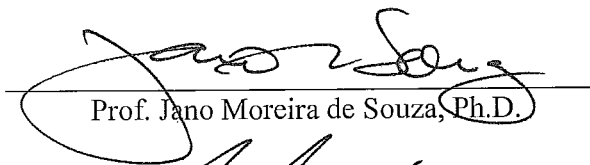


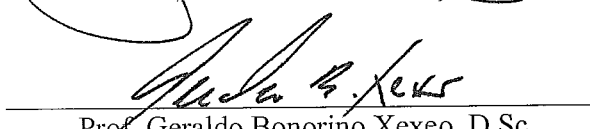
THOTH: UMA FERRAMENTA PARA REUTILIZAÇÃO DE PROCESSOS  
CIENTÍFICOS

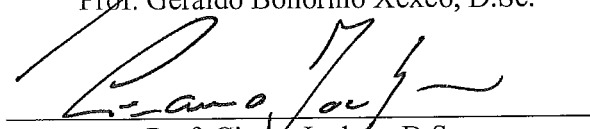
Elder Lopes Pedreira Bomfim

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. Jano Moreira de Souza, Ph.D.

  
Prof. Geraldo Bonorino Xexeo, D.Sc.

  
Prof. Ciraco Iochpe, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

JULHO DE 2005

BOMFIM, ELDER PEDREIRA

THOTH: Uma ferramenta para reutilização de processos científicos [Rio de Janeiro] 2005

X, 123 p. 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia de Sistemas e Computação, 2005).

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. Gestão de Conhecimento
2. Raciocínio Baseado em Casos
3. Sistemas de Workflow

I. COPPE/UFRJ II. Título (série)

*À minha mãe, pessoa chave do meu crescimento.*

*Aos meus irmãos.*

## Agradecimentos

Primeiramente, a Deus, senhor supremo da minha vida, por iluminar meus caminhos e dar a força necessária para eu desenvolver esta dissertação.

À minha família, minhas tias Irene, Julinda, Janete e Darci, além do pessoal de Brotas, pelo carinho durante toda a minha vida e pelas calorosas recepções nas viagens a Salvador durante os últimos três anos.

Ao meu orientador Jano, pela oportunidade de desenvolver este trabalho.

Ao professor Xexéo, pela ajuda providencial na definição de alguns componentes importantíssimos deste trabalho e por se disponibilizar a participar da banca; e ao professor Cirano, por também aceitar fazer parte desta banca.

Ao “co-orientador” Manuel, pela ajuda e amizade, principalmente na fase inicial desta dissertação. E à “co-orientadora”, conselheira, analista, terapeuta, psicóloga... Jonice, pelos conselhos, broncas, palavras de incentivos, dezenas de revisões e idéias que ajudaram e muito o desenvolvimento deste trabalho.

Aos amigos que me ajudaram neste desafio, em especial Get, Barros, Melfry e Adriana Vivacqua, pelos ótimos conselhos, idéias e pelas palavras de apoio. A Jeann, Pablo, Matheus, Juliano, Rogea e ao pessoal de Caxias, pela disponibilidade para ajudar, quando necessário.

A Blaschek, pela oportunidade profissional, a Márcio Duran, que, com sua imensa cordialidade, sanou inúmeras dúvidas, e a Rodrigo Monteiro, pela ajuda no aprendizado do Delphi e Oracle. Ao pessoal da DABM, que me propiciou um excelente ambiente de trabalho.

Ao pessoal do Laboratório de Hidrologia, pela ajuda e dicas sobre a área de Recursos Hídricos e pela disponibilidade para viabilizar a aplicação do trabalho nessa área de pesquisa.

Ao corpo administrativo da COPPE/Sistemas, pela presteza e atenção sempre que precisei, em especial à Patty, pelo sorriso e brincadeiras que amainavam o ambiente durante a espera pelas reuniões com meu orientador.



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

## THOTH: UMA FERRAMENTA PARA REUTILIZAÇÃO DE PROCESSOS CIENTÍFICOS

Elder Lopes Pedreira Bomfim

Julho/2005

Orientador: Jano Moreira de Souza

Programa: Engenharia de Sistemas e Computação

O trabalho científico pode ser caracterizado pelo alto grau de qualificação dos seus executores, pela singularidade dos problemas a serem resolvidos, pela necessidade de serem consultadas inúmeras fontes de informações e pela análise de grande quantidade de dados. Estas características trazem desafios para que o mesmo possa ser desenvolvido com qualidade, já que os seus executores devem ter segurança suficiente para tomar complexas decisões em cada atividade desenvolvida.

Para apoiar esse tipo de trabalho, são utilizados sistemas de informação, como os sistemas de *workflow*, que auxiliam o pesquisador a organizar e coordenar suas atividades, gerenciadores de modelos e sistemas de gestão do conhecimento, que propiciam melhor controle sobre os conhecimentos produzidos, além de facilitarem e agilizarem o acesso aos conhecimentos necessários para a execução adequada das atividades.

Nesse contexto, surgiu a proposta do Thoth, que tem o objetivo de prover meios para que as informações sobre as atividades já executadas possam ser consultadas durante a execução de novas atividades, usando para isso a abordagem de raciocínio baseado em casos (RBC) sobre a base do sistema de *workflow* científico.

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

## THOT: A TOOL FOR SCIENTIFIC PROCESS REUSE

Elder Lopes Pedreira Bomfim

July/2005

Advisors: Jano Moreira de Souza

Department: System and Computing Engineering

Scientific work is characterized by the high-quality background needed by its performers, the uniqueness of the problems being addressed and the necessity to consult large amounts of information and analyze a large volume of data. These characteristics create challenges for conducting this type of activity, since its performers should be confident enough to make complex decisions for each individual case.

Some information system are commonly used to support this kind of work, such as workflow management systems, which aid researchers in organizing and coordinating activities, models management systems and knowledge management systems, which provide better control over generated knowledge, besides facilitating access to required knowledge for adequately executing activities.

In this context, we propose Thot, whose goal is to provide ways for researchers to check information about past activities during the execution of new ones, by using the case base reasoning (CBR) approach over scientific workflow management system database.

# Índice do Texto

Índice do Texto .....	vii
Índice de Figuras .....	x
Capítulo 1 – Introdução.....	1
1.1 – Motivação.....	1
1.2 – Objetivo.....	3
1.3 – Organização da Dissertação .....	4
Capítulo 2 – Revisão de Literatura.....	5
2.1 – O Trabalho Científico .....	5
2.1.1 – Método Científico .....	6
2.1.2 – Características do Trabalho Científico.....	8
2.2 – CSCW e <i>workflow</i> , tecnologias de suporte ao trabalho científico. ....	11
2.2.1 – CSCW .....	12
2.2.2 – Sistemas de <i>workflow</i> .....	13
2.2.2.1 – Tipologia de processos.....	15
2.2.2.2 – Funcionamento dos sistemas de <i>workflow</i> .....	17
2.2.2.3 – <i>Workflows</i> científicos.....	17
2.2.2.4 – Reutilização de processos científicos.....	19
2.2.3 – Percepção ( <i>Awareness</i> ) .....	20
2.2.3.1 – Percepção de processo .....	21
2.2.3.2 – Ferramentas de mensagens instantâneas .....	23
2.3 – Gestão do conhecimento .....	23
2.3.1 – Conhecimento .....	24
2.3.2 – Conceito de gestão do conhecimento.....	26
2.3.3 – Gestão do conhecimento científico.....	29
2.3.4 – Memória organizacional.....	32
2.3.4.1 – RBC - Raciocínio baseado em casos.....	33
2.3.4.2 – Ontologias .....	36
2.3.4.3 – Sistema de <i>workflow</i> e a memória organizacional .....	38
2.4 – Conclusão.....	39
Capítulo 3 – Thoth .....	41

3.1 – Thoth - Introdução.....	41
3.1.1 – Apoiando a reutilização em procedimentos científicos .....	44
3.1.2 – Abordagem RBC no Thoth .....	45
3.2 – A configuração dos elementos do RBC no Thoth.....	46
3.2.1 – Similaridade e utilidade dos casos .....	48
3.2.2 – Ontologia no Thoth .....	49
3.2.2.1 – Criação da ontologia .....	50
3.2.3 – Metodologia de definição de similaridade e recuperação de casos .....	53
3.2.3.1 – Definição da similaridade entre os casos .....	53
3.2.3.2 – Estrutura física e recuperação dos processos .....	58
3.2.3.3 – Etapa de Classificação .....	59
3.2.3.4 – Armazenamento dos índices e realização das busca .....	62
3.2.4 – Módulos do Thoth.....	64
3.2.4.1 – Interfaces com os usuários .....	65
3.2.4.1.1 – Consulta de processos .....	66
3.2.4.1.2 – Interfaces de visualização das informações de processos .....	67
3.2.4.1.3 – Módulo de colaboração .....	69
3.2.4.2 – Busca e classificação.....	69
3.2.4.2.1 – Módulo de classificação.....	70
3.2.4.2.2 – Modulo de busca .....	71
3.2.4.3 – Interfaces com sistemas externos.....	71
3.3 – Implementação do protótipo .....	72
3.3.1 – Implementação das interfaces com usuário.....	72
3.3.2 – Colaboração.....	73
3.3.3 – Busca e classificação.....	74
3.4 – Conclusão .....	75
Capítulo 4 – Estudo de caso .....	77
4.1 – O LABHID.....	77
4.1.1 – Dificuldades enfrentadas.....	79
4.1.2 – Aplicação do Thoth no laboratório .....	80
4.2 – Processo de outorga da água .....	81
4.3 – Utilização do protótipo.....	87

4.3.1 – Planejamento .....	87
4.3.2 – Participantes .....	88
4.3.3 – Procedimento.....	88
4.3.4 – Resultados esperados .....	89
Capítulo 5 – Conclusão .....	90
Contribuições .....	91
Trabalhos Futuros .....	92
Referências Bibliográficas .....	94
APÊNDICE A – Sistema de Gestão do Conhecimento Científico (GCC) ....	102
APÊNDICE B – Ontologia desenvolvida .....	106

## Índice de Figuras

Figura 1 - Método científico (FILHO, 2000).....	7
Figura 2 - Etapas do trabalho científico (SCHUR, KEATING et al., 1998) .....	9
Figura 3 - Hierarquia do conhecimento (Davenport e Prusak) .....	24
Figura 4 - Espiral do conhecimento (NONAKA & TAKEUCHI, 1995) .....	28
Figura 5 - Processos de transformação de conhecimentos no processo científico .....	31
Figura 6 - Ciclo básico de processamento em RBC (adaptado de (AAMODT & PLAZA, 1994)).....	35
Figura 7 - Ciclo RBC aplicado ao sistema de <i>workflow</i> .....	47
Figura 8 - Estrutura das bacias hidrológicas .....	51
Figura 9 - Estrutura dos modelos .....	52
Figura 10 - Ontologia com aspectos atualmente selecionados .....	53
Figura 11 - Representação dos passos para escolha de conceitos similares .....	58
Figura 12 - Processo de classificação (SALTON & MCGILL, 1983) .....	59
Figura 13 - Meta-modelo da definição de processos (HOLLINGSWORTH, 1995). ...	60
Figura 14 - Modelo de busca baseado em ontologia (Adaptado (KHAN, 2000)) .....	62
Figura 15 - Consulta SQL gerada com termos originais da busca.....	63
Figura 16 – Consulta SQL gerada com termos considerados similares.....	63
Figura 17 - Arquitetura do Thoth.....	65
Figura 18 – Interface para realização de busca.....	67
Figura 19 - Visualização de informações de um processo consultado .....	68
Figura 20 - Tela para escolha de conceitos (aspectos de busca).....	71
Figura 21 - Arquitetura do módulo de colaboração .....	73
Figura 22 – Arquitetura dos componentes de classificação e busca.....	74
Figura 23 - Fluxograma da análise de disponibilidade hídrica para Industrias dentro do processo de Outorga.....	85
Figura 24 - Arquitetura do GCC .....	102
Figura 25 - Árvore Hiperbólica.....	105

# Capítulo 1 – Introdução

## 1.1 – Motivação

Uma preocupação atual das organizações é a maneira como devem agir em um ambiente ágil e composto por concorrentes cada vez mais capacitados, e por tecnologias, técnicas e conhecimentos desenvolvidos de forma cada vez mais veloz. Outro fator de preocupação é a diminuição do ciclo de vida dos produtos, sejam eles conhecimentos, serviços ou bens materiais que ficam obsoletos cada vez mais rápido.

Tais fatores determinam que essas organizações necessitem estar sempre se avaliando, modificando-se e melhorando suas práticas na busca pelo controle e melhoria da qualidade dos seus produtos. Parte desses esforços é direcionada justamente ao reconhecimento da forma como suas atividades são desempenhadas e como os conhecimentos necessários para a realização das mesmas são criados, manipulados e trocados pelos profissionais da organização.

Nesse contexto, tópicos como gestão de conhecimento, inteligência competitiva, gestão de qualidade e gestão de processos, entre outros, passaram a fazer parte do dia-a-dia dessas organizações como mais um subsídio para que essas questões sejam tratadas.

Essa situação, embora inicialmente observada no ambiente empresarial, é extensiva também ao ambiente científico, em organizações como universidades, laboratórios e centros de pesquisa e desenvolvimento das organizações. Tal comportamento foi observado por Robert Howard (RATCLIFFE-MARTIN, COAKES et al., 2004), que afirma que essas organizações deverão sofrer mudanças drásticas nos próximos anos devido às demandas da sociedade baseada no conhecimento, e na qual o aprendizado organizado deva se tornar um processo que perdure por toda a vida dos profissionais..

Dessa forma, organizações científicas devem estar atentas aos problemas que ameaçam o bom andamento do trabalho. Um desses problemas é a retenção de conhecimentos pelas pessoas. Os profissionais adquirem boa parte do seu conhecimento durante a execução do seu trabalho. Muitas vezes, esse conhecimento fica restrito a eles próprios, não sendo realizado o seu compartilhamento com outros

membros organização. Por isso, é possível que esses conhecimentos sejam perdidos na medida em que se tem uma alta rotatividade de profissionais, sendo comum trocar de projeto, pesquisar outros problemas ou sair da organização.

Valorizar a troca de conhecimentos e experiências é muito importante no ambiente científico, onde as atividades podem ser multidisciplinares, normalmente possuem alto grau de complexidade e demandam grande capacitação dos pesquisadores. Assim, essa troca supre a possível inexperiência de um profissional ou ainda a necessidade de conhecimentos oriundos das diversas áreas de saber envolvidas.

Outra dificuldade reside na grande quantidade de informações existentes sem estratégias adequadas de consulta. No caso específico dos procedimentos científicos, a disponibilidade de dados é um fator tão importante quanto à noção de como eles foram adquiridos e manuseados anteriormente.

O uso de tecnologias computacionais auxilia na solução desses problemas ao permitir que pessoas troquem informações, mesmo quando geograficamente distribuídas, e que partes das atividades sejam executadas com o auxílio do computador, além de preservar os dados gerados para consultas posteriores.

Uma dessas tecnologias são os sistemas de *workflow*. Eles têm a função de facilitar o planejamento, além de auxiliar no controle e na coordenação de processos.

O sistema de *workflow* auxilia na resolução de parte dos problemas explanados anteriormente, na medida em que as informações nele armazenadas documentam os processos e as atividades. A reutilização dessas informações permite que os conhecimentos explicitados na forma de documentos, dados e observações possam ser consultados na construção de novos conhecimentos ou na resolução de problemas similares.

No entanto, para que seja possível a reutilização de processos científicos, as características dos mesmos precisam ser observadas. Dentre elas está o caráter particular de cada problema estudado, que demanda estratégias, abordagens e decisões específicas de acordo com o problema. Dessa maneira, para que a reutilização seja realizada de forma satisfatória, os sistemas de *workflow* precisam ser combinados com outras estratégias e tecnologias, tais como gestão de conhecimento, raciocínio baseado em casos (RBC) (AAMODT & PLAZA, 1994) e busca de informações (*information retrieval*) (SALTON & MCGILL, 1983), de forma que as informações armazenadas possam ser categorizadas, buscadas e reutilizadas de maneira adequada.



## 1.2 – Objetivo

O objetivo principal deste trabalho é o desenvolvimento de um sistema que permita a localização e o reuso de processos científicos armazenados em um sistema de *workflow*. Ele tem o propósito de auxiliar o pesquisador durante o planejamento e a execução de suas atividades, e para o seu desenvolvimento foram utilizadas técnicas, ferramentas e conceitos desenvolvidos na área de gestão do conhecimento, CSCW, inteligência artificial e busca de informações.

Conforme será visto nos próximos capítulos, a gestão de conhecimento é um tópico de grande interesse nas organizações, inclusive as científicas. Isso porque o registro dos conhecimentos adquiridos durante a realização das atividades é uma das formas de explicitação úteis para a construção de novos conhecimentos organizacionais. Desta forma, enquanto o planejamento e a execução das atividades através do sistema de *workflow* tornam o trabalho mais ágil e controlado, a reutilização das práticas e informações dessas atividades permite que novas atividades sejam planejadas de forma mais segura e precisa, assim como a suas próprias realizações.

O sistema consiste em uma ferramenta *web* direcionada para os pesquisadores, capaz de determinar dentre processos já executados aqueles que podem ser úteis no contexto do trabalho atualmente em execução. Assim, características dos processos científicos, tais quais singularidade dos problemas tratados, multidisciplinaridade e flexibilidade dos métodos de execução, determinam que a localização deve se basear em aspectos semânticos do problema, como, por exemplo, as áreas de conhecimento envolvidas, estratégias ou abordagens utilizadas. O sistema foi chamado de “Thoth” e se baseia na abordagem de RBC, que prevê a busca de soluções de problemas utilizando a similaridade com os casos anteriormente solucionados.

Thoth é a divindade egípcia que trouxe para a Terra a escrita hieroglífica, a alquimia, a matemática, a arquitetura, a medicina, a magia -- enfim a base de todas as ciências. Os mitos ao seu respeito mostram que seus conselhos eram sempre procurados pelas outras divindades. Cabe a ele, como escriba confidencial do deus Osíris, determinar a dignidade dos mortos e registrar criteriosamente o resultado dos julgamentos do juízo final, assinalando se aquele que estava sendo avaliado havia se comportado bem ou não (LAERCIO, 2004). Vem daí a analogia que originou o nome

da ferramenta, que tem como objetivo facilitar as escolhas e decisões por parte dos cientistas a partir da análise de atividades anteriormente realizadas.

### **1.3 – Organização da Dissertação**

Neste trabalho é proposto um sistema para apoiar a reutilização de conhecimentos em processos científicos. Entretanto, para a compreensão do trabalho, faz-se necessária a revisão da literatura, redigida no capítulo 2. Nela, são mostrados os conceitos sobre o trabalho científico, o conceito de processo, os sistemas de *workflow*, conceitos relacionados à área de CSCW e como esses itens auxiliam no trabalho científico. Os conceitos de conhecimento e gestão do conhecimento são apresentados para se compreender a importância dessa prática para a organização, as características do conhecimento científico e o papel dos sistemas de *workflow* na gestão do conhecimento. Ainda no contexto da gestão do conhecimento, a abordagem RBC, que é um importante componente na concepção da proposta, também é apresentada.

O Thoth é apresentado no capítulo 3, sendo descritos a arquitetura do sistema, a forma como o mesmo interage com o sistema de *workflow*, como são feitas as buscas dos processos e a forma como os pesquisadores interagem com o sistema.

O Thoth foi concebido e desenvolvido a partir das observações dos trabalhos de estudantes, professores e pesquisadores do Laboratório de Hidrologia e Estudos do Meio Ambiente (LABHID), da COPPE/UFRJ. No capítulo 4, há uma descrição do ambiente encontrado no laboratório, na forma como o Thoth auxilia na execução desse trabalho, assim como a descrição de como o protótipo é avaliado nesse contexto. Finalmente, o capítulo 5 apresenta a conclusão do trabalho, descrevendo as colaborações, as limitações e apontando o seu desenvolvimento futuro.

## Capítulo 2 – Revisão de Literatura

Neste capítulo será realizada a explicação dos tópicos que foram pesquisados e utilizados no desenvolvimento deste trabalho. Inicialmente, será feita uma explanação sobre o ambiente científico e sobre a maneira como os trabalhos são realizados no mesmo, procurando identificar as suas características e a influência destas na concepção do sistema.

A partir de então, será possível compreender a razão da adoção das tecnologias e como cada uma delas pode ser utilizada no contexto do trabalho científico. Dessa maneira, serão apresentados: os sistemas de *workflow*, o suporte por computador ao trabalho cooperativo (CSCW – *Computer Supported Cooperative Work*) e, finalmente, a gestão de conhecimento, enfatizando os aspectos da memória organizacional e das estratégias de busca e recuperação de conhecimentos.

### 2.1 – O Trabalho Científico

O trabalho científico é aquele realizado com o objetivo de desenvolver a ciência, sendo conduzido em instituições como laboratórios, universidades e centros de pesquisa e desenvolvimento. O cunho científico dessas organizações está não apenas no desenvolvimento de novos conhecimentos, mas também no fato deste ser obtido através do “pesquisar cientificamente”, ou seja, através da utilização de procedimentos capazes de dar confiabilidade aos resultados obtidos.

A ciência pode ser definida como o conjunto de descrições, interpretações, teorias, leis e modelos, dentre outros, que visam a compreensão de uma parcela da realidade, em contínua ampliação e renovação, e que resulta da aplicação deliberada de uma metodologia.(FREIRE-MAIA, 1995).

Pesquisar é o processo de construção do conhecimento, que tem como metas principais a descoberta e a confirmação, ou refutação, de conhecimentos existentes. O simples ato de pesquisar é realizado por qualquer pessoa no momento em que busca a solução de algum problema. Entretanto, a pesquisa científica deve obedecer aos rigores impostos pelo método científico (CLARK & CASTRO, 2002). Dessa forma, o conhecimento científico se diferencia da aplicação do senso comum, pois é produzido através da metodologia científica, surgindo não apenas da necessidade de encontrar

soluções para problemas de ordem prática da vida diária, mas do desejo de fornecer explicações sistemáticas que possam ser testadas e criticadas.

A motivação para realizar uma pesquisa científica surge da necessidade de descobrir algo novo, ou corroborar ou refutar algo já conhecido. Ou seja, quando se descobre que os conhecimentos existentes, sendo eles originários do senso comum ou do conjunto de conhecimentos existentes na ciência, são insuficientes para solucionar algum problema.

### **2.1.1 – Método Científico**

As bases para a realização da pesquisa científica foram instituídas por Galileu Galilei, Francis Bacon e René Descartes, no século XVI (GOLDENBERG S., 2001). Galileu Galilei (1564-1642) introduziu o método científico, ressaltando o valor da experimentação como o mais importante dos procedimentos naturais, e foi o primeiro a combinar a experimentação científica com a matemática. Francis Bacon (1561-1626) estabeleceu as recomendações para realizar experimentos de caráter indutivo, onde a descoberta de fatos verdadeiros depende da observação e da experimentação guiadas pelo raciocínio indutivo. Já Descartes (1596-1650) pregava a fuga do subjetivismo e a dúvida como meio de raciocínio. Para ele, duvidar significava pensar.

Dessa forma, o método científico é constituído por uma série de princípios, utilizados para que o estudo de um fenômeno seja realizado o mais racionalmente possível, de modo a evitar enganos e sempre buscando evidências e provas para as idéias, conclusões e afirmações. Os princípios que são respeitados são o da falsificabilidade, das aproximações sucessivas e o da parcimônia:

O princípio da falsificabilidade, estabelecido por Karl POPPER (POPPER, 1959), assinala que a ciência só pode estudar temas em que o conhecimento adquirido possa ser negado de alguma forma. Se alguma coisa não pode ser negada, ela não pode ser posta em dúvida e, portanto, não podemos testar sua validade. Assim, a ciência exige o confronto da teoria com os dados empíricos.

O princípio das aproximações sucessivas (POPPER, 1959) estabelece que a verdade sobre determinado fato jamais é atingida integralmente, mas vem sendo aperfeiçoada continuamente. POPPER afirma que "uma explicação é algo sempre incompleto; sempre podemos suscitar um outro por quê, que talvez leve a uma nova teoria que não só explique, mas corrija a anterior". Por fim, o princípio da parcimônia

estabelece que se duas hipóteses explicam os dados com igual eficiência, deve prevalecer a mais simples.

O método científico é composto por quatro fases: observação e descrição de um fenômeno, formulação de hipóteses para explicá-lo, uso de hipóteses para prever a existência de outros fenômenos e execução de experimentos para a comprovação das hipóteses. Se as hipóteses não forem refutadas pelos experimentos, pode ser formulada uma teoria ou uma lei (PENNAFORT;WOLFS,1996). A pesquisa é realizada através de uma linha de raciocínio do pesquisador, que intercala as fases do método (SEVERINO, 2002) . FILHO (2000) afirma que o trabalho científico é composto por quatro campos de atuação, que formam o método, e por quatro tipos de operações mentais (dedução, análise, indução e síntese), que constituem uma via de raciocínio em espiral, como apresentado na Figura 1.

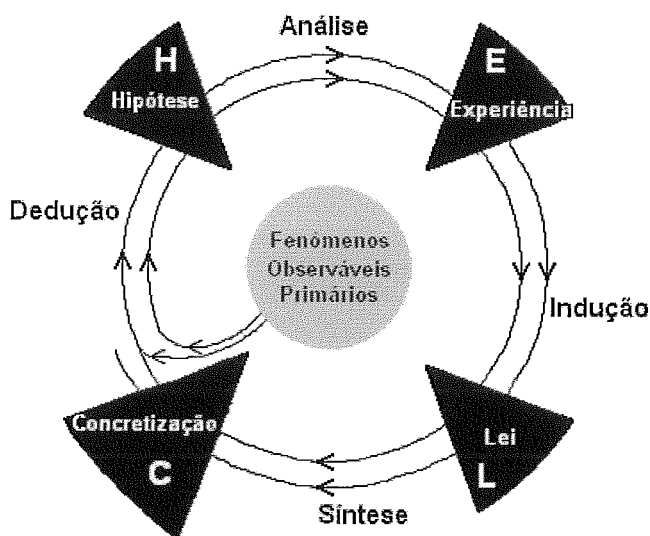


Figura 1 - Método científico (FILHO, 2000)

O primeiro passo é a dedução das hipóteses, que são enunciados particulares correspondentes a um dado fenômeno geral observado. A dedução se caracteriza pela evolução do geral para o particular.

Conhecida pelo menos uma hipótese, prossegue-se para a análise, destinada a especificar os procedimentos e detalhes técnicos necessários para o teste da hipótese. Na análise, ocorre a seleção dos materiais apropriados para a execução dos testes, a definição das metodologias que serão adotadas, a seleção das variáveis a serem consideradas e dos dispositivos para medição das variáveis escolhidas etc. A complexidade desse procedimento analítico é enfatizada por POPPER: "O

experimentador não está principalmente empenhado em fazer observações exatas; seu trabalho é, também, em grande parte, de natureza teórica. A teoria domina o trabalho experimental, desde o seu planejamento inicial até os toques finais, no laboratório".

Concluída a experiência, tem início o processo inverso. E então, através das relações constatadas, são enunciadas as leis ou verificado o inesperado (a menos que a experiência tenha sido conduzida na expectativa de falsear a hipótese). Nessa etapa, é utilizado o raciocínio indutivo, ou seja, partindo do caso particular tenta-se chegar ao caso mais geral.

Finalmente, é efetuada a síntese, onde a lei enunciada é aplicada a uma situação limitada para se chega ao processo geral. O conhecimento assim adquirido acumula-se no campo das concretizações.

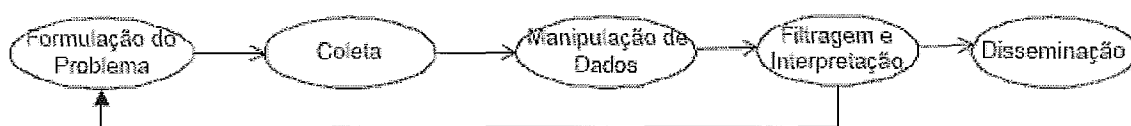
É importante salientar que existem outras propostas para o método científico, mas nenhuma goza de unanimidade. De fato, o processo de construção do conhecimento científico é um tópico em discussão na área de filosofia da ciência. No século XX, por exemplo, podem ser citadas duas correntes principais, a de Karl POPPER (POPPER, 1959) e a de Thomas Kuhn (MAIA, 2005). POPPER estabelece uma regra metodológica associada ao método dedutivo de prova, contrária ao indutivismo e tendo a falsificabilidade como critério a ser adotado pelos cientistas durante o teste de suas teorias. Para ele, o cientista deve ter atitudes revolucionárias permanentes, contrárias à lógica de Kuhn, que considera que o cientista deve se acomodar aos paradigmas vigentes.

Não é objetivo deste trabalho discutir a filosofia do trabalho científico. A explanação do trabalho científico é importante para prover o entendimento básico, ainda que superficial, da maneira como o mesmo é realizado, e, a partir de então, permitir a identificação das suas características e a definição das propriedades necessárias no sistema de informação utilizado para dar suporte às atividades dessa natureza.

### **2.1.2 – Características do Trabalho Científico**

O trabalho científico possui características que muitas vezes não são aplicáveis a outros ambientes, como o empresarial. Uma delas é a de ser constituído como uma seqüência de atividades, estruturadas nas fases que compõem o método científico. Essas atividades possuem uma estruturação lógica possível de ser descrita, como assinalada por Popper: “partirei da suposição que ela consiste apenas em investigar os

métodos empregados na forma sistemática a que toda idéia nova deve ser submetida para que possa ser levada em consideração” (POPPER, 1959). Popper é enfático ao separar a lógica do conhecimento da psicologia do conhecimento, que tem como função compreender os processos envolvidos na concepção das idéias. A lógica, por sua vez, “se preocupa com a reconstituição racional das provas posteriores pelas quais se descobriu que a inspiração era uma descoberta ou veio a ser reconhecida como conhecimento”. Essa seqüência lógica pode ser equiparada às observações realizadas em (CHIN, LEUNG et al., 2002), para o qual os cientistas estruturam os seus experimentos como passos seqüenciais, nos quais são combinados os artefatos utilizados, como ilustrado na Figura 2.



**Figura 2 - Etapas do trabalho científico (SCHUR, KEATING et al., 1998)**

Outro aspecto a ser observado é o foco do trabalho científico, isto é, o conhecimento, que é ao mesmo tempo a sua principal ferramenta e o seu principal produto. As argumentações que justificam a resolução de um problema são realizadas através da articulação de métodos, modelos, dados e idéias, enfim, conhecimentos que possam justificar o que se quer demonstrar. Os conhecimentos utilizados nessas atividades são os acumulados pelos pesquisadores durante a sua vida profissional, os de domínio das organizações e os oriundos das consultas executadas nas diversas fontes de informação estudadas.

Normalmente, durante a realização de uma pesquisa, são analisados grandes volumes de dados e informações, adquiridos a partir de simulações e experimentos anteriores, cálculos e modelos. Esses dados normalmente são mais complexos do que aqueles utilizados em aplicações comerciais e operacionais, porque sua estrutura pode variar de acordo com as fontes dos mesmos e também porque podem estar localizados em bases distribuídas.

Por fim, o trabalho científico é colaborativo, como afirmado por WOLFS (1996): “Em um campo onde há constante experimentação e comunicação aberta entre os membros da comunidade científica, os preconceitos de indivíduos ou grupos podem ser cancelados, porque os experimentos são repetidos por cientistas diferentes que podem ter concepções diferentes sobre o problema estudado. Além disso,

diferentes configurações experimentais vão possuir fontes diferentes de erros sistemáticos. Durante um período em que são realizados uma variedade de experimentos (normalmente, pelo menos alguns anos), um consenso é desenvolvido na comunidade sobre qual resultados experimentais estavam sendo observados.”

Além disso, os problemas estudados em certas áreas científicas não se limitam, geralmente, a um domínio específico, sendo necessário utilizar conhecimentos oriundos de diversos campos do saber. Para adquirir esses conhecimentos, os estudos são realizados, normalmente, com a participação de especialistas das áreas necessárias, com os quais a colaboração tem de ser realizada de forma razoavelmente intensa.

Na prática, as características citadas acima influenciam diretamente o modo como o trabalho científico é realizado, isso porque, para compreender adequadamente um problema e solucioná-lo, o cientista comumente modifica parâmetros, estratégias e abordagens, e repete atividades. Adicionalmente, cada vez que um problema ou fenômeno é estudado, a percepção a respeito do mesmo se modifica na medida em que mais conhecimento a seu respeito é gerado. Isso implica diretamente na forma como um pesquisador planeja, descreve e executa suas atividades, de forma que cada profissional tende a executar cada um desses passos de forma distinta, de acordo com a sua familiaridade com o problema estudado.

O trabalho científico, como representado aqui, vai além dos experimentos. Estes são equivalentes à manipulação dos dados da Figura 2 e representam a fase geralmente mais visível de um trabalho de pesquisa. De fato, as fases anteriores são muito importantes, pois é a partir do momento que um problema é identificado (formulação do problema) e estudado (coleta de dados), que se torna possível a construção de soluções práticas para a sua resolução e a manipulação dos dados de acordo com as observações que precisam ser feitas. A partir do momento que os resultados são gerados e analisados, as informações sobre o problema podem ser disseminadas.

O pensar científico está presente nos trabalhos realizados por profissionais, pesquisadores, professores e estudantes das diversas áreas de saber. Para o contexto deste trabalho, é importante compreender que esse processo ocorre também na área de engenharia e nos processos decisórios ambientais.

