

UMA ABORDAGEM PARA GERÊNCIA DO CONHECIMENTO EM GRUPOS DE PESQUISA

Daniel Lichtnow

Universidade Federal de Santa Catarina - UFSC
CTC – Campus Universitário – Trindade
Caixa Postal, 476 – Florianópolis – SC – Brasil – 88040-900

Universidade Católica de Pelotas - UCPEL
Rua Félix da Cunha, 412
Pelotas – RS – Brasil – 96100-000
lichtnow@inf.ufsc.br

Christiane Gresse von Wangenheim

Universidade do Vale do Itajaí - UNIVALI
Centro de Educação Superior VII - Ciência da Computação
Rod. SC 407, Km 04 - CEP 88122-000
São José/SC – BRASIL
gresse@sj.univali.br

Aldo von Wangenheim

Universidade Federal de Santa Catarina - UFSC
CTC – Campus Universitário – Trindade
Caixa Postal, 476 – Florianópolis – SC – Brasil – 88040-900
awangenh@inf.ufsc.br

RESUMO

Este trabalho apresenta uma abordagem híbrida que utiliza diversas técnicas na construção de um Sistema de Gerência de Memória Corporativa (*Corporate Memory Management System – CMMS*), para auxiliar a Gestão do Conhecimento dentro de grupos de pesquisa. Dentre as técnicas a utilizar estão a Recuperação de Informação, o Raciocínio Baseado em Casos e a Extração de Informação, que podem vir a subsidiar a criação da infra-estrutura necessária. O protótipo do sistema está sendo desenvolvido e implantado dentro do Projeto Cyclops na Universidade Federal de Santa Catarina.

Palavras-chave: Gestão do Conhecimento, Memória Corporativa, Recuperação de Informação

Categoria: Sessão Técnica

Área: Inteligência Artificial (Gestão do Conhecimento)

UMA ABORDAGEM PARA GERÊNCIA DO CONHECIMENTO EM GRUPOS DE PESQUISA

RESUMO

Este trabalho apresenta uma abordagem híbrida que utiliza diversas técnicas na construção de um Sistema de Gerência de Memória Corporativa (*Corporate Memory Management System – CMMS*), para auxiliar a Gestão do Conhecimento dentro de grupos de pesquisa. Dentre as técnicas a utilizar estão a Recuperação de Informação, o Raciocínio Baseado em Casos e a Extração de Informação, que podem vir a subsidiar a criação da infra-estrutura necessária. O protótipo do sistema está sendo desenvolvido e implantado dentro do Projeto Cyclops na Universidade Federal de Santa Catarina.

ABSTRACT

This article presents a hybrid approach that uses several techniques in the construction of a Corporate Memory Management System - CMMS, in order to establish a technical infrastructure for Knowledge Management in research groups. In order to provide comprehensive support, various techniques such as Information Retrieval, Case-Based Reasoning and Information Extraction have been integrated. A prototypical system is currently being developed and applied in The Cyclops Project at the Federal University of Santa Catarina.

Palavras-chave: Gestão do Conhecimento, Memória Corporativa, Recuperação de Informação

1. Introdução

Considerando a natureza de um grupo de pesquisa, onde as atividades desenvolvidas são bastante complexas, que requerem conhecimento altamente especializado e facilidade de acesso às novas informações, constata-se facilmente que um dos mais importantes fatores para o sucesso destas organizações está ligado à sua capacidade de gerenciar o conhecimento, de reter as experiências, o *know-how* e as habilidades adquiridas ao longo do tempo.

Algumas das características presentes em grupos de pesquisa dentro de universidades evidenciam ainda mais a necessidade da gerência do conhecimento: a volatilidade de sua composição, a limitação de recursos financeiros e humanos, a informalidade das relações e o fato de que alguns membros encontram-se geograficamente dispersos.

Assim, assume importância a área denominada Gestão do Conhecimento (*Knowledge Management – KM*) que tem seu foco voltado para questões relacionadas a como as organizações podem tirar maior proveito do conhecimento existente dentro delas próprias, como membros da organização podem distribuir o conhecimento para quem este pode ser útil, como

registrar as soluções adotadas para solução de problemas, como reter o conhecimento de seus especialistas mesmo quando estes deixam a organização, além de discutir formas de gerar novo conhecimento a partir do conhecimento existente dentro da organização ou a partir de fontes externas [1] [2].

