

Sistema de Recomendação para apoio à colaboração em ambientes baseados em conhecimento

Stanley Loh^{1,2}, Daniel Lichtnow¹, Ramiro Saldaña¹, Rodrigo Branco Kickhöfel¹, Gabriel Simões¹, Thyago Borges¹, Tiago Primo¹, Gustavo Piltcher¹, Luiz Carlos Ribeiro Junior¹, Cleber Gouvêa¹

¹Universidade Católica de Pelotas (UCPEL) – Grupo de Pesquisa em Sistemas de Informação R. Félix da Cunha, 412, Pelotas, RS – CEP 96010-000

²Universidade Luterana do Brasil (ULBRA) – Faculdade de Informática R. Miguel Tostes, 101, Canoas, RS – CEP 92420-280

{lichtnow, rsaldana, rodrigok, thyago, lcr}@ucpel.tche.br, sloh@terra.com.br, gsimoos@vetorial.net, tiagoprino@brturbo.com, gustavopil@yahoo.com.br, cleberg@email.com.br

Abstract. *This paper presents a recommender system to support collaboration among people. The system provides a private chat room, where people can exchange textual messages. Analyzing the messages, the system discovers the themes discussed and then it can help people by recommending complementary knowledge sources. Furthermore, chat sessions are stored and can be retrieved later, so that people can review decisions processes or learn from past discussions. Statistical analyses of the sessions can reveal important knowledge about the discussions.*

Resumo. *Este artigo apresenta um sistema de recomendação para apoio a colaboração através de pessoas. O sistema dispõe de uma sala de chat privada, onde pode-se trocar mensagens textuais. Analisando as mensagens, o sistema descobre os temas discutidos e então este poderá ajudar outras pessoas, recomendando fontes complementares de conhecimento. Além disso, as sessões de chat são armazenadas, podendo ser recuperadas futuramente, então, essas pessoas poderão rever processos decisórios ou aprender com as discussões passadas. Análises estatísticas das sessões podem revelar importantes conhecimentos sobre as discussões.*

1. Introdução

O sucesso de uma organização está relacionado em grande parte ao conhecimento que esta possui. O conhecimento de uma organização é formado a partir do conhecimento dos seus membros e pode aumentar à medida que este é armazenado e compartilhado. Esse é o conceito de organizações que aprendem (*learning organization*), conforme SENGE (2001).

Segundo NONAKA e TAKEUCHI (1995) a maior parte do conhecimento organizacional vem da interação entre pessoas. Pessoas tendem a reusar soluções de outras pessoas para ganhar produtividade.

Nesse contexto de valorização do conhecimento como um recurso corporativo, surgiu a chamada Gestão do Conhecimento (*Knowledge Management – KM*). A Gestão

do Conhecimento tem seu foco voltado para questões relacionadas a como as organizações podem tirar maior proveito do conhecimento existente dentro delas, facilitando a distribuição deste conhecimento entre seus membros, encorajando o registro das soluções adotadas para resolução de problemas, evitando a perda do conhecimento de seus especialistas quando estes deixam a organização (DAVENPORT & PRUZAC, 1997).

Quando pessoas se comunicam para trocar informação ou para adquirir conhecimento a fim de realizar um trabalho ou ajudar outra pessoa, o processo é chamado *Colaboração*. A colaboração pode ser realizada através de interações síncronas (por exemplo, um *chat*), assíncronas (como fóruns), contato direto (pessoas falando) ou contato indireto (quando pessoas armazenam conhecimento e outras podem recuperá-lo a qualquer momento).

A colaboração é uma das mais importantes tarefas para a inovação e vantagem competitiva (SENGE, 2001). Ela também é importante no armazenamento do conhecimento para uso posterior. Se o conhecimento não é adequadamente registrado, organizado e recuperado, a conseqüência é o re-trabalho, baixa produtividade e perda de oportunidades.