O engenheiro é também um solucionador de problemas que elabora novos produtos, sistemas ou processos, ou a melhoria destes. É possível afirmar que o



método de trabalho de um engenheiro é similar ao de um cientista (PEREIRA & BAZZO, 2002), no entanto, enquanto este preocupa-se normalmente só com a construção do conhecimento, aquele combina o conhecimento científico com o conjunto de inventos, componentes, materiais e métodos fabris, além das condições econômicas e mercadológicas do ambiente.

Outros processos onde os conhecimentos e a práxis científica são utilizados de forma intensa são os processos decisórios ambientais, nos quais são tomadas decisões sobre a utilização dos recursos naturais de uma região. Exemplo destes é o zoneamento agroclimático, que tem por objetivo definir a viabilidade de uma cultura em determinada área, melhor explicado em (CASTRO, 2003), e a outorga para o uso da água, explicado detalhadamente no capítulo 4 desta dissertação. Esses processos são caracterizados pela necessidade de consultar diversas fontes de informação e analisar diversos tipos de dados, utilizando para isso metodologias e modelos universalmente aceitos e encontrados na literatura especializada do processo decisório em execução.

As características citadas acima sugerem ser vantajosa a utilização de sistemas de *workflow* no auxílio do planejamento e da execução do trabalho científico. Por outro lado, a gestão de conhecimento pode auxiliar na descoberta, localização, criação e gerenciamento dos conhecimentos durante o andamento dos processos, e as ferramentas de *groupware* podem apoiar a colaboração entre os cientistas.

## **2.2 – CSCW e *workflow*, tecnologias de suporte ao trabalho científico.**

A cooperação é essencial para a realização do trabalho científico, normalmente executado por grupos de pesquisadores que precisam de coordenação e sintonia durante a realização das atividades, além do compartilhamento de informações, modelos, ferramentas, recursos e resultados obtidos em cada etapa do trabalho. SCHUR, KEATING et al. (1998) identificam quatro tipos de colaboração em atividades científicas:

- ✓ **Ponto-a-ponto** - pesquisadores empregam treinamentos e vocabulários comuns e trabalham próximos uns dos outros. Esses pesquisadores utilizam ferramentas colaborativas, que permitem controlar o compartilhamento de instrumentos, esboços e dados.

- ✓ **Mentor-Estudante** - mentores utilizam materiais didaticamente preparados e executam demonstrações ao vivo para ensinar tópicos aos estudantes, como aquisição de dados e técnicas de análise. Mentores observam o desenvolvimento dos estudantes e os guiam conforme necessário. Este tipo de ajuda pode ser altamente interativo: falando, mostrando o que necessita ser feito, e executando uma atividade em conjunto.
- ✓ **Interdisciplinar** - pesquisadores não compartilham as mesmas experiências e freqüentemente traduzem resultados de forma que ambos os compreendam.
- ✓ **Produtor-Consumidor** - pesquisadores provêem dados para outros pesquisadores de diferentes disciplinas, que os utilizam para alcançar diferentes objetivos. Os pesquisadores-fornecedores freqüentemente sabem pouco sobre o que outros fazem com os dados fornecidos.

Os estudos realizados na área de CSCW, explicados na próxima seção, quando aplicados aos ambientes científicos, permitem que cada uma dessas interações seja executada da melhor forma possível.

## 2.2.1 – CSCW

O conceito de CSCW foi utilizado inicialmente em 1984, em um seminário organizado por Paul Cashman e Irene Greif que reuniu pesquisadores de várias disciplinas com o interesse de estudar a forma como as pessoas realizam suas tarefas e de como a tecnologia poderia auxiliá-los. Desde então, o interesse pelo assuntos relacionados ao suporte a grupos de trabalho vem crescendo (GRUDIN, 1994).

CSCW é uma área de pesquisa multidisciplinar que congrega pesquisadores de diferentes áreas como sociologia, economia, antropologia e computação. Como descrito por Bannon e Schmidt (BANNON & SCHMIDT, 1989), não existe uma definição das fronteiras aplicáveis à área, sendo a mesma bastante ampla. O autor chega a citar que o termo serve como um “fórum eficiente para uma variedade de pesquisadores com diferentes *backgrounds* e técnicas discutirem seu trabalho”. Esta indefinição chega a estar presente nos termos pertencentes ao acrônimo, mais precisamente na definição do que é trabalho cooperativo e de suas diferenças em relação ao trabalho colaborativo.

Essa diferenciação é feita por TIJIBOY & MAÇADA (1997). Segundo eles, na colaboração ocorre interação entre os envolvidos, existindo ajuda mútua ou

unilateral. No caso da cooperação, existe interação e colaboração, mas também objetivos comuns, além de atividades e ações conjuntas e coordenadas.

Outra diferenciação que precisa ser esclarecida é entre os termos CSCW e *groupware*, freqüentemente considerados sinônimos. O termo CSCW é mais abrangente, representando a área de pesquisa. Já o termo *groupware* representa uma classe de sistemas baseados em computador que apóiam grupos de pessoas engajadas em uma tarefa (ou objetivo) comum através de uma interface que proporciona um ambiente comum (ELLIS, GIBBS et al., 1991).

Dessa forma, sistemas de correio eletrônico, teleconferência e editores de texto colaborativos são exemplos de *groupware*, pois promovem a interação entre os membros das equipes, contribuindo para que o trabalho seja realizado de forma mais eficiente do que quando esses membros trabalham isoladamente. Os estudos de CSCW avaliam como o trabalho de grupo é realizado, incluindo aí o impacto da utilização de novas ferramentas desenvolvidas.

Tanto os estudos da área de CSCW, quanto as ferramentas de *groupware*, despertam o interesse das organizações, pois propiciam melhorias na comunicação e na interação entre os profissionais, facilitando a troca e a disseminação de conhecimentos e, principalmente no contexto científico, satisfazendo um dos requisitos para que as atividades sejam executadas de forma satisfatória.

Além da colaboração, é necessário prover meios para que o trabalho possa ser organizado, coordenado e controlado. Segundo CHIN, LEUNG et al. (2002), o cientista costuma estruturar suas atividades como passos seqüenciais, de forma que elas podem ser visualizadas como um processo. A partir dessa observação, surge a motivação para utilizar sistemas de *workflow* no controle dessas atividades.

### **2.2.2 – Sistemas de *workflow***

A União para Gerência de Workflow (*Workflow Management Coalition* – WfMC) (WAINER, WESKE et al., 1996) define um processo de negócio como sendo um conjunto de uma ou mais atividades ou procedimentos relacionados, que atingem um objetivo coletivamente dentro do contexto de uma estrutura organizacional que define papéis funcionais e relações.

O princípio utilizado como base nos sistemas de *workflow* é que os processos podem ser representados como fluxos de trabalho, isto é, modelos que especificam as

atividades, a ordem e as condições em que estas devem ser executadas, os seus executores, as ferramentas a serem utilizadas e os documentos manipulados.

A utilização desses sistemas permite a automação total ou parcial e o monitoramento dos processos de negócios (representados por fluxos de trabalho). Esses sistemas são constituídos de um conjunto de ferramentas, que permite o projeto e a representação de fluxos de trabalho, sua instanciação e execução controlada, bem como a coordenação e a integração de diversas ferramentas dentro de um mesmo fluxo de trabalho (ARAÚJO & BORGES, 2001).

Os sistemas de *workflow* são originários das pesquisas realizadas para a automação de escritórios a partir do final dos anos 60. O principal objetivo desses trabalhos era prover soluções para o armazenamento, compartilhamento e direcionamento de documentos nas organizações (ARAÚJO & BORGES, 2001). Segundo ELLIS (1999), alguns sistemas desenvolvidos nessa época já podiam ser considerados sistemas de *workflow* primitivos, pois já continham complexas especificações dos procedimentos das corporações.

Nesse primeiro momento, se observou que muito do otimismo atribuído a esses sistemas foi infundado e que o sucesso organizacional dependia da criatividade das pessoas e da capacidade de se adaptarem às situações, construtivamente modificando ou evitando os procedimentos padronizados quando necessário (ELLIS, 1999). Assim, quando esses sistemas eram introduzidos nos escritórios, atrapalhavam ao invés de auxiliarem, na medida em que não permitiam a execução dessas adaptações.

Nos anos 80, a percepção sobre a tecnologia mudou. Começou-se a notar que esses sistemas envolvem pessoas e devem levar em conta o caráter, muitas vezes não estruturado, das tarefas. Boa parte dessa visão foi adquirida através dos estudos realizados na área de CSCW e *groupware*, que influenciaram na definição desses sistemas como ferramentas para a coordenação de trabalho em grupo. Essa mudança de foco pode ser visualizada em (ABBOTT & SARIN, 1994): “A ênfase no gerenciamento de *workflow* é no uso do computador para ajudar a administrar os processos de negócio, que podem ser compostos de muitas tarefas individuais, e não para automatizá-las”.

O sucesso obtido por esses sistemas no meio comercial, aliado às vantagens potenciais que podem ser conquistadas com o seu uso, motivou a sua utilização em outras áreas. A flexibilização realizada nos sistemas permite que sejam utilizados em

processos com diferentes graus de estruturação, como aqueles que se caracterizam por serem complexos e fortemente dependentes dos conhecimentos. Existem ainda iniciativas para a aplicação em diversos campos, como nos ambientes de suporte aos processos decisórios (CASTRO, SOUZA et al., 2003), de suporte à educação à distância (PALAZZO, NICOLAO M. et al., 1998) e, finalmente, nos processos científicos (WAINER, WESKE et al., 1996), (CARDOSO, SOUZA et al., 2002). Dessa forma, a escolha do sistema adequado para uma organização é condicionada ao grau de flexibilidade dos processos que devem ser controlados, como será explicitado abaixo.

### 2.2.2.1 – Tipologia de processos

Como visto acima, a atual aceitação do uso dos sistemas de *workflow* nas organizações foi adquirida graças ao reconhecimento de que os processos não podem ser considerados como estruturas rígidas, eles possuem graus variados de estruturação, que dependem da natureza e área de aplicação dos mesmos.

Assim, os sistemas de *workflow* podem ser classificados com base no grau de estruturação e na frequência de execução dos fluxos de trabalho que apóiam. Nesse caso, são identificadas três categorias de classificação: *ad-hoc*, administrativos e de produção (CHAFFEY, 1998).

Os processos *ad-hoc* são caracterizados por fluxos de trabalho pouco estruturados. Geralmente, correspondem a processos executados uma única vez, com características muito peculiares. Não é possível prever todas as tarefas, nem seu encadeamento em uma fase de modelagem, antes da execução propriamente dita. Os próprios usuários atuam como desenvolvedores e coordenadores de seus próprios processos. É comum envolverem trabalho em grupo, geralmente centrado em algum documento ou artefato, sobre o qual interagem os membros da equipe. São relacionados tipicamente a atividades de criação de conhecimento, onde a criatividade dos executores tem grande influência em seu sucesso, demandando, portanto, flexibilidade em seu tratamento.

Os processos administrativos possuem um grau maior de estruturação, tornando as atividades e o encadeamento das mesmas mais previsíveis, o que facilita o reaproveitamento de fluxos de trabalho. Além disso, são particulares de cada organização e podem ser alterados com frequência.

Nos processos de produção, os fluxos de trabalho podem ser considerados completamente estruturados, podendo ser totalmente modelados antes de sua execução. Ocorrem freqüentemente e, geralmente, são as principais atividades de seus executores, podendo ser comparados a processos fabris. É desejável que possam ser ajustados para que sua execução seja a mais eficiente possível. Pelo fato de terem sua estrutura bem definida, seus executores não precisam usar de sua criatividade para definir etapas seguintes.

Essas categorias não são disjuntas, ou seja, processos não são necessariamente completamente *ad-hoc*, administrativos ou de produção. Quando visto no mundo real, a execução de um processo administrativo, por exemplo, poder necessitar de ações não previstas. Essas situações são tratadas de modo *ad-hoc*.

CHAFFEY(1998) sugere também outra classificação levando em consideração não só o grau de estruturação dos processos, mas também a necessidade de colaboração que é requerida entre os usuários para a realização do fluxo de trabalho. Nessa classificação, o autor observa que os sistemas de produção dão suporte a processos estruturados, onde a granularidade na definição das atividades e o sequenciamento é maior. A alta previsibilidade dos processos evita que seus executores necessitem tomar decisões sobre o direcionamento das atividades. Conseqüentemente, o grau de interação e colaboração entre os envolvidos na execução do processo é baixo.

A situação oposta é observada nos processos *ad-hoc*, onde a seqüência das atividades é feita em um nível de abstração maior, o que determina que os seus executores tenham que tomar decisões sobre o trabalho que realizam e negociar entre si os próximos passos.

Os sistemas de *workflow* são classificados de forma homônima, isto é, existem sistemas de *workflow* de produção, destinados ao suporte de processos predominantemente de produção, assim como sistemas de *workflow* administrativos. Os sistemas de *workflow* utilizados em procedimentos científicos são chamados de sistemas de *workflow* científicos e são *ad-hoc*, já que o trabalho científico possui alguma imprevisibilidade. Estes sistemas serão explicados de forma mais aprofundada na seção 2.2.2.3 —, mas antes, o funcionamento dos sistemas de *workflow* será explanado de forma mais precisa.

### **2.2.2.2 – Funcionamento dos sistemas de *workflow***

Seja qual for o tipo de processo a receber suporte, o sistema de *workflow* possui formas similares de funcionamento, baseadas nos componentes dos processos. Os principais elementos de um fluxo de trabalho são as atividades, que correspondem a uma parcela de trabalho a ser realizada dentro do processo. São caracterizadas pelos seus elementos, como nome, objetivos, instruções, dados, formulários e documentos necessários, que podem ser manipulados durante sua execução.

As atividades são executadas por indivíduos, agentes, grupos ou papéis. Os indivíduos podem ser responsáveis pela execução de determinada atividade dentro de um fluxo de trabalho. Essa não é a estratégia ideal porque as organizações estão sujeitas à rotatividade de pessoas, o que determinaria a necessidade de se modificar os fluxos de trabalho sempre que as pessoas responsáveis pelas execuções das atividades mudassem. Assim, o papel reúne um conjunto de características ou responsabilidades necessárias para a execução de uma ou mais atividades atribuídas a pessoas ou agentes automatizados.

Outro aspecto importante na definição do fluxo de trabalho é o encadeamento de atividades, ou seja, é preciso definir a seqüência de execução das atividades para se compor um processo.

Todas essas características são definidas quando o modelo do processo é criado no sistema de *workflow*, respeitando a observação de como o processo efetivamente ocorre ou foi planejado para ocorrer no mundo real. O modelo contém justamente as informações anteriormente citadas, isto é, atividades, profissionais, grupos, papéis, seqüências entre as atividades etc. Esse modelo é utilizado para criar a instância do processo, quando o mesmo efetivamente ocorre no mundo real.

Comportamento similar ocorre nos sistemas de *workflow* científicos, utilizados para representar processos científicos, assim como controlar a execução de cada instância desses processos, como apresentado na próxima seção.

### **2.2.2.3 – *Workflows* científicos**

Os sistemas de gerenciamento de *workflow* científicos servem para definir processos científicos e acompanhar a sua execução através do controle da manipulação dos dados, modelos, programas e ferramentas utilizados na execução de cada tarefa.

A necessidade dos *workflows* científicos reside nas particularidades que diferenciam o trabalho científico do trabalho operacional. Essas diferenças são assinaladas por WAINER, WESKE et al. (1996). Segundo estes autores, quando se executa o trabalho operacional, a preocupação é prioritariamente alcançar os objetivos definidos pelas regras das empresas. Os dispositivos para tratamento de exceção, existentes em alguns desses sistemas, fazem com que passos planejados sejam negligenciados em prol do cumprimento dos objetivos. Esses desvios atestam que tal flexibilidade, embora aceitável, só deve ser utilizada quando ajudar na execução das atividades.

Já no trabalho científico, o principal elemento é o dado, e o objetivo é gerá-lo, coletá-lo e analisá-lo. Nesse caso, a obtenção e junção de dados oriundos de vários experimentos, a geração de dados a partir de modelos e a análise são as atividades que recebem suporte dos sistemas de *workflow*. Assim, o tratamento de insucessos no ambiente científico é visto como um mecanismo de aprendizado, para evitar a repetição de erros. A reavaliação de um resultado inesperado, por exemplo, pode gerar novas idéias ou ser aplicada na solução de um outro problema em um contexto diferente. No ambiente de escritório, um insucesso pode ser fruto de um problema isolado e nem sempre é importante estudá-lo de forma criteriosa.

No ambiente científico, o planejamento e a execução das atividades são realizados pelo próprio pesquisador. Além disso, é possível que o mesmo realize alterações nas atividades programadas mesmo durante a execução destas. Isso não ocorre em ambientes convencionais, onde geralmente existe um responsável pela definição dos processos que serão posteriormente executados pelos profissionais da organização, sem possibilidade de alterações.

Assim, modificações e adaptações são realizadas comumente na busca pela melhor resposta para os problemas. Ou seja, o grau de flexibilidade que os cientistas têm em seu trabalho é usualmente muito superior ao existente no ambiente operacional das empresas, onde os processos são definidos e executados como rotinas.

As características citadas demandam funcionalidades não contempladas pelos sistemas de *workflow* convencionais. Por isso, os *workflows* científicos podem ser definidos como extensões sobre os convencionais, com o objetivo de suprir essas necessidades. As extensões são as seguintes (WAINER, WESKE et al., 1996):



- **Incompletude:** *workflows* científicos podem ser executados até mesmo quando incompletos, sendo assim construídos progressivamente durante sua execução e não sendo necessário serem totalmente definidos antes de serem executados;

- **Reutilização parcial:** *workflows* científicos diferem dos tradicionais, pois são considerados como blocos em construção para especificação de experimentos. Assim, *workflows* parciais podem ser utilizados para a especificação de novos *workflows*;

- **Modificação dinâmica:** *workflows* científicos permitem re-executar atividades e também retroceder a uma atividade anterior, restabelecendo seu contexto e continuando a execução por um novo curso de ação. Isto é, a especificação de um *workflow* pode ser modificada dinamicamente;

- **Execução de processos inválidos:** no domínio científico, processos de decisão são baseados no mecanismo de tentativa e erro. *Workflows* científicos devem ser flexíveis o suficiente para apoiar esse tipo de abordagem;

#### 2.2.2.4 – Reutilização de processos científicos

HOLLENBACH & FRAKES (1996) definem a reutilização de processos como a utilização da descrição de um processo na criação de outra descrição, diferenciando esta da múltipla execução de um mesmo processo. Eles indagam ainda sobre o que determinaria a propriedade de um processo ser reutilizável e concluem que não existem indicadores definitivos para determinar tal propriedade. Concluem simplesmente afirmando que “se um processo for reutilizado, então ele é reutilizável”. A reutilização pode ser de quaisquer artefatos manuseados no processo. Na proposta de Hollenbach & Frakes, por exemplo, os usuários ficam livres para reutilizar artefatos como o código-fonte, a documentação, os modelos de dados e as diretrizes, bem como outras informações relevantes.

Essa conceituação, embora realizada no contexto da área de engenharia de *software*, é geral o suficiente para abrigar outros domínios de aplicação, como os processos científicos. Nesse caso, o pesquisador também deve ter a possibilidade de selecionar dos itens dos processos aqueles que são realmente importantes para a realização do seu trabalho.

A reutilização no contexto científico está relacionada com o aproveitamento de procedimentos, modelos e práticas. Um exemplo pode ser observado na arquitetura CHIMERA (FOSTER, VOCKLER et al., 2002), que foca a gerência de metadados e

pressupõe que, uma vez que os dados para realização de uma pesquisa foram obtidos, é importante saber as transformações originalmente realizadas. Além disso, se já existe um programa que consulta uma base de dados ou implementa determinado modelo, a sua reconstrução se torna desnecessária. Muitas vezes, se um experimento foi anteriormente realizado, a sua re-execução em um novo projeto pode ser desnecessária, o que propicia economia de tempo e recurso.

Uma das características mais importantes a ser considerada na reutilização é a definição de quais processos podem ser úteis no contexto do problema que está sendo trabalhado. No caso do ambiente científico, essa escolha é dificultada pelo caráter particular de cada problema. Assim, aspectos semânticos dos processos, como os modelos científicos utilizados, os assuntos tratados, as competências necessárias, dentre outras, devem servir de base para a localização dos processos.

Outro aspecto importante que deve ser considerado é a percepção do usuário, que deve identificar a utilidade dos itens resultantes da busca no seu contexto. Por isso, é necessário prover dispositivos que permitam que os itens de busca sejam reconhecidos como úteis por esses usuários. Esses mecanismos são chamados dispositivos de percepção.

### **2.2.3 – Percepção (*Awareness*)**

Ultimamente, o conceito “percepção” tem recebido muita atenção dos pesquisadores da área de CSCW. Esse interesse é originário da consciência sobre a importância das informações contextuais no processo de comunicação e colaboração (HANDEL, 2002), e nos processos executados por uma organização. A percepção é conceituada por DOURISH (1992) como um entendimento das atividades de terceiros que provê um contexto para a própria atividade, sendo essa informação importante para garantir a relevância das contribuições individuais diante das atividades de todo o grupo.

DAVID & BORGES (2001) descrevem os dispositivos de percepção como interfaces desenvolvidas para apresentar informações que podem ser capturadas das interações, filtradas e tratadas para informar os usuários sobre os aspectos do processo de interação.

São diversos os tipos de informação contextual que podem ser providas, sendo possível que os dispositivos sejam classificados de acordo com diversos aspectos. A

classificação proposta por STEINFELD, JANG et al. (1999) leva em consideração o tipo de informação que é disponibilizada. As classes identificadas são as seguintes:

- Percepção de atividade – Informação sobre as ações que outros membros do grupo estão executando no momento, disponibilizada, por exemplo, pelos radares, comuns em ferramentas de edição colaborativa.
- Percepção de disponibilidade – Informação sobre a presença de determinado usuário do sistema e a disponibilidade para interação com o mesmo.
- Percepção de processo – É a percepção normalmente disponibilizada pelos sistemas de *workflow* e permite que usuários identifiquem o estado corrente de um processo (seqüência de tarefas).
- Percepção perspectiva – Informação sobre o ponto de vista de outros membros (crenças e conhecimentos), ajuda na compreensão das idéias, contribuições e decisões dos participantes de um grupo.
- Percepção do ambiente – Prover informações sobre eventos que ocorrem fora do local de trabalho, mas que podem causar implicações nas atividades do grupo.

Devido à importância no trabalho proposto, duas dessas classes serão melhores explicadas: a percepção de processos e a percepção de disponibilidade.

### **2.2.3.1 – Percepção de processo**

A percepção de processo é, basicamente, fornecida pelos sistemas de *workflow*. Ela é importante para tornar possível que cada envolvido tenha ciência da situação em que se encontra um processo e sobre seu papel e responsabilidade para que se alcance a conclusão do mesmo. Dessa forma, o sistema de *workflow* deve prioritariamente prover informações de percepção sobre os objetos compartilhados e sobre o andamento das atividades.

A quantidade das informações fornecidas varia de acordo com o sistema de *workflow*. ARAÚJO & BORGES (2002) descrevem os dispositivos comuns a todos os sistemas:

- Mapa do processo (*Process map*) - O mapa do processo fornece uma visão gráfica do mesmo durante a sua execução, sendo possível

identificar a seqüência de atividades e a situação em que cada uma delas se encontra.

- Lista de trabalho (*worklist*) - Apresenta para cada participante as tarefas que requerem a sua atenção em um momento específico.

Além dessas, outras informações de cunho mais administrativo podem ser analisadas, tais como prazos não respeitados, linhas críticas, profissionais sobrecarregados com muitas tarefas.

Eles também identificam outras informações que, quando fornecidas, permitem uma visão mais global dos processos, viabilizando a colaboração mais intensa entre os profissionais e o maior aprendizado sobre a organização:

- Grupos, papéis e responsabilidades – Fornecimento de formas para que os membros compreendam o perfil do seu grupo e, conseqüentemente, as possibilidades de colaboração.
- Interações – Fornecimento de informações que permitam a identificação de colegas que estão próximos em termos das atividades executadas, os que estão distantes e com quais a colaboração será mais eficiente.
- Nível de participação – Fornecimento de informações sobre as colaborações voluntárias de cada profissional, sendo este um forte indício do comprometimento de cada membro com o processo como um todo.
- Impacto das tarefas – Prover informações sobre a importância das tarefas para o processo; pode aumentar o grau de comprometimento dos profissionais.
- Foco e convergência – Prover informações sobre outros profissionais que executam a mesma atividade em outras instâncias do mesmo processo; pode viabilizar a colaboração entre esses profissionais.
- Conflitos e divergências - Prover informações sobre o nível de conflitos ou divergências entre participantes que executam a mesma tarefa. Essa situação ocorre quando, por exemplo, dois profissionais executam atividades de formas distintas, estando sujeitos às mesmas condições.

Embora normalmente fornecidas com o intuito de viabilizar ou facilitar a execução das atividades, esses dispositivos são importantes no momento em que se deseja reutilizar um processo e, conseqüentemente, são importantes para o sucesso da ferramenta aqui proposta, já que permitem que o pesquisador reconheça as características dos processos candidatos à reutilização e que possa determinar se os mesmos são úteis ou não no seu trabalho.

### **2.2.3.2 – Ferramentas de mensagens instantâneas**

As ferramentas de mensagens instantâneas obtiveram sucesso no suporte à comunicação predominantemente informal. São basicamente sistemas de comunicação entre dois ou mais usuários que agregam mecanismos de percepção sobre a disponibilidade das pessoas cadastradas em uma lista de contatos.

Embora absorvida de forma positiva no uso doméstico, a utilização dessas ferramentas ainda é discreta no ambiente de trabalho. Isso porque as ferramentas enfrentam a resistência de algumas organizações que a consideram motivo de desconcentração de seus funcionários e conseqüente perda de eficiência. Cabe salientar que essa resistência diminui gradativamente, principalmente em empresas onde a comunicação entre empregados ou com clientes geograficamente distribuídos precisa ser realizada com freqüência. Alguns autores mais entusiastas afirmam que, no futuro, essas ferramentas serão aceitas e utilizadas de forma tão natural quanto o *e-mail*.

## **2.3 – Gestão do conhecimento**

A gestão de conhecimento vem ganhando cada vez mais importância em todas as esferas produtivas da sociedade. Essa prática também é necessária no ambiente científico, porque o conhecimento é o item mais importante para a realização do trabalho científico, na medida em que é, ao mesmo tempo, o seu principal instrumento e produto.

Nesta seção, serão apresentados os temas relacionados à gestão do conhecimento. Primeiramente, será explorado o conceito de conhecimento. A seguir, serão expostos os principais conceitos relacionados à Gestão de Conhecimento. Por fim, será descrito como esses conceitos se enquadram no trabalho científico e, mais precisamente, na abordagem para reutilização proposta no Thoth.

### 2.3.1 – Conhecimento

A definição do que é conhecimento não é tão simples quanto parece, pois normalmente seu significado é confundido com outros conceitos, como o de informação. Por isso, tão importante quanto a definição de conhecimento, é a diferenciação desses outros conceitos. Isso é realizado por DAVENPORT & PRUSAK (1998), que fazem a identificação de quatro conceitos distintos: dado, informação, conhecimento e ação. Conforme apresentado na figura 3.

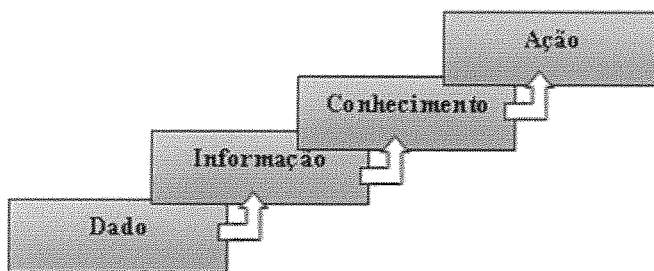


Figura 3 - Hierarquia do conhecimento (Davenport e Prusak)

Na figura, os elementos são ordenados de forma que cada um represente uma base para o outro. Os dados são coleções de fatos, medidas e estatísticas. As informações são dados dotados de relevância e propósito, exercendo alguma influência sobre o julgamento do indivíduo que as utiliza. Já o conhecimento tem embutido em si valores como sabedoria e intuição dos agentes envolvidos. Por fim, a ação é conceituada como algo que se faz com o conhecimento, que pode ser caracterizada pelas decisões e atitudes tomadas para agregar valor ao que é realizado. Em alguns trabalhos, como em (MORESI,2000), a hierarquia é modificada, sendo o conceito “ação” suplantado pelo conceito “sabedoria”, que, segundo o autor, representa o conhecimento contextualizado relevante que permite atuar com vantagem em um determinado ambiente ou situação.

A diferenciação apresentada acima é importante para se visualizar que o conhecimento está intrinsecamente relacionado ao processamento realizado sobre as informações, e é condicionado às experiências, valores e crenças do seu detentor. Essa noção pode ser ratificada através da definição realizada em (DAVENPORT & PRUSAK,1998), para os quais o conhecimento pode ser definido como “uma mistura fluida de experiências, valores, informação contextual, discernimento de especialistas

e intuição embasada, provendo um ambiente e uma estrutura para a avaliação e incorporação de novas experiências e informação”.

Existem diversas classificações distintas para o conhecimento na literatura, sendo que cada uma considera um aspecto diferente. Michael Polanyi (OLIVEIRA, 2003) *adupt* (POLANYI, 1983) diferencia o conhecimento em tácito e explícito. O conhecimento tácito é difícil de ser articulado na linguagem formal. É o conhecimento pessoal incorporado à experiência individual, envolvendo fatores intangíveis como as crenças, perspectivas e o sistema de valores de cada indivíduo. Já o conhecimento explícito, é articulado na linguagem formal mais facilmente pelo uso de expressões gramaticais, expressões matemáticas ou lógicas, manuais, especificações e assim por diante. Dessa forma, é mais facilmente transferido entre os indivíduos.

Uma segunda classificação é realizada por LAKATOS & MARCONDI (1991), que separam o conhecimento em religioso, filosófico, popular e científico.

O conhecimento religioso é o aceito pela fé teológica. Esse conhecimento é apoiado em doutrinas de proposições sagradas e direcionado à compreensão do mundo em sua totalidade. O conhecimento religioso caracteriza-se por ser valorativo, advindo de inspiração, sistemático, não verificável e falível.

O conhecimento filosófico baseia-se no filosofar, é a tentativa de decifrar certa interrogação e exige um método experimental. O objeto da análise da filosofia são idéias, relações conceituais e exigências lógicas, que não são redutíveis a realidades materiais e, por essa razão, não são passíveis de observação sensorial direta ou indireta (por instrumentos), como a que é exigida pelo conhecimento científico.

O conhecimento vulgar ou popular, muitas vezes denominado de senso comum, não se distingue do conhecimento científico nem pela veracidade, nem pela natureza do objeto conhecido. Os dois se diferenciam pela forma, o modo ou o método e os instrumentos do “conhecer”. Isto é, o conhecimento comum é adquirido sem ser procurado, sem reflexão ou aplicação de métodos. As informações são assimiladas por tradição e experiências causais e ingênuas.

O conhecimento científico vai além do empírico, pois se preocupa não só com os efeitos, mas principalmente com causas e leis. Ocorre de forma lenta, sendo um processo contínuo de construção, com um complexo processo de pesquisa, análise e síntese. “É uma busca constante de explicações e soluções, e a reavaliação de seus resultados”. O conhecimento científico é o resultante das atividades científicas, e o

seu objetivo é demonstrar, mediante argumentos, uma solução proposta para um problema relativo a um determinado tema (SEVERINO, 2002).

### **2.3.2 – Conceito de gestão do conhecimento**

A gestão de conhecimento tem se tornando um tema cada vez mais importante nas organizações, consolidando-se como forte tendência na área de gestão empresarial. Segundo Drucker (FREIRE-MAIA, 1995), as organizações perceberam que a dinâmica da sociedade atual é regida pelo conhecimento, o que impõe a necessidade de se estar sempre preparado para responder às mudanças provenientes do ambiente, refletindo-as na própria estrutura organizacional. Assim, o conhecimento organizacional é o fator que permite criar uma vantagem competitiva, de forma a elevar o nível de qualidade dos bens, serviços e métodos.

As empresas estão tornando-se produtoras e consumidoras intensivas de conhecimento, que está rapidamente assumido o papel de principal recurso para a organização ante o capital, os recursos físicos e o trabalho manual. O conhecimento é um importante elemento quando é necessário realizar modificações, cada vez mais frequentes devido à constante diminuição do ciclo de vidas dos produtos.

Segundo HOWARD (2000), as organizações tem de estar preparadas para abandonar estratégias e metodologias a fim de ser possível construir novos produtos. Essa postura é necessária pela volatilidade de um mercado composto por clientes, que têm acesso diariamente a novas informações, tecnologias e produtos, desejando novidades e maior qualidade naquilo que consomem, e por competidores ávidos por fatias cada vez maiores desse mercado. Posturas conservadoras em um ambiente como esse podem fazer com que os produtos fiquem inadequados e, conseqüentemente, que a empresa perca o seu espaço. Esses novos produtos, tecnologias e informações trazem para dentro das organizações uma quantidade crescente de conhecimentos que precisam ser administrados.

Além disso, a troca de informações realizadas pelos profissionais quando aprendem, criam e compartilham conhecimentos é realizada de forma mais intensa que no passado. Essas trocas, por sua vez, são realizadas por meios formais ou informais. Os canais de comunicação informais permitem que conhecimentos tácitos sejam trocados entre colegas. Não prestigiar esses processos pode fazer com que o conhecimento tácito de cada funcionário seja perdido no momento em que os mesmos saiam da empresa. De fato, o sucesso das organizações está associado a sua habilidade



de criar novos conhecimentos de forma consistente, disseminá-los rapidamente, armazená-los e colocá-los em novos produtos e serviços.

