Proceedings of the 6th International Workshop on Knowledge Representation meets Databases (KRDB'99), Linkping, Sweden, 29-30 julho, 1999.

(JACOBSON et al., 1997) Jacobson, I., Griss, M., Jonsson, P., *Software Reuse: Architecture, Process e Organization for Business Success*; Addison Wesley Longman, 1997.

(JAVA, 2002) Disponível em: <http://java.sun.com> (último acesso: 15/02/2003), 2002.

(JOHNSON, 1992) Johnson, R., *Documenting Frameworks Using Patterns*, Proceedings of the Conference on Object-Oriented Programming System, Languages and Interfaces (OOPSLA'92), págs. 63-76, Vancouver, Canadá, Outubro, 1992.

(KLEIN et al., 2000) Klein, M., Fensel, D., van Harmelen, F., Horrocks, I., *The Relation between Ontologies e Schema-Languages: Translating OIL-Specifications to XML-Schema*, In Proceedings of the Workshop on Applications of Ontologies e Problem-solving Methods, 14th European Conference on Artificial Intelligence ECAI'00, Disponível em: <http://delicias.dia.fi.upm.es/WORKSHOP/ECAI00/7.pdf>, Berlin, Alemanha, Agosto, 2000.

(MAEDCHE, 2002) Maedche, A., *Ontology Learning for the Semantic Web*, University of Karlsruhe, Germany, Kluwer Academic Publishers, 2002.

(MAEDCHE e STAAB, 2001) Maedche, A., Staab, S., *Ontology Learning for the Semantic Web*, IEEE, págs. 72-79, 2001.

(MAURO, 1998) Mauro, R. C., *Aspectos de Gerência de Objetos Persistentes: A Implementação do GOA ++*, Dissertação de Mestrado, COPPE/UFRJ, Programa de Engenharia de Sistemas e Computação, Rio de Janeiro, Dez, 1998.

(MCGUINNESS e HARMELEN, 2002) McGuinness, D. L., Harmelen, F., *Feature Synopsis for OWL Lite e OWL*, Disponível em: <http://www.w3.org/TR/2002/WD-owl-features-20020729/> (último acesso: 17/02/2003), 2002.

- (MELLO e HEUSER, 2000) Mello, R., Heuser, C., *Aplicação de Ontologias a Dados Semi-Estruturados*, CLEI 2000 - XXVI Conferência Latinoamericana de Informática, Disponível em: http://metropole.inf.ufrgs.br/grupo_heuser/groupPublications.html; 2000.
- (MITRA et al., 1999) Mitra, P., Wiederhold, G., Jannink, J., *Semi-automatic Integration of Knowledge Sources*, Proceedings of Fusion '99, Sunnyvale CA, Disponível em: <http://www-db.stanford.edu/SKC/publications.html>, Julho, 1999.
- (MOKRANE, 2001) Mokrane, B., *Using Semantics to Improve Schema e Data Integration*, International Workshop on Information Integration on the Web - Technologies e Applications, Disponível em: <http://www.cos.ufrj.br/wiiw/slides/mokrane.pdf>, Rio de Janeiro, Brasil, 2001
- (MOURA, 2002) Moura, A. M., *A Web Semântica: Fundamentos e Tecnologias*, tutorial ministrado no XVII Simpósio Brasileiro de Banco de Dados, Gramado, RS, Brasil, 2002.
- (NIETO, 1998) Nieto, E. M., *OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies*, Tese de Doutorado, Departamento de Informática e Engenharia de Sistema, Universidade de Zaragoza, Nov, 1998.
- (NIETO et al., 2000) Nieto, E. M., Illarramendi, A., Kashyap, V., Sheth, A. P., *OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies*, International Journal on Distributed e Parallel Databases (DAPD), ISSN 0926-8782, 8(2): págs. 223-272, Abril, 2000.
- (NOY et al., 2000) Noy, N., Ferguson, R., Musen, M., *The Knowledge Model of Protégé-2000: Combining Interoperability e Flexibility*, Proc. Knowledge Engineering e Knowledge Management: 12th Int'l Conf. (EKAW-2000), Lecture Notes in Artificial Intelligence, no. 1937, págs.17-32, Springer-Verlag, Berlin, 2000.

