

## 4 Estudo de Caso Preliminar

### 4.1. Desenvolvimento do Portal Semântico do Grupo TecComm: Gestão de Conhecimento em Projetos de Pesquisa utilizando Ontologias e Web Services

#### 4.1.1. Introdução

Como já foi visto, a Gestão de Conhecimento vem se tornando um importante fator crítico de sucesso (Sure et al., 2002c) para diversos tipos de organizações, atingindo desde empresas até universidades e grupos de pesquisa. No entanto, a implementação de uma iniciativa de Gestão de Conhecimento é, ao mesmo tempo, uma tarefa complexa e arriscada. O Grupo TecComm (Teccomm) – ligado ao Laboratório de Engenharia de Software da PUC-Rio (LES) – é composto por estudantes de graduação, pós-graduação, professores e doutores que coordenam e executam diversos projetos de pesquisa e desenvolvimento em Tecnologia da Informação. Esses projetos englobam o estado-da-arte em áreas multi-disciplinares como E-Business, E-Learning, Gestão de Conhecimento e Sistemas Multi-Agentes - apenas para citar alguns exemplos - o que exige um alto grau de criação e inovação.

No entanto, sendo um grupo substancialmente formado por estudantes que estão frequentemente completando seus cursos, dissertações e teses, as equipes de pesquisa podem apresentar uma significativa rotatividade, com novos pesquisadores sendo incorporados ao grupo ou deixando o laboratório para ocupar colocações no mercado. Este é um grande risco a ser gerenciado, uma vez que uma rotatividade elevada pode criar um vácuo de conhecimento difícil de ser suprido na velocidade necessária quando algum integrante deixar o grupo. Para resolver esses problemas foi proposta uma abordagem compreensiva e sistemática de Gestão do Conhecimento em Pesquisa e Desenvolvimento, descrita no estudo de caso a seguir.

O estudo de caso apresentado neste trabalho engloba o projeto e desenvolvimento de um Portal Semântico de Conhecimento para Projetos de Pesquisas (Orlean et al., 2003). Este Portal foi proposto para permitir a Gestão de Conhecimento em Pesquisa e Desenvolvimento no contexto do Grupo TecComm e foi construído com base em ontologias e *web services*. Além de ter sido projetado para

permitir o compartilhamento de conhecimento sobre projetos realizados e em andamento entre os pesquisadores do TecComm, era também um requisito muito importante que este portal pudesse ser acessado por entidades autônomas como agentes de software, web services e aplicações, possibilitando a execução automática de diversas tarefas.

#### **4.1.2. O Projeto e Desenvolvimento do Portal Semântico de Conhecimento do Grupo TecComm**

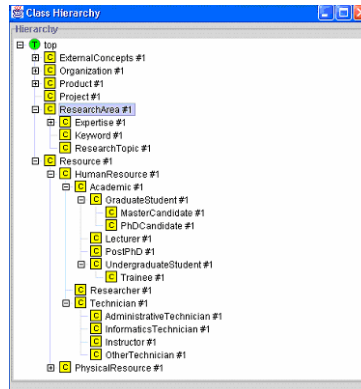
O projeto e o desenvolvimento do Portal Semântico de Conhecimento do Grupo TecComm encontra-se descrito nas próximas sub-seções. A anotação semântica nas páginas do Portal permite que aplicações e agentes de software processem dados e meta-dados sobre seu conteúdo, oferecendo assim serviços aos pesquisadores do Grupo ou mesmo à comunidade de pesquisa em geral. O primeiro passo desse projeto consistiu na criação de uma ontologia que descreve o domínio de Projetos de Pesquisa em Tecnologia da Informação. Esta ontologia foi então mapeada para os modelos que permitiram a instânciação semi-automática de um Portal de Conhecimento através do Framework Portalware (Rocha et al., 2002). Em seguida, foi implementado um Web Service que acessa e disponibiliza as informações semânticas aos agentes e aplicações.

#### **4.1.3. A Ontologia de Projetos de Pesquisa**

O desenvolvimento da ontologia foi iniciado com uma busca por ontologias existentes no domínio da ontologia a ser desenvolvida – Projetos de Pesquisa em Tecnologia da Informação. Esta busca resultou em uma série de conceitos já mapeados por outras ontologias os quais foram reutilizados no contexto da ontologia do TecComm. Os autores dessa ontologia optaram por descrever os conceitos em Português, com o intuito de permitir um melhor entendimento do domínio. Quando foi concluído que a ontologia havia atingido uma versão estável, foi realizada a tradução para o inglês. Todas as versões geradas para a ontologia podem ser consultadas em [www.teccomm.les.inf.puc-rio.br/onts/daml/](http://www.teccomm.les.inf.puc-rio.br/onts/daml/).

