

# Modelos de Domínio e Ontologias: uma comparação através de um estudo de caso prático em hidrologia

Marco Aurélio S. Mangan<sup>1,2</sup>  
Leonardo Gresta Paulino Murta<sup>1</sup>  
Jano Moreira de Souza<sup>1</sup>  
Cláudia Maria Lima Werner<sup>1</sup>  
{mangan, murta, jano, werner}@cos.ufrj.br

<sup>1</sup>COPPE/UFRJ – Programa de Engenharia de Sistemas e Computação  
Universidade Federal do Rio de Janeiro  
Caixa Postal 68511 – CEP. 21945-970  
Rio de Janeiro – Brasil

<sup>2</sup>Faculdade de Informática – Pontifícia Universidade Católica do Rio Grande do Sul  
Av. Ipiranga, 6681 Prédio 30, Bloco 4 – CEP. 90619-9000  
Porto Alegre – Brasil

## Abstract

Ontologies and Domain Models are domain knowledge representations largely accepted by industry and academia. Both techniques share the same object of study (concrete reality) and supply its practitioners with a set of similar abstractions (discrete entities, attributes, and relationships concepts, for instance). This work presents a case study applying ontology and domain modeling abstractions to represent the scientific domain of hydrology. Throughout this case we point out similarities and suggest the necessity of a generalized representation combining strengths of both approaches.

**Palavras-chave:** Modelos de Domínio, Ontologia, Aquisição de Conhecimento, Modelagem do Conhecimento.

## 1 Introdução

A Gestão do Conhecimento é uma abordagem de administração que valoriza os ativos intangíveis de uma organização. O principal ativo intangível de uma organização é o seu conhecimento (STEWART, 1998; SVEIBY, 1997).

O conhecimento e as estratégias de gestão de conhecimento são considerados atualmente como um fator determinante do sucesso de uma empresa que deseja manter-se neste ambiente competitivo gerado pelo fenômeno da globalização.

De uma forma mais ampla, o problema da gestão de conhecimento (STEWART, 1998) não é exclusivo das empresas. A gestão de conhecimento pode ser aplicada em organizações públicas e na gestão de um país (COMITEE ON KNOWLEDGE ASSESSMENT OFFICE OF INTERNATIONAL AFFAIRS NATIONAL RESEARCH COUNCIL, 1996). A própria ciência (do latim: *scientia*, conhecimento) está em busca constante por novos processos de gestão de conhecimento que auxiliem na manipulação do volume de conhecimento e informação acumulado nos dias atuais.

O computador é a principal ferramenta de auxílio disponível na tarefa de registro, organização, transformação e busca desse conhecimento acumulado. A condição para que o computador possa ser utilizado é a existência de representações de conhecimento explícitas que possam armazenar informações de forma acessível aos programas.

### 1.1 O problema da gestão de conhecimento

O problema enfrentado atualmente na Gestão do Conhecimento é que grande parte do conhecimento disponível necessário para a ciência, para a economia nacional e para as empresas encontra-se inacessível aos indivíduos e entidades durante a tomada de decisão e o planejamento, em grande parte, por não se encontrar disponível para o processamento através de programas.

O conhecimento não se encontra disponível por dois fatores: (1) não existe representação computacional disponível para esse conhecimento e (2) a representação disponível não é adequada ao processamento desejado.

Cada fator tem seus próprios desdobramentos. No primeiro, existe a possibilidade de que o conhecimento não seja tratável computacionalmente. Caso seja tratável, representar esse conhecimento depende de um processo de captura adequado. No segundo, a captura já ocorreu, mas a representação escolhida para o armazenamento não é adequada para o tipo de processamento necessário em uma situação específica. Em alguns casos, o conhecimento capturado encontra-se disponível para um número reduzido de usuários locais de um mesmo sistema.

Grande parte das informações disponível atualmente possui uma semântica insuficiente por ser descontextualizada. Armazena-se a forma e não o conteúdo, o significado das coisas. Portanto, uma tarefa importante no domínio tecnológico da gestão de conhecimento é identificar quais as características desejáveis para um modelo de representação de conhecimento.

## **1.2 Motivação e objetivos do trabalho**

Uma visão simplificada do processo de gestão de conhecimento (NONAKA et al., 1997) consiste de apenas quatro etapas: (1) criação e aquisição, (2) organização e armazenamento, (3) distribuição e (4) aplicação do conhecimento.

Ontologias e modelagem de domínios são duas técnicas de grande aceitação no domínio tecnológico da gestão de conhecimento. A motivação deste trabalho está na comparação dos modelos para organização e armazenamento disponíveis nessas duas técnicas. A utilidade dessas representações está em tornar o conhecimento de um domínio disponível em um sistema de computador.

A análise apresentada realiza uma comparação entre as abstrações oferecidas para a modelagem por cada abordagem e busca identificar fatores comuns que nos permitam propor componentes gerais de uma representação do conhecimento de domínio adequada para a manipulação através de programas de computador.

A perspectiva de comparação escolhida é a prática de modelagem através do uso de uma ferramenta adequada para cada abordagem. As ferramentas escolhidas refletem as possibilidades de modelagem práticas de cada abordagem. O estudo de caso apresentado parte do princípio que a aquisição já foi realizada e que o indivíduo que realizará a modelagem já possui algum domínio sobre o conhecimento sendo modelado.

A comparação irá se concentrar em determinar as abstrações disponíveis em cada abordagem destacando abstrações para representação de (1) entidades no domínio do problema, (2) seus atributos e (3) relacionamentos.

O restante deste trabalho encontra-se organizado da seguinte forma. A Seção 2 discute sobre outras comparações existentes entre ontologias e modelos de domínio. A Seção 3 apresenta algumas informações sobre o domínio de ontologia e justifica a escolha deste domínio. A Seção 4 apresenta uma ontologia desenvolvida para este domínio e a Seção 5 apresenta o modelo de domínio que consideramos correspondente. A Seção 6 apresenta a comparação das abstrações utilizadas. A Seção 7 apresenta as conclusões deste trabalho.

## **2 Ontologias e Modelos de Domínio**

Nesta seção, apresentamos o processo geral de aplicação das duas técnicas comparadas, explicando os motivos que nos permitem compará-las. O uso combinado de ontologias e modelos de domínios já foi abordado em outros trabalhos. Ainda nesta seção, vamos relacionar em linhas gerais as duas técnicas e comentar casos conhecidos do seu relacionamento.

Ambas as técnicas foram propostas como uma forma de otimizar a construção de sistemas de computador. Modelos de domínio são usados, principalmente, pela comunidade de reuso de software (ARANGO, 1994). Ontologias são aplicadas, principalmente, pela comunidade de inteligência artificial na perspectiva de modelagem de conhecimento. Entretanto, à medida que se tornam maduras, passam a contribuir para outras áreas do conhecimento, a gestão do conhecimento é uma dessas áreas.