A partir do exposto acima, é possível visualizar que o papel da gestão de conhecimento é propiciar à organização a possibilidade de responder à volatilidade dos ambientes interno e externo, através da valorização da troca de conhecimento entre os profissionais e a alimentação contínua do conhecimento organizacional.

A gestão do conhecimento não é uma disciplina nova no âmbito da administração das organizações, e é objeto de uma variedade de abordagens, definições e percepções nem sempre concordantes, sendo que nenhuma dessas abordagens foi aceita de forma consensual (CLARK & CASTRO, 2002).

Algumas dessas abordagens enfocam a gestão de conhecimento no âmbito meramente técnico. Propiciam o controle do conhecimento explícito através do uso de dispositivos computadorizados, ferramentas *groupware* e ferramentas de apoio à decisão. Essa abordagem pode ser observada em algumas conceituações encontradas na literatura como, por exemplo, a fornecida em (BARROSO & GOMES, 1999): “seria possível definir gestão de conhecimento como o trabalho de gerenciar documentos e outros veículos de informação e de conhecimento, com o objetivo de facilitar a aprendizagem da organização (entendimento)”.

Entretanto, essas abordagens não dão a importância apropriada aos profissionais, principais envolvidos na construção de uma estrutura eficaz para a gestão de conhecimento organizacional. Dessa forma, é correto afirmar que a construção de poderosas redes e o desenvolvimento de *intranets* robustas não é suficiente como uma infra-estrutura eficaz de gestão de conhecimento. Ou como afirmado em (DAVENPORT & PRUSAK, 1998): “A tecnologia da informação é somente um sistema de distribuição e armazenamento para o intercâmbio do conhecimento. Ela não cria conhecimento e não pode garantir nem promover a geração ou o compartilhamento do conhecimento numa cultura corporativa que não favoreça tais atividades”.

O importante é observar que o objetivo da gestão do conhecimento é otimizar os processos organizacionais, que têm como principal fator para serem executados de forma eficaz a capacitação e o envolvimento dos profissionais. O alcance de tais metas depende do desenvolvimento de uma cultura organizacional que incentive e valorize a comunicação, e encoraje as pessoas a compartilharem os seus conhecimentos e a buscarem ajuda quando desconhecem algum assunto.

A gestão de conhecimento deve se preocupar com a criação, a captura e o fornecimento de conhecimento, e permitir que a pessoa correta aplique o conhecimento correto no momento correto. Além disso, a gestão de conhecimento deve ser vista como uma atividade que pode até ter um começo no momento em que a empresa conscientemente toma atitudes para efetuar esse processo, mais que não deve ter um fim (TIWANA, 2000).

A importância da figura humana no processo de gestão de conhecimento pode ser confirmada em (NONAKA & TAKEUCHI, 1995), que descreve o processo de criação do conhecimento, tanto o tácito como o explícito, como interações contínuas e dinâmicas entre essas duas formas de conhecimento. Essas interações podem ser vistas como uma espiral, iniciada pelos indivíduos e que vai evoluindo até o nível organizacional, quando o conhecimento passa a ser um patrimônio da organização.



**Figura 4 - Espiral do conhecimento (NONAKA & TAKEUCHI, 1995)**

A conversão se dá de quatro formas, como pode ser visualizado na Figura 4: do conhecimento tácito para o conhecimento tácito (socialização), do tácito para o explícito (externalização), do explícito para o explícito (combinação) e do explícito para o tácito (internalização).

Na socialização, ocorre a interação e a conseqüente troca de conhecimentos tácitos entre os indivíduos através do compartilhamento das emoções, modelos mentais, idéias e da percepção relacionadas a uma determinada situação, permitindo que práticas sejam observadas e depois repetidas em contextos diferentes do original.

Na externalização, o conhecimento tácito é traduzido em conceitos possíveis de serem justificados, categorizados e contextualizados no âmbito organizacional.

Quando essa atividade é desempenhada adequadamente, o conhecimento subjetivo passa a ser efetivamente compreendido e se torna útil para a organização como um todo, possibilitando a reutilização em desafios futuros.

A combinação consiste na sistematização dos conceitos. Nesse modo de conversão, há a combinação de conjuntos diferentes de conhecimento explícito. Ou seja, os indivíduos trocam e combinam conhecimento através de documentos, manuais, reuniões ou redes de comunicação computadorizada. A reconfiguração das informações se dá através da classificação, acréscimo, combinação e categorização do conhecimento explícito, e leva a novos conhecimentos. Acontece, por exemplo, na criação do conhecimento através dos cursos ou do treinamento nas escolas.

A internalização compreende a incorporação individual dos conhecimentos explícito da organização. Nesse caso, o conhecimento organizacional se torna uma ferramenta para o crescimento de cada um dos seus profissionais. Para que esse processo seja efetuado, são necessárias a verbalização e a diagramação sob a forma de documentos, manuais ou histórias orais.

Tais processos dependem predominantemente da figura humana. Logo, o importante para a organização é fornecer os meios para que as pessoas possam executá-los da forma mais natural possível, através de uma cultura organizacional propícia, constituída de políticas que incentivem a troca de conhecimento aliadas a ferramentas tecnológicas.

### **2.3.3 – Gestão do conhecimento científico**

O interesse pela gestão de conhecimento surgiu das profundas modificações sócio-econômicas que tornaram o conhecimento um importante fator para o processo de produção, ao lado dos fatores clássicos como a força de trabalho, máquinas e matéria-prima.

Essas modificações repercutiram no ambiente científico, que despertou para a importância da gestão de conhecimentos, na medida em que suas atividades dependem basicamente de conhecimentos, muitas vezes complexos. Entretanto, é importante observar as diferenças estruturais existentes entre organizações empresariais convencionais e as instituições científicas. Tais diferenças demandam que sejam utilizadas estratégias e ferramentas diferenciadas no processo de gestão de conhecimento de cada uma destas organizações.

Uma diferença que necessariamente precisa ser analisada é o fato dos dados e sistemas científicos serem mais complexos do que os encontrados nos ambientes operacionais. O dado científico é mais complexo tanto no seu significado quanto na sua estrutura, pois o dado científico pode estar relacionado simultaneamente a múltiplas informações ou perspectivas, como temperatura, clima e tipo de solo. Muitas vezes, é necessário consultar dados originários de diversas organizações, o que demanda um esforço adicional para a compreensão da semântica desses dados, ou para a realização de adaptações ou padronizações na estrutura física do mesmo, a fim de que possa ser utilizado em uma plataforma possivelmente diferente da qual ele foi criado. Não raramente, é necessário se coletar dados e conhecimentos de alto custo, oriundos das outras organizações, o que determina que sejam realizados esforços para a negociação e determinação de políticas de colaboração.

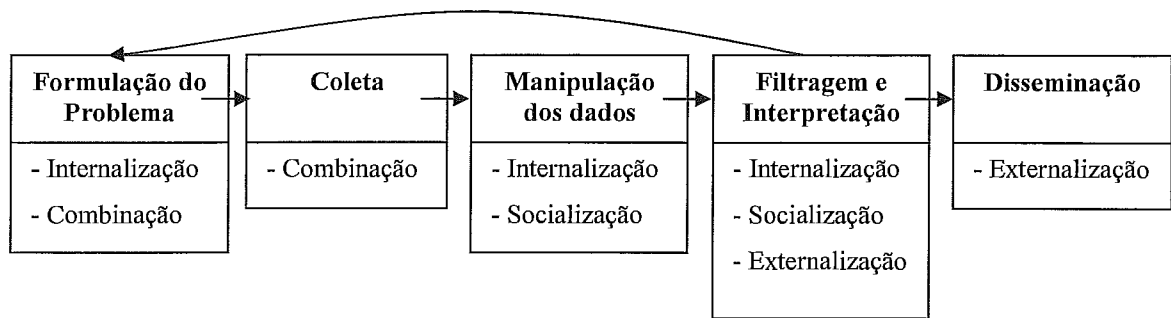
OLIVEIRA(2003) identifica outros fatores motivadores para a definição de uma infra-estrutura para a gestão do conhecimento científico:

- O capital intelectual da instituição não pode ficar associado exclusivamente a pessoas que detêm o conhecimento crítico, mas deve ser distribuído entre os membros de uma equipe de pesquisa.
- Que seja possível a identificação de áreas de conhecimento com escassez de profissionais e, assim, planejar uma forma de adquirir tais conhecimentos, seja por treinamento ou contratação de pesquisadores externos.
- Que se torne possível o constante acompanhamento do nível de conhecimento de cada pesquisador.

Dessa forma, o apoio das atividades com sistemas de gestão de conhecimento propicia uma gama maior de conhecimentos, informações e dados sobre o problema abordado, permitindo que incertezas possam ser minimizadas. Por outro lado, o próprio trabalho que se está realizando aprimora os conhecimentos retidos na base de conhecimento e na memória organizacional.

Para se ter uma compreensão adequada de como ocorre a criação e a utilização do conhecimento nos processos científicos, é necessário identificar as formas de conversão de conhecimento dentro de cada atividade desse processo, como ilustrada na

Figura 5.



**Figura 5 - Processos de transformação de conhecimentos no processo científico**

Quando ocorre a formulação do problema, os pesquisadores pesquisam o problema, aprendem e formam suas próprias opiniões sobre o mesmo (internalização). Além disso, parte do que é aprendido é estruturado de forma integrada (combinação).

Durante a coleta de dados, o pesquisador procura por novas fontes de informações sobre o problema para que suas idéias possam ser confirmadas, ocorrendo predominantemente o processo de combinação de conhecimentos.

Durante a manipulação dos dados, o pesquisador confirma ou refuta suas idéias. Nessa fase, os resultados são geralmente discutidos entre cientistas, mas não em caráter definitivo, ocorrendo predominantemente a internalização e a socialização.

Na filtragem e interpretação, o pesquisador realiza praticamente todos os processos de forma conjunta, isto é, ele avalia os resultados obtidos do trabalho. Dessa interpretação pode surgir novos conhecimentos e idéias (internalização), no entanto, esse trabalho é comumente realizado em consonância com outras pessoas existentes no grupo e com as quais esses resultados são discutidos (socialização). A partir daí, os resultados obtidos já são mais definidos e, assim, são explicitados de maneira mais formal (externalização).

Finalmente, deve ocorrer a disseminação do resultado formal do trabalho para toda a comunidade científica, correspondendo ao processo de externalização.

O papel dos sistemas de informação deve ser otimizar a forma como esse processo é realizado. Para isso, OLIVEIRA(2003) descreve as ferramentas e tecnologias atualmente utilizadas nas organizações de pesquisa para auxiliar a gestão do conhecimento, dividindo-as conforme área de atuação e propósito. São elas: os sistemas de banco de dados distribuídos e heterogêneos (SBDDH), as linguagens para padronização de dados científicos, os sistemas gerenciadores de modelos (SGM) e as bibliotecas digitais (BD).

Para o desenvolvimento deste trabalho, as observações se concentram nos sistemas de *workflow* científicos, nas ferramentas de *groupware* e CSCW, anteriormente apresentados, e nas ontologias, utilizadas para auxiliar a construção e a busca de conhecimentos na memória organizacional.

### **2.3.4 – Memória organizacional**

Um dos objetivos da gestão do conhecimento é incentivar o aprendizado organizacional. NONAKA & TAKEUCHI (1995) conceituam o aprendizado organizacional como um processo de mudança adaptativo e crescente influenciado pelas experiências passadas e focalizado na modificação e no desenvolvimento de rotinas. Ele é uma base para o estabelecimento de estruturas que permitem a criação de uma “organização que aprende” (*knowing organization*), isto é, que analisa de forma crítica seus processos para que de alguma maneira possa melhorá-los. Por outro lado, o sucesso do aprendizado organizacional está fortemente ligado à construção e à utilização de mecanismos para preservar, manter e consultar os conhecimentos, ou seja, o desenvolvimento da memória organizacional. No contexto científico, a forma como os experimentos são executados, assim como os resultados que são obtidos, são elementos importantes na composição da memória, na medida que auxiliam na execução mais eficiente dos trabalhos futuros.

Diversas definições de memória organizacional podem ser encontradas na literatura. Assim como o conceito de gestão de conhecimento, elas são influenciadas por diversas perspectivas e visões. Para CONKLIN(1996), por exemplo, a memória organizacional é o registro de uma organização personificada em seus documentos e artefatos. Por outro lado, EUZENAT (1996) descreve a memória organizacional como uma coleção de dados e recursos de uma organização, incluindo experiência de projetos, experiência na solução de problemas e raciocínios de desenvolvimento, entre outros.

A definição realizada por EUZENAT(1996) parece mais adequada, na medida em que assinala a importância de contextualizar as informações, descrevendo a forma como foram adquiridas, modificadas e aplicadas em novas situações, de maneira que tais práticas possam ser repetidas no cotidiano da organização.

A memória organizacional constitui a representação explícita dos conhecimentos de uma organização e, por isso, é um espaço onde as informações

estão em evolução contínua, já que é alimentado diretamente pelos conhecimentos criados durante a rotina normal de trabalho (FISCHER & OSTWALD, 2001).

Segundo O'LEARY (1998), a memória organizacional é constituída de dois elementos: bases de conhecimento e ontologias. As bases de conhecimento armazenam fisicamente o conhecimento, enquanto as ontologias auxiliam a representação e recuperação do conhecimento nessas bases.

As bases de conhecimento podem ser agrupadas segundo o tipo de conhecimento que armazenam. O'LEARY apresenta alguns tipos de bases de conhecimento, dentre os quais:

- Bases de conhecimento de compromissos - Contém informação sobre diferentes tipos de compromissos assumidos pela organização;
- Bases de conhecimento de melhores práticas - Contém descrições dos processos da empresa que aparentam ser a melhor forma de realizar as atividades;
- Bases de conhecimento de lições aprendidas - Contém informações de experiências de sucesso e fracasso da organização;
- Bases de conhecimento de especialistas - contém descrições das habilidades, conhecimentos e competências dos membros da organização.

#### **2.3.4.1 – RBC - Raciocínio baseado em casos**

Um dos importantes aspectos que deve ser considerado na construção da base de conhecimento é a estratégia de busca utilizada. Essa escolha depende predominantemente do tipo de conhecimento que se deseja buscar e da forma como o mesmo está estruturado. Neste trabalho, a reutilização pode ser traduzida como a busca de práticas, experiências e dados -- enfim, conhecimentos -- contidos nos trabalhos científicos desempenhados em uma organização. Esses trabalhos são estruturados como processos e armazenados em um sistema de *workflow*. Esse conhecimento é factual, na medida que representa como um problema ou situação foi tratado em algum momento. Atualmente, o método RBC vem sendo bastante utilizado para a realização de reutilização de conhecimentos que possuem as características descritas acima.

O RBC é um enfoque para a solução de problemas e aprendizado baseado em experiência passada. A resolução do problema é realizada a partir da recuperação e adaptações sobre experiências passadas, chamadas de casos, armazenadas na base de casos. O princípio básico do RBC é de que “problemas similares possuem soluções similares”. Dessa forma, um novo problema pode ser resolvido a partir da adaptação de uma solução conhecida para um problema similar (LEAKE & KOLODNER, 2000). O RBC é amplamente utilizado nas atividades relacionadas com a recuperação e reutilização de informações em sistemas baseados em conhecimento.

Uma das razões para a aceitação da abordagem é a própria forma como as pessoas tendem usualmente a categorizar as coisas através de ocorrências passadas. Os casos transformam conceitos abstratos em imagens reais, facilitando a descrição das situações (LEAKE & KOLODNER, 2000). Assim, o RBC é muito vantajoso em domínios onde a identificação e a caracterização de regras ou modelos são complexas e imprecisas, o que certamente dificulta a construção de sistemas utilizando abordagens tradicionais. Nesses casos, para um especialista, pode ser mais fácil explicitar seus conhecimentos, adquiridos durante anos de trabalho, através de casos concretos, em vez de tentar descrevê-los através da formulação de regras (VON WANGENHEIM & VON WANGENHEIM, 2003b). Quando se desenvolve um sistema baseado no RBC, quatro elementos prioritariamente devem ser considerados na definição da arquitetura: a representação do conhecimento, a medida de similaridade, a adaptação e o aprendizado (VON WANGENHEIM & VON WANGENHEIM, 2003b).

A **representação do conhecimento** está relacionada à forma da representação dos casos que descrevem as experiências concretas passadas. O caso, no contexto de um sistema RBC, pode ser definido como uma peça de conhecimento contextualizado que registra um problema, ou situação problemática, que foi total ou parcialmente solucionado. O caso é a espinha dorsal do sistema RBC e nele devem estar descritos o problema e a respectiva solução adotada.

A **medida de similaridade** é que determina a capacidade de encontrar um caso relevante para o problema atual e a definição de se esse caso é similar o suficiente e pode ser útil na resolução do novo problema. Esse é o aspecto mais importante a ser considerado, na medida em que está diretamente ligado à premissa básica do RBC de que problemas similares possuem soluções semelhantes. Por isso, o sucesso na aplicação da abordagem depende da escolha de um conceito de





de viagem, alguns descritores seriam o destino e o custo dos pacotes previamente negociados.

Na **reutilização**, a solução do problema (isto é, o novo caso) é construída a partir das soluções dos casos recuperados ou de parte delas, quando então são realizadas adaptações e modificações nessas soluções.

A **revisão** corresponde à avaliação e ao testes realizados para se determinar se a solução adquirida realmente é útil.

Na **retenção** o novo caso é efetivamente armazenado na memória.

A realização das atividades contidas nesse ciclo depende do tipo de implementação realizada. BARTSCH-SPORL, LENZ et al. (1999) descrevem dois tipos básicos de sistema: os totalmente automáticos, onde a resolução dos problemas é realizada de forma autônoma e tem mecanismos de interação com o mundo para avaliar os resultados de suas decisões, e os sistemas de recuperação baseado em casos, que apenas auxiliam pessoas na resolução dos problemas, como uma extensão da memória do usuário, a quem cabe realizar o raciocínio e tomar decisões.

Embora, normalmente, o ciclo ocorre com a intervenção humana, muitas ferramentas RBC agem primordialmente como sistemas de recuperação e reutilização de casos, sendo que os outros passos, principalmente a revisão e a adaptação, são desempenhados pelos usuários do sistema (VON WANGENHEIM & VON WANGENHEIM, 2003b).

Neste trabalho, a abordagem RBC é utilizada integrada com ontologias. Estas têm a função de propiciar conhecimento sobre o domínio de trabalho no momento da identificação dos descritores de cada caso e também na geração das consultas.

#### **2.3.4.2 – Ontologias**

As ontologias foram originalmente pesquisadas por um braço da filosofia que estuda a existência, o ser enquanto ser (*ontos*). Ontologias formais são propostas desde o século XVIII, incluindo contribuições recentes como as propostas por Carnap (1968) e Bunge (1977) (JURISICA, MYLOPOULOS et al., 1999).

Na ciência da computação, o interesse surgiu através dos estudos realizados na área de inteligência artificial. Posteriormente, elas foram utilizadas em trabalhos de outras áreas, como a lingüística computacional e banco de dados. Atualmente, podem ser encontradas no mundo várias iniciativas de desenvolvimento de ontologias em diversas áreas do conhecimento, como na tradução de linguagem natural, na

engenharia, na medicina, no comércio eletrônico e nos sistemas de informações geográficas.

No Brasil, uma iniciativa notável é realizada pela comunidade LMPL, um grupo de instituições que definem ontologias para a Plataforma Lattes (CNPQ, 2004). Essas ontologias são representadas através da linguagem de marcação XML (vindo daí o nome LMPL - Linguagem de Marcação da Plataforma Lattes) e têm por finalidade estabelecer entre as instituições de pesquisa e o CNPq uma forma comum de troca de informações sobre a plataforma.

Do ponto de vista computacional, uma ontologia define os conceitos usados na descrição e na representação de uma área de conhecimento, sendo usada pelas pessoas, banco de dados e aplicações que precisam compartilhar informações. Assim, nela se encontram os conceitos de um domínio e os relacionamentos entre esses conceitos (HEFLIN, 2004). A formalização imposta permite o desenvolvimento de algoritmos que realizam inferências a partir de um conjunto de fatos sobre o mundo ou a verificação da consistência de informações, facilitando o reuso do conhecimento.

A utilização da ontologia propicia diversos benefícios (USCHOLD & GRUNINGER, 1996):

- Ontologia para especificação: pode ser utilizada na produção de um vocabulário para especificar os requisitos para uma ou mais aplicações. Por exemplo, a ontologia pode ser utilizada como parte da especificação de requisitos no desenvolvimento de um sistema.
- Padronização no acesso as informações: a ontologia facilita a comunicação entre pessoas e organizações, assim como a interoperabilidade, na medida em que reduz a confusão conceitual e terminológica oriundas das diferentes necessidades e dos diferentes pontos de vistas gerados a partir do contexto de cada organização ou pessoa.
- Realização de busca baseada em ontologia: o uso da ontologia para busca de recursos desejados (documentos, *web pages*, áudio) em bases de informações melhora a precisão da busca e reduz o tempo total gasto, pois quando não é encontrada uma resposta perfeita à consulta, a estrutura semântica da ontologia capacita o sistema a retornar com respostas que são mais próximas da consulta.

A utilização da ontologia neste trabalho se baseia justamente na facilitação das buscas dos processos que precisam ser realizadas a partir do momento que os pesquisadores de uma organização reconhecem os conceitos existentes na ontologia. A sua utilização durante a busca permite que as pessoas identifiquem mais facilmente os aspectos que levaram à escolha dos itens retornados durante a busca.

### **2.3.4.3 – Sistema de *workflow* e a memória organizacional**

Sistemas de *workflow* têm uma relação estreita com a memória organizacional, já que, embora tenham como objetivo primordial o auxílio à coordenação e à execução das atividades, as informações armazenadas nesses sistemas podem ser úteis em outras situações, visto que representam parte dos processos realizados e, conseqüentemente, guardam uma parte da realidade do dia-a-dia da organização.

Segundo ZHAO (1998), um modelo de fluxo de trabalho pode ser considerado um tipo de metadado que engloba o conhecimento do processo em questão e parte do conhecimento institucional, e o histórico do *workflow* contém conhecimento sobre casos típicos e exceções. Ou seja, no modelo dos processos está representado o modo pelo qual uma organização “percebe” e executa suas atividades, sendo possível identificar os diferentes tipos de conhecimento envolvidos, tais como as descrições de tarefas, os papéis e responsabilidades desempenhados pelos funcionários, regras e rotas. Por outro lado, as instâncias geradas pela execução de fluxos de trabalho constituem importante fonte de conhecimento sobre o negócio da organização, pois armazenam o conhecimento envolvido cada vez que o processo é executado e podem ser utilizadas para treinamento, melhorias de processos, pesquisa, identificação de erros comuns e na descoberta de regras de negócio vigentes na organização ainda não documentadas.

No caso dos procedimentos científicos, a documentação permite a observação da forma pela qual uma pesquisa, ou experimento, foi executada, sendo útil, por exemplo, para evitar que os erros anteriormente cometidos sejam repetidos. Em alguns casos, a simples verificação de como um experimento foi executado pode permitir que dúvidas sejam sanadas sem necessidade de reexecução, economizando tempo e recursos.

Por essa perspectiva, o sistema de *workflow* é vantajoso ao permitir que, ao mesmo tempo em que as atividades são coordenadas e controladas, possam servir de fonte de conhecimento para outros pesquisadores no futuro e que todas essas

operações sejam realizadas sobre uma estrutura conhecida (a do próprio sistema de *workflow*), padronizada e de fácil compreensão.

Embora todo esse conhecimento esteja presente nos sistemas de *workflow*, em geral, eles têm como foco o controle e a coordenação das atividades, não possuindo formas mais adequadas para a exploração desses conhecimentos (ARAÚJO & BORGES, 2002). Por isso, é necessário o desenvolvimento de estruturas que permitam a consulta e a recuperação das informações de forma mais fácil pelos profissionais através de mecanismos de busca, filtragem e apresentação das informações coerentes com as atividades da organização.

Como descrito anteriormente, no ambiente científico, as atividades têm o objetivo de permitir a compreensão dos fenômenos e problemas estudados, sendo comum modificar parâmetros, abordagens e técnicas para posterior re-execução de experimentos e modelos sobre as novas configurações.

Essas premissas foram consideradas e são os principais elementos trabalhados na concepção da ferramenta Thoth. Isto é, cada instância em um sistema de *workflow* científico representa uma situação particular com atributos e execuções próprias específicas, condicionadas ao problema tratado. No caso do Thoth, é feita a analogia entre cada instância armazenada na base e a ocorrência de um caso, de forma que a abordagem RBC possa ser utilizada para explorar as informações contidas no sistema de *workfow*.

## 2.4 – Conclusão

Neste capítulo, foram apresentados uma série de conceitos e definições que motivaram o desenvolvimento do Thoth. Primeiramente, foi apresentada uma descrição do ambiente de trabalho científico, a partir da qual foi possível identificar três fatores que influenciam diretamente na qualidade e na eficiência do trabalho. São eles: o uso do conhecimento de forma intensa, a necessidade de colaboração entre pesquisadores e a necessidade de se organizar e controlar as atividades que precisam ser realizadas.

Para auxiliar a organização e a coordenação do trabalho, podem ser utilizados os sistemas de *workflow* científicos, desenvolvidos especificamente para dar suporte às características das atividades desenvolvidas durante o trabalho de pesquisa. A colaboração, por sua vez, pode ser realizada através ferramentas *groupware*

desenvolvidas a partir dos estudos realizados na área de CSCW, tais como *e-mail*, ferramentas de votação e dispositivos de mensagens instantâneas.

Outro aspecto importante, o conhecimento, assume ao mesmo tempo o papel de principal produto do processo científico e também a principal ferramenta de trabalho do pesquisador. Em outras palavras, o pesquisador combina seus conhecimentos com os provenientes da colaboração com outros pesquisadores e resultantes dos estudos realizados, e propõem uma solução para o problema enfrentado, que, por si só, é um novo conhecimento. A prática da gestão de conhecimento permite que as estratégias e as tecnologias aqui apresentadas sejam aplicadas de forma harmoniosa para o controle e utilização eficiente do conhecimento pelo pesquisador.

Esses conceitos são importantes para demonstrar a viabilidade do Thoth, um sistema que alia a importância dos sistemas de *workflow* com a necessidade de se executar a gestão de conhecimento. O sistema foi idealizado para permitir que o pesquisador explore as atividades anteriormente executadas na organização e busque conhecimentos que lhe possam ser úteis para a execução de suas próprias atividades. A estrutura do Thoth, seus módulos e modo de operação serão apresentados no próximo capítulo.

## Capítulo 3 – Thoth

Neste capítulo, serão demonstradas as características do Thoth, tema desta dissertação. O Thoth tem por objetivo viabilizar a reutilização de processos científicos. Dessa forma, será exposto como as características da prática científica e as ferramentas previamente apresentadas foram estruturadas para a concepção e desenvolvimento desse sistema.

### 3.1 – Thoth - Introdução

Partindo da observação da demanda crescente por conhecimentos por parte dos cientistas durante a execução de suas atividades, este trabalho propõe um sistema para reuso de conhecimentos, na forma de fluxo de atividades, modelos científicos e dados durante a prática de um processo científico. O Thoth é um sistema desenvolvido para a utilização através da *web* e é dotado de interfaces para consulta de uma base de processos científicos. A partir da análise dessa base e com a utilização da abordagem RBC, são fornecidos dados oriundos de atividades anteriormente executadas que podem ser úteis para a execução de uma atividade.

Como apresentado anteriormente, o ato de pesquisar é bastante complexo, envolvendo muitas decisões nas quais o pesquisador precisa possuir ou localizar os recursos necessários para tomá-las em tempo hábil.

As características do trabalho científico apresentadas anteriormente foram consideradas na concepção do Thoth, sendo elas a estruturação das atividades, a colaboração entre pesquisadores e a forte dependência de conhecimentos relacionados ao problema estudado.

O primeiro fator está relacionado à necessidade de estruturar e controlar as atividades que serão executadas. Assim, um dos comportamentos previstos é que as atividades científicas sejam armazenadas e estruturadas através de um sistema de *workflow*. Esse comportamento é o normalmente observado nos sistemas de apoio ao trabalho científico encontrados na literatura como o WOODSS (SEFFINO, MEDEIROS et al., 1999) e o ambiente descrito por CHIN, LEUNG et al. (2002). Adicionalmente, parte da motivação para a construção deste trabalho é oriunda das observações realizadas durante o desenvolvimento dos ambientes *Spatial Decision*

*Support Collaborative System for Environment Design - SPECS* (MEDEIROS, 2002), (MEDEIROS, SOUZA et al., 2001) e *Gestão do Conhecimento Científico – GCC* (SOUZA; OLIVEIRA et al, 2005), sistemas utilizados para apoiar a realização de atividades científicas inseridas em processos decisórios que utilizam sistemas de *workflow* para criação, controle e execução das atividades.

O ambiente SPeCS tem por finalidade ajudar na tomada de decisão colaborativa em problemas ambientais, com ferramentas que podem ajudar o grupo a lidar com todas as atividades desse tipo de projeto, facilitando a troca de conhecimento e disponibilizando ferramentas para ajudar na chegada do consenso.

O GCC está sendo desenvolvido com o objetivo de auxiliar a gestão do conhecimento científico a partir da observação do ambiente da COPPE - UFRJ, ajudando a formação de comunidades de pesquisadores e indicando oportunidades de colaboração entre membros ou equipes da instituição. O ambiente provê um sistema de *workflow* para que os pesquisadores planejem e coordenem as suas atividades ou a do seu grupo.

Em ambos os ambientes, cada processo é definido e executado de acordo com o problema específico tratado. Por isso, os processos decisórios e científicos apoiados não são estanques, como os processos operacionais, que são primeiramente definidos para então serem executados. Assim, quando uma falha ocorre em uma atividade ou é observado um comportamento inesperado, o processo pode ser redefinido com novas atividades para corrigir aquela problemática, ou para efetuar novas observações. Dessa forma, a adoção de sistemas de *workflow* nesses ambientes não prioriza a automatização, mas a manutenção e controle das atividades, bem como a documentação dos processos, gerando um histórico dos mesmos.

Outro fator importante estudado é a dinâmica do conhecimento envolvida nesses processos. Conforme descrito anteriormente, o pesquisador coleta e analisa grande quantidade de dados, executa simulações e desenvolve modelos, entre outras tarefas, que precisam ser planejadas.

Durante o planejamento, os pesquisadores definem a sua estratégia para alcançar respostas e soluções para os problemas que motivaram a pesquisa. Planejar, nesse contexto, significa determinar quais simulações precisam ser realizadas; localizar, construir ou adaptar os modelos e dados, determinar quais documentos serão consultados e definir a seqüência adequada de execução das atividades que precisarão ser realizadas. Essa é uma atividade razoavelmente complexa, dependente de



conhecimentos oriundos de diversas áreas do saber e da habilidade de manuseá-los de forma eficiente. Tal multidisciplinaridade torna necessária a consulta a informações oriundas de diversos domínios.

Além disso, é comum o excesso de dados e informação, o que torna difícil a localização daqueles mais adequados, sendo necessário um esforço extra para filtrá-los. Por outro lado, domínios pouco explorados possuem escassez de informações, o que dificulta a execução do trabalho, tornando necessário adequar os modelos ao pequeno volume e a qualidade dos dados disponíveis. Esse problema ocorre, por exemplo, quando se realiza um estudo sobre regiões geográficas pouco estudadas. Por isso, se faz necessário prover meios para que os dados sejam disponibilizados de forma adequada para os pesquisadores.

Baseado nas informações acima, um processo científico pode ser considerado como consumidor e, ao mesmo tempo, fonte de conhecimentos, na medida em que diversas formas de conhecimento são consultadas para a solução de um problema. Da mesma forma, quando essa solução é encontrada, pode se caracterizar como uma informação útil em novas situações. Por todas estas razões, é importante considerar os aspectos semânticos dos conhecimentos manuseados em cada atividade.

Outro aspecto que deve ser observado é a colaboração. Durante a realização dos trabalhos, o pesquisador internaliza uma grande quantidade de informações, forma suas próprias idéias sobre elas e formula novas idéias e conclusões sobre o problema estudado. Esses conhecimentos são de natureza tácita e, conseqüentemente, de difícil representação (SVEIBY, 2004). Prover meios para que os pesquisadores colaborem entre si pode melhorar o compartilhamento desses conhecimentos.

Em resposta a isso, o Thoth possui como requisitos prover as seguintes funcionalidades:

- Fornecer ferramentas para reutilização de atividades científicas;
- Prever que a reutilização ocorra em ambos os momentos do procedimento (planejamento e execução);
- Mapear os conhecimentos utilizados e usá-los como principal parâmetro na busca por itens a serem reutilizados;
- Permitir a colaboração entre o usuário do sistema e os envolvidos no planejamento ou execução dos trabalhos consultados, permitindo que aspectos não representados possam ser descobertos e discutidos.

### 3.1.1 – Apoiando a reutilização em procedimentos científicos

Como afirmado anteriormente, as características das atividades científicas são diferentes das atividades operacionais usualmente encontradas em organizações empresariais. Essas diferenças influenciam na forma como essas atividades são executadas e como são representadas e armazenadas no sistema de *workflow*.

No ambiente empresarial, o importante é alcançar os objetivos delineados em cada atividade, e eventuais problemas durante a execução são tratados por dispositivos de controle de exceção e causam pouca ou nenhuma mudança na percepção que se tem das atividades, não modificando a forma como elas são executadas.