Apesar de existirem outras ontologias em domínios que possuem alguma interseção com o domínio de Projetos de Pesquisa em TI optou-se por desenvolver a ontologia desde seu início. Essa opção teve dois motivos principais. Um dos

objetivos do estudo de caso era avaliar as diversas alternativas encontradas ao se desenvolver uma ontologia, além de se praticar um processo de engenharia de ontologias, o que por si só já justificava esta abordagem. Por outro lado, as ontologias encontradas aparentemente não atendiam aos requisitos da ontologia sendo desenvolvida, com exceção de alguns conceitos e relações identificadas - o que foi confirmado posteriormente no projeto.



**Figura 9 - Fragmento da Ontologia para Projetos de Pesquisa em TI do TecComm**

Depois da decisão sobre que conceitos externos deveriam ser reutilizados nesta ontologia e a identificação de quais seriam considerados conceitos-chave, foram elaboradas uma série de questões de competência (Uschold and Gruninger, 1996) que permitiriam definir e limitar o escopo da ontologia sendo desenvolvida. Questões de competência são consultas que devem ser respondidas a partir da estrutura, das instâncias ou das possíveis inferências de uma ontologia. Os conceitos-chave identificados foram: Pessoa, Projeto, Recurso Físico, Produção Científica, Produção Tecnológica, Perceiro, Patrocinador e Área de Pesquisa. A partir destes conceitos chave, novos conceitos foram identificados e incorporados. Uma taxonomia descrevendo a hierarquia dos conceitos desenvolvidos também foi construída. Um fragmento da ontologia final pode ser visualizado na Figura 9, estando sua última versão disponível em [http://www.teccomm.les.inf.puc-rio.br/daml/onts/v0\\_9/projeto.daml](http://www.teccomm.les.inf.puc-rio.br/daml/onts/v0_9/projeto.daml). As questões de competência foram úteis na identificação de novos conceitos e relações entre os mesmos. Algumas dessas questões, referentes ao conceito Projeto, são apresentados na

Tabela 1.

| Dado um Projeto:  |
|---|
| - quais são as pessoas associadas a ele?                  |
| - quais são os recursos físicos utilizados e necessários? |
| - quem são seus coordenadores?                            |
| - qual foi a produção científica gerada?                  |
| - qual foi a sua produção tecnológica?                    |
| - quem são seus patrocinadores?                           |
| - quem são os parceiros relacionados?                     |
| - quem são os parceiros relacionados?                     |
| - quem são os participantes?                              |

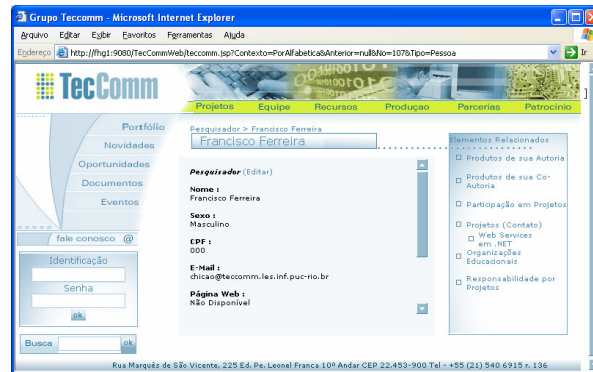
**Tabela 1 - Questões de Competência para o Conceito Projeto da Ontologia**

Concluído o desenvolvimento da ontologia, foi discutido como o Framework Portalware seria instanciado para permitir o acesso e a manipulação das informações na base de conhecimento da ontologia. Na Figura 10 podemos ver um *screenshot* do Portal do Conhecimento gerado pelo Framework Portalware com base nos modelos desenvolvidos a partir da ontologia proposta.

#### **4.1.4.O Framework Portalware**

O Framework Portalware (Rocha et al., 2002) foi concebido e desenvolvido para automatizar o processo de instanciação de Portais de Conhecimento. Portais de Conhecimento têm como objetivo facilitar o acesso a informações relevantes e específicas de um determinado domínio, através de uma interface universal como a WWW (Staab et al., 2001). Frameworks (Fayad et al., 2000), por outro lado, podem ser descritos como geradores de famílias de aplicações em um mesmo domínio. Esta geração é baseada na customização de pontos de flexibilização do Framework, culminando com uma aplicação completa e funcional. No caso específico do

Portalware, os pontos de flexibilização mais importantes são os modelos conceituais e de interface e a própria identidade visual do Portal.