### **2.1 Processo de aplicação de ontologias e modelos de domínio**

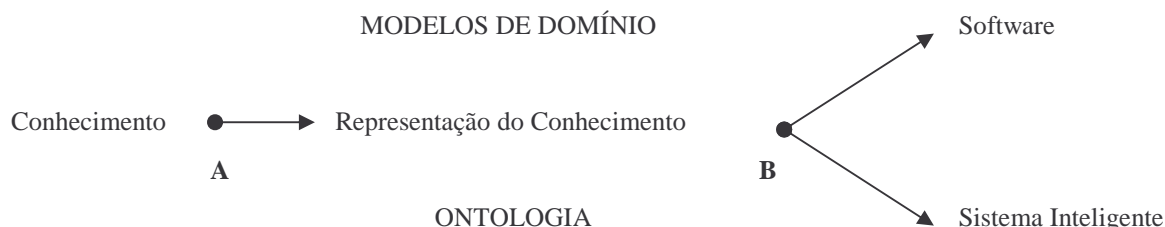
Uma visão geral do processo de aplicação de modelos de domínio e ontologias torna mais clara essa relação com a gestão de conhecimento. A Figura 1 apresenta os processos de modelagem de ambas as abordagens sobrepostos.

A partir do conhecimento disponível sobre uma realidade concreta, um especialista desenvolve uma representação do conhecimento útil e necessária para os seus propósitos. O propósito do engenheiro de domínio é a criação de um programa de computador (*software*). Para o engenheiro de conhecimento, o propósito é a criação de informação necessária para um determinado *sistema inteligente*.

Em ambas as técnicas, o especialista que realiza a modelagem inicia seu trabalho tentando compreender um domínio de problema desconhecido. Nesta etapa inicial, ambas as técnicas determinam que o problema seja representado através de conceitos do próprio espaço de problema, ou seja, não devem ser inseridos conceitos ou relacionamentos que não se apresentam naturalmente no domínio. Essa etapa do processo corresponde ao segmento AB na Figura 1.

Até o ponto B, em ambos os processos, a representação do conhecimento ainda é familiar a um especialista do domínio, o que permite uma validação da modelagem. Após o ponto B, a representação é transformada

através de refinamentos sucessivos em uma nova representação que é significativa apenas para o indivíduo que realiza a modelagem. Em outras palavras, o conhecimento passa a ser representado no domínio da solução (sistemas inteligentes ou programas de computador, dependendo da abordagem).



**Figura 1: O processo de aplicação de Ontologias e de Modelos de Domínio.**

Uma revisão de diferentes processos em ambas as técnicas pode ser encontrada em outros trabalhos (ARANGO, 1994; FERRE et al., 1999; JONES et al., 1998). Neste trabalho, nos concentramos nas representações que são geradas dentro do segmento AB, ou seja, enquanto ainda está sendo realizada uma modelagem dentro do domínio de problema.

## **2.2 Outras comparações entre ontologias e modelos de domínio**

Fouro (FOURO et al., 2001) compara Modelagem de Domínio e Ontologias e chega às seguintes considerações:

Modelos de Domínio e Ontologias são formas de estruturação e organização de conhecimento visando possibilitar a sua reutilização. São o resultado da busca por uma prática de reuso de software sistemática, formal e efetiva.

Ambas estão em busca do armazenamento de informações sobre um domínio que possam ser reutilizadas, ou que auxiliem na tarefa de reutilização durante o desenvolvimento de uma nova aplicação daquele domínio. A principal diferença entre eles, é que a ontologia não assume a pré-existência de nenhum sistema, no domínio a ser modelado. Desta forma, o nível de abstração das ontologias é mais alto que o dos modelos de domínio.

A teoria do domínio, representada por modelos de domínio ou ontologias, é útil tanto na preparação para a execução da atividade de análise quanto na sua efetiva realização. Ela provê, ainda, o mapeamento de quais conceitos são importantes para determinadas tarefas, indicando o que realmente deve ser investigado.

São duas etapas distintas e complementares da modelagem de domínio, que possibilitam a reutilização e que têm como um dos diversos fatores para sua criação, a redução dos custos de desenvolvimento, como consequência do compartilhamento de custos entre as diversas aplicações de uma mesma área ou domínio. Entretanto, ferramentas e técnicas adequadas ainda precisam ser desenvolvidas para o apoio a análise de domínio, e para a construção de ontologias em geral.

Essas considerações indicam a viabilidade em utilizar a Engenharia de Domínio para a conceitualização de um determinado domínio de conhecimento.

O uso de Ontologias destaca-se na aquisição e transmissão de conhecimento (VISSER, 1996). Considerando-se o uso de ontologias por pessoas dentro de uma organização social, o processo de descrição de uma ontologia pode identificar as contradições e diferenças de pensamento entre os participantes. Nesse sentido, ao identificar e definir um universo de discurso, o próprio relacionamento entre as classes e conceitos da ontologia representa um importante processo de aquisição de conhecimento.

A ontologia como produto é um artefato limitado. Paradoxalmente, quanto mais completa uma ontologia, mais restrito será seu uso. Um indivíduo que não participou do processo social da construção da ontologia vai inteirar-se dela através de um processo de aquisição semelhante ao da leitura de um dicionário. As definições contidas na ontologia minimizam problemas de vocabulário, contribuindo para uma melhor comunicação. No contexto de sistemas de gestão de conhecimento, as ontologias são aplicadas para a formação de bases de conhecimentos independentes de um sistema de gerência de conhecimento em particular. Essa decomposição (separação entre sistema e base) é necessária para facilitar a consulta das diversas bases disponíveis e atualmente não interoperáveis (GRUBER, 1991).

A Modelagem de Domínios se aplica na Gestão de Conhecimento pela introdução de técnicas estruturadas ou formais de descrição de domínios. Em particular, na atual adoção da *Unified Modeling Language* em diversos campos do conhecimento (HURLBLUT, 1997). Em contrapartida, a gestão de conhecimento amplia a aplicação dessas técnicas.

De uma maneira geral, as ontologias e modelos de domínios contribuem para definir as fronteiras do conhecimento que são relevantes para o problema de gestão. Como comentário final, a Gestão do Conhecimento é uma abordagem interdisciplinar e, por isso, mais ampla e está experimentando um razoável sucesso à medida que as organizações começam a perceber que tecnologia somente "não é tudo isso". De fato, o uso combinado de ontologias e modelos de domínio já está sendo aplicado com resultados significativos na indústria (BEECHAM, 2001; INTRASPECT, 2001).

O estudo de caso apresentado a seguir é utilizado para evidenciar as vantagens e possibilidades de uma modelagem de conhecimento explícita.