Destaca-se também a chamada Fábrica de Experiências (*Experience Factory – EF*), abordagem que surgiu dentro do domínio da Engenharia de Software. A Fábrica de Experiências é um organização física e/ou lógica que suporta o desenvolvimento do projeto analisando e sintetizando todos os tipos de experiência, atuando como um repositório para cada experiência e oferecendo esta experiência para projetos em demanda [5].

Para operacionalizar esta abordagem na prática é necessária uma infra-estrutura técnica denominada de Sistema de Gerência de Memória Corporativa (*Corporate Memory Management System – CMMS*). Esta infra-estrutura emprega a chamada Memória Corporativa (*Corporate Memory – CM*) que atua como um repositório para a informação e conhecimento armazenado dentro do CMMS. Neste artigo, todo este conteúdo existente será referenciado pelo termo *CM Assets*.

Este artigo apresenta uma abordagem para a construção de um CMMS adaptado às necessidades de grupos de pesquisa. Esta abordagem está sendo implementada e avaliada dentro do Projeto Cyclops [6], que é um projeto de pesquisa internacional que envolve a Universidade Federal de Santa Catarina do Brasil e a Universidade de Kaiserslautern na Alemanha. As atividades do projeto estão ligadas sobretudo às áreas de análise de imagens médicas e telemedicina, sendo que dentro do projeto são desenvolvidas novas ferramentas que fornecem suporte a estas áreas.

2. Requisitos para a infra-estrutura técnica

O desenvolvimento de um CMMS eficaz e eficiente dentro de um grupo de pesquisa não é trivial, sendo necessário considerar o preenchimento de vários requisitos a fim de que a Gestão do Conhecimento possa efetivamente receber o auxílio do CMMS. Baseado em nossas experiências dentro do Projeto Cyclops e na literatura os seguintes requisitos foram obtidos:

Suporte multimodal. O CMMS deve fornecer suporte para viabilizar o acesso a vários tipos de informação ou conhecimento (documentos, FAQ's e sites, por exemplo), relacionados a diversas áreas (programação, processamento de imagens médicas), contemplando vários objetivos (realização de pesquisas, resolução de problemas) e levando em conta pontos de vista diferentes (médicos, cientistas da computação).

Acesso eficiente e eficaz às informações e/ou aos conhecimentos úteis. São necessárias técnicas inteligentes de recuperação que possibilitem a recuperação baseada em similaridade de informações e/ou conhecimentos úteis, sem que sejam fornecidas aos usuários um grande número de informações irrelevantes. O mecanismo deve permitir que sejam feitas consultas em linguagem natural e também possibilitar a recuperação multilíngüe (a partir de uma consulta formulada em inglês, serão recuperadas mesmo as informações em português).

Distribuição pró-ativa do conhecimento. Além de permitir a recuperação, o CMMS deve também dar suporte à recomendação pró-ativa do conhecimento para um usuário com interesses específicos.

Evolução contínua da CM. A CM estará sempre incompleta e sofrendo constantes mudanças. Em função disto, a Gestão do Conhecimento em grupos de pesquisa requer um CMMS que suporte a contínua evolução das informações e do conhecimento e que possa lidar com informações e/ou conhecimentos incompletos ou incoerentes.

Manutenção da CM. Definir uma estrutura inicial para a CM geralmente é apenas o primeiro passo em direção a um CMMS que obtenha sucesso. Deste modo, deverá ser fornecido suporte para permitir a manutenção do conhecimento, bem como do conhecimento geral do domínio.

Melhoria do CMMS. O CMMS deverá ser continuamente melhorado e adaptado durante o seu ciclo de vida, sendo necessária uma contínua avaliação do seu uso.

Estes requisitos mostram que é necessário um suporte sofisticado e abrangente para o desenvolvimento de um CMMS em um grupo de pesquisa.

3. Estado da arte e prática

Hoje, um grande número de Tecnologias da Informação são utilizadas para implementar CMMSs de uso geral, bem como CMMSs voltados para o uso no domínio de Engenharia de Software, por exemplo:

Sistemas Gerenciadores de Banco de Dados (SGBD). Lidam com grandes volumes de dados, mas não possuem recursos específicos para lidar com a Gestão do Conhecimento.

Mapas do conhecimento (*Knowledge Maps*). Tem como função indicar a localização do conhecimento dentro da organização, isto é, as pessoas que possuem determinado conhecimento ou informação, ao invés de armazenar o conhecimento [1].