O processo de fabricação de um de carro, por exemplo, tende a sofrer poucas modificações. Uma anormalidade, como a interrupção no fornecimento de uma peça, implicará em aplicação de medidas paliativas, como a compra da peça de outro fornecedor, para que a produção volte ao normal o mais rápido possível; contudo, não implicará na mudança da forma como o carro é fabricado.

Processos dessa natureza, sobre a ótica do sistema de *workflow*, possuem um ciclo de vida composto por duas fases bem separadas. Primeiro é realizada a modelagem, quando uma equipe é formada para avaliar a forma como o processo é executado no mundo real, sintetizar as suas características e propor um modelo das atividades a serem executadas. Então, para cada ocorrência no mundo real, é criada uma instância fiel a esse modelo (LEYMANN & ROLLER, 2000).

No ambiente científico, se objetiva a compreensão dos fenômenos, a adaptação dos modelos e o entendimento dos dados resultantes de experimentos, sendo comum a modificação dos parâmetros, das estratégias e das abordagens utilizadas e também a reexecução das atividades. Esse comportamento pode ser melhor visualizado em (SEFFINO, MEDEIROS et al., 1999) que descreve um ambiente para suporte a análises ambientais. O processo é complexo porque é preciso realizar diversas escolhas, geralmente condicionadas à região estudada.

Nesse tipo de ambientes, os processos são criados de forma contínua, isto é, as atividades de planejamento e execução são realizadas de forma intercalada pelos próprios executores, quando detectam a necessidade de modificações por causa de uma descoberta ou pela observação de algum comportamento inesperado. Além disso, cada vez que um problema ou fenômeno é estudado, a percepção a respeito do mesmo

tende a se modificar na medida em que mais conhecimentos ao seu respeito são gerados. Isso influencia substancialmente a forma como cada especialista planeja, descreve e executa suas atividades, de maneira tal que cada profissional tende a executar cada um desses passos de forma distinta, de acordo com a sua familiaridade com o problema tratado.

Esses fatores influenciam diretamente o modo como as atividades são armazenadas e, conseqüentemente, a forma como podem ser reutilizadas, já que, enquanto processos operacionais, obedecem fielmente o esquema modelo – instância do processo científico, ou seja, cada ocorrência representa um caso distinto e oferece oportunidades distintas de aprendizado.

Assim, para a reutilização destes procedimentos ser realizada satisfatoriamente, são necessárias estratégias mais apropriadas para a identificação e a busca dos mesmos na base de casos, estratégias estas que possuem maior complexidade devido ao esforço para analisar os aspectos semânticos dos conhecimentos utilizados em cada instância analisada. Para satisfazer esses requisitos, o Thoth utiliza a abordagem RBC.

### **3.1.2 – Abordagem RBC no Thoth**

Um dos benefícios da abordagem RBC é permitir que os problemas sejam resolvidos minimizando a utilização de várias tentativas e a ocorrência de erros. Desta maneira, é dada aos pesquisadores a possibilidade de tomar as suas decisões com maior segurança. Quando esse objetivo é alcançado, o trabalho tende a ser realizado como maior eficiência e correção (DHAR & STEIN, 1996).

Outras razões fundamentam a utilização da abordagem, como o fato dos sistemas RBC normalmente serem mais fáceis de serem construídos do que sistemas baseados em regras, já que estas, quando regulam problemas complexos, podem não ser completamente conhecidas. Mesmo quando conhecidas essas regras podem ser difícil de descrever ou traduzir para uma forma na qual possam ser computacionalmente exploradas (LEAKE & KOLODNER, 2000).

O caso é o principal elemento de um sistema RBC. No Thoth, é feita a analogia entre as ocorrências dos processos executados pelos pesquisadores (instâncias) e os casos. A visualização da unidade de reutilização como caso provê uma forma de filtragem de conhecimentos, diminuindo a sobrecarga de informações,

já que a consulta é contextualizada sobre o universo restrito do problema tratado na ocorrência similar consultada.

### 3.2 – A configuração dos elementos do RBC no Thoth

Conforme descrito previamente, o projeto de um sistema RBC deve considerar a fixação de quatro elementos básicos: a maneira como o caso é representado, a medida de similaridade utilizada durante a busca, a adaptação da solução para o problema atual e o aprendizado com o tratamento de cada novo problema. Outra questão é a forma como esses elementos se interligam e são utilizados e operacionalizados no ciclo de atividades do RBC. O projeto dessas características foi realizado tendo como alvo a adequação de cada elemento e de cada passo do ciclo à estrutura de um sistema de *workflow*.

No Thoth, o **caso** é basicamente uma instância do sistema de *workflow* correspondente a uma análise previamente executada. É importante salientar que cada uma dessas instâncias possui uma quantidade considerável de conhecimento com potencial de ser reutilizado. Cada documento, artefato manuseado, profissional envolvido, assim como as próprias informações do fluxo (roteamento das atividades, tempo de execução, datas) podem ser, de forma isolada ou combinada, úteis para a solução de um novo problema.

A utilização de uma abordagem automática na **adaptação** de soluções se mostrou inviável nesse primeiro momento, quando características e aspectos mais importantes, tais como a aceitação do usuário, a facilidade de uso e o reconhecimento da utilidade dos casos sugeridos, precisam ser avaliados. Adicionalmente, o sistema foi projetado para ser utilizado em um ambiente de alta complexidade, onde a definição das regras para adaptação precisaria de um tempo dilatado para ser determinada. Dessa forma, se optou pela adaptação manual de soluções, onde o pesquisador ficaria responsável pela seleção e alteração dos atributos dos casos que efetivamente deseja reutilizar.

O **aprendizado** está relacionado com a maneira como as ocorrências no mundo real são traduzidas em casos no momento do armazenamento na base. Para a configuração desse elemento, é necessário realizar uma filtragem dos aspectos, das características e dos atributos da ocorrência no mundo real que serão descritos no caso. Na abordagem aqui proposta, seria equivalente à modificação da definição do modelo do processo que representa cada caso. Um modelo rico é importante para

fornecer uma contextualização do trabalho e, muitas vezes, já possui diversas características omitidas durante a sua construção pelo pesquisador. Por isso, concluiu-se ser inadequada qualquer modificação.

O principal elemento a ser definido é o **critério de similaridade**, responsável pela determinação do grau de correspondência entre os casos armazenados na base com o problema atual. Esse elemento está envolvido com diversos aspectos e componentes da arquitetura do Thoth e será explicado com maior detalhe na seção 0.

Além dos elementos, as quatro tarefas conceituais componentes do ciclo RBC precisaram ser adequadas para o sistema de *workflow*, conforme apresentado na Figura 7.

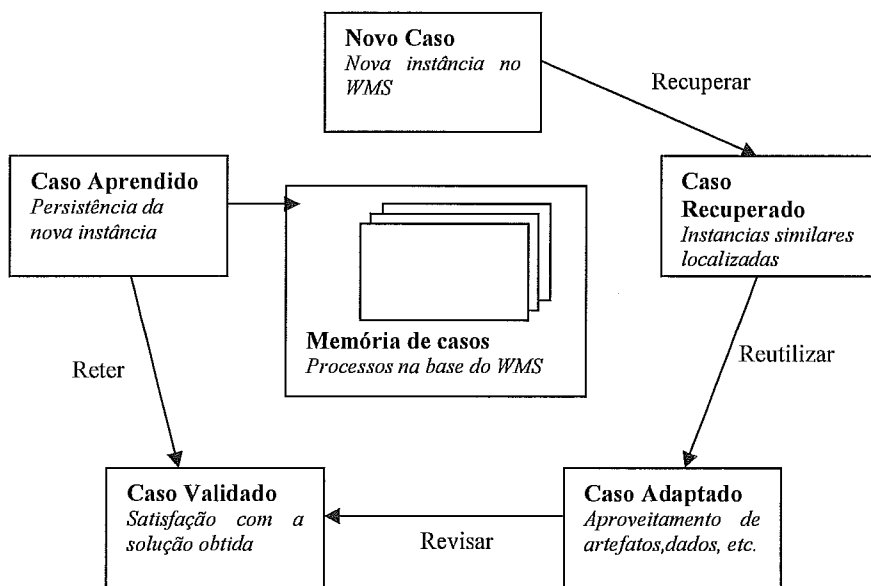


Figura 7 - Ciclo RBC aplicado ao sistema de *workflow*

A **ocorrência do problema**: no momento em que as atividades são planejadas, a consulta a outros trabalhos relacionados pode ser útil na escolha de abordagens, técnicas e ferramentas corretas. Por outro lado, no momento da execução, dúvidas sobre os resultados obtidos ou sobre o modo de se executar tarefas podem ser sanadas a partir da consulta às informações de uma pesquisa similar anterior, ou ao conhecimento dos participantes dessa pesquisa.

**Busca de casos similares**: o pesquisador, após sentir a necessidade de informações adicionais, busca na base de processos aqueles mais aplicáveis à situação atual.

**Adaptação da solução:** ao verificar as informações dessas instâncias, o pesquisador poderá reaproveitar a definição destas na definição do seu próprio processo de pesquisa, podendo definir as modificações desejadas, ou selecionar documentos, dados, modelos e outros conhecimentos para serem reutilizados.

**Retenção de casos:** em um sistema RBC a retenção de casos tem um papel fundamental por garantir que a base de casos seja sempre realimentada e atualizada. No caso do Thoth, um benefício esperado com a abordagem é que esse passo seja realizado com menor esforço, pois a fonte das informações analisadas é a base do sistema de *workflow*, a mesma utilizada para o controle das atividades de planejamento e execução.

### 3.2.1 – Similaridade e utilidade dos casos

O aspecto mais importante a ser analisado na utilização da abordagem de RBC é a definição dos critérios utilizados para determinar a similaridade entre casos. A definição de uma abordagem adequada para a definição de similaridade é importante porque, sem ela, a receptividade ao sistema pode ser comprometida, já que o usuário pode não reconhecer a razão clara para a definição de um caso como adequado ao contexto do seu problema. Assim, podemos concluir que o conceito de similaridade está relacionado à utilidade dos casos na solução de um problema.

Segundo WOLFS (1996), é muito difícil definir *a priori* a utilidade dos casos. Tanto a definição de similaridade, quando a determinação de utilidade é intuitiva e sensível ao contexto, dependendo das características do problema a ser resolvido, da área de pesquisa estudada e da percepção do usuário sobre o problema e o caso. Para exemplificar, voltaremos à comparação entre os processos do ambiente empresarial e do ambiente científico. No ambiente empresarial, as instâncias dos processos sofrem poucas variações e estão sempre relacionadas ao mesmo serviço ou produto. Nesses casos, a similaridade poderia ser definida sobre critérios como prazos, custos, períodos do ano etc.

No ambiente científico, a definição de similaridade depende de fatores bem mais variados e subjetivos. Dificilmente dois trabalhos são realizados da mesma forma. Se isso ocorrer, provavelmente será para verificações ou correções de problemas da primeira execução, ou para teste de novas abordagens, técnicas e ferramentas. Entretanto, a percepção sobre o problema já será diferente, sendo necessário redefinir as atividades que serão executadas.

A reutilização de conhecimento desejada no ambiente científico pode ser observada nas perguntas geralmente executadas pelos pesquisadores quando estão trabalhando: “Você sabe que já estudou esse assunto? Quem sabe usar esse *software*? Como se usa esse equipamento? Será que já não fizeram isso antes?”. Essas perguntas, freqüentemente escutadas em um laboratório, sintetizam o desejo de se localizar uma pessoa que passou por uma situação similar à estudada e que possua experiência para ajudar na solução do problema atual.

No ambiente científico, é possível observar que a utilidade está vinculada aos conhecimentos utilizados e produzidos durante as tarefas, ou seja, à área do saber, aos modelos utilizados e aos aspectos semânticos dos dados manipulados, entre outros fatores.

Para que seja possível o reconhecimento destas características, o processo de raciocínio do Thoth é realizado com uma ontologia, utilizada para a classificação e posterior localização dos processos. A ontologia permite a representação dos conceitos existentes em um domínio do conhecimento de uma maneira que pode ser reconhecida pelos especialistas da área (GUARINO, 1998), e também permite que os aspectos semânticos que precisam ser mapeados sejam identificados e analisados de uma forma computacionalmente viável.

### **3.2.2 – Ontologia no Thoth**

A ontologia tem um papel fundamental na arquitetura do Thoth. Ela viabiliza a definição de uma estratégia de definição de similaridade que propicia a compreensão, por parte do usuário, dos aspectos semânticos que levaram à sugestão de um processo.

Dois desafios precisaram ser enfrentados para que fosse possível a utilização da ontologia. A primeira questão se refere à definição de um mecanismo capaz de determinar a similaridade entre os itens classificados pelos termos ontológicos. A segunda questão se refere à necessidade desse mecanismo também permitir o armazenamento e a recuperação dos itens classificados em tempo satisfatório.

Na ontologia, estão definidos os conceitos e os relacionamentos entre eles. A forma como a ontologia é construída depende da aplicação a que ela se destina. A ontologia pode ser construída de duas formas: dependente do domínio ou genérica. Ontologias genéricas são mais generalizadas e têm como proposta a estruturação de diversas categorias. CYC (CYCORP, 2004) e WordNet (PRINCETON UNIVERSITY, 2004) são exemplos de ontologias genéricas. Elas são geralmente

grandes (em relação à quantidade de termos), mas não são muito detalhadas. No entanto, devido ao caráter universal que possuem, são de difícil construção.

Para esta proposta, é mais interessante utilizar ontologias dependentes do domínio. Geralmente menores, elas provêm conceitos em um nível mais específico. Além disso, o conhecimento nelas existente sobre o domínio permite maior exatidão nas inferências e computações realizadas pela aplicação que a utiliza.

Assim, a ontologia no Thoth permite que os processos sejam classificados com um vocabulário próprio ao dos profissionais que utilizarão o sistema, o que facilita a compreensão dos resultados da buscas. A ontologia também potencializa a busca e a definição da similaridade, como será melhor explicado na seção 0.

O Thoth foi desenvolvido a partir dos requisitos obtidos com a observação do trabalho dos alunos, professores e pesquisadores do LABHID. Os estudos desenvolvidos por esses pesquisadores estão relacionados aos fenômenos hidrológicos e sociais das bacias hidrológicas brasileiras (ver capítulo 4). Por isso, foi necessário criar uma ontologia de domínio com conceitos relacionados a essa área.

### **3.2.2.1 – Criação da ontologia**

Segundo NOY & MCGUINNESS (2001), uma ontologia pode ser construída seguindo os passos descritos abaixo:

- Definição das classes da ontologia;
- Organização dessas classes em uma hierarquia taxonômica;
- Definição dos atributos das classes;
- Alimentação das classes com suas respectivas instâncias.

Nesse caso, produto resultante é não só a ontologia, mas também uma base de conhecimento, já que existem instâncias das classes.

No contexto deste trabalho, a criação da ontologia se iniciou com a definição dos conceitos que precisavam ser representados. Para tanto, foi realizado um trabalho de identificação dos aspectos mais representativos do trabalho na área de hidrologia. Essa identificação ocorreu através da observação do dia-a-dia de trabalho dos pesquisadores e das interações realizadas com eles.

O trabalho de um pesquisador na hidrologia geralmente está vinculado à observação dos fenômenos hidrológicos que ocorrem em corpos de água, como os rios e os lagos. Por isso, as pesquisas normalmente são contextualizadas através da bacia



ou sub-bacia hidrográfica da qual este corpo massa faz parte. Logo, esse é um dos aspectos observados e categorizados.

A bacia hidrográfica representa a área total drenada por um rio e seus afluentes (UNIVALE, 2004). Da mesma forma, cada afluente possui sua própria bacia hidrográfica, que é considerada uma sub-bacia da bacia principal. Essa estrutura foi mapeada da maneira apresentada na Figura 8.

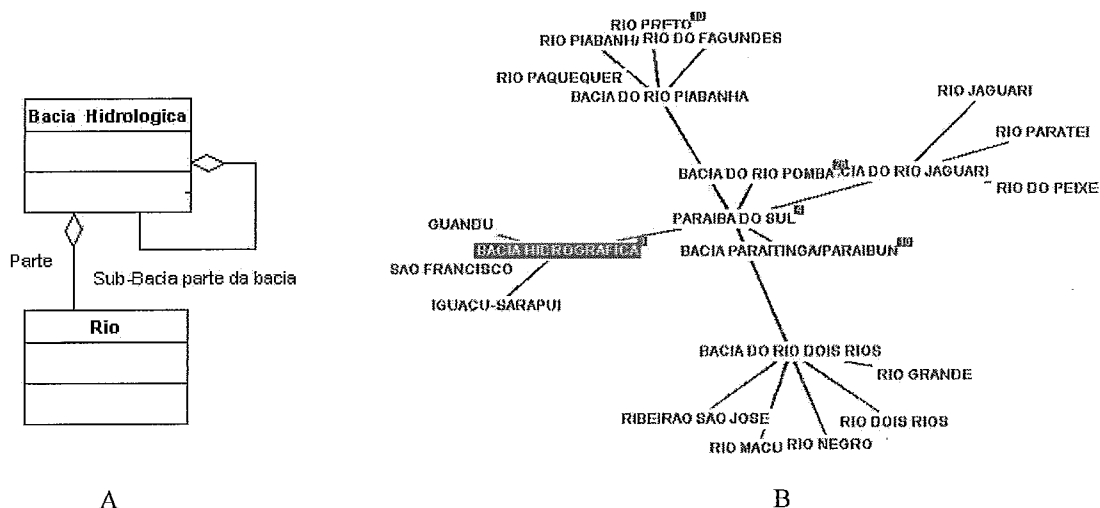


Figura 8 - Estrutura das bacias hidrológicas

Na Figura 8.a, as bacias hidrográficas são organizadas hierarquicamente através do auto-relacionamento *part-of* na classe “Bacia hidrográfica”. Esse relacionamento simboliza a característica das bacias hidrológicas de serem compostas por suas sub-bacias e pelo próprio rio, mapeado através do relacionamento *part-of* entre as duas classes. Na Figura 8.b, é apresentada uma parte da representação dessa hierarquia na ontologia desenvolvida.

Outro item tomado como referência são os modelos utilizados. Os modelos são abstrações simplificadas da realidade, construídos com o intuito de facilitar a resolução dos problemas através da representação das porções importantes daquilo que é modelado e ocultação de detalhes irrelevantes (PEREIRA & BAZZO, 2002).

Na ontologia desenvolvida, a organização do modelo foi realizada obedecendo ao aspecto funcional (ex.: modelos de fluxo, modelos de transporte etc.), como ilustrado na Figura 9.

Outro aspecto privilegiado são os fenômenos analisados na bacia estudada. Exemplos desses fenômenos são a seca, a voçoroca e as inundações, entre outros.

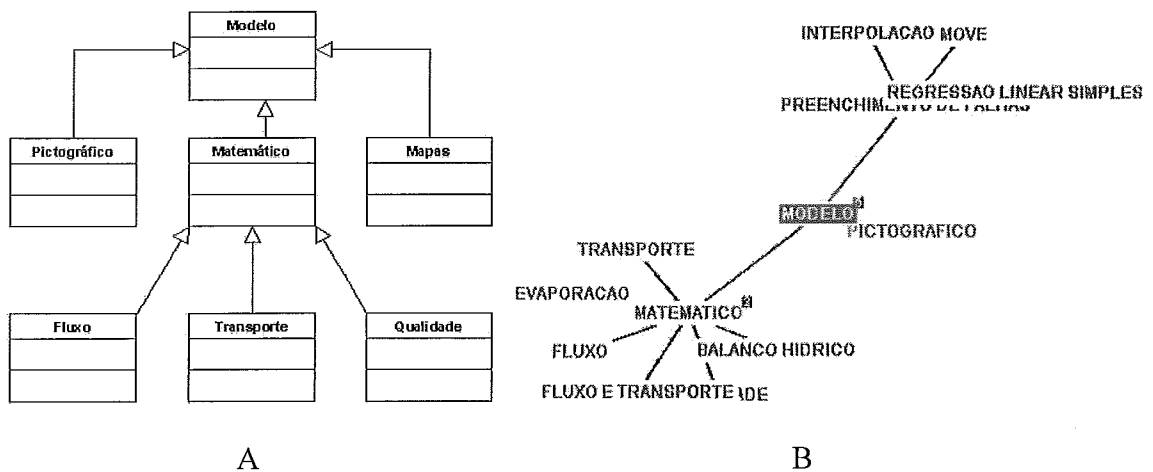


Figura 9 - Estrutura dos modelos

Os conceitos foram coletados primeiramente de outras ontologias relacionadas à área de hidrologia. São exemplos destas a *Ontology for Geographic Information* (ISLUM, BERMUDEZ et al., 2004), que possui informações sobre metadados geográficos; a *Ontology for Surface Water and Water Quality Models* (ISLUM & PIASECKI, 2004), uma ontologia constituída com modelos numéricos para vazão e qualidade da água; e a ontologia com entidades hidrológicas desenvolvida por Luiz Bermudez (BERMUDEZ, 2004). Outra fonte utilizada foi o glossário de termos hidrológicos, disponibilizado pela Agência Nacional de Águas - ANA. Os dados sobre as bacias hidrológicas foram obtidos a partir de informações da própria ANA e de outras instituições, como a SERLA, órgão gestor de recursos hídricos do Rio de Janeiro, e o Instituto Brasileiro de Geografia e Estatística - IBGE. Entretanto, a maior parte da ontologia foi constituída a partir da colaboração com profissionais do LABHID.

Como explicitado anteriormente, ontologias são estruturas flexíveis e extensíveis. O Thoth foi concebido respeitando esse preceito. Dessa forma, é possível realizar alterações na ontologia sem nenhum prejuízo ao funcionamento do sistema. Além disso, é importante observar que o desenvolvimento da ontologia não se limitou aos conceitos relacionados aos três aspectos selecionados para a definição da similaridade. Nela foram inseridos outros conceitos importantes no domínio da engenharia e hidrologia.

A utilização desses conceitos no processo de busca pode ser realizada em uma segunda etapa, quando a aplicabilidade da ferramenta for efetivamente avaliada, assim

como a sua receptividade. De fato, esses aspectos são configurados, e a seleção de um novo aspecto para ser considerado durante a busca corresponde à seleção de um nó na ontologia, que passará a funcionar como raiz de mais uma hierarquia, como ilustrado na , que apresenta a configuração atual, com os aspectos modelo, bacia hidrográfica e fenômenos selecionados.

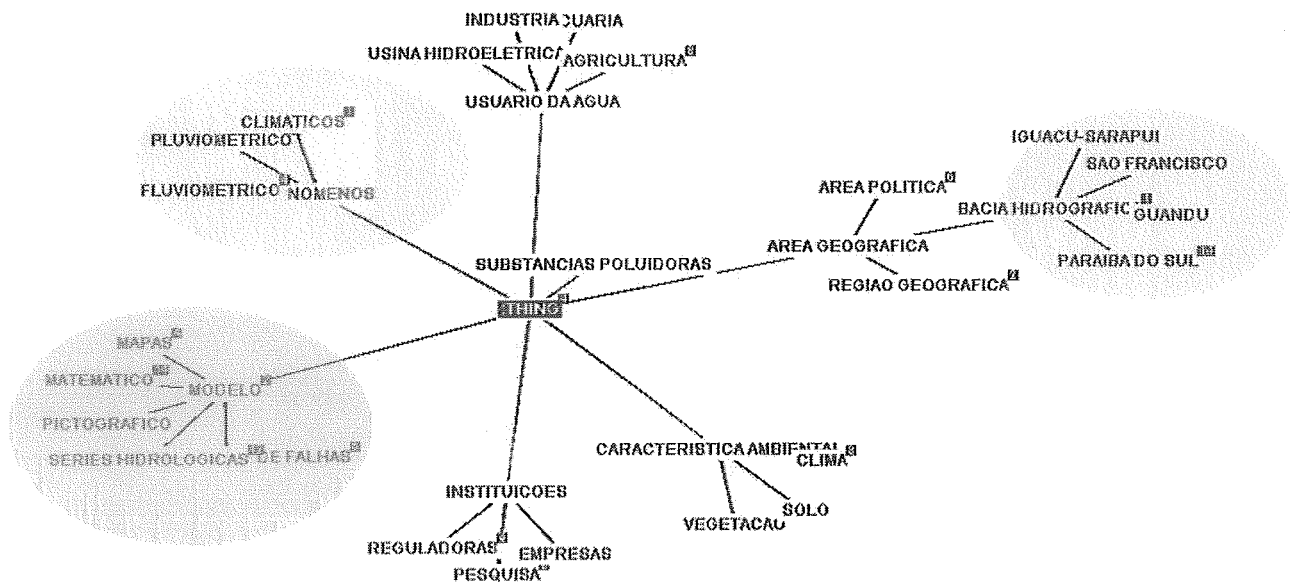


Figura 10 - Ontologia com aspectos atualmente selecionados

### 3.2.3 – Metodologia de definição de similaridade e recuperação de casos

Como apresentado anteriormente, a ontologia desenvolvida foi utilizada como base para a definição de similaridade entre os casos (processos armazenados) e também para a localização e busca desses itens na base. A forma como cada uma dessas atividades é desempenhada no Thoth será explanada nas seções 3.2.3.1 e 3.2.3.2

#### 3.2.3.1 – Definição da similaridade entre os casos

Para definir a abordagem de similaridade, dois fatores foram observados: as características das atividades científicas e o tipo de resultado esperado com a busca.

Inicialmente, se cogitou a utilização de estratégias menos refinadas, inclusive sem a utilização de ontologias. Nesse caso, seria utilizado um processo baseado na igualdade entre os termos. Por exemplo, um processo referente à *Bacia do Rio Piabanha*, só seria considerado similar a outro processo referente à mesma bacia. Essa

estratégia se mostrou inadequada, pois como afirmado anteriormente, o número de casos relacionados a um tópico pode ser bastante restrito.

Com a utilização da ontologia, é possível aumentar a quantidade de itens retornados a partir do momento que se considera, durante a busca, os itens categorizados com conceitos semanticamente próximos. Voltando ao exemplo do parágrafo anterior, a utilização da ontologia tornaria possível a percepção que a *Bacia do Rio Preto* e a *Bacia do Rio dos Fagundes* (ver Figura 8) compartilham propriedades com a *Bacia do Rio Piabanha* por serem componentes da bacia hidrográfica do Paraíba do Sul; conseqüentemente, itens relativos a esses dois rios podem ser retornados durante a busca.

Definido o papel para a ontologia, o próximo passo foi analisar metodologias que auxiliassem na definição da similaridade entre os itens classificados a partir da ontologia. Assim, foram revisadas diversas abordagens para definição de similaridade, dentre elas a apresentada por JIANG & CONRATH (1998), utilizada para medir a similaridade semântica entre palavras e conceitos; a definida em (RODRIGUEZ & EGENHOFER, 2003) e também a desenvolvida por BERGMANN & STAHL (1988).

No trabalho de RODRIGUEZ & EGENHOFER, a similaridade faz parte de uma metodologia para integração de ontologias, onde as classes similares são definidas através da comparação entre os seus nomes (se são sinônimos), da correspondência entre os seus atributos e da localização na ontologia.

Já BERGMANN e STAHL determinam a similaridade a fim de estipular a similaridade entre objetos em uma hierarquia de classes. Embora o trabalho não tenha sido focalizado sobre uma ontologia, é importante observar que o desenvolvimento de uma ontologia utiliza idéias originárias do desenvolvimento dos modelos de objetos, se diferenciando pelo foco de desenvolvimento. Isto é, enquanto os programadores focalizam o aspecto operacional das classes, a construção da ontologia focaliza as propriedades estruturais das classes (NOY & MCGUINNESS, 2001).

Embora o objetivo dos dois trabalhos apresentados seja distinto, as metodologias utilizadas seguem as mesmas prerrogativas. Para ambos, a definição da similaridade entre duas classes ou dois objetos pode ser realizada considerando a distância entre eles em uma hierarquia e também pelos atributos que os qualificam. No entanto, dado o propósito de cada metodologia, tais atributos são explorados de forma distinta. Quando se faz a integração, a verificação realizada tem o objetivo de

identificar a correspondência entre as classes, a partir da correspondência entre os seus atributos nas ontologias. Quando o objetivo é determinar a similaridade entre dois objetos, a análise é realizada sobre os valores dos atributos em cada objeto comparado.

No caso do Thoth, o segundo enfoque é o mais adequado. Assim, a metodologia de BERGMANN e STAHL serviu como base para o desenvolvimento do sistema. Nessa metodologia, a similaridade é definida através de dois passos: a definição das similaridades interclasse e intraclasse.

A similaridade intraclasse é calculada a partir das propriedades comuns dos objetos comparados. Para capturar essas propriedades, a superclasse comum mais específica entre os dois objetos é identificada e a similaridade definida a partir dos atributos dessa classe. Nesse caso, uma função de comparação deve ser definida para cada atributo e a similaridade é obtida através de uma função de agregação aplicada sobre os valores de todas as comparações locais realizadas, como descrito na Equação 1:

$$\text{SIM intra}(q,c) = \Phi(\text{Sim}_{A1}(q.A1,c.A1), \dots, \text{sim}_{An}(q.An,c.An)),$$

**Equação 1 – Similaridade Intra-Classe**

Onde  $\Phi$  representa a função de agregação,  $q.Ai, c.Ai$  denotam os valores do atributo  $Ai$ , nos objetos  $q$  e  $c$ , que estão sendo comparados (se aplica também para consultas e instâncias), e  $\text{Sim}_{Ai}$  representa a similaridade local para o atributo  $Ai$ .

A Equação 2 é um exemplo simples dessas funções. Nesse exemplo, a similaridade seria binária, isto é, os objetos só seriam similares quando os valores comparados fossem os mesmos.

$$\text{Sim}_{A1} = \begin{cases} 1 & q.A1 = c.A1 \\ 0 & q.A1 \neq c.A1 \end{cases}$$

**Equação 2 – função de similaridade local**

A similaridade interclasse tem o objetivo de explorar a semântica contida na estrutura hierárquica das classes. Ela é importante porque evita que valores resultantes da comparação intraclasse sejam mal interpretados. Essa situação ocorre quando uma comparação entre dois objetos de classes distintas obtém um valor máximo de

similaridade devido aos valores dos atributos comuns serem iguais. Nesse caso, a similaridade interclasse certamente atenuará esse valor de similaridade e tornará o resultado mais apropriado.

A similaridade interclasse é definida através da atribuição para cada classe na hierarquia de um valor de similaridade  $S_i \in [0,1]$ , respeitada a condição de que uma classe mais profunda na hierarquia necessariamente possua um valor maior de  $S_i$  do que qualquer uma das suas superclasses.

A similaridade total entre as instâncias comparadas é computada pelo produto da similaridade interclasse pela similaridade intraclasse:

$$\text{sim}(q,c) = \text{SIM intra}(q,c) \cdot \text{SIMinter}(\text{class}(q),\text{class}(c))$$

### Equação 3 – Similaridade Total

onde  $\text{class}(q)$  e  $\text{class}(c)$  denotam as classe de  $q$  e  $c$ , que representam o par objeto-objeto ou objeto-consulta.

A revisão das metodologias supracitadas forneceu subsídios para o desenvolvimento da metodologia utilizada no Thoth. No entanto, outros fatores avaliados determinaram a simplificação da estratégia utilizada.

O primeiro fator diz respeito à dificuldade na definição da similaridade interclasse. Assim como a própria construção da ontologia, a definição e a compreensão das propriedades de cada classes não é simples, mesmo para especialistas da área. Essa dificuldade também ocorre no momento da definição da forma ou função de comparação entre os atributos das classes.

Além disso, a forma de definição de similaridade tem de ser harmônica com a estratégia de recuperação e busca utilizada (que será explicada na próxima seção), prestigiando também a performance satisfatória das buscas. É importante observar que explorar exhaustivamente toda a hierarquia de classes, definindo as duas similaridades para todos os casos armazenados, pode tornar o método ineficiente.

Por isso, a estratégia de busca teve de ser modificada, dispensando a **correção** e a **completeza** do método de recuperação. Esses conceitos são introduzidos em (VON WANGENHEIM & VON WANGENHEIM, 2003a). O método de recuperação é considerado correto, se a relação de similaridade definida pelo método entre o caso e o problema atual também existe no conceito de similaridade implementado. Por sua vez, ele é considerado completo, se toda a relação de similaridade representada no modelo do sistema se encontra no método de recuperação.

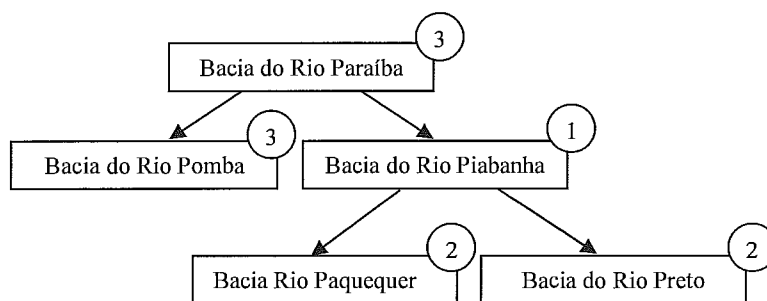
Exemplificando, uma recuperação que analisasse todos os casos e também privilegiasse toda a definição de similaridade (no nosso caso, tanto a similaridade interclasse quando a similaridade entre os atributos) seria completa e correta. No entanto, ainda segundo (VON WANGENHEIM & VON WANGENHEIM, 2003a), a prova da correção e da completeza podem ser dispensadas em prol da eficiência do método. Assim, na abordagem utilizada, foi privilegiada a percepção cognitiva associada às hierarquias dos modelos, fenômenos e bacias hidrográficas.