(NOY et al., 2001) Noy, N., Sintek, M., Decker, S., Crubézy, M., Ferguson, R., Musen, M., *Creating Semantic Web Contents with Protégé-2000*, IEEE Intelligent Systems, mar/abr, 16(2): págs. 60-71, 2001.

(ODYSSEYSHARE, 2002) Disponível em: <http://www.cos.ufrj.br/~odyssey/share/> (último acesso: 15/02/2003), 2002.

(OLIVEIRA, 1999) Oliveira, K. M., *Modelo para Construção de Ambientes de Desenvolvimento de Software Orientados a Domínio*, Tese de Doutorado, COPPE/UFRJ, Programa de Engenharia de Sistemas, Rio de Janeiro, Out, 1999.

(OLIVEIRA et al., 2002) Oliveira, A. M., Souza, R. P., Braga, R. M. M., Mattoso, M. L. Q., Werner, C. M. L., *Uso de Mediadores e Ontologias para Recuperação de Componentes de Software*, WDBC 2002, Workshop de Desenvolvimento Baseado em Componentes, Itaipava, Rio de Janeiro, 2002.

(OMG, 2002) Disponível em: <http://www.omg.org> (último acesso: 15/02/2003), 2002.

(ONTOBROKER, 2002) *How to Use Ontobroker - Users e Developers Guide for the Ontobroker System*, Tutorial Version 1.4, covers Ontobroker Version 3.1, Disponível em: http://www.ontoprise.de/download/ontobroker_tutorial.pdf, (último acesso: 18/11/2002), 2002

(ONTOEDIT, 2002a) *OntoEdit Datasheet 2002*, Disponível em: http://www.ontoprise.de/documents/ontoedit_data_sheet.pdf, (último acesso: 18/11/2002), 2002

(ONTOEDIT, 2002b) *OntoEdit Manual, OntoEdit V2.6*, Disponível em: http://www.ontoprise.de/documents/ontoedit_tutorial.zip, (último acesso: 18/11/2002), 2002

- (OUKSEL e SHETH, 1999) Ouksel, A. M., Sheth, A., *Semantic Interoperability in Systems Information Global*, SIGMOD Record. Vol 28, Número 1, Disponível em: <http://www.acm.org/sigmod/record/issues/9903/>, Março, 1999.
- (PATEL-SCHNEIDER et al., 2002) Patel-Schneider, P. F., Hayes, P., Horrocks, I., Harmelen, F., *Web Ontology Language (OWL) Abstract Syntax e Semantics*, Disponível em: <http://www.w3.org/TR/2002/WD-owl-semantics-20021108/> (último acesso: 11/01/2003), 2002.
- (RDF, 2002) *RDF Vocabulary Description Language 1.0: RDF Schema*, Disponível em: <http://www.w3.org/TR/rdf-schema>, (último acesso: 15/02/2003), 2002
- (RDF, 2003) Disponível em: <http://www.w3.org/RDF/> (último acesso: 15/02/2003).
- (RUSSEL e NORVIG, 1995) Russel, S., Norvig, P., *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.
- (SAMETINGER, 1997) Sametinger, J., *Software Engineering with Reusable Components*, Springer-Verlag, ISBN 3-540-62695-6, 1997.
- (SCHREIBER et al., 1995) Schreiber, G., Wielinga, B., Jansweijer, W., *The Kactus View on the 'O' Word*, In: Workshop on Basic Ontological Issues in Knowledge Sharing/IJCAI95, Disponível em: <http://www.swi.psy.uva.nl/usr/Schreiber/papers/Schreiber95a.pdf>, Montreal, Canadá, 1995.
- (SHAW e GARLAN, 1996) M. Shaw, D. Garlan, *Software Architecture, Perspectives on an Emerging Discipline*, 1ª Ed., Prentice Hall, 1996.
- (SMITH et al., 2002) Smith, M. K., McGuinness, D., Volz, R., Welty, C., *Web Ontology Language (OWL) Guide Version 1.0*, Disponível em: <http://www.w3.org/TR/2002/WD-owl-guide-20021104/> (último acesso em: 11/01/2003), 2002.