Groupware. Fornecem links de comunicação entre os membros e possuem a capacidade de gerenciar documentos, como por exemplo Lotus/IBM [7], mas não suportam diretamente a Gestão do Conhecimento adaptada as necessidades de grupos de pesquisa.

Sistemas de hipertexto. Possibilitam a navegação entre documentos por meio de links. Mas apesar de seu enorme sucesso, carecem de mecanismos de recuperação mais elaborados quando o volume de documentos aumenta.

Recuperação de Informação (*Information Retrieval*). Esta área ocupa-se principalmente com a recuperação da informação em documentos de texto [3]. Fornece mecanismos, cuja a exatidão é limitada, o que faz com que freqüentemente sejam recuperados um grande número de documentos irrelevantes.

Raciocínio Baseado em Casos (*Case-Based Reasoning*) [4]. Fornece suporte para o desenvolvimento de sistemas de aprendizado baseados em conhecimento. Suas maiores vantagens são o uso da recuperação baseada na similaridade, o seu foco no conhecimento experimental e no aprendizado incremental contínuo.

Filtragem de Informações (*Information Filtering*). Tem por objetivo classificar grande volumes de informações de forma a comunicar automaticamente a um usuário específico aquilo que for do seu interesse.

Extração de Informações (*Information Extraction*). Esta área ocupa-se da extração de informações de um texto. Assim, conforme destacado em [9] esta tecnologia tem seu enfoque voltado para a retirada automática de informações relevantes dos textos.

Para que o CMMS auxilie de fato na Gestão do Conhecimento dos grupos de pesquisa, várias abordagens deverão ser integradas. Não existem até o presente momento abordagens que preencham completamente os requisitos descritos na seção 2, necessários à implementação de CMMSs no contexto de um grupo de pesquisa.

4. Uma abordagem híbrida para o CMMS

Nesta seção será apresentada uma abordagem híbrida para construção de um CMMS dentro do contexto de um grupo de pesquisa. O objetivo é criar uma infra-estrutura técnica que permita:

- modelar e armazenar vários tipos de informação e conhecimento dentro da CM de forma que estes possam ser reutilizados,
- oferecer acesso à informação e ao conhecimento armazenados na CM, por meio de procedimentos de recuperação e distribuição,
- a evolução contínua da CM, mediante a aquisição e integração de novas informações e conhecimentos e
- a manutenção e a evolução do CMMS.

A abordagem integra tecnologias de diferentes áreas: técnicas de Raciocínio Baseado em Casos são usadas para a representação do conhecimento, recuperação baseada em similaridade e aprendizagem incremental; técnicas de Recuperação da Informação e de Processamento de Linguagem Natural são as bases para extração da informação de consultas formuladas em linguagem natural e documentos de forma a dar suporte a recuperação, assim como à evolução do conhecimento geral do domínio. Sistemas de hipertexto são usados para possibilitar a exploração interativa dos *CM Assets* recuperados por uma consulta; e técnicas de Filtragem de Informações são usadas para recomendar informações ou conhecimentos que possam ser de interesse do usuário.

Nas próximas seções serão descritas cada um dos principais componentes do CMMS.

4.1. Representação do conhecimento

O principal objetivo do CMMS é reunir o conhecimento existente em um grupo. Este conhecimento existe na forma explícita, isto é, está presente especialmente nos documentos produzidos pelo grupo, mas também na forma tácita na forma de experiências pessoais. Todo o conteúdo da CM é representado na forma de casos, denominados *CM Assets*. Vários tipos de *CM Assets* são armazenados:

Registro de documentos. É feito o registro de todos os livros, artigos, manuais e revistas relevantes para os membros do projeto. Sobre cada documento serão armazenadas informações básicas como título, autor(es), ano de publicação e idioma, além de resumos e comentários elaborados pelos membros do grupo. Será possível armazenar o documento dentro da CM quando este estiver disponível em arquivo.

Tutoriais on-line. Possibilitam aos usuários aprender por conta própria, quando isto for necessário.

Frequently Asked Questions (FAQ). Aqui estão registradas perguntas e respostas fornecidas por especialistas relacionadas à área de pesquisa. Este tipo de recurso evitará que as mesmas perguntas precisem ser respondidas novamente, deixando os especialistas liberados para atividades mais complexas.

Receitas How-to-do. Estes documentos são elaborados por membros do projeto a partir de suas experiências pessoais, descrevem como executar determinadas tarefas ligadas as áreas do grupo.