Em uma equipe de trabalho, cada indivíduo carrega consigo uma série de conhecimentos tácitos que, se disseminados para os outros membros, podem contribuir de forma expressiva para uma melhor eficácia na execução dos processos, resultando assim em sucesso organizacional. A capacidade de mapear o conhecimento tácito de cada membro da equipe gera valor para a organização, já que assim, pode-se evitar a necessidade de contratar pessoas ou serviços buscando tal conhecimento, pois o mesmo foi identificado em alguém que faz parte do time.

Segundo DAVENPORT e PRUZAC (1997), gigantes como HP e IBM não possuem a menor idéia de como gerenciar o valor agregado das informações e conhecimentos de suas empresas. Sabendo disso, a proposta de uma ferramenta que auxilie na resolução destes problemas, sem duvida, seria de grande valia para muitas organizações de médio e grande porte.

Indo mais além, um desafio para Gestão do Conhecimento é usar o conhecimento de uma forma pró-ativa, ou seja, se antecipando às ações do usuário. Uma tecnologia que pode auxiliar nessa tarefa são os sistemas de recomendação. O objetivo principal desses sistemas é localizar fontes de informação relacionadas ao interesse ou necessidade do usuário (MONTANER et al., 2002), ou seja, eles fornecem informação sem que o usuário precise realizar uma busca.

Visto a grande necessidade que as empresas têm de gerenciar sua memória organizacional e aumentar a produtividade através de um melhor aproveitamento de seus recursos, este trabalho apresenta um sistema de recomendação que apóia a colaboração entre pessoas. O sistema constitui uma Biblioteca Digital que armazena os recursos da empresa, e um *chat web* onde os usuários realizam interações síncronas e recebem recomendações conforme os assuntos abordados. Essas recomendações podem ser de itens da Biblioteca Digital, de discussões anteriores (sessões de *chat* já realizadas) e usuários com certo grau de conhecimento no assunto abordado.

Na seção 2 deste artigo, são aprofundados alguns aspectos relacionados aos Sistemas de Recomendação e sistemas existentes. A seção 3 apresentada a arquitetura do sistema proposto. Por fim, na seção 4, as conclusões discutem contribuições e limitações

da proposta, bem como informações referentes a desempenho e precisão do protótipo implementado, salientando suas possíveis aplicações. São listados, também nas conclusões, desafios encontrados na implementação e implantação do protótipo.

2. Sistemas de Recomendação e trabalhos correlatos

Um grande problema na maioria dos softwares que gerenciam conhecimento é que eles não fazem isso de uma forma pró-ativa. Dessa forma o usuário precisa realizar uma consulta na base de conhecimento para encontrar o que interessa.

Por exemplo, as Bibliotecas Digitais são importantes no processo de gestão do conhecimento, entretanto as pessoas precisam procurar em grandes volumes de itens. Feigenbaum (*apud* DAVIES, 1989) compara as bibliotecas atuais com as do futuro. As atuais consistem de um repositório de objetos passivos. Por outro lado, as bibliotecas do futuro serão coleções de documentos ativos, ajudando pessoas a descobrir novos conhecimentos, fornecendo associações previamente desconhecidas, analogias e novos conceitos sem que as pessoas tenham que indicar claramente suas necessidades de informação.

Nesse contexto os sistemas de recomendação são uma tecnologia emergente. Seu objetivo é fornecer informação útil na tomada de decisão. Essas informações pode ser sobre livros, documentos, CDs, restaurantes e outros (RESNICK & VARIAN, 1997).

TERVEEN e HILL (2001) discutem vários sistemas de recomendação. Por exemplo, o sistema PHOAKS extrai páginas *Web* que foram mencionadas em mensagens do *newsgroup* Usenet para futuramente permitir a recuperação destas indicações. Já o sistema proposto por Donath et al. (*apud* TERVEEN & HILL, 2001) analisa discussões do Usenet e de *chats*, atribuindo-lhes propriedades (por exemplo, número de participações). Isto permite recuperar discussões com certas propriedades. Estas ferramentas são úteis para analisar as mensagens, mas não geram recomendações durante a discussão. Neste caso, a função é apenas a de registrar as mensagens para permitir a posterior recuperação de informações.