**Figura 10 - Portal do Conhecimento Gerado a partir do Framework Portalware**

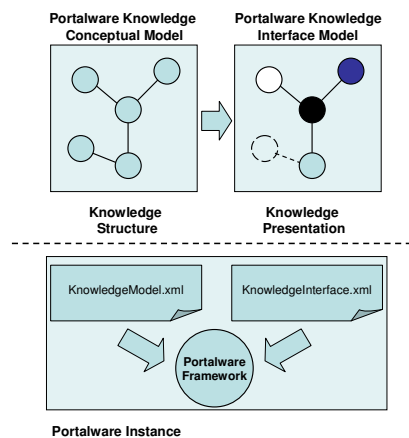
Como já foi dito, ontologias são um mecanismo bastante efetivo para se disponibilizar informação semanticamente, permitindo seu processamento de forma automática por agentes de software e aplicações. No entanto, a maior parte do conhecimento manipulado no domínio em questão – Projetos de Pesquisa em TI – é criado, acessado e disseminado por membros do grupo, entre professores, estudantes e patrocinadores de projetos. Com isso, torna-se importante não apenas a criação de uma ontologia, mas a existência de mecanismos que facilitem o acesso, inserção, manutenção e edição da base de conhecimento – em outras palavras, as informações descritas através da ontologia. Infelizmente, linguagens de representação de ontologias como RDF (RDF), DAML+OIL (DAML+OIL) ou OWL (OWL) não são as mais indicadas para este fim, uma vez que estas não foram desenvolvidas com o intuito de serem facilmente interpretadas por seres humanos.

Além disso, de acordo com Thomas Davenport e Lawrence Prusak (Davenport and Prusak, 1998), um dos mais importantes desafios enfrentados por projetos de Gestão de Conhecimento reside na dificuldade em se motivar os geradores de conhecimento – os funcionários ou membros de uma organização – a contribuírem com uma base de conhecimento corporativa, principalmente se isso representar um esforço adicional ao seu trabalho diário.

Esperar, portanto, que os usuários sintam-se motivados a contribuir trabalhando diretamente com a representação formal da ontologia e da base de

conhecimento irá certamente fadar o projeto de Gestão de Conhecimento ao fracasso, uma vez que o conteúdo disponível ficará desatualizado em um piscar de olhos. Então, um fator crítico de sucesso neste tipo de abordagem é prover uma interface simples a partir da qual os usuários podem adicionar novas informações, modificar informações antigas e indicar a relação entre informações de forma bem rápida e intuitiva.

Aplicações hipermédia – quando bem projetadas - têm um importante papel neste contexto. Estas podem ser encaradas como partes de um time homem-máquina no qual parte do problema deve ser resolvido pela máquina e parte pelo usuário. Do ponto de vista da máquina, diversas técnicas podem ser aplicadas, como Bancos de Dados, Bases de Conhecimento, Redes Neurais, entre outras. Do outro lado, o usuário humano utiliza o paradigma hipermédia para gerenciar – acessar, manter, editar – o conhecimento armazenado. Esse paradigma é usado no contexto do Portalware para integrar a representação de conhecimento formal – computável pelas máquinas – à representação do conhecimento informal – interpretável pelo usuário humano (Schwabe e Salim, 2002).



**Figura 11 - A relação entre o Modelos utilizados para a instanciação do Framework Portalware**

Como foi visto, para a geração de uma instância do Framework Portalware é necessário o desenvolvimento de dois modelos que serão usados como configuração do Framework, o Modelo Conceitual de Conhecimento e o Modelo da Interface de Conhecimento. Esses modelos são fortemente inspirados nos modelos conceitual e navegacional do método OOHD, descritos em (Schwabe e Rossi, 1998), podendo

ser gerados a partir deste método. A relação entre esses dois modelos e o Portalware está ilustrada na Figura 11. O Modelo Conceitual de Conhecimento descreve a estrutura do conhecimento que será armazenado – de acordo com o domínio de conhecimento para o qual está sendo gerado o portal – através de classes e propriedades – que podem representar tanto atributos simples quanto relações entre classes - segundo a abordagem Orientada a Objetos. Este modelo contém a estrutura conceitual do domínio e não possui nenhuma informação sobre como esses conceitos devem ser apresentados ao usuário. A estrutura conceitual pode ser vista como a informação semântica sobre o domínio enquanto que o Modelo da Interface de Conhecimento informa como a base de conhecimento será acessada pela aplicação hipermídia.