- (SOUZA, 2002) Souza, R. P., *Compublish: Um Sistema para a Publicação, Busca e Recuperação de Componentes de Software na Internet*, Dissertação de Mestrado, COPPE/UFRJ, Programa de Engenharia de Sistemas e Computação, Rio de Janeiro, Dez, 2002.
- (STAAB e MAEDCHE, 2000) Staab, S., Maedche, A., *Ontology Engineering beyond the Modeling of Concepts e Relations*, ECAI'2000, Workshop on Applications of Ontologies e Problem-Solving Methods, Disponível em: <http://delicias.dia.fi.upm.es/WORKSHOP/ECAI00/cfp.html>, Berlin, 2000.
- (SWARTOUT et al., 1996) Swartout, B., Patil, R., Knight, K., e Russ, T. *Toward Distributed Use of Large-Scale Ontologies*, In Proceedings of 10th Knowledge Acquisition for Knowledge-Based Systems Worskhop, Disponível em: http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html, Banff, Canadá, 1996.
- (USCHOLD e GRÜNINGER, 1996) Uschold, M., Grüninger, M., *Ontologies: Principles, Methods e Applications*, AIAI – TR- 191, Disponível em: <ftp://ftp.aiai.ed.ac.uk/pub/documents/1996/96-ker-intro-ontologies.ps.gz>, 1996,
- (WERNER et al., 2000a) Werner, C. M. L., Braga, R. M. M., Roseti, M. Z., Barros, M. O., Murta, L. G. P. *Odyssey-LE: Uma Infra-Estrutura de Reutilização para o Domínio de Processamento Legislativo*, V ENIAL - Encontro Nacional de Informática Aplicada ao Legislativo, Disponível em: <http://www.cos.ufij.br/~odyssey/publicacoes/enial00.pdf> Vitória, ES, agosto 2000.
- (WERNER et al, 2000b) Werner, C. M. L., Braga, R.M.M, Mattoso, M.L.Q., Murta, L.G.P, Miler, N., Costa, M. N., Pinheiro, R.; Oliveira, A.M., *Infra-estrutura Odyssey: Estágio Atual*, SBES'2000, Caderno de Ferramentas, págs. 366-369, João Pessoa, 2000.
- (WERNER et al., 2002) Werner, C. M. L., Mangan, M. A. S., Murta, L. G. P., Pinheiro, R.; Oliveira, A. M., Mattoso, M. L. Q., Braga, R. M. M., Borges, M., *OdysseyShare:*

Um ambiente para o desenvolvimento Cooperativo de Componentes, XVI Simpósio Brasileiro de Engenharia de Software, Caderno de Ferramentas, págs. 444-449, Gramado, RS, outubro 2002.

(WIEDERHOLD, 1994) Wiederhold, G., *Interoperation, Mediation, e Ontologies*, Proceedings International Symposium on Fifth Generation Computer Systems (FGCS94), Workshop on Heterogeneous Cooperative Knowledge-Bases, Vol.W3, pages 33-48, ICOT, Tóquio, Japão, Dezembro, 1994.

(WIEDERHOLD, 1999) Wiederhold, G., *Mediation to Deal with Heterogeneous Data Sources*, Proc. Interop'99, págs. 1-16, Disponível em: <http://www-b.stanford.edu/pub/gio/1999/Interopdocfigs.html>, (último acesso em: 08/04/2003), Zurich, Março, 1999.

(WIEDERHOLD e JANNINK, 1998) Wiederhold, G., Jannink, J., *Composing Diverse Ontologies*, Relatório Técnico, August 1998.

(WOOLDRIDGE et al., 2000) Wooldridge, M., Jennings, N. R., Kinny, D., *The Gaia Methodology for Agent-Oriented Analysis e Design*, In Journal of Autonomous Agents e Multi-Agent Systems, págs. 285-312, Setembro, 2000.