Já o sistema Tapestry permite que as pessoas registrem suas avaliações sobre certas mensagens postadas em uma discussão. Entretanto, neste caso, os processos de armazenar e recuperar conhecimento acontecem por iniciativa do usuário.

Um exemplo de sistema pró-ativo (sistema que realiza alguma tarefa sem que o usuário precise solicitá-la) voltado para apoio a interações é apresentado em KOMOSINSKI et al. (2000). Este sistema identifica termos presentes nas mensagens trocadas entre os participantes de um *chat* e apresenta a definição de tais termos durante a interação. Entretanto, esta ferramenta não oferece nenhum tipo de recomendação, mas somente explicações dos termos usados.

3. Sistema de Recomendação para apoio a Colaboração

Para auxiliar na colaboração entre pessoas e contribuir com a Gestão de Conhecimento nas organizações foi desenvolvido um sistema chamado de SisRecCol.

O SisRecCol (Sistema de Recomendação para apoio a Colaboração) consiste em um *chat Web* onde os usuários trocam mensagens e recebem recomendações conforme o assunto identificado nas mensagens. Essa identificação é realizada através de técnicas de *text mining* apoiadas em uma ontologia desenvolvida para o sistema. As recomendações

são de documentos eletrônicos e *links* para páginas da *Web* (armazenados numa Biblioteca Digital privada do sistema), de autoridades nos assuntos (identificados pelo próprio sistema) e de discussões anteriores sobre o mesmo tema.

O objetivo do SisRecCol é apoiar a troca de conhecimento entre os usuários, e permitir o armazenamento e recuperação de conhecimento explícito e também tácito através do armazenamento das sessões de *chat*. O sistema pode ser adaptado para ser utilizado em diferentes áreas, bastando incluir uma nova ontologia, referente ao novo domínio. O protótipo do sistema está disponível em <http://gpsi.ucpel.tche.br>. Na figura 1 é possível ver a tela do *chat* com um exemplo de discussão e a lista de recomendações referentes ao assunto abordado.

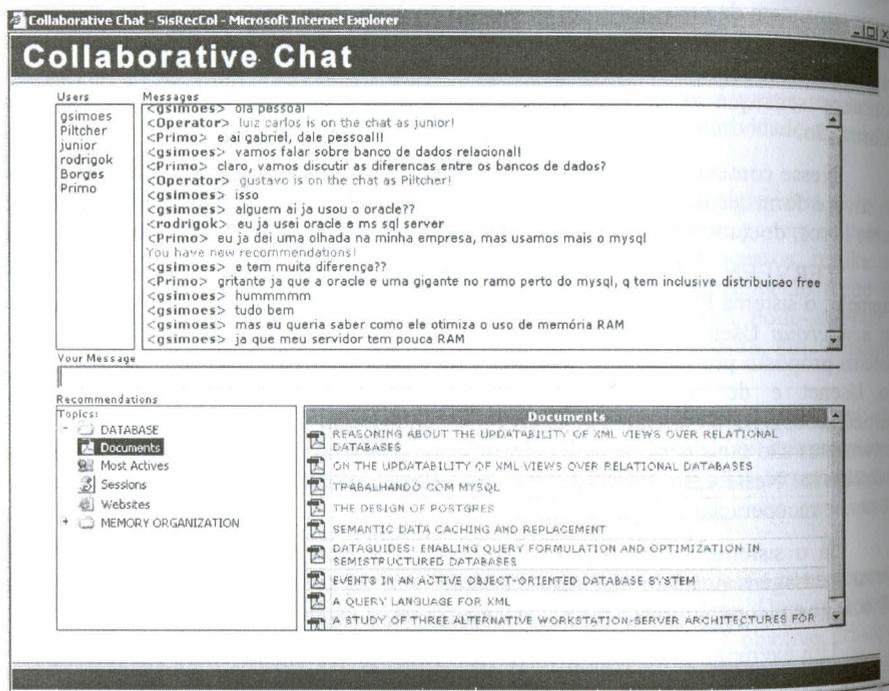


Figura 1. A interface do *chat*, mostrando recomendações

3.1. A Ontologia

A base do sistema é uma ontologia de domínio utilizada para classificar itens na biblioteca digital, para identificar temas nas mensagens e para traçar o perfil dos usuários.