Dada essa explicação, é possível demonstrar como essas hierarquias são exploradas para a definição de similaridade. No caso dos modelos e fenômenos, a própria hierarquia da estrutura de classes é explorada. No entanto, a hierarquia das bacias hidrográficas é obtida a partir do auto-relacionamento apresentado na Figura 8.

Os passos realizados para a recuperação dos processos são descritos abaixo e podem ser visualizados na Figura 11. A execução é limitada por um parâmetro que indica a quantidade de casos que devem ser retornados em cada busca:

- O primeiro passo é localizar os casos classificados com os mesmos conceitos fornecidos na consulta. Supondo que a consulta tenha sido realizada com o conceito “Bacia do rio Piabanha”, este seria o conceito utilizado na primeira busca. Se a quantidade de casos for suficiente, o processamento termina e os casos localizados são retornados ao usuário.
- Caso a quantidade de casos não seja suficiente, uma nova busca é realizada a fim de se localizar os casos classificados com os conceitos filhos, sinalizados na figura com o número 2.
- Caso a quantidade ainda não seja suficiente, são buscados os casos classificados com o conceito pai e com os conceitos irmãos, assinalados com o número 3 na figura.

Nesse ponto, os casos localizados são apresentados para o usuário e o procedimento é finalizado, mesmo se a quantidade de casos retornados for inferior ao parâmetro da quantidade de retorno de casos desejada. Os processos são ordenados de acordo com o passo em que foram encontrados.



**Figura 11 - Representação dos passos para escolha de conceitos similares**

A busca foi implementada tendo em vista uma posterior utilização da similaridade intraclasse para a ordenação dentro de cada passo ou eliminação de casos. De fato, que as duas similaridades podem ser computadas de forma independentes e depois integradas. No caso específico da hidrologia, os atributos hidromorfológicos de uma bacia ou de um rio, como vazões máxima e mínima, índice de salinidade e a profundidade máxima, podem ser considerados durante a busca. No entanto, critérios tão específicos que só poderiam ser explorados no nível da similaridade intraclasse, não foram avaliados como críticos para a aceitabilidade do sistema em sua fase inicial.

### **3.2.3.2 – Estrutura física e recuperação dos processos**

Como afirmado na seção anterior, uma preocupação no desenvolvimento do sistema é a definição de uma estratégia de similaridade eficiente, mesmo com o crescimento contínuo da base de casos. Essa preocupação permeou a definição da estrutura física do armazenamento, que deve ser adequada para utilização com a metodologia de similaridade adotada.

Na definição dessa estrutura, alguns aspectos precisaram ser analisados. Primeiro, os casos são os processos ou atividades armazenados na base do sistema de *workflow* e a estrutura utilizada para identificar os casos tem de ser suficientemente independente para não necessitar da alteração da estrutura desse sistema. Além disso, com o crescimento das bases, torna-se cada vez maior o esforço para se acessar os casos e analisá-los durante a busca. Finalmente, existe a necessidade de se categorizar os processos com a ontologia, o que certamente não pode ser realizado no momento em que a busca é realizada.



Esses três fatores determinaram a escolha de uma estrutura indexada de armazenamento e busca. Tal estrutura permite que todo o processo de categorização dos processos seja realizado previamente e que, no momento da busca, apenas os índices sejam consultados. Dessa forma, também é possível se obter um grau adequado de autonomia em relação à estrutura do sistema de *workflow* utilizado, já que os índices podem ser mantidos em um local distinto ao das informações indexadas, o que propicia também um menor esforço na execução da consulta. Outro benefício reside na possibilidade de se realizar cálculos de similaridade com níveis maiores de complexidade.

Definido o tipo de estrutura utilizada, o próximo passo foi escolher o modelo de indexação a ser utilizado, isto é, a forma como os indexadores devem ser gerados, armazenados e consultados.

Como explanado em (SALTON & MCGILL, 1983), um processo de busca utilizando a indexação começa com a classificação dos itens indexados, isto é, a categorização desses itens com os índices, de acordo com a regra de indexação utilizada. Após a classificação, os índices são armazenados e finalmente são consultados, como demonstrado na Figura 12.

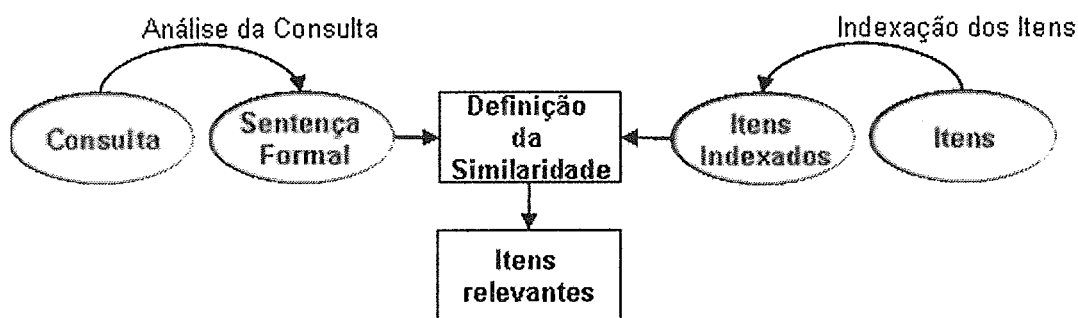


Figura 12 - Processo de classificação (SALTON & MCGILL, 1983)

Neste trabalho, a indexação é realizada a partir da ontologia. Assim, processos e atividades são categorizados com um ou mais conceitos ontológicos, como será descrito a seguir.

### 3.2.3.3 – Etapa de Classificação

A primeira etapa do processo de indexação é a realização da classificação dos itens, no caso do Thoth, os processos armazenados no sistema de *workflow*. Normalmente, estratégias de indexação são utilizadas para itens isolados, como texto,

áudio ou imagens, o que não é o caso desse sistema. Por isso, foi preciso avaliar a estrutura dos processos para se definir a maneira como os mesmos poderiam ser classificados.

O propósito da reutilização aqui utilizada não se restringe apenas à estrutura do processo em si, mas primordialmente aos assuntos que ele se relaciona. Dessa maneira, a classificação executada tem como função permitir a localização desses assuntos, sendo necessário definir quais são os componentes do processo que podem ser utilizados para esse propósito.

Os componentes básicos de um processo são ilustrados na Figura 13, que representa o meta-modelo proposto pela *WfMC* (HOLLNGSWORTH, 1995). Desses componentes, os que efetivamente fornecem indícios dos assuntos tratados são as pessoas e grupos que desempenham as atividades (*role*), os dados utilizados (*workflow relevant data*) e os aplicativos utilizados (*invoked applications*).

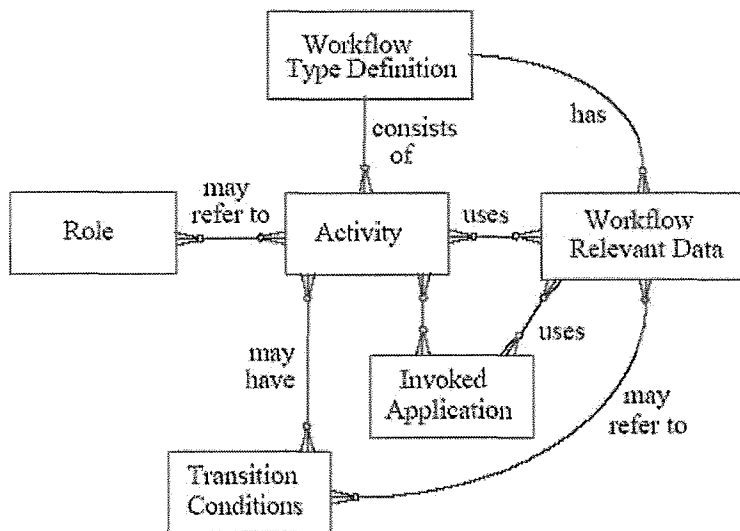


Figura 13 - Meta-modelo da definição de processos (HOLLNGSWORTH, 1995).

No protótipo atual, a classificação é realizada apenas sobre os dados utilizados. Essa escolha foi feita porque a utilização de pessoas e grupos demandaria a consulta ao perfil de ambos, o que, atualmente, não é possível. Além disso, considerou-se que a construção desse perfil seria realizada a partir dos dados manuseados pelos executores durante a execução das suas atividades. Por outro lado, a utilização de um aplicativo pode fornecer indícios de modelos que são utilizados ou outros atributos mapeados na ontologia. No entanto, essa classificação é bastante complexa, dependendo do aplicativo utilizado e do grau de automação da ferramenta de *workflow*, sendo também descartada no primeiro protótipo.

Os dados correspondem a quaisquer informações utilizadas durante a realização das atividades. No caso do LABHID, por exemplo, os dados que são utilizados durante um trabalho são de vários tipos, dentre eles planilha de dados geográficos, documentos de texto com prospectos de uma região, mapas e dados resultantes da execução de modelos.

A classificação ocorre no momento em que os processos são criados e/ou salvos no sistema de *workflow*. Três tipos de classificação são possíveis: a manual, a semi-automatizada e automatizada. Na manual, o usuário define quais conceitos são relacionados aos artefatos. A abordagem semi-automatizada pode ser utilizada a partir do momento que existam analisadores automáticos disponíveis para algum tipo de dado. Por exemplo, documentos de texto podem ser analisados por processadores de texto e o resultado disponibilizado para o usuário, que definiria quais conceitos poderiam efetivamente ser aplicados ao texto.

A diferença entre a utilização de estratégias automáticas e semi-automatizadas está na exatidão obtida pela ferramenta que executa a extração de termos. Por exemplo, dados armazenados em tabelas de banco de dados possuem estruturas conhecidas e é provável que os conceitos possam ser facilmente identificados. Assim, é possível concluir que dados bem estruturados são fortes candidatos para utilização de estratégias automáticas de classificação.

O protótipo atual não conta com nenhum dispositivo semi-automático ou automático de classificação. Entretanto, algumas ferramentas de processamento de texto, tais como a Judge (LIMI, RUNYAN et al., 2005) e a ConceptNet (LIU & SINGH, 2004), foram avaliadas em cooperação com outra tese em fase de desenvolvimento (MORENO, VIVACQUA et al., 2003) e podem ser utilizadas futuramente.

A associação dos conceitos com processos e atividades ocorre de forma simples, respeitando a estrutura do meta-modelo (Figura 13). Após a identificação dos conceitos associados aos artefatos utilizados em cada atividade, esses mesmos conceitos são automaticamente relacionados às atividades nas quais esses artefatos são utilizados. Por sua vez, o processo do qual a atividade faz parte também é classificado com esses conceitos.

### 3.2.3.4 – Armazenamento dos índices e realização das busca

Para a definição da forma de armazenamento dos índices, assim como da forma de consulta sobre eles, foi utilizado o modelo de busca proposto por KHAN(2000). O propósito original do modelo é viabilizar a busca de áudio, fazendo uso de ontologias para alcançar esse objetivo.

Embora o modelo tenha sido inicialmente aplicado para áudio, segundo o próprio autor, a sua aplicabilidade é extensível a outros tipos de informações e se mostra adequado para a indexação também dos processos.

O modelo foi concebido para utilização sobre um banco de dados relacional e são necessárias duas tabelas. A primeira tabela é a de ligação entre a ontologia e o objeto classificado, e armazena o conjunto de índices e o identificador desse objeto. A outra tabela se refere às informações do próprio objeto indexado, como apresentado na Figura 14. Assim, após a etapa de classificação, os conceitos ontológicos são armazenados em tuplas junto com um identificador do processo ou da atividade que foi classificado.

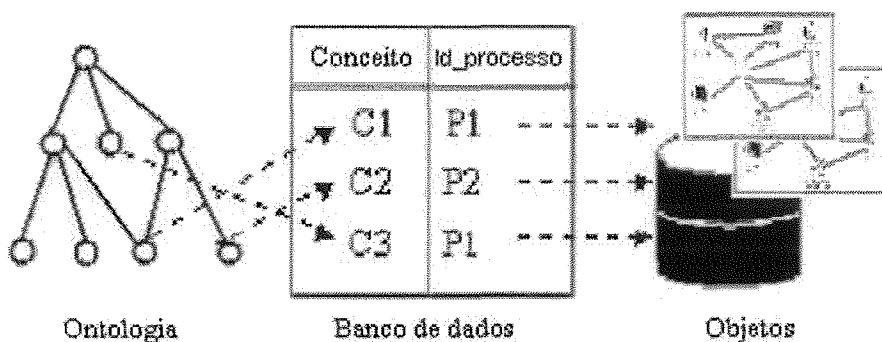


Figura 14 - Modelo de busca baseado em ontologia (Adaptado (KHAN, 2000))

No momento em que uma busca é requisitada, é executada uma consulta SQL sobre a tabela de índices. A consulta é gerada automaticamente a partir dos critérios fornecidos na requisição de busca. Por critérios, entenda-se um conjunto de conceitos ontológicos que, direta ou indiretamente, são extraídos da busca. A consulta SQL é gerada a partir da estratégia de similaridade adotada, que foi explicada anteriormente. Nesse ponto, é seguindo um modelo booleano de busca.

Como explicado anteriormente, os primeiros processos que são retornados são aqueles classificados apenas com os conceitos fornecidos na busca. Para exemplificar, suponhamos que um usuário trabalha em um processo relacionado com a bacia do rio Paraíba do Sul e esteja utilizando o modelo QTEC, de transporte de sedimentos em rios. Neste caso, seu trabalho seria identificado como o conceito “BACIA DO

PARAIBA DO SUL”, da hierarquia de bacias hidrológicas e “QTEC”, uma instância dos modelos de transporte. Para este caso, inicialmente só estes dois conceitos serão explorados e a consulta gerada é similar a demonstrada na

```
select m1.id_processo
from meta_table m1, meta_table m2
where p.Id = m1.Id and p.Id = m2.Id
and m1.TERM_LABEL = "PARAIBA DO SUL"
and m2.TERM_LABEL = "OTEQ"
```

**Figura 15 - Consulta SQL gerada com termos originais da busca**

Essa consulta corresponde ao primeiro passo da definição de similaridade, explicado anteriormente. Caso não sejam identificados processos suficientes, os conceitos considerados similares são adicionados, de forma que a consulta SQL realizada fica similar a demonstrada na Figura 16.

```
Select m1.id_processo
from meta_table m1, meta_table m2
where p.Id = m1.Id and p.Id = m2.Id
and (m1.TERM_LABEL = "PARAIBA DO SUL" or m1.TERM_LABEL = "
    BACIA DO RIO JAGUARI" or .. m1.TERM_LABEL = " BACIA DO RIO
    PIABANHA" )
and (m2.TERM_LABEL = "OTEQ" or m2.TERM.LABEL = "OTIS" OR ... OR
    m2.TERM.LABEL = "WASP5" )
```

**Figura 16 – Consulta SQL gerada com termos considerados similares**

Caso necessário, o terceiro passo é executado, com mais termos adicionados. Existe a possibilidade do crescimento excessivo do número de termos nas consulta. Em um primeiro momento, não foi desenvolvido nenhum dispositivo para evitar tal comportamento, sendo recomendado aos usuários que utilizassem conceitos mais específicos para evitá-lo.

Também é importante ressaltar que, como resultado da estratégia de classificação utilizada, a estrutura de índices é gerada não só para os processos, mas também para as atividades e para os artefatos que compõem os processos. Dessa forma, são utilizadas três tabelas de indexação / classificação. Uma para os processos, outra para atividades e outras para artefatos manipulados.

### 3.2.4 – Módulos do Thoth

Explicado o modo como ocorre a definição de similaridade e como a busca é realizada, esta seção apresentará a arquitetura do Thoth, isto é, seus módulos e a forma como se dá a interação com o ambiente.

Uma das características desejadas no Thoth é o isolamento com os outros sistemas utilizados, como, o sistema de *workflow*. Esse fator deve ser considerado por causa da possibilidade, bastante provável, de não ser possível acessar diretamente a base de dados desses sistemas.

Por isso, o Thoth fornece interfaces para que o usuário visualize os processos buscados. Por outro lado, ele interage com o sistema de *workflow* no momento em que precisa das informações para a realização das buscas, mas essas operações são realizadas de forma a não interferir na comunicação *workflow* – usuário e o funcionamento do sistema.

Estruturalmente, o Thoth foi concebido como um conjunto de módulos organizados em três grupos funcionais: as interfaces com o usuário, as interfaces com os sistemas externos e os componentes de busca e recuperação, como ilustrado na Figura 17.

As interfaces com os usuários são o meio de comunicação entre os usuários e o sistema onde são apresentadas as informações buscadas. O grupo de busca e classificação consiste no principal elemento do sistema e realiza a indexação e execução das consultas, explicadas anteriormente. Finalmente, as interfaces com os sistemas externos realizam a consulta aos dados dos processos e dados oriundos de outros sistemas, como gerenciadores de documentos e gerenciadores de dados espaciais.

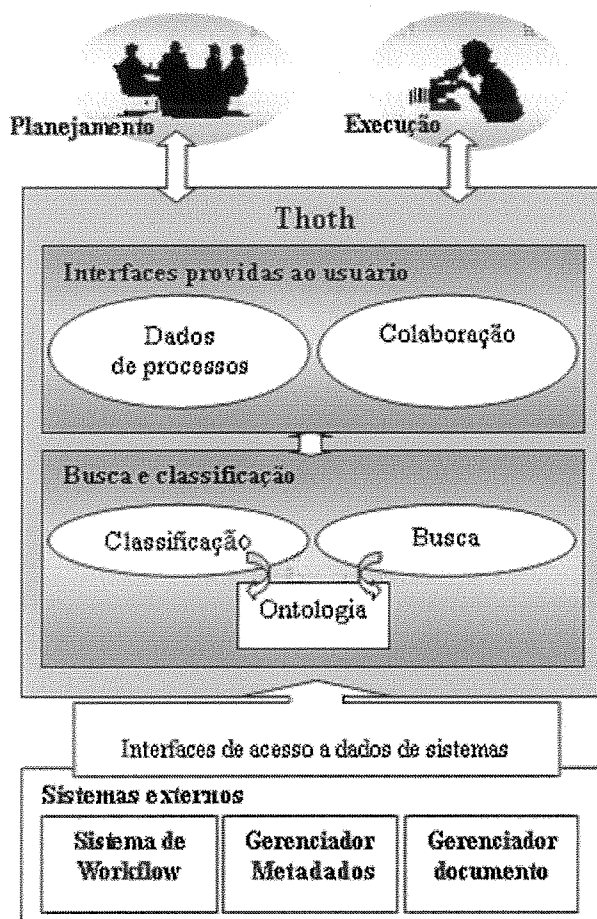


Figura 17 - Arquitetura do Thoth

### 3.2.4.1 – Interfaces com os usuários

As interfaces com o usuário representam os meios pelos quais as informações dos processos recuperados são visualizadas. O Thoth foi planejado para ajudar o usuário nos dois momentos do ciclo de vida das atividades: o planejamento e a execução.

O apoio a ambas as fases é necessário porque os processos científicos tendem a ser sempre diferentes, ou seja, mesmo procedimentos bastante comuns, como a análise da qualidade da água do trecho de um rio, são executados de forma distinta, de acordo com as características geológicas do rio e da disponibilidade de dados da região. Nesse caso, torna-se útil ao cientista, durante o planejamento, saber detalhes de trabalhos que outrora foram realizados sobre condições similares a sua. Por outro lado, os parâmetros de configuração de modelos podem ser reaproveitados das execuções prévias, o que pode eliminar o desperdício de tempo nas tentativas de ajustes em execuções subsequentes.

Em termos práticos, o Thoth é um dispositivo de busca diferenciando-se pelo tipo de objeto que é localizado. Enquanto os dispositivos de busca são normalmente utilizados para itens como documentos, imagens e metadados, o Thoth localiza processos. Essa localização é complexa, pois um processo é composto por diversos itens que podem influenciar na percepção de utilidade de um caso para cada usuário.

Assim, quando um dispositivo de busca de documentos ou imagens é projetado, as interfaces têm como único propósito disponibilizar o item que foi buscado. No caso de um sistema RBC, a interface tem um papel importante no reconhecimento, por parte do usuário, da utilidade do processo. No entanto, este é caracterizado por uma infinidade de informações, sendo necessário definir quais destas realmente precisam ser apresentadas para o usuário.

No caso do Thoth, seria necessário definir quais das informações do processo precisariam ser apresentadas para o usuário. No entanto, tal escolha depende basicamente do propósito de utilização da ferramenta, que pode ser diferente de organização para organização. Por isso, as interfaces aqui apresentadas podem ser modificadas e novas podem ser construídas, de acordo com a identificação de novas demandas.

#### ***3.2.4.1.1 – Consulta de processos***

As consultas constituem a forma básica de solicitar a recuperação de casos (processos) similares. Elas podem ser executadas de duas formas.

No primeiro modo, o usuário configura os parâmetros a serem utilizados, isto é, define os conceitos ontológicos relacionados com os processos que deseja localizar, como ilustrado na Figura 18.

O segundo tipo de consulta pode ser realizado quando o processo (ou a atividade) em que se está trabalhando já foi classificada. Neste caso é possível buscar os processos similares de acordo com as características deste processo.

Embora atualmente o resultado das duas formas de consulta seja apresentado da mesma maneira, a última pode servir para a construção de interfaces mais sofisticadas, como dispositivos de lições aprendidas, dispositivos de ajuda, sensíveis ao contexto e acionados automaticamente.



**Busca de Processos:**

Selecione para cada um dos aspectos um ou mais conceitos que se relacionam, com a atividade que está sendo desempenhada, caso nenhum conceito relacionado a um aspecto seja selecionado, este aspecto será desprezado durante a busca.

**Conceitos relacionados**

- BACIA HIDROGRAFICA
- GUANDU
- IGUAÇU-SARAPUI
- PARAIBA DO SUL
- SAO FRANCISCO
- FENOMENOS
- MODELO

**Conceitos selecionados**

PARAIBA DO SUL

Nome do processo	Chadui	Data de criação
Atividades realizadas na fase inicial da tese (Panorama geral da bacia do Paraíba)	Augusto Getirana	2004-06-10
Definição do perfil de usuário da água no rio Jaguarí	Paulo Cameiro Magalhães	2004-09-05
Condições sócio-econômicas da população ribeirinha do Rio Carangola	Jander Duarte Campos	2004-10-22
Diagnóstico da poluição no Rio dos Peixes	Paulo Cameiro Magalhães	2005-01-01

Figura 18 – Interface para realização de busca

### 3.2.4.1.2 – Interfaces de visualização das informações de processos

As interfaces de visualização permitem ao usuário visualizar as informações pertinentes dos processos consultados. As informações que devem ser disponibilizadas dependem das necessidades da organização em que o sistema venha a ser utilizado. No entanto, os elementos mais importantes dos processos científicos são os dados e conhecimentos manuseados, sendo estes o seu principal insumo e produto. Além disso, é natural colaborar em atividades científicas, portanto, é importante saber quem participou de uma atividade. Baseado nesses preceitos, as interfaces primárias desenvolvidas focalizam o acesso aos dados manuseados nas atividades e informações sobre os seus participantes.

Atributos do processo		Fase Inicial da Tese (Panorama geral da bacia do Paraíba)	
<b>Artefatos utilizados</b> Família dos dados de vazão da bacia Dados censitários fornecidos pelo IBGE Mapas da região Texto da tese Observações realizadas a partir da visita		<b>1</b> Descrição: Estudo realizado sobre a problemática de cobrança da água no rio Paraíba do Sul; necessária a definição de vanos usados para cada usuário de	
<b>Pesquisadores participantes</b> Benedetto Ferreira de Oliveira Paulo Canedo Magalhães José Paulo Soares de Azevedo			
<b>Outros processos similares</b> Gestão ambiental da bacia hidrográfica Definição do perfil de usuário da água no rio Paraíba do Sul Condições socio-econômicas da população Diagnóstico da poluição no Rio dos Peixes		<b>2</b>	
<b>Dados do artefato</b> Nome: Planilha dos dados de vazão da bacia		<b>Dados da atividade</b> Nome: Gestão de modelos hidrodinâmicos	
<b>5</b> Descrição: Dados da vazão da bacia, planilhas conseguidas na SEPLA e no sistema de informação desenvolvido pelo Laboratório até o ano 2000		<b>3</b> Descrição: Trabalho de estruturação dos dados relacionados à vazão, ao perfil do rio, cálculo do balanço hídrico e da demanda necessária para formação de um parecer mais adequado para a	
Baixar o arquivo		Ver mais detalhes da atividade	

Figura 19 - Visualização de informações de um processo consultado

A Figura 19 ilustra a principal interface de visualização do Thoth. Nessa interface, é possível visualizar o fluxo de atividade e as informações sobre dados e artefatos manuseados, bem como sobre os participantes das atividades. As informações estão dispostas em quadros que serão explicados abaixo.

No quadro 1, estão os **dados do processo**, ou seja, o nome, a descrição, a data de criação e o pesquisador responsável pelo processo selecionado.

No quadro 2, se visualiza o **fluxo de atividades**, isto é, o diagrama representando o modelo do processo selecionado. O fluxo permite o acesso às informações de cada uma das atividades, que são detalhadas no quadro 3.

No quadro 3, ocorre o **detalhamento de atividades**, onde as informações detalhadas das atividades podem ser verificadas.

No quadro 4, ficam listados os **artefatos do processo**. Por artefato compreendemos qualquer documento, mapa, planilha, tabela, entre outros tipos de insumos, manuseados durante a realização do processo. A seleção de um dos artefatos faz com que o mesmo seja detalhado no quadro 5.

A partir do **detalhamento do artefato**, é possível visualizar as informações cadastradas do artefato selecionado e também efetuar o *download* do mesmo, caso ele esteja disponibilizado e possa ser utilizado.

O quadro 6 apresenta os **participantes do processo**. Através das funcionalidades providas pelo módulo de colaboração é possível se comunicar com esses participantes.

Caso as informações coletadas em um processo não sejam adequadas, o usuário pode visualizar **outros processos similares**, que são os itens listados no quadro 7.

#### **3.2.4.1.3 – Módulo de colaboração**

Um modelo de processo é uma abstração da forma como o processo efetivamente ocorre no mundo real. Dessa forma, muitos detalhes e fatores que influenciam o processo e sua forma de execução não são guardados. Mesmo sem esses detalhes, um profissional que participou do mesmo terá maior possibilidade de compreendê-lo, porque tem ciência dos detalhes omitidos.

O mesmo comportamento não pode ser esperado daqueles que não participaram da execução das atividades. Esses podem ter diversas dúvidas e, conseqüentemente, a compreensão do processo fica comprometida por não saberem das informações que influenciaram a evolução do processo, mas foram omitidas no histórico do mesmo.

O módulo de colaboração tem o objetivo de amenizar essa situação, ao permitir que os usuários interajam com os participantes dos processos consultados. Assim, espera-se que as dúvidas sejam elucidadas, propiciando uma maior compreensão, por parte do usuário, daquilo que consultou (BOMFIM, CASTRO et al., 2003).

O usuário pode se comunicar com os participantes através de dispositivos de mensagens instantâneas. No entanto, tal recurso só pode ser utilizado quando os dois pesquisadores estão utilizando tal dispositivo. Quando o participante não está disponível, a comunicação pode ser iniciada através de um *e-mail*.

#### **3.2.4.2 – Busca e classificação**

Os módulos deste grupo têm a função de realizar a classificação, indexação e busca dos processos e atividades. Essas atividades foram explicadas anteriormente na seção 0.

### 3.2.4.2.1 – Módulo de classificação

O módulo de classificação é responsável pela definição dos conceitos que devem ser atribuídos a artefatos, atividades e processos.

A utilização de uma abordagem baseada em índices propicia diversos benefícios. Como o processo de classificação é realizado no momento em que os processos são salvos, essa abordagem permite que a classificação seja realizada sem maiores preocupações com a performance, já que os índices dos processos não precisam ser definidos no momento em que o usuário realizar uma consulta. Além disso, esse processamento pode ser realizado em *segundo plano*.

Desta maneira, na classificação podem ser utilizados métodos com diferentes graus de complexidade e que demandem grande tempo de processamento. No caso do Thoth, isso é importante porque novas regras de classificação podem ser adicionadas a partir do momento que são descobertas.

Uma característica importante dessa etapa é que os processos podem ser classificados com qualquer conceito da ontologia, não se restringindo aos aspectos considerados na busca (bacias hidrológicas, modelos utilizados e fenômenos). Isso auxilia a parametrização da busca ao permitir que esses aspectos possam ser modificados sem grandes impactos no sistema.

Tal característica é importante, porque, como foi assinalado anteriormente, o enfoque na utilização do sistema pode mudar, surgindo novos aspectos a serem explorados, assim como a própria ontologia pode ser modificada. Nesses casos, o administrador do sistema pode realizar a adequação dos aspectos através da navegação e seleção desse aspecto na ontologia, como demonstrado na Figura 21. A realização dessa operação é equivalente à seleção de outra hierarquia, como demonstrado anteriormente (ver ).

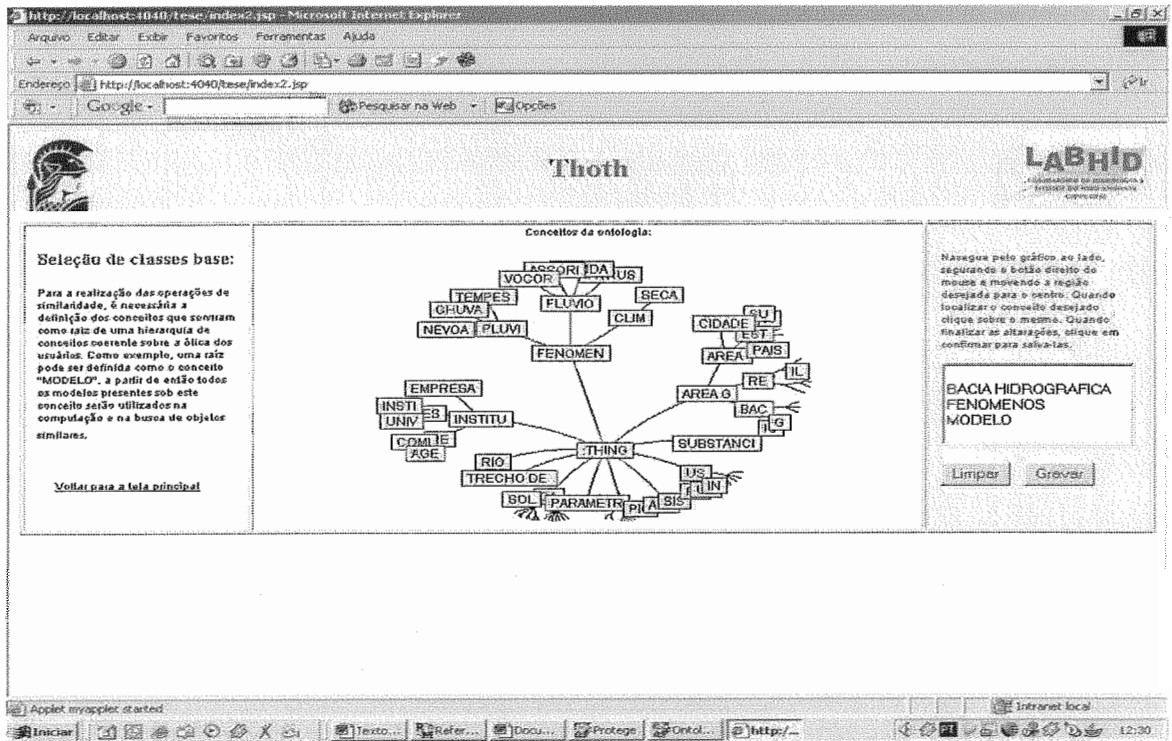


Figura 20 - Tela para escolha de conceitos (aspectos de busca)

### 3.2.4.2.2 – Modulo de busca

O módulo de busca é responsável pela recepção da consulta realizada pelo usuário e pela busca dos processos similares, sendo também o que realiza a definição de similaridade entre os processos.

Ao contrário do que ocorre na classificação, todo o procedimento de busca é realizado no momento em que a consulta é executada. Dessa forma, o método de definição de similaridade deve ter sua complexidade equilibrada com a eficiência dessa operação.

### 3.2.4.3 – Interfaces com sistemas externos

As interfaces com os sistemas externos são responsáveis pela comunicação com os sistemas que controlam as informações utilizadas no processo e que precisam ser avaliadas durante a busca ou apresentadas ao usuário.

O desenvolvimento dessas interfaces facilita a utilização do Thoth em outras organizações e a execução de manutenções, pois as modificações nos sistemas externos acarretam modificações apenas nessas interfaces.

A principal interface desenvolvida é a responsável pela comunicação com o sistema de *workflow*. As outras interfaces podem ser necessárias porque os artefatos

utilizados nos processos geralmente não são controlados pelo sistema de *workflow*. Destacam-se nessa classe os sistemas gerenciadores de documentos, sistemas gerenciadores de metadados científicos e bancos de dados convencionais (SGBDs).

### 3.3 – Implementação do protótipo

Embora não possa ser tratado necessariamente como um requisito, a implementação do sistema está voltada para a internet e intranet. Essa escolha foi feita a partir da observação de que é comum os pesquisadores trabalharem em lugares diversos e não só em sua mesa de trabalho, sendo útil o acesso a todas as informações relevantes em qualquer lugar em que se esteja.

Em termos gerais, a plataforma de implementação do Thoth é composta pelo sistema gerenciador de banco de dados, Microsoft SQL Server 2000 (MICROSOFT, 2003), com uso da linguagem padrão *Structured Query Language* (SQL ) e da linguagem de programação Java (SUN, 2003b). O ambiente utilizado para desenvolvimento foi o Microsoft Windows 2000 Server (MICROSOFT, 2001) e o servidor de *internet* Apache Tomcat (APACHE SOFTWARE FOUNDATION, 2004). Para geração de páginas *web* com conteúdo dinâmico no servidor, a tecnologia escolhida foi *Java Server Pages* (JSP) (SUN, 2003c), por sua total integração com a linguagem Java. Além disso, para a operação com a ontologia foi utilizado o software Protégé (STANDFORD UNIVERSITY, 2003).