Mapas de WWW. Armazenar *sites* com informações relevantes às áreas de atuação do projeto assume grande importância no momento em que existe muito conhecimento disponível na Web, mas que por vezes requer um gasto de tempo considerável na sua busca.

Páginas amarelas. Servem para localizar as pessoas que possuem determinados conhecimentos.

Kit para principiantes. Fornecem informações aos usuários iniciantes sobre quais documentos e especialistas deverão ser procurados no início do seu trabalho.

Mensagens de novidades (*News Messages*). São geradas mensagens para os usuários indicando, por exemplo, a existência de algum novo material que merece ser consultado. A idéia é também armazenar mensagens relativas a informações trocadas entre os usuários, bem como comunicados importantes que objetivam manter os usuários informados.

Biblioteca de software. Muitas vezes os membros do projeto precisarão ter acesso ou reutilizar algum software. Assim, o CMMS deverá dar suporte a armazenar softwares relevantes à área de pesquisa específica do grupo. Isto inclui softwares de uso mais geral e também bibliotecas e classes desenvolvidas pelo próprio grupo de pesquisa.

4.2. Conhecimento geral do domínio

Além destes recursos, dentro da CM é necessário que esteja presente a terminologia específica de cada área a ser utilizada no processo de recuperação. Para isto, será criada uma estrutura composta basicamente por vocabulários que possuem os termos descritivos do domínio de cada área de pesquisa (p.ex. nome das classes da linguagem de programação Smalltalk), *Thesauri* que mostram como estes termos se relacionam e um Dicionário que contem a tradução de termos para outro idioma (p.ex. Inglês) permitindo que as consultas sejam feitas em vários idiomas.

4.3. Recuperação de *CM Assets*

O objetivo da recuperação é encontrar *CM Assets* da CM que sejam úteis para o trabalho de pesquisa ou desenvolvimento que está sendo realizado por um membro. Baseado na consulta formulada por um usuário, a CM é pesquisada e aqueles *CM Assets* que forem mais relevantes são retornados. O resultado é apresentado como uma lista de links de hipertexto que permite ao usuário explorar em detalhes cada um dos *CM Assets* retornados. Se o CMMS não fornecer uma resposta satisfatória, o usuário pode automaticamente direcionar sua consulta para um especialista da área. O especialista recebe a pergunta via e-mail. Uma vez que a resposta esteja disponível, ela é automaticamente enviada ao usuário que encaminhou a consulta. Pela composição da consulta do usuário e da resposta fornecida pelo especialista, um novo *CM Asset* é criado e integrado à CM.

De forma a permitir o acesso aos vários tipos de informação e conhecimento considerando os mais diversos objetivos (ver Seção 2) o CMMS oferece várias técnicas de recuperação. Dentre as diversas formas possíveis de um usuário expressar suas consultas estão por exemplo as consultas por navegação (usadas freqüentemente na Web), consultas utilizando o par atributo-valor (como por exemplo utilizadas em SGBDs), consultas por termos (utilizadas em sistemas de Recuperação de Informação), consultas em linguagem natural e as consultas por exemplo (neste tipo de consulta o usuário fornece um exemplo do que ele quer recuperar, recebendo documentos semelhantes). A descrição de outras formas de consulta podem ser encontradas em [11].

4.4. Distribuição de *CM Assets*

Outro objetivo do CMMS é auxiliar os pesquisadores a encontrar informações ou conhecimentos que satisfaçam seus objetivos de longo prazo. Para assegurar isto, a abordagem de distribuição do conhecimento é integrada dentro do CMMS, evitando que seja necessário que um usuário procure ativamente por informação, e possibilitando que os *CM-Assets* sejam enviados ao usuário, considerando o seu possível interesse.

Para que possa fazer tais recomendações, no CMMS são usadas três diferentes abordagens: Baseada em Conteúdo (*Content-Based*), Colaborativa (*Collaborative*) e Demográfica (*Demographic*) [10]. Na primeira Baseada em Conteúdo, o usuário irá receber informações a partir da análise do seu comportamento, isto é, caso ele tenha demonstrado interesse por algum material, serão fornecidos a ele materiais semelhantes. Na abordagem Colaborativa o usuário recebe informações a partir das consultas feitas por pessoas com interesses semelhantes aos seus, na abordagem Demográfica são estabelecidas características que possam determinar interesses semelhantes, como a idade, formação dentre outros.