Uma ontologia é uma definição formal e explícita de conceitos (classes ou categorias) e seus atributos e relações (NOY & MCGUINNES, 2002). Uma ontologia do domínio – *domain ontology* é uma descrição de “coisas” que existem ou podem existir em um domínio (SOWA, 1997).

No SisRecCol a ontologia é apresentada como uma hierarquia de conceitos e cada conceito contém uma lista de termos associada. Ela deve ser criada para o domínio de aplicação do sistema, ou seja, deve conter os conceitos e termos da área que será utilizada.

3.2. Módulo de Text Mining

O módulo de *Text Mining* é responsável pela identificação dos assuntos das mensagens do *chat*. Ele realiza o batimento entre as palavras da mensagem com os termos presentes na ontologia e, através de métodos probabilísticos, identifica o assunto referente. Essa identificação não ocorre apenas individualmente nas mensagens, mas sim, analisando um contexto (se várias mensagens tratam do mesmo assunto e, com um grau de importância acima de um determinado parâmetro, a probabilidade desse assunto ser o tópico da discussão é alta, sendo esse então, identificado como assunto principal desse conjunto de mensagens).

O módulo de *Text Mining* também é utilizado para a indexação dos documentos presentes na biblioteca digital.

3.3. A Biblioteca Digital

Biblioteca Digital é uma coleção de recursos digitais, organizados sob certa lógica, acessíveis para recuperação sobre uma rede de computadores (KOCHTANEK et al., 2001).

A Biblioteca Digital do SisRecCol contém documentos eletrônicos e *links* para páginas *Web*, no momento. Este conteúdo é previamente indexado com ferramentas de *software* que realizam o processo de forma automática com base nos mesmos métodos usados para identificar os assuntos das mensagens do *chat*. O processo de indexação determina o grau de relacionamento dos documentos, (artigo ou *site*) aos conceitos existentes na ontologia.

A inclusão dos itens na Biblioteca Digital pode ser feita pelos usuários, num processo independente da sessão do *chat*.

3.4. A Base de Perfis

A base de perfis contém a identificação dos usuários e suas áreas de conhecimento. A base se assemelha aos chamados Mapas do Conhecimento ou Páginas Amarelas (*Yellow Pages*), que servem para indicar que pessoas possuem determinados conhecimentos. Um Mapa do Conhecimento é, portanto, um guia e não um repositório de conhecimento (STEWART, 1998).

A indicação dos conhecimentos que as pessoas possuem é feita através de uma relação de conceitos (dentre os presentes na ontologia) e um grau associado a cada conceito. Este grau determina o quanto uma pessoa está relacionada à um certo assunto ou tema do domínio. O grau não é absoluto, mas relativo às demais pessoas do grupo. Este valor pode ser usado para indicar o grau de competência da pessoa na área ou para referenciar se uma pessoa está mais ativa do que as outras em determinada área. Além disso é dinâmico, sendo modelado conforme o usuário utiliza o sistema.

3.5. Módulo de Recomendação

A partir do assunto identificado nas mensagens do *chat*, o módulo de recomendação proativamente sugere a cada usuário uma lista de itens relevantes vindos da Biblioteca Digital (documentos, páginas da *Web*), sessões anteriores e da base de perfis (usuários com conhecimento no assunto).

Essa recomendação é apresentada em uma área separada, não interferindo na discussão, e permite que o usuário acesse o conteúdo, por exemplo, abrir um documento para leitura.