Como assinalado anteriormente, uma das premissas no desenvolvimento do sistema é a fácil utilização em diversas configurações de ambientes. Esse objetivo também permeou a implementação do protótipo, realizada sempre com a tecnologia Java, que permite que o sistema possa ser utilizado no Linux, por exemplo. Dos produtos utilizados, o banco de dados SQL Server é o único que só opera no ambiente Windows. No entanto, só foram utilizados comandos SQL padrões, o que torna possível a migração para outros bancos de dados, como, por exemplo, PostgreSQL (POSTGRESQL, 2003) ou Firebird (FIREBIRDSQL FOUNDATION, 2003), que são bancos de dados gratuitos. A seguir, alguns detalhes da implementação de cada um dos módulos do Thoth serão descritos.

#### 3.3.1 – Implementação das interfaces com usuário

Seguindo a premissa de que o sistema deveria ser para utilização na intranet/internet, as interfaces de consulta e visualização de processos foram

desenvolvidas como páginas *web*, utilizando a tecnologia Java /JSP. A escolha da linguagem Java se deve principalmente a necessidades do módulo de busca e recuperação, que serão explicadas a frente.

Para que fosse possível a visualização do diagrama de atividades, foi necessária a utilização de Applets (SUN, 2003a) de visualização de fluxos, desenvolvida pela empresa Lassale (2004).

### 3.3.2 – Colaboração

O módulo de colaboração conta com duas funcionalidades básicas, o envio de *e-mails* e o dispositivo de mensagens instantâneas. Para permitir o envio do *e-mail* a partir das páginas JSP, foi utilizada a API JavaMail (SUN, 2004).

Já os dispositivos de mensagens instantâneas são constituídos de dois elementos: o servidor e o cliente. Cada servidor possui seu próprio protocolo, que pode ser compatível com diversas ferramentas clientes distintas, conforme ilustra a Figura 21.

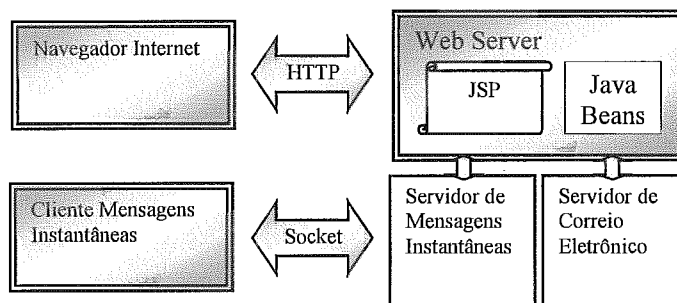


Figura 21 - Arquitetura do módulo de colaboração

A escolha das ferramentas utilizadas foi pautada pela necessidade de se acessar as informações de disponibilidade de cada pessoa nela cadastrada. Além disso, a primeira mensagem enviada a uma pessoa consultada pode ser enviada a partir da tela de visualização de processos. Dessa forma, as ferramentas devem ser integradas com as outras tecnologias utilizadas no desenvolvimento, isto é, o Java e o JSP. Conseqüentemente, optou-se pela utilização do servidor de mensagens instantâneas Jabber (JABBER INC, 2004) e pela utilização do cliente Exodus (MILLARD, 2004). Também foi necessária a utilização do Smack (TUCKER & DOMBIK, 2004), uma API utilizada para a integração entre as páginas JSP e o servidor de mensagens instantâneas.

### 3.3.3 – Busca e classificação

Diversas tecnologias distintas foram integradas para a construção dos componentes responsáveis pela busca e classificação. A classificação, atualmente realizada de forma predominantemente manual, é feita pelo usuário que categoriza seus artefatos com os termos da ontologia, utilizando para isso páginas desenvolvidas em JSP.

Como afirmado anteriormente, os termos utilizados para classificar os artefatos são automaticamente associados às atividades e processos em que tal artefato é usado. As associações entre os conceitos e esses itens persistem no banco de dados escolhido, no caso, o Microsoft SQL Server (MICROSOFT, 2004). Assim, *triggers* desenvolvidos sobre o banco propagam os termos dos artefatos para atividades e processos.

Adicionalmente, quando as interfaces foram explicitadas, foi afirmado que, por características dos componentes de busca, a plataforma JSP – Java deveria ser utilizada. Essa escolha foi originada principalmente pela utilização do Protégé para o controle da ontologia. Essa arquitetura é ilustrada na Figura 22.

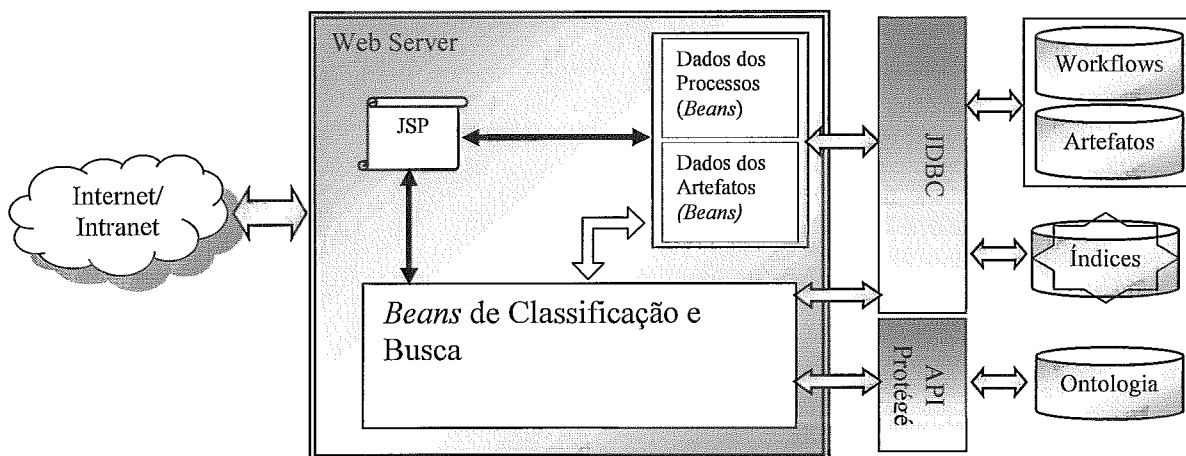


Figura 22 – Arquitetura dos componentes de classificação e busca

O Protégé é um editor de ontologias de código aberto, baseada em Java e desenvolvido pelo Stanford Medical Informatics e a Escola de Medicina de Stanford. Ele foi escolhido inicialmente para a edição da ontologia por ser gratuito. Além disto, sua interface é de fácil utilização, mesmo para usuários com pouca familiaridade com ontologias e sistemas de computador, como normalmente são os especialistas de áreas



diferentes da computação. Outro recurso importante do Protege é o provimento de uma API de desenvolvimento bastante robusta, que permite customizar as operações realizadas sobre a ontologia.

A API foi utilizada para acessar classes e instâncias, e também para navegar na ontologia durante a definição de similaridade. O desenvolvimento foi realizado mais rapidamente e de forma mais simples, dado que os métodos disponibilizados isolam a complexidade inerente à operação em hierarquias, como, por exemplo, a estratégia mais adequada de visitas aos nós. No entanto, para ser possível a utilização dessa API, os componentes de busca foram desenvolvidos em Java e encapsulados como *beans*, permitindo o acesso a partir das páginas JSP.

O componente mais suscetível a modificações são as interfaces com os sistemas externos, pois sua construção está condicionada à estrutura computacional existente na organização em que o Thoth é utilizado.

Geralmente, é interessante utilizar as API disponibilizadas pelos fabricantes do sistema. Tal situação ocorreu quando o sistema foi avaliado com o Oracle Workflow (ORACLE, 2001). Embora útil na fase de testes, a utilização desse sistema no auxílio das atividades científicas é dificultada por ela ser desenvolvida para o suporte dos processos de produção das organizações.

Assim, o protótipo e seus testes foram realizados sobre o sistema de *workflow* e controle de artefatos disponibilizados no sistema GCC, desenvolvido pelo grupo de pesquisa de banco de dados da COPPE. Tal opção se mostrou viável principalmente pelo fato da ferramenta de *workflow* prover suporte a atividades não estruturadas e semi-estruturadas, como são as atividades científicas. As APIs de acesso aos dados do GCC estão em fase de desenvolvimento e, por isso, não puderam ser utilizadas. O acesso aos dados dos processos e dos artefatos foi realizado diretamente na base de dados do sistema. Esse procedimento, no entanto, foi completamente realizado através dos *beans*, o que garantiu o isolamento adequado entre os dois sistemas.

### **3.4 – Conclusão**

Neste capítulo, foram descritas as características do Thoth, tema desta dissertação. O Thoth tem por objetivo a reutilização de processos científicos. Dessa forma, também foi descrita a forma como as características do ambiente científico influenciaram e se traduziram nos requisitos considerados na concepção do sistema.

Ante essas características, a metodologia RBC se mostrou apropriada para a localização dos processos candidatos à reutilização. Portanto, também foi descrita a analogia entre os processos e os casos, e como outros elementos e operações necessárias em uma ferramenta RBC foram executados sobre um sistema de *workflow*.

O desenvolvimento foi realizado a partir da observação do trabalho dos profissionais do LABHID. No próximo capítulo, serão apresentados a maneira como os trabalhos são desenvolvidos no laboratório e as possibilidades de utilização da ferramenta, bem como os testes executados.

## Capítulo 4 – Estudo de caso

Realizada a explicação da arquitetura do Thoth e de suas funcionalidades no capítulo anterior, será agora apresentado como ele pode ser utilizado no mundo real.

O desenvolvimento do Thoth foi motivado (e em grande parte abalizado) pelas observações realizadas no LABHID. Neste capítulo, será realizada uma descrição do LABHID, dos problemas relacionados à reutilização de conhecimentos enfrentados e de como o Thoth pode auxiliar a resolução desse problema.

### 4.1 – O LABHID

O LABHID é vinculado à área de recursos hídricos do Programa de Engenharia Civil da mesma instituição, sendo composto por professores, engenheiros, doutorandos, mestrandos e estudantes de graduação que realizam diversos estudos relacionados à hidrologia, hidráulica, gestão de recursos hídricos e sistemas de informação geográficos, entre outros tópicos.

O principal objetivo desses estudos é a identificação de potencialidades, limitações e problemas das bacias hidrográficas brasileiras, a fim de ajudar na definição de políticas apropriadas para utilização dos recursos existentes nessas bacias. Essas políticas representam uma ferramenta de controle sobre a utilização da água, recurso cada vez mais escasso e fundamental para preservação da qualidade de vida da população. Os estudos são geralmente realizados em colaboração com as instituições governamentais nacionais responsáveis pela definição dessas políticas.

Diversas questões relacionadas à água precisam ser analisadas, discutidas e solucionadas. Algumas delas têm a ver com a definição da maneira mais apropriada de cobrar pelo uso da água, a escolha de metodologias para tratamento do esgoto, a elaboração de estratégia para controle de erosão e enchentes, e o estudo de viabilidade de implantação de culturas agrícolas ou indústrias diante da utilização dos recursos hídricos disponíveis.

O primeiro passo para a realização desses estudos é a coleta de dados utilizados nas análises que precisam ser realizadas, como os dados climatológicos da região estudada, além de dados sobre o solo, a vegetação, índices de precipitação e rios que compõem a bacia.

Além disso, dados sócio-econômicos são necessários para se avaliar o impacto das decisões tomadas na qualidade de vida da população ribeirinha e no desenvolvimento econômico dessa população, das cidades, das indústrias e dos produtores agrícolas, entre outras organizações existentes na área de influência dessas decisões. Entre estes dados, estão o perfil da área (se urbana ou rural), número e perfil dos habitantes, número e tipo das indústrias, e quantidade de área destinada à agropecuária.

Todos esses dados podem eventualmente estar disponíveis no laboratório, mas muitas vezes é necessário consultar fontes externas de informação, como a ANA a SERLA, o IBGE e a Agência Nacional de Energia Elétrica - ANEEL, entre outras.

Para analisar esses dados, são necessários conhecimentos interdisciplinares. Por isso, o laboratório conta com pessoas oriundas de diversas áreas, como geologia, biologia, matemática, engenharia civil, cartografia, sociologia e geografia, que colaboram durante a realização das atividades.

A forma como essa análise é realizada depende do objetivo do estudo. Para exemplificar, será descrito o estudo conduzido para definir uma estratégia de controle de fluxo d'água para irrigação e drenagem de canais através do uso de comportas. Esses canais estão localizados na região da Lagoa Feia, no norte do Estado do Rio de Janeiro.

O controle de fluxo d'água é necessário porque, em certos períodos do ano, plantações existentes ao redor dos canais podem ser inundadas se o volume de água não for controlado. Por outro lado, durante o período de seca, a água armazenada deve ser distribuída de forma equânime para atender o consumo dessas plantações. Assim, a análise tem o objetivo de definir uma forma apropriada para abertura e fechamento da comporta, além de definir como os proprietários de terra e beneficiados podem pagar pela utilização da água armazenada durante o período de seca. Nessa operação é preciso se considerar não só a necessidade daqueles que tem suas plantações inundadas ou sofrem com a seca, mas também o impacto das mudanças no regime natural para os produtores que esperam que o canal tenha seu fluxo normal e que podem ser prejudicados no momento em que as comportas são utilizadas.

Logo, definir a política que vai ser aplicada não depende só do conhecimento técnico-científico, utilizado para analisar os dados coletados, definir e aplicar os modelos estatísticos, matemáticos e hidrológicos utilizados na definição das características do canal, e na demanda hídrica de cada terreno, entre outros fatores.

Isso deve levar em conta também as perspectivas e opiniões, nem sempre abalizadas nos critérios técnicos, dos responsáveis pelas áreas influenciadas pelas modificações realizadas, pois, dos interesses individuais de cada um desses responsáveis, podem surgir conflitos que precisam ser observados na definição da abordagem adequada.

A maioria dos estudos realizados no laboratório possui perfil similar, isto é, são constituídos de atividades onde modelos e abordagens são combinados para a análise das características físicas da bacia do corpo d'água (rio, lagoa, canal) e intercaladas com atividades de avaliação e solução dos conflitos oriundos da situação original do canal, bem como da resultante das intervenções propostas ou realizadas sobre o sistema.

#### **4.1.1 – Dificuldades enfrentadas**

Para realizar as suas atividades de forma satisfatória, um pesquisador do LABHID enfrenta diversas dificuldades, que estão relacionadas abaixo. Elas foram diagnosticadas durante a observação do trabalho ou por entrevistas realizadas no laboratório. Adicionalmente, é possível observar que elas são similares às dificuldades relatadas em outros ambientes descritos na literatura (CASTRO, OLIVEIRA et al., 2003):

- **Nenhuma informação sobre o andamento do processo é armazenada:** em uma eventual ausência de algum dos envolvidos, não será possível definir como está o andamento das atividades do ausente e pode não ser possível dar continuidade aos trabalhos até a comunicação entre os participantes ser restabelecida. Outra possibilidade é a ocorrência de retrabalho;
- **O histórico das atividades não é armazenado:** não ocorre o registro das atividades realizadas, não sendo possível definir como dados e artefatos foram utilizados. Geralmente, são feitos relatórios no fim de um estudo, no entanto, eles se concentram na descrição das decisões tomadas e dos impactos das mesmas. A metodologia, os dados e os artefatos utilizados são normalmente omitidos ou descritos de maneira superficial;
- **O conhecimento gerado não é armazenado adequadamente:** Grande parte do conhecimento utilizado ou gerado durante a realização dos trabalhos não é documentado, o que prejudica o desenvolvimento da

memória organizacional e faz com os novos conhecimentos gerados dentro do processo não possam ser armazenados para futuro acesso;

- **Busca de documentos e dados difícil de ser executada:** quando dados e documentos são pesquisados em fontes externas, algum tempo de trabalho é perdido para entender e transformar esses dados em um formato adequado. Nesse caso, tanto as transformações que foram executadas quanto a fonte original da informação podem ser perdidas. Além disso, não há nenhum controle de qualidade sobre a coerência e usabilidade dos mesmos;
- **Excesso de informação:** para determinar os dados de solo, clima e censitários relevantes, há uma pesquisa dos registros bibliográficos sobre a região estudada no acervo do laboratório ou em fontes externas. Nesse caso, o especialista não possui nenhum parâmetro inicial de busca de informação, e, às vezes, encontra-se diante uma imensa quantidade de informação a ser analisada.

#### 4.1.2 – Aplicação do Thoth no laboratório

A finalidade do Thoth é prover um sistema de reutilização de processos científicos, privilegiando os aspectos semânticos dos processos candidatos ao reuso, definido a similaridade entre eles e o estudo que está sendo realizado (BOMFIM, OLIVEIRA et al., 2005).

No LABHID, o Thoth é utilizado integrado ao GCC para que as dificuldades descritas acima possam ser eliminadas ou reduzidas. A seguir, é mostrado como os dois sistemas auxiliam no gerenciamento do conhecimento e podem ajudar na melhoria do processo de trabalho do laboratório.

Problema	Aplicação
Nenhuma informação sobre o andamento do processo é armazenada	O GCC provê uma ferramenta de <i>workflow</i> , tornando possível a modelagem, o acompanhamento, o controle e o registro das tarefas.
O histórico das atividades não é armazenado	Uma base com modelos dos processos científicos é gerada a partir da criação e execução destes, fornecendo uma fonte de informações para a memória organizacional.

<p>O conhecimento gerado não é armazenado adequadamente</p>	<p>Em paralelo com o sistema de <i>workflow</i>, outras ferramentas providas no GCC permitem que os diversos conhecimentos utilizados ou criados na execução dos processos possam ser armazenados na base de conhecimento do sistema.</p>
<p>- Busca de documentos e dados difícil de ser executada - Excesso de informação</p>	<p>A partir do momento que os conhecimentos são armazenados, o Thoth provê meios para que as informações possam ser localizadas e reutilizadas durante a execução ou o planejamento de novos processos.</p> <p>O fato da busca ser baseada em casos similares torna possível que sejam localizadas informações já filtradas e contextualizadas em problemas e soluções úteis ao estudo atual.</p>

## 4.2 – Processo de outorga da água

Na seção anterior, foi demonstrado como o Thoth auxilia a realização dos processos do LABHID. Outro cenário analisado para a utilização da ferramenta surgiu da colaboração entre o laboratório e as agências reguladoras do uso das águas. Estas existem no âmbito federal, estadual e municipal, e são responsáveis pela regulação do uso de água dos rios, lagos e reservatórios subterrâneos sob controle da sua respectiva Unidade Federativa.

Um importante elemento de regulação utilizado por essas instituições é a concessão de outorga de direitos de uso da água, que consiste na definição de como e por quem os corpos de água podem ser explorados, através da concessão a usuários solicitantes (ANA, 2005). Assim, a outorga assegura o aproveitamento de forma eficiente da água, de acordo com os padrões de qualidade e quantidade satisfatórios para os usuários atuais e também para as gerações futuras (SRH-RN, 2004). Esse processo é melhor detalhado abaixo a partir dos procedimentos descritos na documentação da ANA, órgão responsável pela administração dos corpos de água federais.

Através do processo de outorga, pode-se avaliar se um candidato a usuário pode utilizar a água de forma a não causar prejuízos para a região do

empreendimento. Para isso, o processo é composto de três análises: a técnica (viabilidade do empreendimento), a hidrológica (a quantidade e a qualidade do corpo de água) e a jurídica (o preenchimento dos requisitos legais impostos) (ANA, 2005). Dessa forma, é necessário analisar diversas informações, entre elas (ANA, 2004):

- Dados do corpo d'água - vazão, volume de água e localização, entre outros.
- O enquadramento do corpo d'água – Os rios são distribuídos em classes, que definem os níveis de permissão de emissão de poluentes, a fim de se garantir a qualidade do mesmo.
- O tipo de uso – A utilização da água, para efeito de cobrança, se dá de três formas: captação, consumo e diluição. Na captação, o usuário utiliza a água mais a devolve à fonte após a utilização. No consumo, não ocorre devolução da água ao meio. Já a diluição está relacionada com o lançamento de poluentes e ao volume de água necessário para suportá-lo.
- O perfil do solicitante – A análise realizada é condicionada ao tipo de usuário solicitante, a partir da identificação desse perfil, é possível definir quais parâmetros devem ser avaliados, bem com as obrigações e direitos do órgão outorgante e do usuário. Em sistemas de abastecimento público, por exemplo, onde o usuário seria um município, a avaliação deverá considerar as características físicas do sistema, a população atendida e as parcelas referentes aos setores comercial e industrial, entre outras variáveis. Outros tipos de solicitantes são indústrias, usinas hidrelétricas, agricultores e pecuaristas.
- O porte do solicitante – O tamanho de uma indústria influencia no volume de utilização que será realizada. Neste caso, dados como área construída e número de funcionários precisam ser avaliados para se estimar a demanda hídrica necessária. No caso de outros tipos de usos, como o agrícola, a área plantada serve para determinar essa demanda.

Como assinalado acima, o processo de outorga é composto de várias atividades que podem ser separadas em um contexto administrativo e jurídico, onde os documentos e os dados dos usuários são recebidos, seguindo as normas protocolares da ANA. Além disso, existe a análise técnica, que é a focalizada nesse trabalho por



ser a que contém as atividades técnicas conhecidas pelos profissionais do laboratório e pelos especialistas consultados.

O procedimento técnico para expedição de outorga visa definir se as demandas informadas na requisição não provocam degradação ambiental irreversível ao meio e segue basicamente as seguintes etapas (ANA, 2004): (i) avaliação da compatibilidade entre a demanda apresentada pelo usuário e os usos para os quais se destina; (ii) avaliação da disponibilidade hídrica em termos quantitativos e qualitativos no local do empreendimento; (iii) avaliação do impacto do novo uso no recurso hídrico; e (iv) elaboração de recomendações para o uso a serem expressas no ato de outorga.

Embora o objetivo do processamento seja semelhante para todos os perfis de usuários, as variáveis que devem ser mapeadas e analisadas na definição da demanda hídrica dependem desse perfil e influenciam a forma como as atividades são executadas.

Na ANA, os usuários são separados em nove perfis de uso: o abastecimento público e lançamento de esgotos domésticos, a criação animal, industrial, irrigação, mineração, obras hidráulicas, construção de pequenos açudes no semi-árido da região Nordeste, aquicultura e outros usos. Cada um desses perfis demanda um ou mais tipos de uso (captação, diluição, consumo), sendo que modelos próprios são aplicados sobre os dados fornecidos para que se estime a demanda adequada. Por exemplo, variáveis censitárias e modelos estatísticos utilizados para avaliar o comportamento delas no tempo devem ser considerados para avaliar a demanda hídrica solicitada por uma cidade.

**A Erro! A origem da referência não foi encontrada.** apresenta o processamento básico realizado para a outorga industrial em rios. Usuários industriais precisam captar água para utilizar na produção e operações administrativas, e também efetuar lançamento dos dejetos industriais no rio.

Inicialmente, é analisada a demanda de captação, onde os dados da vazão requerida pela indústria, o teor de água no produto e o montante de água utilizado no processo fabril são analisados com os índices de produtividade, a fim de se estimar quanto de água é utilizada no uso consutivo (diretamente na produção) e no uso não-consutivo (higiene, subsistência dos funcionários etc.). Esses passos são representados respectivamente pelas atividades A0 e A1.

A partir do momento em que essas duas vazões são definidas, verifica-se a consistência da captação solicitada (atividade A2). Nessa atividade, os especialistas

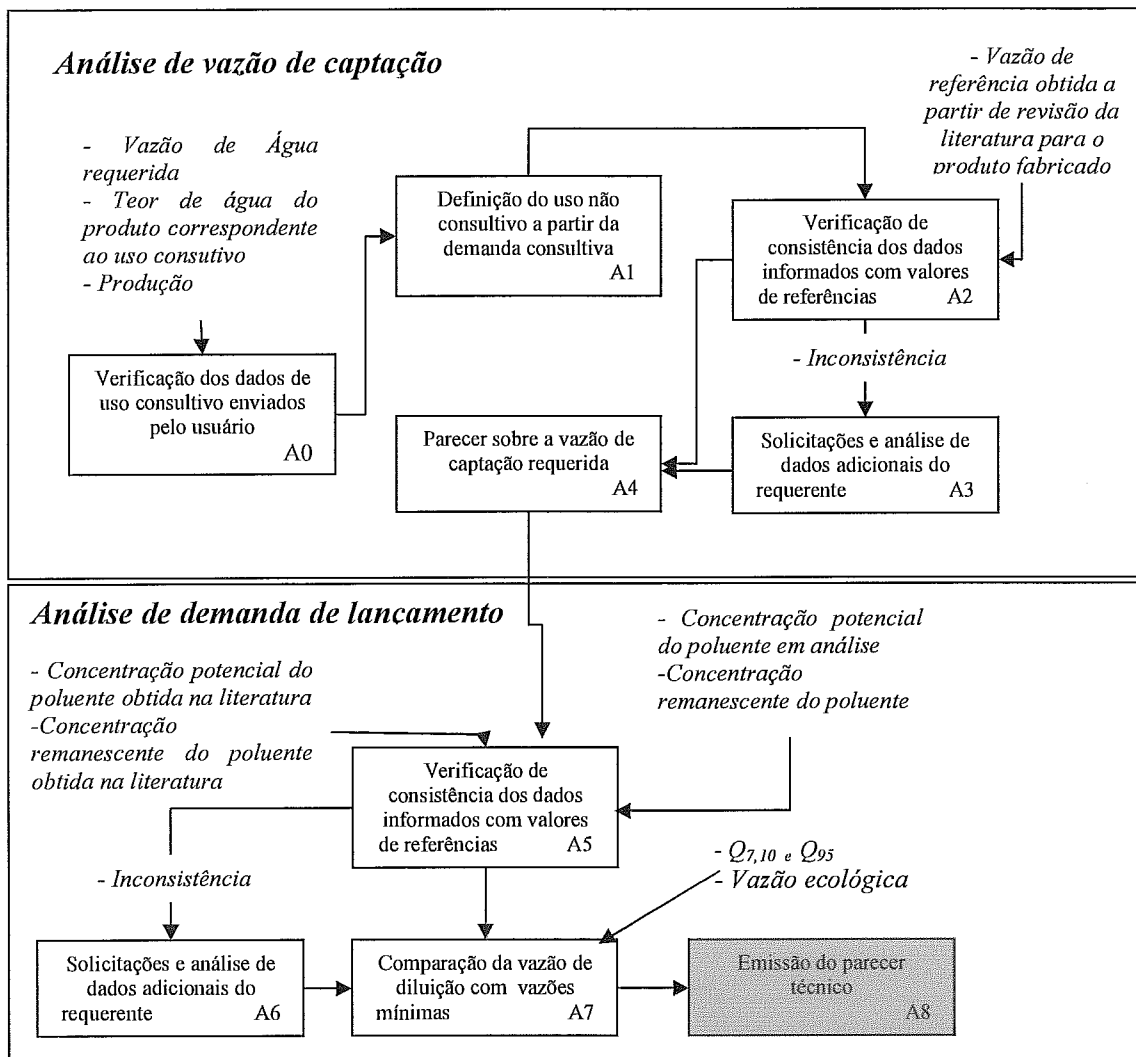
buscam na literatura os valores de referência para utilização de água para o processo fabril conduzido pelo requerente. Modelos também são aplicados para definir, através da quantidade de funcionários e do porte da empresa, o volume não-consutiva de referência. Os valores obtidos são comparados com os solicitados pelo requerente e, caso não sejam satisfatórios, novas informações são solicitadas e analisadas (atividade A3).

Após o fim da verificação dos dados, é concedido o parecer sobre a vazão de captação solicitada (atividade A4). Nesse ponto, a vazão solicitada é comparada com a disponível no rio.

Após a análise da captação, é realizada a análise da demanda de lançamento, onde se avalia a capacidade do rio de absorver os efluentes poluidores da indústria. Primeiramente, os valores de concentração de poluentes são analisados e comparados com valores de referência consultados na literatura (atividade 5). Duas concentrações são analisadas: a potencial e a remanescente. A concentração remanescente é obtida após o tratamento realizado antes do lançamento do efluente no meio. Já a concentração potencial é obtida antes desse tratamento, com a mistura de água e dos poluentes em seu estado bruto. Essa comparação visa avaliar se os valores fornecidos pelo usuário são condizentes com o que é fabricado, com o volume de produção e com o método de tratamento utilizado.

Após a comparação, caso seja necessário, novos dados são solicitados ao requerente e analisados para eliminar as inconsistências identificadas (Atividade 6). Na atividade 7, a concentração estipulada como adequada é utilizada para definir qual vazão de água será necessária para realizar a diluição dos poluentes no rio. Também são obtidos os valores da  $Q_{7,10}$  e  $Q_{95}$  do trecho do rio (LABHID & ANA, 2003). Esses valores de vazão são obtidos em séries históricas e representam, respectivamente, a média das vazões mínimas em sete dias consecutivos em dez anos de recorrência e a vazão mensal com probabilidade de 95% de ser excedida, e servem para se estimar a disponibilidade hídrica mínima do rio. Esses valores são decrescidos da vazão ecológica, que é a mínima vazão que garante as condições ecológicas naturais do rio. Com esses dados, é verificado se a vazão solicitada pelo usuário é inferior à mínima disponível. Se isso ocorrer, existe a viabilidade técnica para a cessão do direito de uso.

Finalmente, as observações, recomendações e resultados são estruturados no parecer técnico, que é o documento utilizado como base pelo corpo administrativo da ANA para a concessão ou não da outorga.



**Figura 23 - Fluxograma da análise de disponibilidade hídrica para Indústrias dentro do processo de Outorga**

Assim como nos trabalhos do LABHID, a definição da outorga deve considerar também os conflitos que podem surgir pela modificação da disponibilidade hídrica a jusante ou a montante do ponto do rio no qual a solicitação foi feita, sendo necessário intercalar atividades técnicas com aquelas destinadas à eliminação de conflitos que possam surgir após a cessão da outorga.

As dificuldades enfrentadas também são similares: a localização de dados de referência na literatura nem sempre é possível, pois a tecnologia utilizada pelos usuários está em constante evolução e também é crítica a disponibilidade de dados sobre algumas regiões. Dessa forma, qualquer informação anteriormente encontrada e

utilizada sobre essa região pode ser valiosa para que outorgas sejam dadas com informações mais próximas do cenário real.

Nesse contexto, o Thoth pode ser integrado com os sistemas de suporte à decisão utilizados. Com a localização realizada através da ferramenta, é possível verificar em processos de outorga anteriores, filtrados pelos critérios de similaridade, os impactos, as práticas, os dados e as informações utilizadas ou resultantes dessas decisões.

Os critérios de similaridade que podem ser utilizados são vários. Abaixo, são apresentados exemplos ilustrativos desses critérios, definidos a partir das entrevistas com os pesquisadores do laboratório e também com um especialista em regulação de recursos hídricos, que trabalha nas agências reguladoras:

- Comparação com processos prévios do mesmo rio – este é o critério mais simples e consiste na apresentação ao usuário de processos localizados no mesmo rio de onde se originou a solicitação.
- Rios da mesma bacia ou em bacias vizinhas a do rio em questão – Este critério ganha em complexidade e, por isso, é preciso conhecimento mais profundo do domínio para que seja possível a identificação das configurações das bacias hidrológicas existentes na região em que a outorga é cedida. Nesse caso, a construção de diques, desvios ou canalizações, mesmo que realizada em uma bacia vizinha à do rio no qual a análise é feita, pode fazer com que a demanda hídrica sobre o rio aumente e esse fator deve ser avaliado pelos outorgantes.
- Comparação entre processos em rios ou bacias com características similares -- Se em um nível alto de abstração, o processo de outorga pode ser descrito de uma forma única, quando analisado em um nível de abstração mais baixo, são observadas muitas diferenças entre cada processo, já que estes são condicionados às bacias onde são feitas as análises e às suas características físicas, hidrológicas e hidráulicas, além dos dados disponíveis sobre a mesma. Dessa forma, a comparação com processos de bacias similares pode fornecer parâmetros e métodos para a avaliação de um novo pedido.

## **4.3 – Utilização do protótipo**

Como explicitado na seção 4.1.2 –, o Thoth atualmente se encontra em fase de teste no LABHID. Como um sistema RBC, O Thoth tem o objetivo de prover, em um tempo razoável, casos similares a um problema enfrentado. Assim, o sistema deve se concentrar em dois pontos: o primeiro é a eficiência (tempo de resposta das buscas) com as estruturas de armazenamento e indexação utilizadas. Por outro lado, o fator mais importante é o reconhecimento pelo usuário da utilidade dos casos retornados diante do seu problema.

A avaliação desses fatores é condicionada à prévia existência ou à construção de uma base de casos, o que atualmente não existente no laboratório. Além disso, a instalação do ambiente que permite que os pesquisadores armazenem os seus trabalhos (o sistema GCC), a capacitação deles para a utilização do sistema, assim como o tempo para que fosse armazenada uma quantidade adequada de casos, supera o tempo de desenvolvimento de uma tese de mestrado.

Por isso, embora o ambiente composto pelo GCC e o Thoth já esteja em utilização no laboratório, existe a dificuldade inerente aos sistemas RBC da população de uma base inicial de casos, o que possibilitaria a execução dos testes para o refinamento da abordagem e avaliação da aceitabilidade do sistema. Assim, a seção seguinte apresenta o estudo experimental planejado e atualmente em execução no LABHID, que propiciará o crescimento da base e a avaliação da ferramenta.