### 3 Hidrologia: Um Estudo de Caso

A hidrologia é um domínio de conhecimento científico e de interesse estratégico. No país, duas agências são responsáveis pela coordenação do uso de recursos hídricos. A ANA (Agência Nacional de Águas) trata de questões como a outorga de exploração de recursos hídricos e da gestão de interesses conflitantes no consumo desses recursos. A outorga concede a uma entidade o direito de explorar o recurso hídrico de uma região, por exemplo, para captação de água em processos industriais. Os interesses envolvem a partilha de um recurso hídrico entre diferentes usos: abastecimento urbano, agricultura, geração de energia, uso industrial. Conforme previsto na *Lei Federal de Gestão de Recursos Hídricos*, a gestão desses recursos será delegada pela ANA para associações em cada micro-bacia hidrográfica. O uso de recursos hídricos será tarifado.

A ANEEL (Agência Nacional de Energia Elétrica) está interessada no potencial de geração de energia hidroelétrica (90% da energia elétrica nacional). A instalação de novas hidroelétricas e a previsão da energia disponível em um dado período dependem do conhecimento desses dados.

#### 3.1 HidroWeb

O HidroWeb (ANEEL, 2001) é um sistema desenvolvido pela Superintendência de Estudos e Informações Hidrológicas (SIH) da ANEEL em conjunto com a ANA. O sistema disponibiliza, mediante consulta, dados sobre recursos hídricos através da Web.

A interface do HidroWeb permite a consulta de dados sobre:

- inventário de bacias, sub-bacias, rios, estados, municípios, entidades e estações de monitoramento de recursos hídricos.
- séries históricas sobre o inventário incluindo cotas, vazões, chuvas, clima, qualidade da água, resumo de descarga, sedimentos, curva de descarga e perfil transversal.
- séries em tempo real coletadas de concessionárias, do Instituto Nacional de Pesquisas Espaciais (INPE) e do Sistema de Vigilância da Amazônia (SIVAM).

Em maio de 2001, o HidroWeb apresentava dados sobre 9 bacias, 80 sub-bacias, 12574 rios, 41 estados, 5068 municípios e 102 entidades.

Não é objetivo deste trabalho avaliar a qualidade do HidroWeb. Acreditamos que o desenvolvimento do HidroWeb envolva uma série de desafios. O domínio hidrográfico apresenta dados geográficos de manipulação não-trivial. O sistema precisa agregar dados de diversas entidades distribuídas geograficamente e também permitir a consulta desses dados por entidades também dispersas geograficamente, através da Web.

Nesse contexto, é natural que o esforço de desenvolvimento se concentre em problemas tecnológicos e se afaste da modelagem semântica do problema. Esse "desvio" está se tornando cada vez mais freqüente à medida que os sistemas de computação aumentam em complexidade e envolvem um número cada vez maior de técnicas.

Portanto, o HidroWeb foi escolhido por ser um domínio estratégico e representativo do desenvolvimento de sistemas.

A semântica dos termos do HidroWeb é definida por um glossário de termos com base na Classificação Decimal Universal (*Universal Decimal Classification* - UDC), que é um esquema de classificação hierárquico através de facetas (*facets*) utilizado principalmente para a organização de publicações nas mais diversas áreas do conhecimento. O UDC representa uma taxonomia de definições organizadas através de um código decimal composto. A Tabela 1 apresenta alguns termos do UDC.

O uso de um glossário geral não alcança o mesmo resultado de um modelo de domínio ou de uma ontologia. Primeiro, nem todos os termos de um glossário geral são de interesse ao sistema. O UDC possui diversas categorias que não são relevantes ao sistema. Por exemplo, no UDC, os códigos iniciados em 0.1 são do domínio da Filosofia e Psicologia.

Por ser de uso geral, algumas definições do glossário não concordam com a semântica do sistema. Durante a análise do HidroWeb foram encontradas algumas inconsistências. Por exemplo, o glossário do HidroWeb recupera 11 definições para o termo "bacia". Apenas o termo qualificado "bacia hidrográfica" apresenta a definição coerente com a interface e os dados oferecidos nessa opção de consulta, mesmo assim, essa é uma

suposição baseada na interpretação de um indivíduo apenas. Não existe uma formalização dessa associação, a qual seria obtida através da construção de uma ontologia ou modelo de domínio.

Código UDC	Área de conhecimento
0.5	Ciência pura
0.55	Geociência geologia-geofísica
0.556	Hidrologia geral
0.556.5	Hidrologia da terra - Águas superficiais
0.556.53	Rios, correntes, estuários
0.556.531	Propriedades da água dos rios
0.556.531.4	Propriedades químicas dos rios
0.556.531.45	Salinidade dos rios ou estuários

**Tabela 1: Exemplo de classificação UDC.**

De uma maneira geral, os termos incluem entidades de naturezas diferentes. O termo “estado” inclui as unidades federativas e o distrito federal, assim como os países que fazem fronteira com o Brasil. Entre os municípios encontram-se unidades administrativas de países vizinhos, como o Uruguai que possui departamentos no lugar de municípios. O termo entidade envolve companhias públicas e privadas, secretarias estaduais, institutos de pesquisa e outras entidades de constituição e porte diferentes.

Tanto em ontologia quanto em modelos de domínio, é necessário ter um vocabulário formalizado entre o usuário e o sistema. Um dos indicadores de que essa formalização foi realizada é o uso coerente de termos na interface do sistema.

Nas próximas seções apresentaremos uma ontologia e a seguir um modelo de domínio que foi gerado através da organização de alguns dos termos que constam da interface e da documentação do HidroWeb. Em ambas abordagens, os conceitos associados com cada passo da modelagem são apresentados a medida que são necessários.

## 4 Uma ontologia hidrológica

O desenvolvimento de sistemas especialistas e bases de conhecimento começa pela descrição dos elementos do espaço do problema através de uma representação que será utilizada pela aplicação como um modelo da realidade. A ontologia é uma das técnicas para criar essa representação. Para exemplificar a construção de ontologias, utilizaremos a abordagem proposta pelo Projeto Protégé (NOY et al., 2000; PROTÉGÉ, 2001).

### 4.1 Descrevendo Entidades Discretas no Domínio

A ontologia hidrológica descreve as principais entidades do inventário do HidroWeb (Figura 2). A descrição de uma ontologia inicia com a criação de um projeto (*project*): uma coleção de definições relacionadas.

A inclusão de termos em uma ontologia ocorre pela externalização de termos implicitamente definidos. A descrição das entidades discretas do domínio é realizada através de classes (*class*). A partir de uma classe abstrata inicial (*thing*) são inseridas novas classes que representam conjuntos de entidades do universo de discurso do problema.

As classes formam uma hierarquia com a classe inicial como raiz e classes mais genéricas próximas dessa raiz. Uma classe é chamada de classe concreta se possui uma instância concreta no domínio do problema. Caso contrário, será chamada de classe abstrata.

No exemplo da Figura 2, foram definidas 12 classes. Três classes abstratas foram criadas para organizar a ontologia. Inventário é uma classe abstrata usada para agrupar os conceitos relacionados com este exemplo. As classes Unidade Administrativa e Recurso Hídrico, ambas abstratas, também foram criadas apenas para fins de organização.