4.5. Coleta e integração contínua

No contexto dos grupos de pesquisa, nem toda informação e conhecimento estão completamente disponíveis na CM quando da implantação do CMMS. Assim, a manutenção da CM deve envolver o acompanhamento contínuo dos projetos e avanços tecnológicos.

Novas informações e conhecimentos devem ser coletados e integrados à CM no momento em que estiverem disponíveis. Isto inclui a coleta de novos *CM Assets*, sua apropriada indexação e validação da qualidade e a sua integração na CM.

Os *CM Assets* serão coletados pelos membros do projeto de pesquisa. Quanto ao que se refere ao momento em que será feita esta integração, serão adotadas as seguintes estratégias:

- Coleta espontânea. Um membro do projeto encontra, por exemplo, um novo site e, usando o CMMS, faz o seu cadastro.
- Coleta dirigida por eventos: ocorrerá quando em função de algum fato novo (o desenvolvimento de uma nova pesquisa) os membros buscam por novas informações sobre uma área.
- Coleta planejada. Os membros do projeto de pesquisa irão ter como missão reunir informações sobre determinada área sobre a qual não existem informações.
- Coleta baseada na recuperação. Irão ocorrer situações em que o membro do projeto não encontrará aquilo que satisfaz suas necessidades. Desta forma, será necessário que ele busque esta informação com um especialista que poderá fornecer-lhe ajuda. Esta ajuda deverá ser convertida em um novo conhecimento a ser acrescentado à CM.

Toda informação será coletada por meio de formulários que serão preenchidos pelos membros do projeto de pesquisa. A tarefa de propor a integralização dos *CM Assets* será de todos os membros do projeto. Em algumas situações (casos da coleta planejada e dirigida por eventos) um Engenheiro do Conhecimento deverá coordenar o processo.

4.6. Manutenção do CMMS

Em função da área de software ser caracterizada por constantes mudanças e avanços tecnológicos, a manutenção do CMMS torna-se especialmente importante. Isto abrange a revisão e a adaptação das informações armazenadas na CM, bem como, os mecanismos de coleta e de acesso. A maioria das atividades de manutenção deverão ser feitas manualmente pelo Engenheiro do Conhecimento (auxiliado pelos especialistas das áreas de pesquisa). O objetivo da infra-estrutura técnica é fornecer suporte através do pré-processamento dos dados de entrada.

5. Arquitetura da Ferramenta

O CMMS é implementado em uma arquitetura cliente/servidor de 3 camadas consistindo das camadas de interface, aplicação e armazenamento. Esta arquitetura é mostrada na figura 1.

Os *CM Assets* são armazenados em um SGBD. O conhecimento geral do domínio e os arquivos referenciados nos registros de documentos são armazenados no sistema de arquivos. As partes centrais da camada de aplicação do CMMS são as ferramentas de recuperação, distribuição, coleta e manutenção. A comunicação dos usuários é feita via intranet através de browser e e-mail. A ferramenta está sendo implementada com o uso de Smalltalk.