### **4.3.1 – Planejamento**

No contexto do Laboratório, o Thoth é utilizado quando um pesquisador reutiliza dados, documentos, mapas e outras fontes de informação vinculadas aos processos considerados similares e, conseqüentemente, úteis aos problemas estudados. O conceito de utilidade está associado à percepção de cada indivíduo e, por isso, a avaliação da viabilidade da ferramenta e de que a definição de similaridade é realizada de forma adequada passa pela avaliação dos pesquisadores que utilizam a ferramenta.

Como assinalado anteriormente, a principal dificuldade enfrentada para a realização dos testes é a concepção de uma base de casos (conjunto de processos). Para sanar esse problema, observou-se o perfil dos profissionais e atividades desenvolvidas no laboratório, e optou-se pelo encorajamento dos mestrandos e doutorandos a registrarem suas atividades de forma integral ou parcial no sistema.

O estudo tomou como hipótese que a utilização da ferramenta reduziria o tempo na localização de informações relacionadas a um problema estudado, o que propicia a resolução mais rápida do mesmo. Adicionalmente, este estudo é focalizado na avaliação das informações subjetivas, relacionadas à satisfação dos pesquisadores no uso da ferramenta.

#### **4.3.2 – Participantes**

Podem ser discriminados dois grupos de participantes no estudo. O primeiro grupo foi explicado anteriormente e consiste nos mestrandos e doutorandos do laboratório, convidados para descrever o fluxograma das suas atividades no GCC.

Já o segundo grupo é composto por três estudantes, sendo um de mestrado e dois de doutorado, que serão os responsáveis pela avaliação do sistema, efetuando as buscas e avaliando se os resultados obtidos são satisfatórios.

#### **4.3.3 – Procedimento**

O procedimento executado para a realização do estudo pode ser sintetizado pelos passos apresentados abaixo. A ordem cronológica de execução está de acordo com a ordem de apresentação dos mesmos:

- Coleta de dados que caracterizam o pesquisador
- Explicação para cada pesquisador do ambiente GCC / Thoth
- Explicação para cada participante dos procedimentos para cadastro de suas atividades no sistema
- Seleção dos pesquisadores para participação na etapa de busca utilizando o Thoth
- Entrevistas com os envolvidos na etapa de busca

Atualmente, está sendo realizado o cadastro da base de processos com as teses. Para realizar essa tarefa, foi solicitada aos orientadores das pesquisas do laboratório uma listagem dos seus alunos. A partir dela, está sendo estabelecido contato com os estudantes, a fim de se realizar reuniões individuais para serem explanadas o propósito dos sistemas utilizados, já que eles podem continuar utilizando a ferramenta, mesmo após o teste. Após essa reunião, o participante pode criar ou editar o modelo de suas atividades no *workflow* do GCC.

Após um período inicialmente estipulado em três meses (contados a partir de maio de 2005), será realizada a segunda etapa, onde os voluntários responsáveis pela avaliação realizarão as buscas e a análise da ferramenta.

Para efetuar a análise, esses voluntários responderão a um questionário elaborado para avaliar: (i) Eficiência, (ii) Facilidade de utilização, (iii) Reconhecimento dos critérios de similaridade, (iv) Aumento de produtividade individual e (v) Satisfação.

#### **4.3.4 – Resultados esperados**

A partir da realização do estudo, espera-se a criação de oportunidades de melhoria do protótipo do Thoth, através da identificação de pontos positivos e negativos da ferramenta pelo o ponto de vista dos pesquisadores do LABHID.

Adicionalmente, o sistema já está disponibilizado para utilização por todos os profissionais do laboratório, o que torna possível que cada um deles possa realizar sua própria avaliação e sugerir melhorias tanto no GCC quanto no Thoth.

## Capítulo 5 – Conclusão

A realização do trabalho científico pode ser caracterizada pela necessidade de um alto grau de qualificação dos seus executores e por ser constituído de atividades intensamente dependentes do conhecimento. Essas características trazem desafios para que possa ser desenvolvido com qualidade, já que o cientista deve ter segurança suficiente para tomar complexas decisões, como a escolha dos métodos e das ferramentas que serão utilizadas, e a configuração e definição dos dados que serão consultados, além de interpretar resultados obtidos dos experimentos e análises.

Assim, para superar esses desafios, o cientista normalmente consulta diversas fontes de informação, colabora com outros profissionais e controla criteriosamente o que executa. Os sistemas de *workflow* auxiliam o pesquisador permitindo a organização e coordenação das atividades. Por outro lado, sistemas de gestão do conhecimento propiciam melhor controle sobre os conhecimentos que são produzidos, além de facilitar e agilizar o acesso aos conhecimentos necessários para a execução adequada das atividades.

Dentro desse contexto, surgiu a proposta do Thoth, tema desta dissertação. O Thoth parte do princípio de que o conhecimento contido em experiências anteriores é um importante elemento da memória organizacional e, dessa forma, deve ser utilizado pelos membros da organização. Assim, o sistema tem o objetivo de fornecer meios para prover reutilização de processos.

A motivação para a construção do sistema surgiu a partir das análises realizadas durante o desenvolvimento do SPeCS (MEDEIROS, SOUZA et al., 2000) e do GCC. No SPeCS, são empregadas noções de trabalho colaborativo, apoio à decisão, gestão do conhecimento, gerência de experimentos científicos, gestão de documentos e integração de dados para fornecer auxílio computacional a processos decisórios e processos científicos colaborativos com dados espaciais. Já o GCC tem como objetivo auxiliar a gestão do conhecimento científico e está sendo desenvolvido para apoiar as atividades dos pesquisadores da COPPE. Nesses sistemas foi possível notar os potenciais benefícios de prover aos usuários acesso às informações de atividades previamente executadas. Além disso, como em sistemas similares descritos



na literatura, ambas as atividades são estruturadas de forma seqüencial e controladas por um sistema de *workflow*.

O objetivo do Thoth é alcançado através da análise das informações contidas no histórico de sistemas de *workflow*. Dada a estrutura dos processos e as características do trabalho científico, optou-se pela utilização da abordagem de solução de problemas RBC, realizando a analogia entre os casos e as instâncias armazenadas no sistema.

A definição da similaridade entre os casos é realizada com o auxílio de uma ontologia para se contemplar os aspectos semânticos do domínio em que o sistema é utilizado, e propiciar maior reconhecimento, por parte dos usuários, da correlação entre os casos buscados e as atividades em execução.

A proposta dessa dissertação é validada através da implementação de um protótipo integrado ao GCC e testado numa área de aplicação real, a hidrologia, sendo a utilização realizada pelos pesquisadores, estudantes e colaboradores do LABHID. Para ser possível a utilização do sistema, foi necessário desenvolver uma ontologia com os conceitos pertinentes à área, sendo ela construída a partir do reaproveitamento de outras ontologias existentes e também através de colaboração e entrevistas com profissionais do próprio laboratório.

## **Contribuições**

A principal contribuição desta dissertação é o desenvolvimento do sistema de reutilização de conhecimentos científicos baseada na abordagem de RBC. Pelo fato da abordagem ser aplicada com os conceitos gerais e padronizados de processos e sistemas de *workflow*, a ferramenta se mostra adaptável a diferentes situações, podendo ser utilizada em diversos tipos de processos relacionados à área de pesquisa ou de processos decisórios apoiados em práticas científicas, como o planejamento ambiental, exemplificado nesta dissertação pelo processo de outorga da água.

Não menos importante, a ontologia produzida constitui um importante elemento para o reconhecimento das características e nuances da área de hidrologia, estando disponível para futuros trabalhos acadêmicos do grupo de banco de dados da COPPE e para desenvolvimento de novos sistemas de gestão de conhecimentos no LABHID.

Adicionalmente, é possível utilizar a abordagem e a ferramenta de busca para outros tipos de objetos, como documentos, planilhas e mapas, realizando pequenas adaptações e desenvolvendo interfaces específicas para cada tipo de objeto.

## **Trabalhos Futuros**

Como trabalho futuro, será realizada a observação mais precisa do comportamento do sistema ante o crescimento gradual da base a fim de confirmar a eficácia do GCC e do Thoth no Laboratório.

O aumento da utilização permitirá também coletar mais informações sobre o domínio, o que tornará possível o enriquecimento da ontologia e também a exploração de outros relacionamentos entre os conceitos ontológicos na definição de similaridade. Além disto, a partir do momento em que o sistema se torne mais utilizado, será possível otimizar o cálculo de similaridade e desenvolver e explorar as propriedades das classes.

Algumas características da abordagem RBC podem ser melhor exploradas, dentre elas a categorização dos casos como de sucesso e insucessos e a adaptação dos casos.

A explicitação de casos de sucesso e insucesso está relacionada com o fornecimento ao usuário da avaliação se os resultados obtidos em um caso anterior foram satisfatórios ou não. Atualmente cabe ao usuário inferir, de acordo com as informações que lhe são apresentados se o caso é de sucesso ou de falha. Para suplantarmos a carência deste tipo de informação, as opiniões e observações daqueles que consultarem o sistema podem ser armazenadas, consideradas e apresentadas durante as novas buscas.

A adaptação de casos não é automática, sendo realizada manualmente pelos usuários. A implementação de adaptações automáticas é fortemente dependente das regras do domínio e podem ser implementadas futuramente de acordo com a necessidade dos profissionais do LABHID ou de outras organizações onde a abordagem for utilizada.

O sistema pode ser utilizado em outros domínios ou no auxílio de processos decisórios, como o de Outorga. Nestes casos, novas formas de representação do processo podem ser avaliadas, como “atos de fala” (MEDINA-MORA, WINOGRAD et al., 2002) ou modelos de argumentação como o IBIS (CONKLIN & BEGEMAN,

1998), a fim de obter maior controle dos trabalhos durante a execução e também uma representação com informações mais detalhadas e organizadas para reutilização.

Adicionalmente, estes processos podem ser mais estruturados que os do LABHID, esta característica permite que os recursos utilizados possam ser previamente identificados e conseqüentemente que a classificação dos casos possa ser realizada com o auxílio de dispositivos para processamento automático. A exceção dos processadores de texto e imagem, outros dispositivos não são viáveis no LABHID porque normalmente os pesquisadores e os grupos trabalham de forma distribuída e independente, não sendo possível determinar a estrutura de todos os artefatos que são utilizados.

## Referências Bibliográficas

- AAMODT, A., PLAZA, E., 1994, "Case-based reasoning: foundational issues, methodological variations and system approaches", *AI Communications*, v. 7, n. 1, pp. 39-59.
- ABBOTT, K. R., SARIN, S. K., 1994, "Experiences with Workflow Management: Issues for the Next Generation".
- ANA, 2004, "Resolução ANA nº 707", December.
- ANA, 2005, "Outorga de direito de uso", Accessed in 10/10/2004.
- APACHE SOFTWARE FOUNDATION, 2004, "Apache Jakarta Tomcat". In: <http://jakarta.apache.org/tomcat/>, Accessed in 10/03/2004.
- ARAÚJO R.M., BORGES M.R.S., 2001, "Sistemas de Workflow". In: *XX Jornada de Atualização em Informática, Congresso da SBC*.
- ARAÚJO, R. M., BORGES, M. R. S., 2002, "Awareness Extensions in Workflow Management Systems - Elements for Collaboration and Process Learning". In: *The Seventh International Conference on Computer Supported Cooperative Work in Design*, pp. 375-380, Brasil.
- BANNON, L. J., SCHMIDT, K., 1989, "Four Characters in Search of a Context", Proceedings of the First European Conference on CSCW, Gatwick, UK.
- BARROSO, A. C., GOMES, E., 1999, "Tentando Entender a Gestão de Conhecimento".
- BARTSCH-SPORL, B., LENZ, M., HUBNER, A., 1999, "Case-based reasoning: Survey and Future Directions". In: *Proceedings of XPS-99*, v. 1570, pp. 67-89.
- BERGMANN, R., STAHL, A., 1988, "Similarity Measures for Object-Oriented Case Representations". In: *Proceedings of the 4th European Workshop on Advances in Case-Based Reasoning*, v. 1488, pp. 25-36, London, UK.
- BERMUDEZ, L., 2004, "Hydrological Units". In: <http://loki.cae.drexel.edu/%7Ehow/HydrologicUnits/hu>, Accessed in 15/01/2004.
- BOMFIM, E. L. P., CASTRO, M., SOUZA, J. M., et al, 2003, "DecisioFlow: Auxílio Colaborativo na Execução de Atividades em Processos Decisórios". In: *Webmidia 2003*, Brasil.
- BOMFIM, E. L. P., OLIVEIRA, J., SOUZA, J. M., 2005, "Thoth: Improving Experiences Reuses in the Scientific Environment through Workflow Management System". In: *Proceeding of the 9th International Conference on CSCW in Design*, Coventry, UK.

- CARDOSO, L. F., SOUZA, J. M., MARQUES, C., 2002, "A Collaborative Approach to the Reuse of Scientific Experiments in the Bill of Experiments Tool". In: *Seventh International Conference on Computer Supported Cooperative Work in Design*, pp. 296-301, Rio de Janeiro, Brazil.
- CASTRO, M., 2003, *DECISIO: UM SISTEMA DE APOIO À DECISÃO PARA O PLANEJAMENTO AMBIENTAL*, COPPE / UFRJ.
- CASTRO, M., OLIVEIRA, J., STRAUCH, J., et al, 2003, "Improving Collaborative Decision Process Quality on Geographic Environments". In: *Proceeding of International Workshop on Data Quality in Cooperative Information Systems*, Italy.
- CASTRO, M., SOUZA, J. M., STRAUCH J.C.M., 2003, "Decisio: A Collaborative Decision Support System for Environmental Planning". In: *International Conference on Enterprise Information Systems*, Angers, France.
- CHAFFEY, D., 1998, "Workflow and Intranets - Reengineering the Enterprise with Collaborative Software". *Digital Press*.
- CHIN, G., LEUNG, L. R., SCHUCHARDT, K., et al, 2002, "New Paradigms in Problem Solving Environments for Scientific Computing". In: *ACM International Conference on Intelligent User Interfaces*, San Francisco, CA.
- CLARK, O. A. C., CASTRO, A. A., 2002, "Planejamento da pesquisa.". In: [http://www.evidencias.com/planejamento/pdf/lv4\\_01\\_planeja.PDF](http://www.evidencias.com/planejamento/pdf/lv4_01_planeja.PDF), Accessed in 10/06/2003.
- CNPQ, 2004, "LMPL - Linguagem de Marcação da Plataforma Lattes". In: <http://www.cnpq.br/lmpl>, Accessed in 10/04/2004.
- CONKLIN, E. J. W. W., 1996, "Wicked Problems: naming the pain in organizations". In: [http://www.wwdemocracy.nildram.co.uk/new\\_visions/systems\\_thinking/wicked\\_problems.htm](http://www.wwdemocracy.nildram.co.uk/new_visions/systems_thinking/wicked_problems.htm), Accessed in 10/06/2004.
- CONKLIN, J., BEGEMAN, M., 1998, "gIBIS: A hypertext tool for exploratory policy discussion".
- CYCOP, I., 2004, "OpenCyc". In: <http://www.cyc.com/cyc/opencyc/overview>, Accessed in 18/06/2004.
- DAVENPORT, T. H., PRUSAK, L., 1998, *Conhecimento Empresarial: Como as organizações gerenciam o seu capital intelectual*, Rio de Janeiro.
- DAVID J.M.N., BORGES, M. R. S., 2001, "Improving the Selectivity of Awareness Information in Groupware Applications". *IEEE Computer Press*, Darmstadt, Germany.
- DHAR, V., STEIN, R., 1996, "SOLVING PROBLEMS BY ANALOGY: CASE-BASED REASONING", *Seven Methods for Transforming Corporate Data Into Business Intelligence*, Upper Saddle River, NJ, Prentice Hall.

- DOURISH, P. E. B., V, 1992, "Awareness and Coordination in Shared Work Spaces". In: *Proceedings of ACM CSCW'92 Conference on ComputerSupported Cooperative Work*, pp. 107-114, Toronto, Canada.
- ELLIS, C. A., 1999, "Workflow Technology". In: <http://www.lri.fr/~mbl/cgi-bin/getpdf?Trends-CSCW/chap2.pdf>, Accessed in 06/06/2003.
- ELLIS, C. A., GIBBS, S. J., REIN, J. L., 1991, "Groupware: Some issues and experiences".
- EUZENAT, J., 1996, "Corporate memory through cooperative creation of knowledge bases and hyper-documents", Tenth Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW96), Alberta, Canada.
- FILHO, A. M., 2000, "Ensaio sobre a Filosofia da Ciência". In: <http://www.ecientificocultural.com/ECC2/FilCien/cap01.htm>, Accessed in 18/03/2004.
- FIREBIRDSQL FOUNDATION, 2003, "Firebird - Relational Database for the New Millenium". In: <http://firebird.sourceforge.net/>, Accessed in 10/10/2003.
- FISCHER, G., OSTWALD, J., 2001, "Knowledge Management: Problems, Promises, realities, and Challenges", *IEEE Intelligent Systems*, v. 16, n. 1, pp. 60-72.
- FOSTER, I., VOCKLER, J., WILDE, M., 2002, "Chimera: A Virtual Data System for Representing, Querying, and Automating Data Derivation". In: *14th International Conference on Scientific and Statistical Database Management (SSDBM'02)*, pp. 37-46, Edinburgh, Scotland.
- FREIRE-MAIA, N., 1995, *A ciência por dentro*, Ed. Vozes.
- GOLDENBERG S., 2001, "Orientação normativa para elaboração de tese.". In: <http://www.metodologia.org/man96.pdf>, Accessed in 2003.
- GRUDIN, J., 1994, "CSCW: History and Focus".
- GUARINO, N., 1998, "Formal Ontology and Information Systems". In: *International Conference on Formal Ontology in Information Systems*, pp. 6-8, Trento, Italy.
- HANDEL, M., 2002, "Presence Awareness & Instant Messaging: Multiple devices, Multiple Endpoints". In: <http://www-personal.si.umich.edu/~handel/research/rvm/documents/Handel-DocConsort.pdf>, Accessed in 28/10/2002.
- HEFLIN, J., 2004, "OWL Web Ontology Language Use Cases and Requirements - W3C Recommendation". In: <http://www.w3.org/TR/webont-req/#section-requirements>, Accessed in 05/05/2004.
- HOLLENBACH, C., FRAKES, W., 1996, "Software Process Reuse in an Industrial Setting". In: *Fourth International Conference on Software Reuse*, pp. 22-30, Orlando, Florida.

- HOLLINGSWORTH, D., 1995, "The Workflow Reference Model. TC00-1003".  
*Workflow Management Coalition.*
- HOWARD, R. E., 2000, *Aprendizado Organizacional Gestao De Pessoas Para A Inovacao*, Rio de Janeiro, EDITORAS CAMPUS.
- ISLUM, A. S., BERMUDEZ, L., BERAN, B., et al., 2004, "Ontology for Geographic Information".
- ISLUM, A. S.,PIASECKI, M., 2004, "Ontology for Surface Water and Water Quality Models". In: <http://loki.cae.drexel.edu/%7Ewbs/ontology/model.htm>, Accessed in 01/03/2004.
- JABBER INC, 2004, "jabberd project". In: <http://jabberd.jabberstudio.org/2/>, Accessed in 10/10/2004.
- JIANG, J. J.,CONRATH, D. W., 1998, "Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy". In: *International Conference on Research in Computational Linguistics*, Taiwan.
- JURISICA, I., MYLOPOULOS, J., YU, E., 1999, "Using Ontologies for Knowledge Management: A Computational Perspective". In: *Annual Conference of the American Society for Information Science*, Washington, DC.
- KHAN, L. R., 2000, *Ontology-based information selection*, Tese de Ph.D, University of South California.
- LABHID,ANA, 2003, *Sistema de Informações de Recursos Hídricos para a Bacia do Paraíba do Sul*.
- LAERCIO, J, 2004, "Thoth, Divindade Egípcia". In: [http://www.joselaerciodoegito.com.br/site\\_hermes\\_bio.htm](http://www.joselaerciodoegito.com.br/site_hermes_bio.htm), Accessed in 11/10/2004.
- LAKATOS, E. M.,MARCONDI, M. D. A., 1991, *Metodologia Científica*. 3, São Paulo, Atlas.
- LASSALLE, 2004, "Lassalle addflow". In: <http://www.lassalle.com/products.htm>, Accessed in 12/08/2004.
- LEAKE, D. B.,KOLODNER, J. L., 2000, "A Tutorial Introduction to Case-Based Reasoning", *Case-Based Reasoning: Experiences, Lessons, and Future Directions*, chapter 2, AAAI Press/MIT Press.
- LEYMANN, F.,ROLLER, D., 2000, "Workflow Managment System Basics", *Production Workflow - Concepts and Techniques*, chapter 3, New Jersey, Prentice Hall Books.
- LIMI , A., RUNYAN, A., ANDERSEN, V., 2005, "Judge", <http://www3.dfki.uni-kl.de/judge>.

- LIU, H., SINGH, P., 2004, "Focusing on ConceptNet's natural language knowledge representation". In: *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES'2004)*, Wellington, New Zealand.
- MAIA, I. M. R. L., 2005, "O desenvolvimento da Ciência em Thomas Kuhn". In: <http://www.consciencia.org/contemporanea/kuhnisabel.shtml>, Accessed in 08/08/2004.
- MEDEIROS, S., 2002, *SPeCS - Sistema de Suporte à Decisão Espacial Colaborativa*, COPPE / UFRJ, Rio de Janeiro.
- MEDEIROS, S., SOUZA, J., STRAUCH, J., et al., 2001, "Coordination Aspects in a Spatial Group Decision Support Collaborative System", *Proceedings of ACM/SAC*, Las Vegas.
- MEDEIROS, S., SOUZA, J., STRAUCH, J., et al., 2000, "SPECS - A Spatial Decision Support Collaborative System for Environment Design". In: *Fifth International Conference in CSCW in Design*, Hong Kong.
- MEDINA-MORA, R., WINOGRAD, T., FLORES, R., et al., 2002, "The Action Workflow Approach to Workflow Management Technology". In: *Proceedings of the 4th Conference on CSCW*, pp. 281-297, New York, NY.
- MICROSOFT, 2004, "Microsoft SQL Server". In: <http://www.microsoft.com/sql/default.asp>, Accessed in 01/06/2004.
- MICROSOFT, 2003, "Microsoft SQL Server". In: <http://www.microsoft.com/sql/default.asp>, Accessed in 2003.
- MICROSOFT, 2001, "Microsoft Windows 2000". In: <http://www.microsoft.com/windows2000/default.asp>, Accessed in 10/05/2004.
- MILLARD, P., 2004, "Exodus". In: <http://exodus.jabberstudio.org/>, Accessed in 15/10/2004.
- MORENO, M., VIVACQUA, A., SOUZA, J., 2003, "An Agent Framework to Detect Opportunities for Collaboration". In: *9th International Workshop on Groupware*, v. 224, Grenoble, France.
- MORESI, E., 2000, "Delineando o valor do sistema de informação de uma organização", Brasil.
- NARDI, B. A., WHITTAKER, S., BRADNER, E., 2000, "Interaction and Outeraction: Instant Messaging in Action". In: *Computer Supported Cooperative Work*, pp. 79-88.
- NONAKA, I., TAKEUCHI, H., 1995, "The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation". *Oxford University Press*.



- NOY, N. F., MCGUINNESS, D. L., 2001, *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory.
- O'LEARY, D. E., 1998, "Using AI in Knowledge management: Knowledge Bases and Ontologies", *IEEE Intelligent Systems*, v. 13, n. 3, pp. 34-39.
- OLIVEIRA, J., 2003, *Epistheme: Um Ambiente de Gestão do Conhecimento Científico*, Tese de M.Sc, COPPE / UFRJ, Rio de Janeiro.
- ORACLE, 2001, "Oracle Workflow Guide". In: <http://www.ic.leidenuniv.nl/awcourse/oracle/workflow.920/a95265/toc.htm>, Accessed in 06/05/2003.
- PALAZZO, M. O. J., NICOLAO M., EDELWEISS, N., 1998, "Conceptual Workflow Modeling for Remote Courses". In: *Proceedings of the XV World Computer Congress, IFIP*, pp. 789-797, Vienna - Austria.
- PENNAFORT, M., "Método Científico".
- PEREIRA, L. T. V., BAZZO, W. A., 2002, "Introdução à Engenharia". In UFSC, 6 ed., chapter 5, Santa Catarina, UFSC.
- POLANYI, M., 1983, "The Tacit Dimension". *Peter Smith Publication*, London.
- POPPER, K., 1959, *A lógica da pesquisa científica*, São Paulo, Cultrix.
- POSTGRESQL, 2003, "PostgreSQL". In: <http://www.postgresql.org/>, Accessed in 2003.
- PRINCETON UNIVERSITY, 2004, "WordNet". In: <http://wordnet.princeton.edu/>, Accessed in 10/02/2004.
- RATCLIFFE-MARTIN, V., COAKES, E., SUGDEN, G., 2004, "Enhancing Knowledge Acquisition and Transfer in the University Sector". In: <http://users.wmin.ac.uk/~coakese/knowledge/bit2000.htm>.
- RODRIGUEZ, M. A., EGENHOFER, M. J., 2003, "Determining Semantic Similarity Among Entity Classes from Different Ontologies", *IEEE Transactions on Knowledge and Data Engineering*, v. 15, n. 2, pp. 442-456.
- SALTON, G. M., MCGILL, M., 1983, "Introduction to Modern Information Retrieval". *McGraw-Hill Book Co.*, New York.
- SCHUR, A., KEATING, K. A., PAYNE, D. A., et al., 1998, "Collaborative Suites for Experiment-Oriented Scientific Research".
- SEFFINO, L. A., MEDEIROS, C. B., ROCHA, J. V., et al, 1999, "WOODSS - a spatial decision support system based on workflows", v. 27, pp. 105-123.
- SEVERINO, A. J., 2002, *Metodologia do Trabalho Científico*. 22a. edição, Cortez.

- SOUZA, J. M; OLIVEIRA, J. ; RODRIGUES, S. ; MIRANDA, R. . GCC: An Environment for Knowledge Management in Scientific Research and Higher Education Centers. In: I - Know ' 05, 2005, Graz. Proceedings of I - Know ' 05, 2005.
- SRH-RN, 2004, "Outorga de Águas - Estado do Rio Grande do Norte". In: [http://www.ambientebrasil.com.br/composer.php3?base=./estadual/index.html&conteudo=./estadual/rn\\_aguas.html](http://www.ambientebrasil.com.br/composer.php3?base=./estadual/index.html&conteudo=./estadual/rn_aguas.html), Accessed in 11/10/2004.
- STANDFORD UNIVERSITY, 2003, "The Protégé Ontology Editor and Knowledge Acquisition System". In: <http://protege.stanford.edu/>, Accessed in 2003.
- STEINFELD, C., JANG, C., PTAFF, B., 1999, "Supporting virtual team collaboration: The TeamSCOPE system", Phoenix.
- SUN, 2003a, "APPLETS". In: <http://java.sun.com/applets/>, Accessed in 2003a.
- SUN, 2004, "JavaMail". In: <http://java.sun.com/products/javamail/>, Accessed in 2005.
- SUN, 2003b, "Java". In: <http://java.sun.com/>, Accessed in 2003b.
- SUN, 2003c, "Java Server Pages (JSP)". In: <http://java.sun.com/products/jsp/>, Accessed in 2003c.
- SVEIBY, K. E., 2004, "Tacit Knowledge". In: <http://www.sveiby.com/articles/Polanyi.html>, Accessed in 03/12/2003.
- TIJBOY, A. V.,MAÇADA, D. L., 1997, "Cooperação e Colaboração: O Nosso Conceito". In: <http://www.niee.ufrgs.br/cursos/topicos-ie/ana/conceito.htm>, Accessed in 02/02/2004.
- TIWANA, A., 2000, *The Knowledge Management Toolkit, A Practical Techniques for Building a Knowledge Management System. 2*, Prentice Hall.
- TUCKER, M.,DOMBIAK, G., 2004, "Open Source Jivi Software: Smack API". In: <http://www.jivesoftware.org/smack/>, Accessed in 2004.
- UNIVALE, 2004, "Bacia hidrográfica". In: <http://g10.cttmar.univali.br/tijucas/baciahidrografica.htm>, Accessed in 2004.
- USCHOLD, M.,GRUNINGER, M., 1996, "Ontologies: Principles, Methods and Applications", *Knowledge Engineering Review*, v. 11, n. 2, pp. 93-155.
- VON WANGENHEIM, A.,VON WANGENHEIM, C. G., 2003a, "Recuperação de Casos", *Raciocínio Baseado em Casos*, 1 ed., chapter 6, Brasil, Manole.
- VON WANGENHEIM, A.,VON WANGENHEIM, C. G., 2003b, "Representação de Casos", *Raciocínio Baseado em Casos*, 1 ed., chapter 4, Brasil, Manole.
- WAINER, J., WESKE, M., MEDEIROS, C. M. B., 1996, "Scientific workflow Management". In: *Proceedings of the NSF Workshop on Workflow Process Automation: State-of-the-art and Future Directions*, Georgia.

WOLFS, F., 1996, "Introduction to the Scientific Method". In:  
[http://teacher.nsr1.rochester.edu/PHY\\_LABS/Labs.html](http://teacher.nsr1.rochester.edu/PHY_LABS/Labs.html), Accessed in 1996.

ZHAO, J. L., 1998, "Knowledge Management and Organizational Learning in Workflow Systems". In: *AIS Americas Conference on Information Systems*, Maryland, USA.

## APÊNDICE A – Sistema de Gestão do Conhecimento Científico (GCC)

A motivação para o desenvolvimento do GCC – Sistema de Gestão de Conhecimento Científico (SOUZA; OLIVEIRA et al, 2005) é originária da observação das características de instituições de pesquisa, como Universidades e Centros de Pesquisa, que possuem estruturas descentralizadas, atividades fortemente associadas ao manuseio, transformação e criação de conhecimentos e onde são utilizados dados com alta complexidade. O GCC é um ambiente WEB que tem como objetivo prover meios para a gestão de conhecimentos nestas organizações, amenizando estas dificuldades.

### Arquitetura do GCC

A arquitetura do GCC é apresentada na Figura 1 e será detalhada nas seções seguintes.

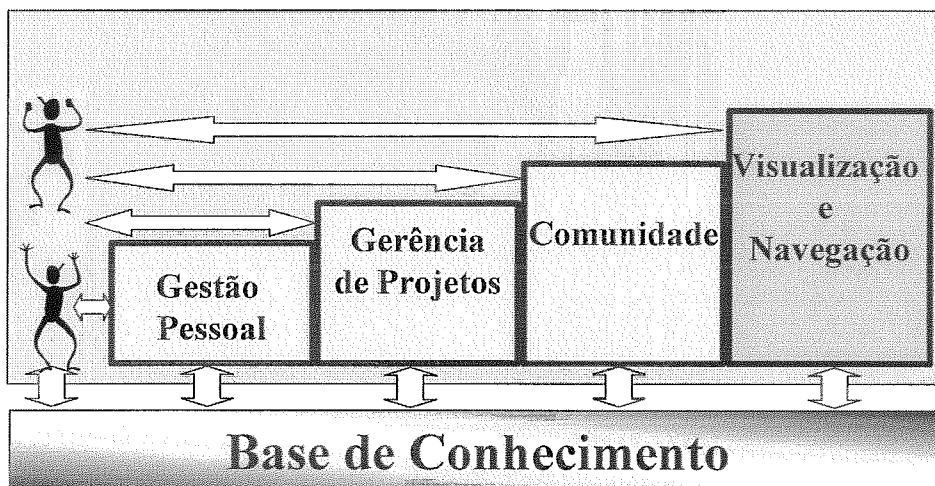


Figura 24 - Arquitetura do GCC (SOUZA; OLIVEIRA et al, 2005)

### Serviços de Gestão de Conhecimento Pessoal

Este módulo provê funções para um pesquisador gerenciar seu próprio conhecimento, assim como suas informações pessoais. Abaixo alguns dos serviços disponibilizados.

**Curriculum Vitae** – É uma das formas de manter informações sobre os profissionais da organização. Dentre as informações armazenadas no GCC estão o

nome, endereço, telefone, *homepage*, experiência profissional e acadêmica, proficiência em idiomas, produção científica, dentre outras. Adicionalmente, o pesquisador deve indicar quais as áreas do conhecimento em que atua e quais são suas competências e habilidades dentro de cada área. Além disto, o pesquisador pode identificar áreas nas quais possui interesse, mas que não se considera ainda especialista.

**Blog Pessoal** – *Weblogs* podem ser vistos como páginas pessoais, no GCC ele atua como mais uma ferramenta para propiciar a gestão do conhecimento pessoal. Em geral, estas ferramentas possuem as seguintes características:

- Edição pessoal: O conteúdo do site é de responsabilidade da própria pessoa, no contexto do GCC um pesquisador, sendo que o *Weblog* reflete alguns tópicos do perfil deste pesquisador;

- Estrutura de *Hyperlinks* - O conteúdo de um *Weblog* consiste tipicamente em pequenas submissões com *links* associados a conteúdo externo ao site. A seleção do material a ser vinculado também é de responsabilidade do editor. No GCC, geralmente são feitas associações com a descrição de experimentos executados, descrição de lições aprendidas dentre outros tipos de informações importantes para a prática científica;

- Um primeiro passo antes da criação de uma comunidade – A verificação de *Weblogs* permite a identificação de grupos de pesquisadores com interesses comuns.