As classes Estação Pluviométrica e Estação Fluviométrica foram criadas para diferenciar os tipos de estações presentes no cadastro do HidroWeb.

### 4.2 Descrevendo atributos

Os atributos de uma entidade são representados através de *slots*. O *slot* pode representar um valor simples como números inteiros e reais, cadeias de caracteres, valores lógicos ou uma referência a uma instância. O último tipo é útil para a definição de relacionamentos entre classes e entre instâncias.

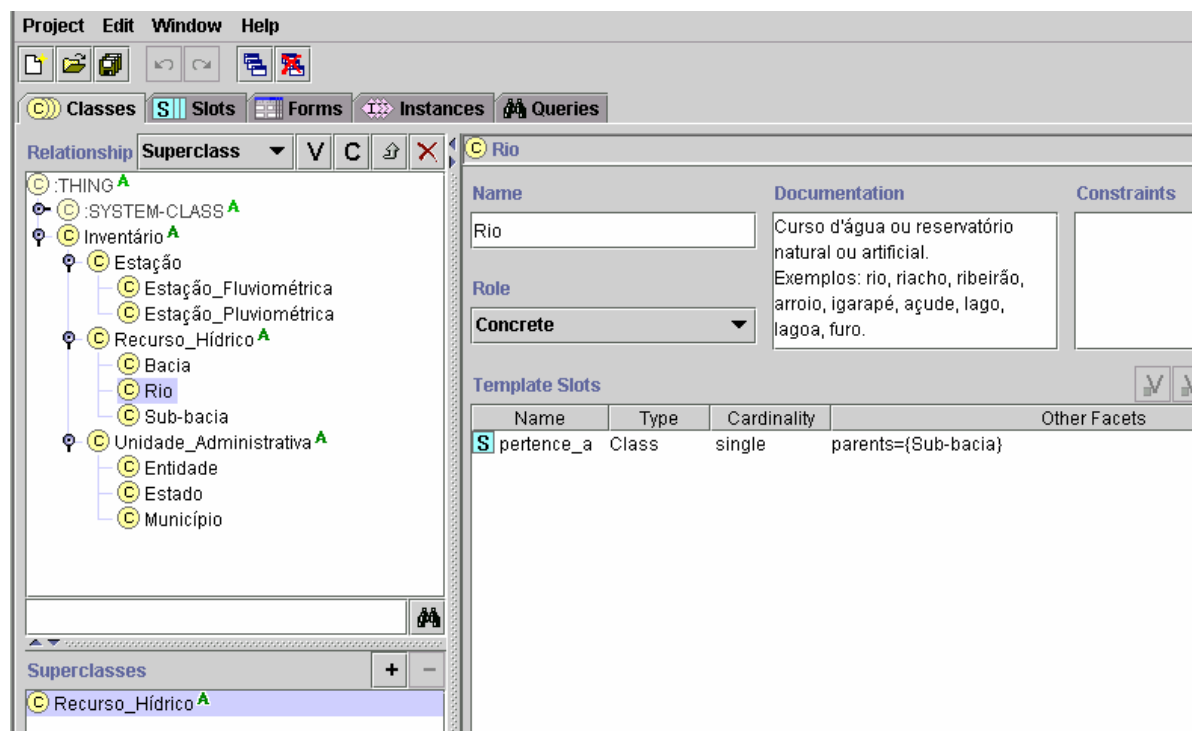
O tipo qualquer (*any*) permite criar uma referência a qualquer outra entidade discreta na ontologia. Com isso é possível trabalhar com referências sem tipo determinado ou criar enumerações de valores com tipos diversos (uma referência a uma instância, uma referência a uma classe e um número).



O valor em um *slot* pode apresentar cardinalidade simples múltipla (conjunto), especialmente útil para a construção de relacionamentos. O relacionamento inverso do apresentado: “uma bacia possui vários rios” é um exemplo do uso da cardinalidade múltipla.

O tipo símbolo (*symbol*) permite a referência eficiente a valores com semântica definida pela própria aplicação. Por exemplo, em uma dada aplicação, os símbolos: “masculino” e “feminino” poderiam ser utilizados para definir o domínio de um atributo “sexo” em uma classe pessoa.

A Figura 2 apresenta o relacionamento entre a classe Rio e a classe Sub-bacia. Um rio pertence a uma e somente uma sub-bacia. Os *slots* pré-definidos nome (*name*) e documentação (*documentation*) foram os únicos atributos utilizados no exemplo.



**Figura 2: Uma ontologia para os termos da interface do HidroWeb.**

Teoricamente, uma classe pode possuir mais de uma superclasse (herança múltipla). O conflito de herança é evitado, pois um *slot* é uma abstração no mesmo nível da classe. Em outras palavras, não pode haver dois *slots* com o mesmo nome em todo o projeto. Entretanto, um mesmo *slot* pode ser associado a mais de uma classe. Por exemplo, é possível definir um atributo “autor” e associá-lo a todas as classes do projeto. Não há conflito de herança, pois todas as classes possuem uma mesma referência a todos os valores.

Não existem atributos no nível de classes em ontologias, mas uma instância pode fazer referência a uma classe.

### 4.3 Outros recursos disponíveis

A ferramenta utilizada permite que o especialista visualize e, em alguns casos, interfira nas definições de classe, *slot*, faceta e restrições. As abstrações usadas na modelagem estão descritas no mesmo nível das abstrações sendo modeladas (meta-modelo exibido na Figura 3). Essa característica permite a adaptação da estrutura de modelagem, por exemplo, pode ser inserido em *class* um atributo adicional para identificar o autor e a data de criação da estrutura.

A utilização de restrições (*constraints*) aumenta a correção de um modelo. Através dos mecanismos que determinam a cardinalidade, tipo, opcionalidade e domínio de um atributo é possível descrever restrições que considerem valores em diferentes classes. Por exemplo, seria teoricamente possível estabelecer a restrição de que “a soma da vazão de todos os rios em uma bacia deve ser igual ou superior ao volume estimado de consumo nessa bacia”. Entretanto, essas restrições não são normalmente utilizadas na descrição de uma ontologia, por uma das situações: (1) o especialista não conhece essas restrições durante a modelagem, (2) o especialista não domina a linguagem de definição de restrições ou (3) as restrições serão adicionadas posteriormente, de acordo com as necessidades de cada aplicação.