(XAVIER, 2001) Xavier, J., *Uma Ferramenta de Apoio à Definição e Instanciação de Arquiteturas Específicas de Domínio*, Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, Brasil, 2001.

(XML, 2000) *Extensible Markup Language (XML) 1.0 (Second Edition)*, Disponível em: <http://www.w3.org/TR/2000/REC-xml-20001006>,. (último acesso: 15/01/2003), 2000.

Anexo I – Projetos Relacionados: Quadro Comparativo

| Projeto | Conceito de Ontologia | Funcionalidade/Como e/ou para que é utilizada ontologia? | Por que usar ontologia? | Tecnologia | Contribuições |
|---|---|--|---|--|---|
| GOS - Universidade Federal do Rio de Janeiro – COPPE/Sistemas | Especificação de um vocabulário de termos e das relações entre estes termos, existindo também a preocupação de que essa descrição seja compartilhada dentro da comunidade que concorda com sua definição. | - Usar ontologias como metadados para descrever repositório de dados distribuídos e heterogêneos; - Ontologias são consultadas pelos usuários e, por isso apresentamos um mecanismo que, utilizando-se de conhecimento e de um processo de geração e recuperação de regras baseadas em agregações sobre ontologias, permite obter as respostas acessando os repositórios de dados subjacentes | - Auxiliar o usuário que encontra dificuldade em efetuar a busca de um conjunto de informações, quando estas estão espalhadas pelos mediadores; - Descobrir termos relacionados, ligações existentes e estabelecer novas ligações. | - Regras para cuidar da representação das ontologias - Algoritmo de Forward Chaining para gerar novos fatos a partir de um conjunto inicial de dados - Relacionamentos semânticos - Ontologias de domínio - Metadados - OQL | - Gerenciamento de várias ontologias com vocabulários diferentes (navegação) - Uso de relacionamentos semânticos - Gerenciamento de respostas com perda de informação associada (limitada pelo usuário) - Dimensionamento da perda de informação |
| Odyssey Search - Universidade Federal do Rio de Janeiro – COPPE/Sistemas - Regina Braga | - Vocabulário de termos e relações entre eles. Uma definição deve ser criada para cada termo através de uma descrição informal bem como os relacionamentos entre os termos, formando uma rede semântica. | - Odyssey-Search utiliza ontologia de domínio para a busca por informações do domínio, no contexto de um dado domínio ou entre domínios correlatos. - Busca por componentes reutilizáveis em outros domínios de aplicação, onde um conjunto de relacionamentos básicos entre termos de diferentes ontologias foi implementado. | - Auxiliar o usuário na busca de um conjunto de informações; - Descobrir termos relacionados, ligações existentes e estabelecer novas ligações. | - Relacionamentos semânticos - Ontologias de domínio - Metadados - OQL | - Gerenciamento de várias ontologias com vocabulários diferentes (navegação) - Uso de relacionamentos semânticos |
| OBSERVER - Universidade de Zaragoza - Eduardo Mena Nieto | - Conjunto de termos de interesse em um domínio particular e os relacionamentos entre estes termos | - Usar ontologias como metadados para descrever repositório de dados distribuídos e heterogêneos; - Ontologias são consultadas pelos usuários e por isso apresentam um mecanismo que permite obter as respostas acessando os repositórios de dados subjacentes. | - Tornar explícito o conteúdo da informação independente das estruturas de dados subjacentes que podem ser usadas para armazenar a informação em um repositório de dados. | - Metadados; - Lógica Descritiva; - Relacionamentos semânticos; - Ontologias de domínio. | - Gerenciamento de várias ontologias com vocabulários diferentes (navegação); - Uso de relacionamentos semânticos; - Gerenciamento de respostas com perda de informação associada e dimensionamento da perda. |
| Kess - Universidade de Maryland - Henrique Andrade | - Meio de formalizar conhecimento e os relacionamentos entre objetos em um domínio de interesse. | - Kess é um servidor SQL com um modelo de dados semântico baseado em ontologia que permite a execução de inferências limitadas; - Projetado para aplicações que requerem a utilização de ontologias ou cadeias semânticas para responder consultas que necessitam de algum conhecimento extra (tarefas de exploração de dados). | - Compartilhar conhecimento para solucionar problemas pela troca de mensagens de acordo com a consulta; - Verificação da base de conhecimento: Validação dos dados com nos sistemas tradicionais | - Servidor SQL com capacidade de armazenar conhecimento; - SQL estendido. | - Prover conhecimento para um SGBD convencional; - Otimização de consulta. |