No GCC, cada pesquisador pode ter um *Weblog*, e os trechos de texto adicionados podem ser privativos (acesso apenas para o autor) ou públicos. Além disto, as ferramentas de mineração de competência também analisa os dados do *Weblog* para alimentação automática do perfil de cada um dos usuários.

**Mapas Mentais** – No GCC, os pesquisadores podem construir mapas mentais que são utilizados para elucidação de conceitos, auxiliar as sessões de Brainstorm e simplificar as discussões realizadas entre pesquisadores geograficamente distribuídos.

## Serviços para o gerenciamento de processos

Neste módulo são providos serviços para a definição e execução de projetos. Os responsáveis pela coordenação de um projeto criam o modelo de processo, com a seqüência de atividades e os insumos necessários para cada uma destas atividades: áreas de conhecimento do CNPq, competências, documentos, artefatos. Esta atividade

é desempenhada através da ferramenta gráfica de *workflow*. Já a execução dos processos é controlada pela máquina de *workflow*.

## **Serviços de administração de comunidades**

Um dos focos do desenvolvimento do GCC é a criação de comunidades, isto é, grupos de pesquisadores com interesses comuns que então podem trocar informações e trabalhar de forma colaborativa. Este módulo é constituído de um conjunto de ferramentas desenvolvidas com o propósito de melhorar a interação entre as pessoas de uma comunidade, como: votação, fórum, portal de novidades e encontros virtuais pré-agendados. Todas as informações e conversações realizadas são armazenadas, já que são que representam uma forma de explicitação de conhecimento que pode ser consultada e analisada no futuro.

## **Ferramentas para a visualização do conhecimento e serviços de navegação**

Além das ferramentas usuais para visualização de informações, como relatórios e formulários, o sistema conta com duas ferramentas adicionais para navegação e visualização, desenvolvidas com propósito de facilitar a utilização do sistema, tornando o seu uso mais intuitivo. São elas: A Arvore Hiperbólica e o Mapa Conceitual dos projetos.

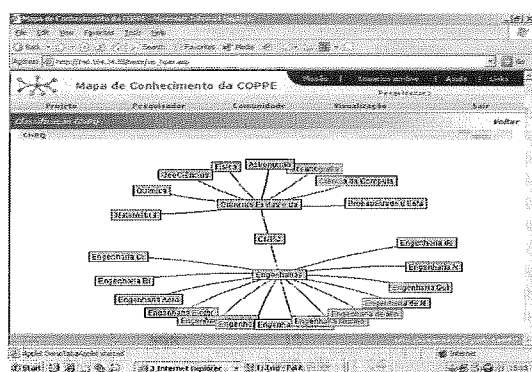
**Arvore Hiperbólica** – Esta é uma interface de visualização de uma estrutura hierárquica em árvore e possui as seguintes características:

Os nós ou componentes da árvore ficam menores na medida em que se afastam do centro, ressaltando as informações que estão sendo consultadas pelo usuário.

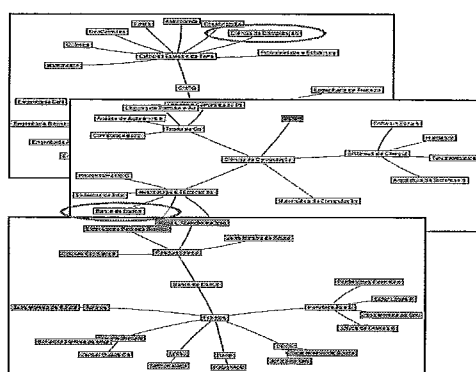
O número de nodos ou componentes cresce exponencialmente ao se navegar dos nodos pais para os nodos filhos.

A figura 25 apresenta a árvore hiperbólica do GCC, onde se visualiza as áreas de conhecimentos de acordo com o CNPq. Um usuário pode navegar pela classificação e consultar os projetos, competências e outras informações associadas com a área de conhecimento selecionada.

**Mapa Conceitual do projeto** - Prove uma visualização de um projeto como um mapa mental facilitando a visualização das informações.



*Árvore Hiperbólica com Áreas de Conhecimento do CNPq*



*Navegação pela Árvore Hiperbólica*

**Figura 25 - Árvore Hiperbólica**

## Serviços para a descoberta de competências

As ferramentas providas por este modelo são úteis para permitir uma melhor identificação das competências de cada pesquisador e também estratégias de busca para a localização de especialistas com competências específicas para participar de um projeto ou substituir outro especialista indisponível em uma atividade.

Uma forma de se conseguir este objetivo é através da análise dos documentos criados, editados ou acessados pelos pesquisadores. Os serviços providos são:

**SMiner** – A função deste dispositivo é minerar as competências baseado em textos e publicações. Atualmente o Sminer trata palavras em inglês e em português.

**Busca de Competências** – A busca tem o objetivo de identificar pesquisadores com competências desejadas ou com competências similares a estas. O processo de busca respeita a seguinte prioridade:

- Competências declaradas pelo próprio pesquisador;
- Competências associadas aos projetos em que os pesquisadores participam;
- Competências extraídas dos textos através da ferramenta Sminer;
- Competências associadas às comunidades que os pesquisadores participam.

## APÊNDICE B – Ontologia desenvolvida

Neste Apêndice é apresentada a Ontologia desenvolvida. Para gerar esta representação foi utilizada a linguagem OWL (Web Ontology Language). A geração se deu através de *plugin* anexado ao software protégé.

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/unnamed.owl#"
  xml:base="http://www.owl-ontologies.com/unnamed.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="MANN-WHITNEY">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="TESTE_DE_CONSISTENCIA"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="PARADA">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="AGUA_DOCE"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="MATA_ATLANTICA">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="VEGETACAO"/>
    </rdfs:subClassOf>
    <rdfs:label
      rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >MATA ATLANTICA</rdfs:label>
  </owl:Class>
  <owl:Class rdf:ID="MATEMATICO">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="MODELO"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="TEMPO_DE_RETORNO">
    <rdfs:label
      rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >TEMPO DE RETORNO</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="SERIES_HIDROLOGICAS"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="PECUARIA">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="USUARIO_DA_AGUA"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="FLUVIOMETRICO">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="FENOMENOS"/>
    </rdfs:subClassOf>
  </owl:Class>
</rdf:RDF>
```



```

<owl:Class rdf:ID="SUPER_UMIDO">
  <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Sem seca</rdfs:comment>
  <rdfs:subClassOf>
    <owl:Class rdf:ID="QUENTE"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Class rdf:ID="MESOTERMICO_BRANDO"/>
  </rdfs:subClassOf>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >SUPER UMIDO</rdfs:label>
  <rdfs:subClassOf>
    <owl:Class rdf:ID="SUB_QUENTE"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="SERIE_DE_DURACAO_PARCIAL">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >SERIE DE DURACAO PARCIAL</rdfs:label>
  <rdfs:subClassOf>
    <owl:Class rdf:about="#SERIES_HIDROLOGICAS"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="INTERPOLACAO">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="PREENCHIMENTO_DE_FALHAS"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="TRECHO_DE_RIO">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >TRECHO DE RIO</rdfs:label>
</owl:Class>
<owl:Class rdf:about="#AGUA_DOCE">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="REGIAO_AQUATICA"/>
  </rdfs:subClassOf>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >AGUA DOCE</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="CAMPINARAMA">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#VEGETACAO"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="SOLO">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="CARACTERISTICA_AMBIENTAL"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="REGIAO_TERRESTRE">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >REGIAO TERRESTRE</rdfs:label>
  <rdfs:subClassOf>
    <owl:Class rdf:ID="REGIAO_GEOGRAFICA"/>
  </rdfs:subClassOf>
</owl:Class>

```

```

<owl:Class rdf:ID="REGRESSAO_LINEAR_SIMPLES">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#PREENCHIMENTO_DE_FALHAS"/>
  </rdfs:subClassOf>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >REGRESSAO LINEAR SIMPLES</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="UNIVERSIDADES">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="PESQUISA"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="LAGOA">
  <rdfs:subClassOf rdf:resource="#PARADA"/>
</owl:Class>
<owl:Class rdf:about="#MESOTERMICO_BRANDO">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="CLIMA"/>
  </rdfs:subClassOf>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >MESOTERMICO BRANDO</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="ILHA">
  <rdfs:subClassOf rdf:resource="#REGIAO_TERRESTRE"/>
</owl:Class>
<owl:Class rdf:ID="OCEANO">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="AGUA_SALGADA"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="#CLIMA">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#CARACTERISTICA_AMBIENTAL"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="HIDROLOGICO">
  <rdfs:subClassOf rdf:resource="#MATEMATICO"/>
</owl:Class>
<owl:Class rdf:ID="TEMPESTADE">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="PLUVIOMETRICO"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="METODO_DOS_MOMENTOS">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >METODO DOS MOMENTOS</rdfs:label>
  <rdfs:subClassOf>
    <owl:Class rdf:ID="DISTRIBUICAO_DE_PROBABILIDADE_FDP"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="VOCOROCA">
  <rdfs:subClassOf rdf:resource="#FLUVIOMETRICO"/>
</owl:Class>
<owl:Class rdf:ID="TRANSPORTE">
  <rdfs:subClassOf rdf:resource="#MATEMATICO"/>
</owl:Class>
<owl:Class rdf:ID="CENTRO-OESTE">
  <rdfs:subClassOf>

```

```

    <owl:Class rdf:ID="ESTADO"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="CLIMATICOS">
  <rdfs:subClassOf rdf:resource="#FENOMENOS"/>
</owl:Class>
<owl:Class rdf:about="#USUARIO_DA_AGUA">
  <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Pessoas físicas ou jurídicas que utilizam a água de vertedouros
de água</rdfs:comment>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >USUARIO DA AGUA</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="USINA_HIDROELETRICA">
  <rdfs:subClassOf rdf:resource="#USUARIO_DA_AGUA"/>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >USINA HIDROELETRICA</rdfs:label>
</owl:Class>
<owl:Class rdf:about="#QUENTE">
  <rdfs:subClassOf rdf:resource="#CLIMA"/>
  <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >MEDIA MAIRO QUE 18 GRAUS</rdfs:comment>
</owl:Class>
<owl:Class rdf:ID="SERIE_DE_DURACAO_ANUAL">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#SERIES_HIDROLOGICAS"/>
  </rdfs:subClassOf>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >SERIE DE DURACAO ANUAL</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="AREA_GEOGRAFICA">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >AREA GEOGRAFICA</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="PAIS">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="AREA_POLITICA"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="INDUSTRIA">
  <rdfs:subClassOf rdf:resource="#USUARIO_DA_AGUA"/>
</owl:Class>
<owl:Class rdf:ID="SECA">
  <rdfs:subClassOf rdf:resource="#CLIMATICOS"/>
</owl:Class>
<owl:Class rdf:ID="BACIA_HIDROGRAFICA">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >BACIA HIDROGRAFICA</rdfs:label>
  <rdfs:subClassOf rdf:resource="#AREA_GEOGRAFICA"/>
</owl:Class>
<owl:Class rdf:ID="EVAPORACAO">
  <rdfs:subClassOf rdf:resource="#MATEMATICO"/>
</owl:Class>
<owl:Class rdf:ID="SUDESTE">

```

```

    <rdfs:subClassOf>
      <owl:Class rdf:about="#ESTADO"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="VALORES_DUPLAMENTE_ACUMULADOS">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >VALORES DUPLAMENTE ACUMULADOS</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#TESTE_DE_CONSISTENCIA"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="SEMI_ARIDO">
    <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >6 A 8 MESES SECOS</rdfs:comment>
    <rdfs:subClassOf rdf:resource="#QUENTE"/>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >SEMI ARIDO</rdfs:label>
  </owl:Class>
  <owl:Class rdf:about="#TESTE_DE_CONSISTENCIA">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >TESTE DE CONSISTENCIA</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#SERIES_HIDROLOGICAS"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="CERRADO">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#VEGETACAO"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="MAPAS">
    <rdfs:subClassOf rdf:resource="#MODELO"/>
  </owl:Class>
  <owl:Class rdf:ID="FERRAMENTAS"/>
  <owl:Class rdf:about="#PREENCHIMENTO_DE_FALHAS">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >PREENCHIMENTO DE FALHAS</rdfs:label>
    <rdfs:subClassOf rdf:resource="#MODELO"/>
  </owl:Class>
  <owl:Class rdf:ID="SISTEMAS_DE_COMPUTADOR">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >SISTEMAS DE COMPUTADOR</rdfs:label>
    <rdfs:subClassOf rdf:resource="#FERRAMENTAS"/>
  </owl:Class>
  <owl:Class rdf:ID="NEVOA">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#PLUVIOMETRICO"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="HIDROGRAFIA">
    <rdfs:subClassOf rdf:resource="#MAPAS"/>
  </owl:Class>
  <owl:Class rdf:ID="INSTITUTOS">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#PESQUISA"/>
    </rdfs:subClassOf>
  </owl:Class>

```

```

    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="AGENCIAS_REGULADORAS">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="REGULADORAS"/>
    </rdfs:subClassOf>
    <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Orgaos publicos que tem como objetivo regular o uso da agua em
rios, lagos e oceanos</rdfs:comment>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >AGENCIAS REGULADORAS</rdfs:label>
  </owl:Class>
  <owl:Class rdf:ID="RIO">
    <rdfs:subClassOf
rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="CORRENTE"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="FLORESTA_ABERTA">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >FLORESTA ABERTA</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#VEGETACAO"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#REGULADORAS">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="INSTITUICOES"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="APARELHOS_DE_MEDICAO">
    <rdfs:subClassOf rdf:resource="#FERRAMENTAS"/>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >APARELHOS DE MEDICAO</rdfs:label>
  </owl:Class>
  <owl:Class rdf:ID="INDICE_DE_CHEIAS">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >INDICE DE CHEIAS</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#SERIES_HIDROLOGICAS"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#REGIAO_AQUATICA">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >REGIAO AQUATICA</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#REGIAO_GEOGRAFICA"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="CIDADE">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#AREA_POLITICA"/>
    </rdfs:subClassOf>
  </owl:Class>

```

```

<owl:Class rdf:about="#SUB_QUENTE">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >SUB QUENTE</rdfs:label>
  <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >MEDIA ENTRE 15 E 18 GRAUS</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#CLIMA"/>
</owl:Class>
<owl:Class rdf:ID="CHUVA">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#PLUVIOMETRICO"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="INTERIOR">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="CONTINENTE"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="MOMENTO_LINEARES_E_MISTOS">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#DISTRIBUICAO_DE_PROBABILIDADE__FDP_"/>
  </rdfs:subClassOf>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >MOMENTO LINEARES E MISTOS</rdfs:label>
</owl:Class>
<owl:Class rdf:about="#CORRENTE">
  <rdfs:subClassOf rdf:resource="#AGUA_DOCE"/>
</owl:Class>
<owl:Class rdf:ID="FLUXO">
  <rdfs:subClassOf rdf:resource="#MATEMATICO"/>
</owl:Class>
<owl:Class rdf:ID="COMITE_DE_BACIA">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >COMITE DE BACIA</rdfs:label>
  <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Grupo de pessoas responsavel pela gestao e controle dos recursos
e problemas existentes em uma bacia</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#REGULADORAS"/>
</owl:Class>
<owl:Class rdf:about="#VEGETACAO">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#CARACTERISTICA_AMBIENTAL"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="MAPA_DO_SOLO">
  <rdfs:subClassOf rdf:resource="#MAPAS"/>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >MAPA DO SOLO</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="RESIDUOS_ACUMULADOS">
  <rdfs:subClassOf rdf:resource="#TESTE_DE_CONSISTENCIA"/>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >RESIDUOS ACUMULADOS</rdfs:label>
</owl:Class>
<owl:Class rdf:about="#PLUVIOMETRICO">

```

```

    <rdfs:subClassOf rdf:resource="#FENOMENOS"/>
  </owl:Class>
  <owl:Class rdf:ID="AGRICULTURA">
    <rdfs:subClassOf rdf:resource="#USUARIO_DA_AGUA"/>
  </owl:Class>
  <owl:Class rdf:ID="TAU_DE_KENDALL">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >TAU DE KENDALL</rdfs:label>
    <rdfs:subClassOf rdf:resource="#TESTE_DE_CONSISTENCIA"/>
  </owl:Class>
  <owl:Class rdf:ID="UMIDO">
    <rdfs:subClassOf rdf:resource="#QUENTE"/>
    <rdfs:subClassOf rdf:resource="#MESOTERMICO_BRANDO"/>
    <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >UM A TRES MESES SECO</rdfs:comment>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="MESOTERMICO_MEDIANO"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#AGUA_SALGADA">
    <rdfs:subClassOf rdf:resource="#REGIAO_AQUATICA"/>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >AGUA SALGADA</rdfs:label>
  </owl:Class>
  <owl:Class rdf:about="#CONTINENTE">
    <rdfs:subClassOf rdf:resource="#REGIAO_TERRESTRE"/>
  </owl:Class>
  <owl:Class rdf:ID="FLORESTA_MISTA">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >FLORESTA MISTA</rdfs:label>
    <rdfs:subClassOf rdf:resource="#VEGETACAO"/>
  </owl:Class>
  <owl:Class rdf:ID="MAXIMA_VEROSIMELHANCA">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >MAXIMA VEROSIMELHANCA</rdfs:label>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#DISTRIBUICAO_DE_PROBABILIDADE_FDP"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="NORTE">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ESTADO"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="SUBSTANCIAS_POLUIDORAS">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >SUBSTANCIAS POLUIDORAS</rdfs:label>
  </owl:Class>
  <owl:Class rdf:about="#DISTRIBUICAO_DE_PROBABILIDADE_FDP">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#SERIES_HIDROLOGICAS"/>
    </rdfs:subClassOf>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >DISTRIBUICAO DE PROBABILIDADE (FDP)</rdfs:label>
  </owl:Class>

```

```

</owl:Class>
<owl:Class rdf:ID="SEMI_UMIDO">
  <rdfs:subClassOf rdf:resource="#MESOTERMICO_BRANDO"/>
  <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >4 A 5 MESES SECOS</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#SUB_QUENTE"/>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >SEMI_UMIDO</rdfs:label>
  <rdfs:subClassOf rdf:resource="#QUENTE"/>
</owl:Class>
<owl:Class rdf:ID="NORDESTE">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#ESTADO"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="PICTOGRAFICO">
  <rdfs:subClassOf rdf:resource="#MODELO"/>
</owl:Class>
<owl:Class rdf:ID="INUNDACAO">
  <rdfs:subClassOf rdf:resource="#FLUVIOMETRICO"/>
</owl:Class>
<owl:Class rdf:ID="CAATINGA">
  <rdfs:subClassOf rdf:resource="#VEGETACAO"/>
</owl:Class>
<owl:Class rdf:about="#PESQUISA">
  <rdfs:subClassOf rdf:resource="#INSTITUICOES"/>

</owl:Class>
<owl:Class rdf:ID="INTRUSAO_SALINA">
  <rdfs:subClassOf rdf:resource="#FLUVIOMETRICO"/>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >INTRUSAO_SALINA</rdfs:label>
</owl:Class>
<owl:Class rdf:about="#MESOTERMICO_MEDIANO">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >MESOTERMICO_MEDIANO</rdfs:label>
  <rdfs:subClassOf rdf:resource="#CLIMA"/>
</owl:Class>
<owl:Class rdf:ID="FLUXO_E_TRANSPORTE">
  <rdfs:subClassOf rdf:resource="#MATEMATICO"/>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >FLUXO E TRANSPORTE</rdfs:label>
</owl:Class>
<owl:Class rdf:about="#REGIAO_GEOGRAFICA">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >REGIAO_GEOGRAFICA</rdfs:label>
  <rdfs:subClassOf rdf:resource="#AREA_GEOGRAFICA"/>
</owl:Class>
<owl:Class rdf:ID="FLORESTA_DENSA">
  <rdfs:subClassOf rdf:resource="#VEGETACAO"/>
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >FLORESTA_DENSA</rdfs:label>
</owl:Class>
<owl:Class rdf:ID="PARAMETROS_QUALIDADE_DAGUA">

```



```

    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >PARAMETROS QUALIDADE DAGUA</rdfs:label>
  </owl:Class>
  <owl:Class rdf:ID="BALANCO_HIDRICO">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >BALANCO HIDRICO</rdfs:label>
    <rdfs:subClassOf rdf:resource="#MATEMATICO"/>
  </owl:Class>
  <owl:Class rdf:ID="LITORAL">
    <rdfs:subClassOf rdf:resource="#CONTINENTE"/>
  </owl:Class>
  <owl:Class rdf:ID="SUL">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ESTADO"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="ASSORIAMENTO">
    <rdfs:subClassOf rdf:resource="#FLUVIOMETRICO"/>
  </owl:Class>
  <owl:Class rdf:ID="QUALIDADE">
    <rdfs:subClassOf rdf:resource="#MATEMATICO"/>
  </owl:Class>
  <owl:Class rdf:ID="LAGO">
    <rdfs:subClassOf rdf:resource="#PARADA"/>
  </owl:Class>
  <owl:Class rdf:about="#SERIES_HIDROLOGICAS">
    <rdfs:subClassOf rdf:resource="#MODELO"/>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >SERIES HIDROLOGICAS</rdfs:label>
  </owl:Class>
  <owl:Class rdf:about="#AREA_POLITICA">
    <rdfs:subClassOf rdf:resource="#AREA_GEOGRAFICA"/>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >AREA POLITICA</rdfs:label>
  </owl:Class>
  <owl:Class rdf:about="#ESTADO">
    <rdfs:subClassOf rdf:resource="#AREA_POLITICA"/>
  </owl:Class>
  <owl:Class rdf:about="#CARACTERISTICA_AMBIENTAL">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >CARACTERISTICA AMBIENTAL</rdfs:label>
  </owl:Class>
  <owl:Class rdf:ID="MOVE">
    <rdfs:subClassOf rdf:resource="#PREENCHIMENTO_DE_FALHAS"/>
  </owl:Class>
  <owl:Class rdf:ID="EMPRESAS">
    <rdfs:subClassOf rdf:resource="#INSTITUICOES"/>
  </owl:Class>
  <owl:ObjectProperty rdf:ID="AFLUENTES">
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
          <owl:Class rdf:about="#RIO"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
  </owl:ObjectProperty>

```

```

    </owl:Class>
  </rdfs:domain>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="PARTE"/>
<owl:ObjectProperty rdf:ID="PARTE_DE_RIO">
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
          <owl:Class rdf:about="#TRECHO_DE_RIO"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >PARTE DE RIO</rdfs:label>
  </owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="PARTE_DE_BACIA">
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
          <owl:Class rdf:about="#RIO"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >PARTE DE BACIA</rdfs:label>
  </owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="AMONTANTE">
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
          <owl:Class rdf:about="#TRECHO_DE_RIO"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >MODELO IMPLEMENTADO</rdfs:label>
  </owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="MODELO_IMPLEMENTADO">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >MODELO IMPLEMENTADO</rdfs:label>
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
          <owl:Class rdf:about="#SISTEMAS_DE_COMPUTADOR"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
  </owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="PARTE_DE">
  <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >PARTE DE</rdfs:label>

```

```

</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="AJUSANTE">
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
          <owl:Class rdf:about="#TRECHO_DE_RIO"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID="PASSA_POR_CIDADE">
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
            <owl:Class rdf:about="#TRECHO_DE_RIO"/>
          </owl:unionOf>
        </owl:Class>
      </rdfs:domain>
      <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
>PASSA POR CIDADE</rdfs:label>
    </owl:ObjectProperty>
    <owl:DatatypeProperty rdf:ID="CAPTACAO">
      <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
      <rdfs:domain>
        <owl:Class>
          <owl:unionOf rdf:parseType="Collection">
            <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
              <owl:Class rdf:about="#USUARIO_DA_AGUA"/>
            </owl:unionOf>
          </owl:Class>
        </rdfs:domain>
      </owl:DatatypeProperty>
      <owl:DatatypeProperty rdf:ID="CONSUMO">
        <rdfs:domain>
          <owl:Class>
            <owl:unionOf rdf:parseType="Collection">
              <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
                <owl:Class rdf:about="#USUARIO_DA_AGUA"/>
              </owl:unionOf>
            </owl:Class>
          </rdfs:domain>
          <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
        </owl:DatatypeProperty>
        <owl:DatatypeProperty rdf:ID="EXTENSAO">
          <rdfs:domain>
            <owl:Class>
              <owl:unionOf rdf:parseType="Collection">
                <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
                  <owl:Class rdf:about="#RIO"/>
                  <owl:Class rdf:about="#TRECHO_DE_RIO"/>
                </owl:unionOf>
              </owl:Class>
            </rdfs:domain>
          </owl:DatatypeProperty>
        </owl:DatatypeProperty>
      </owl:DatatypeProperty>
    </owl:DatatypeProperty>
  </owl:DatatypeProperty>

```

```

<rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
>EXTENSÃO DA NASCENTE A FOZ DO RIO EM KILOMETROS</rdfs:comment>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="PROFUNDIDADE">
<rdfs:domain>
<owl:Class>
<owl:unionOf rdf:parseType="Collection">
<rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
<owl:Class rdf:about="#TRECHO_DE_RIO"/>
</owl:unionOf>
</owl:Class>
</rdfs:domain>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="CRIADOR">
<rdfs:domain>
<owl:Class>
<owl:unionOf rdf:parseType="Collection">
<rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
<owl:Class rdf:about="#MODELO"/>
<owl:Class rdf:about="#FERRAMENTAS"/>
</owl:unionOf>
</owl:Class>
</rdfs:domain>
<rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="NOME">
<rdfs:domain>
<owl:Class>
<owl:unionOf rdf:parseType="Collection">
<rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
<owl:Class rdf:about="#AREA_GEOGRAFICA"/>
<owl:Class rdf:about="#TRECHO_DE_RIO"/>
<owl:Class rdf:about="#MODELO"/>
<owl:Class rdf:about="#FENOMENOS"/>
<owl:Class rdf:about="#FERRAMENTAS"/>
<owl:Class rdf:about="#USUARIO_DA_AGUA"/>
<owl:Class rdf:about="#PARAMETROS_QUALIDADE_DAGUA"/>
<owl:Class rdf:about="#SUBSTANCIAS_POLUIDORAS"/>
</owl:unionOf>
</owl:Class>
</rdfs:domain>
<rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="AREA">
<rdfs:domain>
<owl:Class>
<owl:unionOf rdf:parseType="Collection">
<rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
<owl:Class rdf:about="#PECUARIA"/>
</owl:unionOf>
</owl:Class>
</rdfs:domain>

```

```

    <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Area que cop</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="Ontologia_Slot_2">
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    </owl:DatatypeProperty>
    <owl:DatatypeProperty rdf:ID="Ontologia_Slot_3">
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    </owl:DatatypeProperty>
    <owl:DatatypeProperty rdf:ID="QUANTIDADE_FUNCIONARIOS">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >QUANTIDADE_FUNCIONARIOS</rdfs:label>
    <rdfs:domain>
    <owl:Class>
    <owl:unionOf rdf:parseType="Collection">
    <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
    <owl:Class rdf:about="#INDUSTRIA"/>
    </owl:unionOf>
    </owl:Class>
    </rdfs:domain>

    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="AREA_PLANTADA">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >AREA_PLANTADA</rdfs:label>
    <rdfs:domain>
    <owl:Class>
    <owl:unionOf rdf:parseType="Collection">
    <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
    <owl:Class rdf:about="#AGRICULTURA"/>
    </owl:unionOf>
    </owl:Class>
    </rdfs:domain>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="PRECO">
    <rdfs:domain>
    <owl:Class>
    <owl:unionOf rdf:parseType="Collection">
    <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
    <owl:Class rdf:about="#FERRAMENTAS"/>
    </owl:unionOf>
    </owl:Class>
    </rdfs:domain>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="DATA_DE_CRIACAO">
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"

```

```

>DATA DE CRIACAO</rdfs:label>
<rdfs:domain>
  <owl:Class>
    <owl:unionOf rdf:parseType="Collection">
      <rdfs:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
      <owl:Class rdf:about="#MODELO"/>
    </owl:unionOf>
  </owl:Class>
</rdfs:domain>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="VAZAO">
  <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  <rdfs:domain>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <rdfs:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
        <owl:Class rdf:about="#TRECHO_DE_RIO"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:domain>
  <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >A VAZAO VARIA NA EXTENSAO DO RIO, NOS TRECHOS E POSSSIVEL
ESTIPULAR UMA VAZAO VALIDA PELO MENOS PARA AQUELE
TRECHO</rdfs:comment>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="PIB">
    <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Valores em dolares do PIB da regioao</rdfs:comment>
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <rdfs:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
          <owl:Class rdf:about="#AREA_POLITICA"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#float"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="POPULACAO">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <rdfs:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
          <owl:Class rdf:about="#AREA_POLITICA"/>
          <owl:Class rdf:about="#REGIAO_TERRESTRE"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="AREA_TOTAL">
    <rdfs:comment
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"

```

```

    >Area geografica da entidade serve para o reconhecimento do
tamanho da entidade</rdfs:comment>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >AREA TOTAL</rdfs:label>
    <rdfs:domain>
        <owl:Class>
            <owl:unionOf rdf:parseType="Collection">
                <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
                <owl:Class rdf:about="#AREA_POLITICA"/>
                <owl:Class rdf:about="#REGIAO_TERRESTRE"/>
            </owl:unionOf>
        </owl:Class>
    </rdfs:domain>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:ID="PROPOSITO">
    <rdfs:domain>
        <owl:Class>
            <owl:unionOf rdf:parseType="Collection">
                <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
                <owl:Class rdf:about="#MODELO"/>
            </owl:unionOf>
        </owl:Class>
    </rdfs:domain>
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    </owl:DatatypeProperty>
    <owl:DatatypeProperty rdf:ID="DILUICAO">
    <rdfs:domain>
        <owl:Class>
            <owl:unionOf rdf:parseType="Collection">
                <rdf:Description
rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
                <owl:Class rdf:about="#USUARIO_DA_AGUA"/>
            </owl:unionOf>
        </owl:Class>
    </rdfs:domain>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
</owl:DatatypeProperty>
<FLUXO rdf:ID="DYNHYD5"/>
<FLUXO_E_TRANSPORTE rdf:ID="HSPF"/>
<TRANSPORTE rdf:ID="SED-2D"/>
<TRANSPORTE rdf:ID="HEC-HMS"/>
<FLUXO_E_TRANSPORTE rdf:ID="MIKE_3">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >MIKE 3</rdfs:label>
</FLUXO_E_TRANSPORTE>
<FLUXO_E_TRANSPORTE rdf:ID="QUAL2E"/>
<FLUXO rdf:ID="FEQ"/>
<FLUXO rdf:ID="Ontologia_Instance_1">
    <NOME rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >RACIONAL</NOME>
</FLUXO>
<TRANSPORTE rdf:ID="WASP5"/>
<FLUXO rdf:ID="HEC-RAS"/>
<FLUXO_E_TRANSPORTE rdf:ID="PRMS"/>
<FLUXO_E_TRANSPORTE rdf:ID="MIKE_11">

```

```

    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >MIKE 11</rdfs:label>
</FLUXO_E_TRANSPORTE>
<TRANSPORTE rdf:ID="RMA4"/>
<FLUXO rdf:ID="CH3D-WES"/>
<PARAMETROS_QUALIDADE_DAGUA rdf:ID="ph">
    <NOME rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >ph</NOME>
</PARAMETROS_QUALIDADE_DAGUA>
<PARAMETROS_QUALIDADE_DAGUA rdf:ID="DBO">
    <NOME rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >DBO</NOME>
</PARAMETROS_QUALIDADE_DAGUA>
<FLUXO_E_TRANSPORTE rdf:ID="CE-QUAL-RIV1"/>
<FLUXO_E_TRANSPORTE rdf:ID="EFDC_HEM3D">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >EFDC/HEM3D</rdfs:label>
</FLUXO_E_TRANSPORTE>
<FLUXO_E_TRANSPORTE rdf:ID="CE-QUAL-W2"/>
<TRANSPORTE rdf:ID="CE-QUAL-R1"/>
<TRANSPORTE rdf:ID="OTEQ"/>
<FLUXO rdf:ID="DR3M"/>
<FLUXO rdf:ID="BRANCH">
    <NOME rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >METODO DA CURVA DE PERMANENCIA</NOME>
</FLUXO>
<FLUXO_E_TRANSPORTE rdf:ID="MIKE_SHE">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >MIKE SHE</rdfs:label>
</FLUXO_E_TRANSPORTE>
<TRANSPORTE rdf:ID="BLTM"/>
<EVAPORACAO rdf:ID="Ontologia_Instance_5">
    <NOME rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >TANQUE CLASSE A</NOME>
</EVAPORACAO>
<TRANSPORTE rdf:ID="CE-QUAL-ICM"/>
<FLUXO_E_TRANSPORTE rdf:ID="MIKE_21">
    <rdfs:label
rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >MIKE 21</rdfs:label>
</FLUXO_E_TRANSPORTE>
<FLUXO rdf:ID="Ontologia_Instance_2">
    <NOME rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >METODO DOS FLUTUADORES</NOME>
</FLUXO>
<FLUXO rdf:ID="FESWMS"/>
<TRANSPORTE rdf:ID="OTIS"/>
<FLUXO rdf:ID="Ontologia_Instance_3">
    <NOME rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >METODO DOS MOLINETES</NOME>
</FLUXO>
<FLUXO rdf:ID="DAFLOW"/>
<PARAMETROS_QUALIDADE_DAGUA rdf:ID="DQO">
    <NOME rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >DQO</NOME>
</PARAMETROS_QUALIDADE_DAGUA>
<FLUXO rdf:ID="RMA2"/>
<FLUXO rdf:ID="FourPt"/>

```



</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 2.1, Build 284)  
<http://protege.stanford.edu> -->