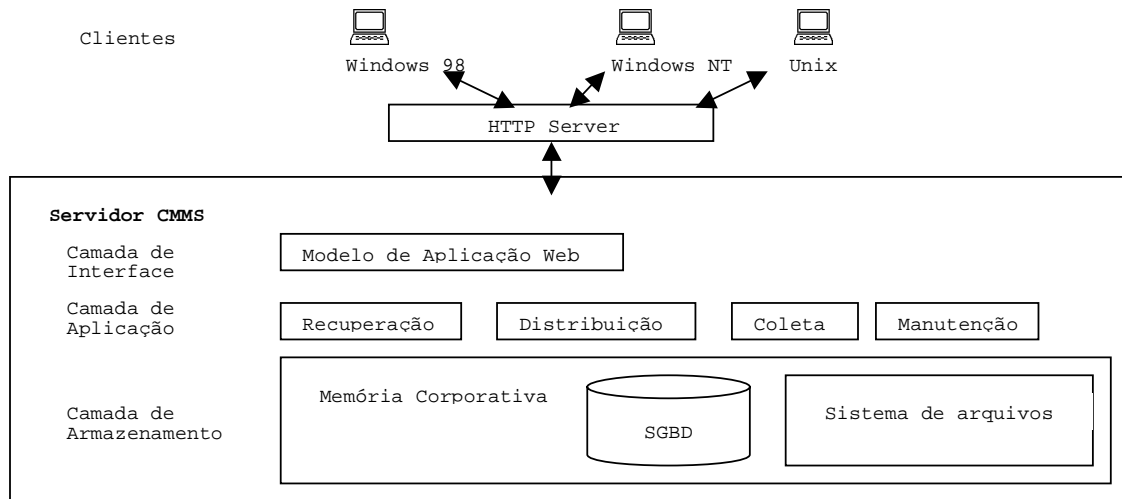


Figura 1- Arquitetura da Ferramenta

6. Resultados Iniciais

Uma primeira versão da ferramenta foi implantada no projeto Cyclops na Universidade Federal de Santa Catarina, tendo por objetivo tentar avaliar a aceitação da ferramenta por parte dos usuários bem como recolher às observações destes usuários quanto a novas funcionalidades a serem implantadas. Posteriormente ainda foi incorporado ao protótipo um módulo de FAQ direcionado a linguagem *Smalltalk* [8].

Durante os dois meses de uso (março e abril de 2001) foram incluídos 326 *CM Assets* referentes a sites e a bibliografia. No primeiro mês, os usuários efetuaram 28 pesquisas sendo que no segundo mês este número cresceu para 65. Este aumento no número de consultas parece demonstrar que os usuários passaram a se interessar mais pela ferramenta. Uma das razões está certamente relacionada ao aumento do conteúdo armazenado na CM, que começou a despertar um maior interesse dos usuários.

7. Conclusões

Neste artigo foi apresentada uma abordagem para o desenvolvimento de uma infraestrutura que fornece suporte à Gestão do Conhecimento dentro de grupos de pesquisa. A utilização integrada de diversas técnicas fornecerá uma maior capacidade de fornecer o esperado auxílio aos membros de um grupo de pesquisa.

Uma primeira versão da ferramenta foi implantada dentro do projeto Cyclops na Universidade Federal de Santa Catarina, tendo sido feita também uma avaliação deste protótipo, na qual os membros do projeto, mostraram-se cientes da importância da implantação da ferramenta, destacando várias melhorias a serem implementadas.

Com base nos resultados da avaliação de desempenho e da utilização da ferramenta, pretende-se continuar a implementação melhorando as técnicas aplicadas e ampliando o escopo das áreas de pesquisa cobertas.

Agradecimentos

Os autores gostariam de agradecer a André Bortolon, Daniel D. Abdala, Euclides M. Barros, Paulo Dellani e Fernando Secco por seu suporte na implementação desta ferramenta. Gostariam também de agradecer a José Eduardo De Lucca pela revisão deste artigo.

Bibliografia

- [1] Davenport, T.; H, Pruzac, L. **Working Knowledge – How Organizations Manage What They Know**. Harvard Business School Press, 1998.
- [2] Nonaka, I.; Takeuchi, T. **Criação de Conhecimento na Empresa**. Rio de Janeiro:Campus,1997. 2ª Edição.
- [3] Salton, G.; McGill, M. J. **Introduction to Modern Information Retrieval**. New York:McGraw Hill, 1983. 1ª Edição.
- [4] Watson, I. Applying **Cased-Based Reasoning: Techniques for Enterprise Systems**. San Francisco:Morgan Kaufmann, 1997. 1ª Edição.
- [5] Basili. V. R., Caldiera G., Rombach, H. D. **The Experience Factory**. Local: John Wiley & Sons, 1994. 1ª Edição.
- [6] The Cyclops Project – Departamento de Informática e Estatística da Universidade Federal de Santa Catarina (<http://www.inf.ufsc.br/cyclops>)
- [7] IBM/Lotus (www.lotus.com/km)
- [8] Bortolon, A. **Desenvolvimento de uma Abordagem Híbrida para o Gerenciamento de Documentos FAQ em Português** Florianópolis: CPGEPS/UFSC, 2001. 80p. (Dissertação de Mestrado – Mestrado em Engenharia da Produção e Sistemas)
- [9] Cunningham, H. **Information Extraction: a user guide**. Research Memo CS-97-02. University of Sheffield, Sheffield, UK, janeiro de 1997.
- [10] Pazzani, M. J. A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence Review*, 1999.
- [11] Loh, S.; Oliveira J. P. M. V. (orientador). **Descoberta de Conhecimento em Textos**. Exame de Qualificação. Porto Alegre, CPGCC/UFRGS, 1999