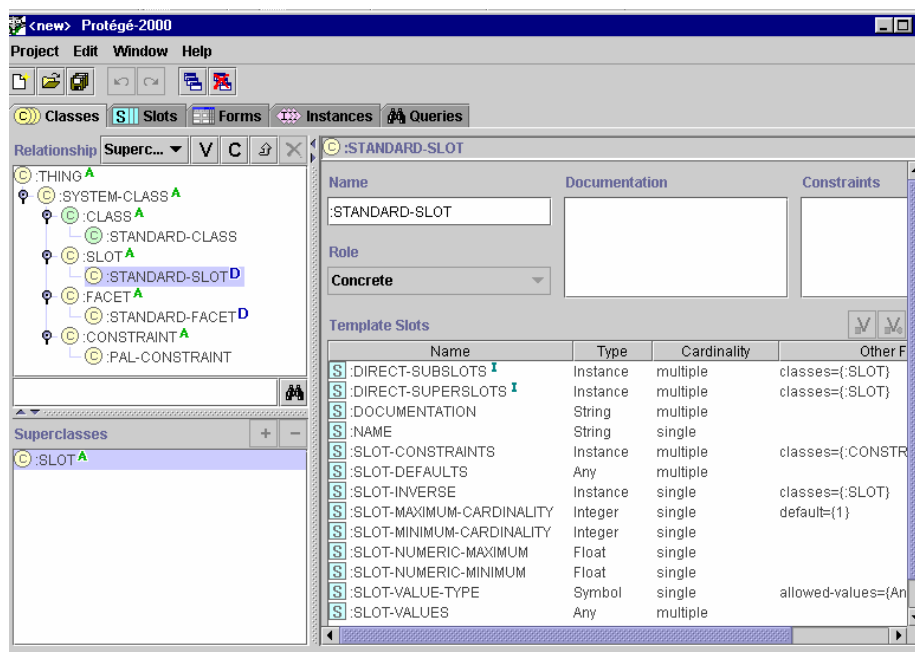


Figura 3: A definição do meta-modelo da ontologia.

## 5 Um modelo de domínio hidrológico

O desenvolvimento de software clássico, seguindo o ciclo de vida cascata, começa pela análise de requisitos do sistema e se estende para projeto, implementação e testes (PRESSMAN, 1997). Entretanto, com o passar do tempo, os Engenheiros de Software constataram que muitos sistemas compartilhavam semelhanças, e que o retrabalho do desenvolvimento de software poderia ser diminuído através da reutilização de especificações, modelos e código-fonte. A prática de reutilização implica em um aumento da produtividade da equipe de desenvolvimento e da qualidade e confiabilidade do produto, além de contribuir para a diminuição dos custos. Contudo, Para que a reutilização seja possível, o desenvolvimento de software precisa passar por uma etapa anterior, onde são analisadas, projetadas e implementadas as características comuns em relação ao domínio em questão. Essa etapa chama-se engenharia de domínio.

Dessa forma, podemos definir dois pontos de vista para o desenvolvimento de software: o desenvolvimento *para* reuso e o desenvolvimento *com* reuso (MOORE et al., 1991). O desenvolvimento *para* reuso, ou Engenharia de Domínio, se preocupa com a geração de componentes reutilizáveis em um determinado domínio. O desenvolvimento *com* reuso, ou Engenharia da Aplicação, se preocupa em construir aplicações reutilizando os componentes criados na Engenharia de Domínio. A Engenharia da Aplicação serve também como agente motivador para a construção de novos componentes inexistentes na base de componentes do domínio.

Para representar o domínio hidrológico utilizando modelos de domínio, em um nível de abstração que facilite a aquisição de conhecimento, construímos inicialmente um modelo de *features* do domínio (BRAGA, 2001). O modelo de *features* tem como objetivo representar graficamente os conceitos e funcionalidades do domínio. Ele é composto por dois sub-modelos (conceitos e funcionalidades), que podem ser combinados em uma visão global. O seu nível de abstração é alto, se preocupando em representar o conhecimento descrito nos casos de uso existentes no domínio.

Para exemplificar a construção de modelos de domínio, utilizamos a abordagem proposta pelo Projeto Odyssey (WERNER, 2001), que usa extensões da UML para representar o conhecimento de um domínio, como por exemplo diagramas de Contexto e de *Features*.

No contexto do Projeto Odyssey, foram definidos os processos de engenharia de domínio (Odyssey-DE) (BRAGA, 2001) e de engenharia de aplicação (Odyssey-AE) (MILER, 2000). Além disso, foi criada a Infraestrutura Odyssey (WERNER et al., 2000), que contém o ferramental necessário para a execução dos processos propostos.

### 5.1 Descrevendo Entidades Discretas no Domínio

O processo de Engenharia de Domínio proposto pelo Projeto Odyssey consiste, inicialmente, em situar o domínio em relação ao seu escopo, limites, relacionamento com outros domínios e principais atores envolvidos. Isto é feito através de diagramas de contexto. Também deve ser construído o diagrama de *features* (a partir de

generalizações dos cenários do domínio), que tem como objetivo facilitar o entendimento dos principais conceitos e atividades do domínio.

Desta maneira, é possível modelar o conhecimento de um domínio através do uso de diagramas de *features*, que são criados baseando-se em casos de uso do domínio. Se o objetivo for definir um arcabouço de componentes dentro do domínio, os casos de uso podem ser detalhados através de diagramas de classes, estados e sequência.

A etapa de *análise de domínio*, dentro da engenharia de domínio, é a que mais nos interessa no contexto deste trabalho. Braga (BRAGA, 2001) detalha a etapa de *análise de domínio* em duas atividades: *descrição do domínio no contexto da organização* e *elicitação do conhecimento do domínio*. A atividade de *elicitação do conhecimento do domínio* ainda é detalhada em *aquisição do conhecimento* e *representação do conhecimento*.

A sub-atividade de *aquisição do conhecimento* é efetuada através da descrição de cenários, que serão transformados em casos de uso. Braga (BRAGA, 2001) define algumas diretivas para a construção de cenários que contém atores, conceitos e ações relevantes no domínio. Cada elemento mencionado deve manter um mesmo nome em todos os cenários que é citado. A origem e as funcionalidades apresentadas no domínio também devem ser explicitadas. Uma mesma funcionalidade pode ser descrita em mais de um cenário.

Para a descrição dos cenários, pode ser utilizado em conjunto a abordagem proposta por Roseti (ROSETI et al., 1999), que define uma sequência de atividades para a execução do processo de aquisição de conhecimento e sugere o uso de algumas técnicas de aquisição de conhecimento com uso consolidado em outras áreas do conhecimento.

A sub-atividade de *representação do conhecimento* utiliza os cenários descritos anteriormente para a construção dos casos de uso. Utilizando esses casos de uso, os analistas de domínio deverão abstrair os conceitos e funcionalidades e construir o diagrama de *features*. Também utilizando os casos de uso, podem ser definidos diagramas UML de sequência, classes (tipos) e estados (todos em nível conceitual). O diagrama de sequência citado não visa a descoberta de métodos das classes. O objetivo é representar a interação entre atores na execução de casos de uso. A Figura 4 exhibe o modelo de *features* do domínio hidrológico construído a partir desse processo, baseando-se nas informações contidas no HidroWeb.