Assim que esses dois modelos estão concebidos e implementados, torna-se necessário apenas iniciar o processo de geração da instância. Toda a infra-estrutura para realização de buscas e manutenção da base de conhecimento – através de funcionalidades de alimentação e edição de dados – já faz parte do framework sendo automaticamente incorporada na aplicação hipermídia gerada.

#### ***4.1.5. Instanciando os Modelos do Portalware a partir da Ontologia***

A principal questão enfrentada nesta etapa foi como extrair o Modelo Conceitual do Conhecimento de uma ontologia, para que o framework pudesse ser usado na geração do Portal de Conhecimento. O primeiro passo foi encarar a ontologia não como o arquivo DAM+OIL ou RDF(S) que o descreve, mas sim como sua definição original – uma especificação formal de uma conceitualização compartilhada, descrita pela estrutura O mencionada anteriormente.

Uma vez que o Modelo Conceitual do Conhecimento possui uma estrutura muito similar ao modelo Orientado a Objetos – independente de algumas particularidades que não trazem impacto a esta abordagem – uma heurística foi definida para guiar o processo de tradução de uma ontologia descrita na estrutura O para a estrutura do Modelo Conceitual do Conhecimento necessário para a instanciação do Portalware.

Esta heurística é descrita resumidamente a seguir:

1. Inicialmente, todos os conceitos da estrutura O devem ser mapeados diretamente em classes no Modelo Conceitual.
2. No passo seguinte, a taxonomia do Modelo Conceitual deve ser construída. Esta taxonomia pode ser inteiramente extraída da hierarquia Hc pertinente a estrutura O e deve ser descrita como uma série de relações de especialização e generalização entre as classes do Modelo. Uma vez que no mundo real relações de herança múltipla são extremamente comuns e podem ser percebidas em diversos domínios de conhecimento, todas as relações taxonômicas desse tipo devem permanecer descritas no Modelo Conceitual.
3. Todas as relações não taxonômicas entre os conceitos devem ser mapeadas como relações entre classes ou serem transformadas em atributos. A decisão acerca da transformação do conceito em relação entre classes ou em atributo da classe depende de uma série de fatores e da decisão final do projetista:

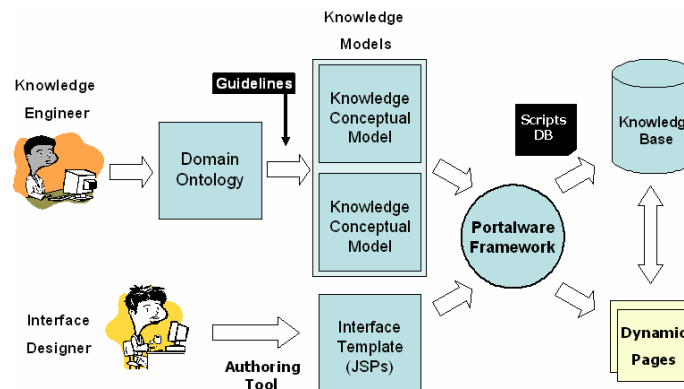
Finalizado o processo de tradução entre a ontologia e o Modelo Conceitual de Conhecimento, inicia-se o processo de construção do Modelo de Interface do Conhecimento. Como foi citado anteriormente, este modelo descreve como a base de conhecimento – agora descrita estruturalmente pelo modelo conceitual – será apresentada pela aplicação hipermídia ao usuário. O Modelo de Interface mapeia cada conceito em um nó – o que significa que cada instância desse conceito terá uma página gerada dinamicamente para a exibição de seus atributos – e cada relação não taxonômica em uma aresta – que será apresentada como um *link* tanto no conceito origem quando no conceito destino. A estrutura taxonômica é usada para gerar os índices de acesso relativos a cada conjunto de nós presentes em um ramo desta hierarquia. Tanto os *links* quanto os atributos deverão ser apresentados através de uma legenda apropriada, a qual também deve ser informada no Modelo de Interface.

Ambos os modelos são codificados em XML, de acordo com suas respectivas DTDs. O Framework Portalware gera *queries* SQL baseadas no Modelo Conceitual para que seja possível gerar a base de dados que irá armazenar a base de conhecimento em si. Também é necessário gerar os templates HTML ou XML que serão utilizados para apresentar as informações descritas no Modelo de Interface. Na versão atual do Portalware os templates HTML devem ser fundidos com *tags* de JSP especialmente criadas com esse fim. Essas *tags* indicam a posição na qual os atributos, links, legendas e índices devem ser renderizados na página, e são implementados



através de Taglibs JSP (Keogh, 2002). Essas Taglibs encontram-se descritas em (Rocha et al., 2002).