Os itens que serão recomendados são definidos conforme o perfil de cada usuário. Se um usuário já leu determinado documento ele não será recomendado e também um documento básico não é recomendado para um usuário avançado.

Além das recomendações geradas durante as sessões de *chat*, o sistema ainda apresenta métodos de recomendação *off-line* que estão disponíveis ao usuário a qualquer momento. Esses métodos utilizam informações como “os documentos lidos pelos usuários mais ativos” e “os documentos lidos por todos os usuários”, gerando uma lista, por exemplo, dos documentos mais lidos pelos usuários mais ativos em todo o sistema nos últimos 30 dias ou os documentos mais lidos no assunto de maior interesse do usuário que recebe essa recomendação. O interesse do usuário é definido através de seu perfil, conforme citado anteriormente.

3.6. Módulo de Estatísticas

O sistema possui um módulo de estatísticas que, conforme exibido em exemplo resumido na figura 2, consiste de um conjunto de ferramentas utilizadas na análise das discussões ocorridas durante as sessões de *chat*.

As estatísticas permitem verificar a interação de cada usuário durante as sessões, descobrindo seu tipo de interesse. Além disso, é possível verificar quem são os usuários mais ativos, os menos ativos (apenas assistem a discussão), se os usuários falam mais sobre assuntos importantes (presentes na ontologia), ou discutem assuntos irrelevantes, entre outros.

A análise da ordem em que os conceitos vão sendo identificados permite descobrir como a discussão se desenrolou, ou seja, a rota da discussão pelos conceitos, permitindo também verificar os desvios do tema central.

Este tipo de análise também permite inferir a *expertise* do grupo como um todo, ou seja, que áreas são de maior conhecimento do grupo e que sub-áreas de um assunto não foram discutidas.

| Statistics of Session #233257 | Additional Informations | Recommendations by Concept |
|--|--|---|
| Number of Users: 10 Number of Messages: 249 Number of Words with Stopwords: 1032 Number of Words without Stopwords: 592 | Words by Message (with Stopwords): 4.453 Words by Message (without Stopwords): 0.492 Recommendations by Message: 0.136 Number of Words without Stopwords: 592 | Concept Weight Count INFORMATION SYSTEMS 0.095672 13 DATABASE 0.020050 22 DISTRIBUTED DATABASE 0.014100 4 NEURAL NETWORKS 0.007820 4 Total: 4 concept(s) |

Figura 2. Algumas métricas do módulo de estatísticas

4. Conclusões

Este trabalho apresentou um sistema para apoiar a colaboração entre pessoas. Este sistema auxilia no processo de transferência do conhecimento entre os membros de uma organização, facilitando o aprendizado organizacional. Além disso, ele dá suporte ao armazenamento do conhecimento, tanto através da Biblioteca Digital, alimentada pelos próprios usuários, quanto pela base de sessões de *chat* que torna explícito o conhecimento tácito dos usuários.

Atualmente o SisRecCol possui apenas uma ontologia de domínio voltada para a área de Ciência da Computação. Dessa forma foram realizados experimentos com grupos de pesquisa em Universidades. Com os testes pôde-se constatar que pessoas conseguem compartilhar melhor o conhecimento com a ajuda das recomendações, pois o sistema mesmo lembra que documentos e *sites* existem sobre o assunto. O compartilhamento também ocorre indiretamente, pois os usuários podem armazenar na Biblioteca Digital documentos e *links* para *sites*, os quais poderão ser recomendados para outras pessoas. Além disso, o conhecimento presente nas discussões fica armazenado e pode ser recuperado através de consultas (ex: que discussões anteriores trataram do tema X) ou poderão ser recomendadas nas próprias sessões de *chat*.

O método utilizado para identificar os assuntos das mensagens do *chat* foi avaliado formalmente, tendo-se medido que a precisão na identificação de conceitos ultrapassou os 85% de acerto.