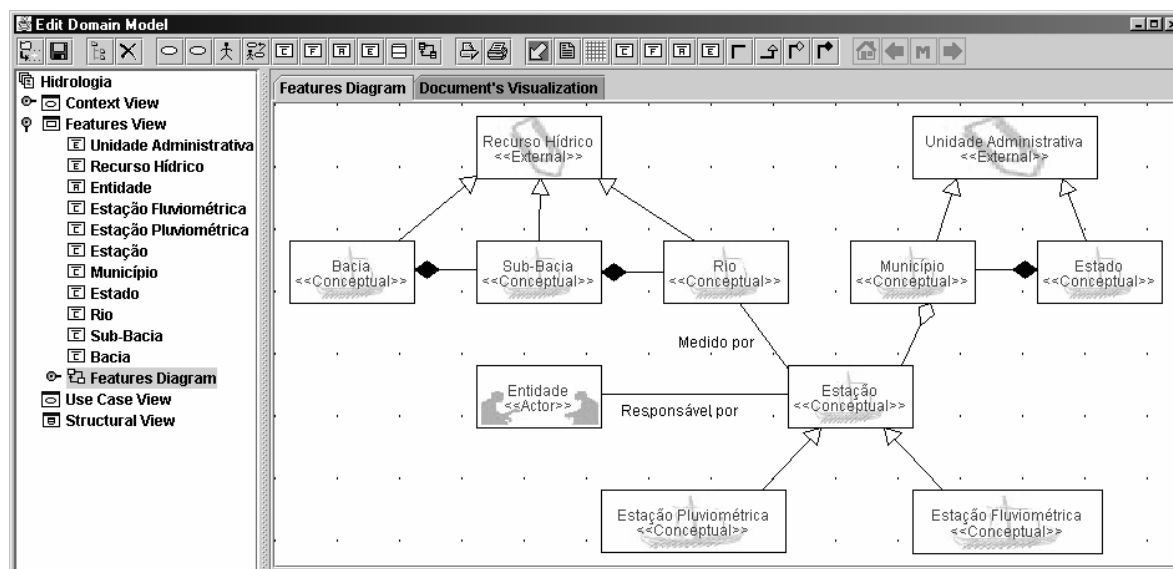
















Figura 4: Modelo de *features* do domínio de Hidrologia.

## 5.2 Descrevendo atributos

O diagrama de *features*, descrito em Miler (MILER, 2000), é composto por elementos que representam funcionalidade ou conceitos do domínio e relacionamentos entre esses elementos. Também é utilizada uma notação de dependência entre os elementos. Todos esses tipos de caracterização das *features* têm uma notação gráfica para a sua representação, como exibido na Tabela 2.

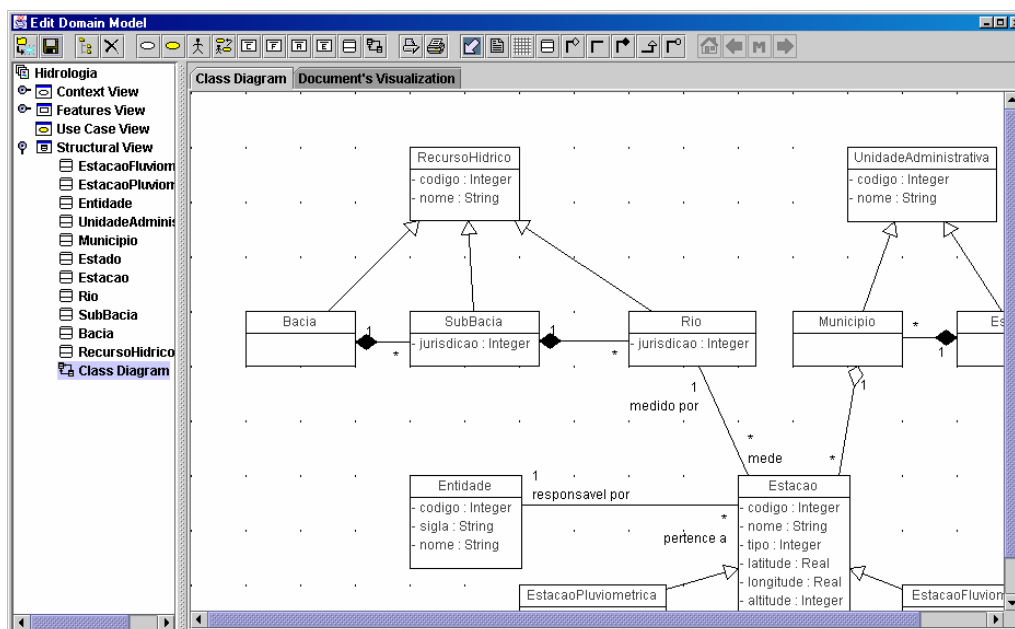
Para detalhar o modelo de *feature*, podemos utilizar o modelo de classes. Nesse modelo, serão introduzidos atributos (características) e métodos (funcionalidades) às entidades, além das cardinalidades dos relacionamentos entre entidades. O modelo de classes tem como objetivo representar graficamente as interações estáticas entre as entidades (tipos) e as suas interfaces de comunicação. Essa abordagem é usualmente utilizada por métodos de desenvolvimento de software baseados em componentes, como por exemplo o *Catalysis* (D'SOUZA et al., 1998).



	<b>Features Essenciais</b> – Características fundamentais do modelo. <i>Features</i> intimamente ligadas à essência do domínio. Características que representam as funcionalidades/conceitos do modelo e correspondem a casos de uso concretos.
	<b>Features Organizacionais</b> – Características do modelo que têm apenas o intuito de facilitar o entendimento ou organizar o domínio. Não possuem ligações concretas com o uso real do domínio.
	<b>Features Entidade</b> – São os atores do modelo. Entidades do mundo real que atuam sobre o domínio. Podem, por exemplo, expor a necessidade de uma interface ou de procedimentos de controle.
	<b>Features Externas</b> – Pertencem a outros domínios. Podem ou não ser refinadas pelo modelo. Mostram a fronteira do domínio e como ela se comporta.
	<b>Features Não definidas</b> – Características já levantadas de um domínio, porém ainda não definidas através de <i>use-cases</i> e/ou modelos conceituais.
	<b>Features Adicionais</b> – Características adicionais, mas importantes para o entendimento do domínio.
	<b>Relacionamento Composição</b> – Relação onde uma <i>feature</i> é composta de várias outras. É uma relação onde uma <i>feature</i> é parte fundamental de outra, de forma que uma não existe sem a outra.
	<b>Relacionamento Agregação</b> – Relação onde uma <i>feature</i> representa o todo, e as outras as partes. Similar a composição, porém sem a relação de dependência entre seus membros.
	<b>Relacionamento Herança (ou Tipificação)</b> – Relação onde as <i>features</i> filhas são tipos da <i>feature</i> mãe. Este tipo de relação indica que as <i>features</i> filhas herdam as características de seus antecessores.
	<b>Relacionamento Associação</b> – Ligação simples entre duas características. Denota algum tipo de relacionamento entre seus membros. Pode ser nomeada, indicando um tipo específico de ligação.
	<b>Notação de Exclusividade</b> – Tipo de ligação cujas <i>features</i> filhas não podem ser usadas ao mesmo tempo. Pode denotar variações, problemas de incompatibilidade ou de impossibilidade de implementação de uma extensão.
	<b>Notação de Opcionalidade</b> – Denota uma característica ( <i>feature</i> ) não mandatória do domínio.
	<b>Notação de Restrição</b> – Problemas ou necessidade de uso conjunto de duas <i>features</i> , que não necessariamente possuem relacionamento direto.
	<b>Notação de Expansibilidade</b> – Permite ao diagrama a flexibilidade de navegação em diferentes níveis de detalhamento. Expressa a existência de um refinamento da característica em questão.