A Figura 12 ilustra o processo completo de geração de um Portal de Conhecimento através do Portalware.



**Figura 12 - Partindo de uma Ontologia até uma Aplicação Hipermídia Utilizando o Framework Portalware**

#### 4.1.6. A Camada Semântica

No entanto, como foi visto, o Portal de Conhecimento gerado com a utilização do Portalware é, basicamente, uma aplicação hipermídia com interface HTML ou XML, o que não garante a semântica dos dados armazenados – apesar do modelo conceitual possuir esta semântica, ela não é disponibilizada externamente. Para que os dados fiquem disponíveis externamente para outras aplicações e agentes de software, estes devem ser descritos em uma linguagem de marcação semântica, como DAML+OIL ou OWL.

A criação de uma Camada Semântica tem como objetivo solucionar este problema, através do desenvolvimento de um *web service* que consulta a estrutura do Modelo Conceitual e o repositório que contém as instâncias da base de conhecimento e publica essas informações como documentos DAML+OIL, seguindo a ontologia previamente definida.

Diversos tipos de consulta podem ser realizadas, tanto na estrutura da ontologia quanto em suas instâncias. Podem ser identificadas, por exemplo, todas as

instâncias do conceito Pessoa, os valores de suas propriedades e outras instâncias de conceitos relacionadas, como os Projeto nos quais essa Pessoa participa ou sua Produção Acadêmica. Essas informações estão possibilitando o desenvolvimento de agentes de software e serviços (Cunha et al., 2003) que irão realizar tarefas como a geração de relatórios acadêmicos, atualização de currículos, geração de programas de desenvolvimento pessoal para novos integrantes, a formação de grupos de trabalho, apenas para citar alguns exemplos.

#### **4.1.7. Conclusões do Estudo de Caso**

A infra-estrutura de Portal de Conhecimento Semântico desenvolvida no contexto deste estudo de caso pode ser aplicada para outros grupos de pesquisa que atuem na área de TI ou mesmo em outras áreas, quando novas ontologias terão que ser desenvolvidas. Isso permitirá a análise de como diferentes grupos utilizam um Portal de Conhecimento para a identificação de pontos fortes e fracos na abordagem apresentada. Questões relacionadas a escalabilidade, performance e segurança do Portal Semântico ainda precisam ser bem estudadas.

Neste estudo de caso identificou-se a importância de se implementar uma iniciativa de Gestão de Conhecimento para o Grupo de Pesquisas do TecComm, evitando-se assim a formação de vácuos de conhecimento causados pela alta rotatividade da equipe – principalmente pela integração de novos membros - e da grande velocidade de inovação nesta área. Foi descrito como o projeto e o desenvolvimento do Portal do Conhecimento foi baseado em tecnologias da Web Semântica, como ontologias e *web services* e como esse portal foi gerado a partir de um framework de geração de portais através de Modelos de Conhecimento, o Portalware.

Também foi possível identificar a necessidade de se implantar não apenas uma interface semântica – ou seja, conteúdo semanticamente anotado – para agentes de software mas também uma aplicação hipermídia na web que permitisse aos usuários – pesquisadores e estudantes – o acesso à base de conhecimento do grupo.

Espera-se que, com a implantação do Portal Semântico de Conhecimento do Grupo TecComm os membros deste grupo intensifiquem o compartilhamento de conhecimento, provendo informações atualizadas sobre seus projetos, seus interesses suas pesquisas, sua produção acadêmica e tecnológica.

Os resultados deste estudo de caso também forneceram subsídios para o desenvolvimento do KUP, uma vez que tratou tanto do desenvolvimento de uma ontologia quanto de aplicações na Web Semântica que utilizam esta ontologia. O fato da ontologia e a aplicação ter sido desenvolvida por equipes diferentes sem a existência de uma metodologia ou processo bem definido, por exemplo, permitiu a identificação de diversas omissões no artefato final, que poderiam ter sido evitadas antecipadamente. Por outro lado, uma forma mais integrada e automática de se instanciar o Portalware a partir da ontologia também se mostrou bastante necessária, uma vez que a heurística de tradução utilizada - por ser manual e altamente dependente do projetista - pode produzir e propagar erros e concepções equivocadas.