| <i>Projeto</i> | <i>Conceito de Ontologia</i> | <i>Funcionalidade/Como e/ou para que é utilizada ontologia?</i> | <i>Por que usar ontologia?</i> | <i>Tecnologia</i> | <i>Contribuições</i> |
|--|--|---|--|---|---|
| OntoSeek - Stefano Borgo (Ladseb /CNR) - Nicolas Guarino (Corinto) | - Conjunto de termos e os relacionamentos entre estes termos | - Ferramenta para recuperação de componentes orientado a objetos baseada em ontologia, com alguma capacidade de inferência e um processo de recuperação baseado em palavras chave; - Projetado para codificar uma informação através de um grafo simples de conceitos e relações rotuladas por palavras em inglês; consultar uma base de dados previamente codificados e executar o casamento semântico baseado em ontologia. | - Executar o “casamento” entre dados e programas apoiado na ontologia lingüística na codificação do componente. | - Mecanismo de inferência; - Linguagem de expressividade limitada, priorizando a simplicidade dos requisitos mais importantes; - Estrutura de grafos para representar as consultas e os componentes; - Processo de busca que implementa um algoritmo de casamento de grafo | - Recuperação de informação baseada em ontologia. |
| THETIS - Universidade de Creta – Grécia - Vassilis Christophides | - Descrição de conceitos e relacionamentos entre estes conceitos relacionados a um domínio particular, tarefa ou aplicação | - Produzir combinações significativas de conjuntos de dados e programas científicos disponíveis para gerar dados precisos para uma tarefa particular (por exemplo planejamento ambiental); - Sistema Baseado em Conhecimento que captura a relação semântica entre dados e programas apoiados em uma ontologia específica da aplicação, e a partir desta informação, dinamicamente deduz caminhos de produção de dados válidos sem a intervenção do usuário. | Compartilhar conhecimento sobre aspectos específicos do mundo real. | - KBS; - Cláusulas de Horn; - Metadados; - Ontologias de aplicação. | Produzir combinações significativas de conjuntos de dados e programas científicos disponíveis. |
| InfoSleuth - Microelectronics e Computer Technology Corporation - Austin, Texas | - Especificação de um vocabulário para um domínio de discurso compartilhado que pode incluir definições de classes, relacionamentos, funções e outros objetos. | - Usadas para capturar esquema de BDs, modelos conceituais e aspectos da arquitetura de agentes InfoSleuth; - Ontologia InfoSleuth é um corpo de metadados que descreve conhecimento dos agentes e os seus relacionamentos com cada um dos outros agentes; - Decisões sobre decomposição de consultas são baseadas em ontologia de domínio, escolhida pelo usuário que descreve o conhecimento sobre os relacionamentos dos dados armazenados. | - Determinar a importância de uma informação sem acessar os dados subjacentes e capturar novas e diferentes visões em ambientes abertos; - Especificar a informação manipulada pelos agentes e o relacionamento entre eles. | - Tecnologia de agentes. - Ontologias de domínio; - KQML: Linguagem de comunicação dos agentes; - KIF: Representar conhecimento; - SQL: Usada internamente para representar consultas sobre ontologias específicas. | - Infra-estrutura baseada em agente, cooperando para ajudar o usuário a encontrar e recuperar a informação necessária; - Processamento de consulta baseada em conceitos, por isso ontologias são usadas para prover especificação declarativa e precisa da informação semântica. |