**Tabela 2: Caracterização de Features.**

A Figura 5 exibe o modelo de classes do domínio hidrológico, baseando-se nas informações contidas no HidroWeb. Este modelo pode ser visto como um detalhamento do modelo de *features* exibido na Figura 4, entretanto, os métodos não foram definidos, pois não existe conceito semelhante fornecido pelo HidroWeb.



**Figura 5: Modelo de classes do domínio de Hidrologia.**

### 5.3 Outros recursos disponíveis

Os modelos de domínio são documentos variados, cada qual com o objetivo de representar o conhecimento do domínio sob um determinado ponto de vista. Além dos modelos de *features* e classes, anteriormente citado, podemos apontar (BRAGA, 2001): modelo de contexto do domínio, descrição estruturada de casos de uso do domínio e modelos UML de casos de uso, sequência e estado.

O *modelo de contexto do domínio* tem como objetivo representar graficamente o domínio em questão em relação ao seu escopo, limites, relacionamentos com outros domínios e principais atores envolvidos.

A *descrição estruturada de casos de uso do domínio* tem como objetivo servir como o meio de comunicação inicial entre os analistas de domínio e os usuários. A partir dessa descrição, os demais modelos são construídos. A representação dos casos de uso é textual, utilizando padrões de documentação. Descrição de casos de uso são usualmente utilizados na elaboração de contratos de desenvolvimento de software.

O *modelo de casos de uso* tem como objetivo representar graficamente as interações estáticas descritas nos casos de uso do domínio.

O *modelo de sequência* tem como objetivo representar graficamente as interações dinâmicas descritas nos casos de uso do domínio (entre conceitos e funcionalidades). Essa representação pode ser útil para a unificação dos nomes utilizados nos casos de uso. O uso de diagramas de sequência nesse contexto difere do uso inicialmente proposto pela UML, que é para a representação dinâmica da interação entre objetos.

O *modelo de estado* tem como objetivo representar graficamente as transições de estados de entidades em um caso de uso. Esse diagrama é muito útil para representar variações em processos complexos.

## 6 Comparação e Comentários

A comparação entre as técnicas será realizada em duas etapas. Primeiro são apresentados comentários sobre as abstrações disponíveis para a modelagem de conhecimento substantivo (Tabela 3). Em seguida, são comparados recursos específicos oferecidos pelas ferramentas utilizadas (Tabela 4).

As abstrações utilizadas para representar entidades discretas são semelhantes. A principal diferença é a definição de interfaces no modelo de domínio. A interface permite fatorar a descrição de uma classe. De forma similar, a promoção de atributos ao nível de classes cumpre uma função similar na ontologia. A vantagem da interface é a capacidade de agrupar diversas propriedades sob um único identificador.

A estrutura flexível dos *slots* é usada para modelar tanto atributos quanto relacionamentos. Modelos de domínio possuem duas abstrações distintas para atributos e relacionamentos. O polimorfismo previsto em UML cumpre a mesma função do operador *any*, ou seja, permitir referências e conjuntos a tipos indeterminados.

Entidades		Ontologia	Modelo de Domínio
Conjunto de instâncias	Classe		<i>features</i> e classe
Classe inicial	<i>Thing</i>		<i>Object</i>
Organização	hierarquia (classificação)		classificação [ <i>features</i> ] e herança [ <i>features</i> e classes]
Tipos	concreta e abstrata		concreta, abstrata e interface
Herança de definição	Múltipla		Múltipla
Atributos		Ontologia	Modelo de Domínio
Denominação	<i>Slot</i>		Atributo
Cardinalidades	simples e múltipla		simples e múltipla
Domínio de atributos	tipos escalares, cadeias de caracteres, símbolos e tipo especial <i>any</i>		tipos primitivos (numéricos, lógicos e alfanuméricos) e classes
Relacionamentos		Ontologia	Modelo de Domínio
Entre instâncias	<i>através slot</i> do tipo <i>instance</i>		relacionamentos no modelo de classes

**Tabela 3: Abstrações para entidades, atributos e relacionamentos em Ontologia e Modelos de Domínio.**

A comparação das ferramentas evidencia necessidades práticas dos especialistas. A visualização do modelo sendo construído é a principal delas. A comunicação visual é favorecida nos modelos de domínio, principalmente a percepção de relacionamentos. Por sua vez, a ontologia oferece facilidades para encontrar classes e *slots* através de consultas.

Visualização		Ontologia	Modelo de Domínio
Visão geral do modelo		hierarquia de classes e formulários	diagrama gráfico utilizando ligações de rastreabilidade
Visão dos relacionamentos		apenas para uma dada classe através de formulário específico.	diagrama gráfico
Níveis de abstração		Um único nível	Contexto, <i>features</i> , casos de uso, classes, sequência e estados.
Outros recursos		Ontologia	Modelo de Domínio
Aquisição de dados		formulários e representação de instâncias	(postergada para o momento da execução)
Restrições		sobre <i>slots</i> (domínio, cardinalidade, opcionalidade) ou sobre o projeto (em uma aplicação)	exclusividade, opcionalidade, restrição e expansibilidade
Consultas		por referências a uma entidade, nome ou valor de atributo.	por referências ao nome da entidade
Funcionalidades		(postergada para a aplicação)	Métodos

**Tabela 4: Recursos de modelagem.**

As diferentes representações dos modelos de domínio favorecem a descrição de diferentes perspectivas de modelagem. Se por um lado essa especialização facilita o trabalho de diversos especialistas, o modelo único da ontologia favorece o treinamento.

## 7 Conclusão

Ontologias e modelos de domínios são duas representações de conhecimento em uso por pesquisadores de diferentes áreas da computação. Ambas técnicas tratam do conhecimento sobre a realidade concreta ou abstrata de um universo de discurso. De certa forma, essa orientação é diferente da que originou a ciência da computação pois se afasta da descrição de máquinas abstratas, computadores e programas. E se concentra na descrição da realidade.