Trabalhando as informações adquiridas com as sessões de *chat*, o sistema proposto permite, além de aumentar a base de conhecimentos da organização, a criação de um mapa do conhecimento organizacional, desde que seus membros utilizem a ferramenta. Este mapa possibilita a criação de *Yellow-Pages*, o que agilizará os processos no momento de buscar membros qualificados, ou até mesmo de delegar funções.

Embora hajam benefícios previstos, é preciso considerar que a aplicação do sistema poderá encontrar algumas barreiras. Uma delas reside no fato da base de conhecimento estar inicialmente com um volume pequeno, o que irá fazer com que não sejam feitas recomendações ou que estas sejam irrelevantes no início da aplicação da ferramenta. Neste sentido, é necessário incentivar o uso do sistema e o armazenamento prévio de documentos e *sites* relacionados aos assuntos de interesse da organização, até que o mesmo atinja um nível satisfatório de relevância.

Um desafio gerado pelo uso do sistema proposto é a necessidade de modificar o comportamento das pessoas envolvidas. Isso se dá já que as pessoas precisam entender que, se não utilizarem o sistema, o mesmo deixará de trazer qualquer benefício a elas próprias e a organização.

Uma outra dificuldade é a criação da ontologia. Atualmente parte do trabalho é manual, pois requer que um especialista informe os conceitos da área de aplicação e selecione manualmente documentos referentes a cada conceito para realizar a configuração das palavras na ontologia (através de ferramentas do próprio sistema). Uma alternativa que já está em estudo seria a criação de uma ferramenta que identifique automaticamente os assuntos e ainda tenha capacidade de buscar na *Web*, por exemplo, os documentos de cada assunto para realizar a criação da ontologia.

Por permitir a recuperação de sessões anteriores de *chat*, o sistema proposto colabora de forma ativa na disseminação de informações. Além disso, contribui para o crescimento da base de conhecimentos da organização, já que, analisando as mensagens das sessões, é possível identificar os assuntos tratados, o que permite a identificação de *expertises* dos usuários.

Agradecimentos

O presente trabalho foi realizado com o apoio do CNPq, uma entidade do Governo Brasileiro voltada ao desenvolvimento científico e tecnológico.

Referências

- Davenport, T. H. & Pruzac, L. (1997) "Working knowledge – How organizations manage what they know", Harvard Business School Press, Harvard.
- Davies, R. (1989) The creation of new knowledge by information retrieval and classification, *Journal of Documentation* 45 p.273-301.
- Kochtanek, T. R. et al. (2001) "A digital library resource Web site: Project DL", *Online Information Review*, v.25, p.29-40.
- Komosinski, L. J. et al. (2000) Uso de Agentes para Apoio à Mediação de Diálogo Entre Estudantes via Internet. In: Congresso Iberoamericano de Informática na Educação (RIBIE 2000), V, Dez, 2000. Proceedings... Viña del Mar (publicado em CD).
- Montaner, M. et al. (2002) Improving case representation and case base maintenance in recommender agents, in: Proc. 6th European Conference on Case Based Reasoning.
- Nonaka, I. & Takeuchi, T. (1995) "The knowledge-creating company: how japanese companies create the dynamics of innovation", Oxford University Press, Cambridge.
- Noy N. F. & McGuiness, D. L. (2002) "Ontology Development 101: a guide to creating your first ontology". Disponível em <http://protege.stanford.edu/publications/>.
- Resnick, P. & Varian, H. (1997) "Recommender Systems" *Communications of the ACM* v.40 p.56-58.
- Senge, Peter M. (2001) A quinta disciplina: arte e prática da organização que aprende. 9ª ed. Best Seller, 444p.
- Sowa, J. F. (1997) "Building, sharing, and merging ontologies", AAAI Press / MIT press, p3-41.
- Stewart, T. A. (1998) Intellectual capital: the new wealth of organizations, Bantam Books, 320p.
- Terveen, L. & Hill, W. (2001) "Beyond recommender systems: helping people help each other", in: J. CARROLL, ed., Human computer interaction in the new millennium, Addison-Wesley.