Neste trabalho, comparamos essas duas técnicas em uma perspectiva prática; a partir das facilidades oferecidas por duas ferramentas específicas. Cada técnica possui diversas variações em metodologia e uso das abstrações. Foi comparada apenas uma instância de cada técnica, o que limita a nossa capacidade em fazer comparações diretas entre as duas abordagens. A modelagem apresentada poderia ser diferente se fosse aplicada por outros indivíduos em outras condições. Embora exista um conjunto de técnicas e processos bem conhecidos, ainda existe muito mais arte do que ciência na tarefa de modelar um universo de discurso.

Não obstante, acreditamos que alguns pontos em comum possam ser destacados: (1) a orientação para uma modelagem dentro do espaço do problema, (2) o uso de abstrações similares para representar o conhecimento substantivo de um domínio: entidades, atributos e relacionamentos.

A contribuição deste trabalho está em confrontar duas técnicas que trabalham sobre o mesmo universo de discurso, e que possuem o mesmo objetivo: modelar o conhecimento disponível de forma acessível ao computador.

A modelagem dentro do domínio do problema apresenta as seguintes vantagens: (1) independência da técnica, (2) o conhecimento substantivo do domínio pode ser aproveitado para diferentes áreas e (3) a representação pode ser mais facilmente compreendida pelo especialista do domínio. Essas vantagens combinadas resultam em um melhor entendimento das forças e das entidades em um domínio.

A ontologia possui uma vantagem relativa na modelagem do domínio do problema pela distinção clara entre a ontologia e a aplicação que será desenvolvida posteriormente. Na modelagem de domínio não existe um ponto definido que separe a aplicação e o problema; essa distinção fica por conta do especialista. Entretanto, a modelagem de domínio fornece várias representações para o mesmo conhecimento, utilizando uma notação gráfica, que contribui para o entendimento global.

Dando continuidade a este trabalho, indicamos trabalhos futuros em três linhas de atuação. Primeiro, existe a necessidade de dispor de um número maior de exemplos de modelagem. O modelo domínio já apresenta seis representações específicas. Devido a facilidade da ferramenta utilizada, seria possível investigar a integração da ontologia como um perspectiva adicional de modelagem. Essa é uma opção para a integração das vantagens das duas propostas.

Também seria interessante desenvolver uma aplicação completa dentro de cada abordagem e comparar os resultados obtidos. Essa tarefa logicamente está além do escopo deste trabalho.

Outra linha de atuação seria encontrar um maior número de exemplos de modelos de domínio e ontologias que tratem de universos de discursos semelhantes, fornecendo mais informações para as tabelas de comparação exibidas na Seção 6. O levantamento bibliográfico de experiências de associação entre engenharia de domínios e engenharia de ontologias seria outra contribuição importante para traçar um paralelo entre as duas abordagens.

## Referências Bibliográficas

- ANEEL, 2001, "HidroWeb", disponível na Internet em <http://hidroweb.aneel.gov.br>, acessado em maio de 2001.
- ARANGO, G., 1994, "Software Reusability - Domain analysis methods", *Ellis HorWood*, pp. 17-49.
- BEECHAM, S. K., 2001, "Smith Kline Beecham", disponível na Internet em <http://www.sb.com>, acessado em maio de 2001.
- BRAGA, R. M. M., 2001, "Busca e Recuperação de Componentes em um Ambiente de Reuso", tese de D.Sc., *COPPE, UFRJ*.
- COMITEE ON KNOWLEDGE ASSESSMENT OFFICE OF INTERNATIONAL AFFAIRS NATIONAL RESEARCH COUNCIL, 1996, "Prospectus for National Knowledge Assessment", *National Academy*, Washington D. C., disponível na Internet em <http://www.nap.edu/books/NI000957/html/index.html>.

- D'SOUZA, D., WILLS, A., 1998, "Objects, components, and frameworks with UML: The catalysis approach", *Addison Wesley*.
- FERRE, X., VEGAS, S., 1999, "An Evaluation of Domain Analysis Methods", *4th CAiSE/IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design (EMMSAD'99)*, pp. 14-15.
- FOURO, A. M. M., WERNER, C. M. L., 2001, "Modelos de domínios ou Ontologias?", *RTInfo - Revista Tecnologia da Informação*.
- GRUBER, T. R., 1991, "The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases", *Morgan Kaufmann*, pp. 601-602.
- HURLBLUT, R., 1997, "A survey on approaches for describing and formalizing use-cases", disponível na Internet em <http://www.iit.edu/~rhurlbut/xpt-tr-97-03.html>, acessado em 30/05/2001.
- INTRASPECT, 2001, "Intraspect", disponível na Internet em <http://www.intraspect.com>, acessado em maio de 2001.
- JONES, D., BENCH-CAPON, T., VISSER, P., 1998, "Methodologies for Ontology Development", *15<sup>th</sup> IFPI World Computer Congress*, pp. 62-75, Londres, UK.
- MILER, N., 2000, "A Engenharia de Aplicações no Contexto da Reutilização baseada em Modelos de Domínio", tese de M.Sc., *COPPE/UFRJ*.
- MOORE, J. M., BAILIN, S. C., 1991, "Domain Analysis: Framework for reuse", *IEEE Computer Society Press Tutorial*, pp. 179-202.
- NONAKA, I., TAKEUCHI, H., 1997, "Criação de Conhecimento na Empresa", *Campus*, Rio de Janeiro.
- NOY, N. F., FERGERSON, R. W., MUSEN, M. A., 2000, "The knowledge model of Protégé-2000: Combining interoperability and flexibility", *2th International Conference on Knowledge Engineering and Knowledge Management*, Juan-les-Pins, France.
- PRESSMAN, R. S., 1997, "Software Engineering: A Practitioner's Approach", quarta edição, *McGraw-Hill*.
- PROTÉGÉ, 2001, "Protégé Homepage", disponível na Internet em <http://protege.stanford.edu>, acessado em maio de 2001.
- ROSETI, M. Z., WERNER, C. M. L., 1999, "A Knowledge Acquisition Systematic within the Domain Analysis Context", *II Ibero-American Workshop on Requirements Engineering*, Buenos Aires.
- STEWART, T. A., 1998, "Capital Intelectual: a nova vantagem competitiva das empresas", *Campus*, Rio de Janeiro.
- SVEIBY, K. E., 1997, "A Nova Riqueza das Organizações", *Campus*, Rio de Janeiro.
- VISSER, P., 1996, "The Formal Specification of a Legal Ontology", *Proceedings of the Ninth International Conference on Legal Knowledge Based Systems (JURIX'96)*, Tilburg University Press, Tilburg, Holanda.
- WERNER, C. M. L., 2001, "Projeto Odyssey", disponível na Internet em <http://www.cos.ufrj.br/~odyssey>, acessado em maio de 2001.
- WERNER, C. M. L., BRAGA, R. M. M., MATTOSO, M., et al., 2000, "Infra-estrutura Odyssey: estágio atual", *XIV Simpósio Brasileiro de Engenharia de Software, Sessão de Ferramentas*, João Pessoa, outubro.