

ARQUITETURA DE UM SISTEMA DE RECOMENDAÇÃO PARA APOIO À COLABORAÇÃO

Ramiro Saldaña Garin¹, Daniel Lichtnow¹, Stanley Loh^{1,2}

¹Universidade Católica de Pelotas
Rua Félix da Cunha, 412 - Pelotas/RS – Brasil - 96010-000
{rsaldana,lichtnow,loh}@atlas.ucpel.tche.br

²Universidade Luterana do Brasil
Rua Miguel Tostes, 101 - Canoas/RS, Brasil - CEP 92420-280

Resumo

Nas organizações, as pessoas aprendem compartilhando conhecimento. Este tipo de tarefa (chamada colaboração) é importante para a aprendizagem organizacional. A colaboração pode ser apoiada através de ferramentas de tecnologia da informação como *chats*, *newsgroups*, fóruns e listas de discussão por e-mail. Porém, este tipo de apoio só permite a comunicação, não ajudando as pessoas no processo de aprendizagem. Para minimizar este problema, há sistemas de recomendação cuja meta é ajudar as pessoas a adquirir conhecimento. Os sistemas de recomendação agem sugerindo o que é novo ou útil.

Este trabalho apresenta a arquitetura de um sistema de recomendação para apoiar a colaboração entre pessoas em uma organização. O artigo discute a funcionalidade do sistema proposto, apresentando seus principais componentes e interações. O sistema analisa mensagens textuais enviadas durante a sessão, identifica o contexto da discussão e sugere documentos, autoridades (as pessoas com competência em um assunto) e discussões passadas dentro do mesmo contexto.

Palavras-chave: colaboração, organizações que aprendem, mineração em textos, *groupware*, sistemas de recomendação.

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 conhecimento é armazenado e compartilhado. Este é o conceito de organizações que aprendem (*learning organization*), conforme SENGE (2001).

Neste 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 o conhecimento é compartilhado, ele pode ser aproveitado por mais pessoas dentro da organização, reduz-se o tempo gasto no desenvolvimento dos trabalhos realizados, a qualidade dos trabalhos cresce e também ocorre uma significativa redução no tempo gasto com o treinamento de novos membros.

Para auxiliar o processo de compartilhamento de conhecimento pode ser empregada a Tecnologia de Informação (TI). As tecnologias projetadas para facilitar a realização do trabalho em grupos são denominadas de tecnologias de Trabalho Cooperativo Auxiliado por Computador, ou *CSCW - Computer-Supported Cooperative Work* (GREENBERG, 1999). *CSCW* é uma área que envolve o estudo de como as pessoas trabalham em grupo e de como o computador e as tecnologias relacionadas afetam o comportamento dos grupos. O suporte computacional relacionado à realização de trabalhos em grupo é denominado *Groupware*.

A troca de conhecimento pode-se dar de forma direta, quando as pessoas mantêm uma interação, ou indireta, sem que seja necessária a interação. Por exemplo, *chats*, fóruns e listas de discussão permitem a interação direta através de diálogos ou mensagens, sendo possível identificar com quem se está interagindo. Por outro lado, é possível trocar conhecimento armazenando-o de forma que futuramente outras pessoas possam recuperá-lo, e isto se dá sem que seja necessário saber quem fará uso do conhecimento nem quando. Um exemplo deste último tipo são as bibliotecas digitais.

Recentemente, surgiram os sistemas de recomendação (*Recommender Systems*). Um sistema de recomendação tem por objetivo auxiliar no processo social de indicar ou receber indicação, seja esta indicação referente a livros, artigos, discos, restaurantes ou informações (RESNICK et al., 1997). Sistemas de recomendação são largamente usados em comércio e marketing, para sugerir produtos ou fornecer informações para ajudar o cliente a decidir a compra (SCHAFER et al., 2001; LAWRENCE et al., 2001). Nestas situações, as pessoas não precisam solicitar ao sistema que forneça as recomendações, mas este decide o que e quando sugerir.

Este trabalho apresenta a arquitetura de um sistema de recomendação para apoiar a colaboração entre pessoas. A colaboração se dá na forma de interações síncronas (como *chats* na Internet), e as recomendações são feitas pelo próprio sistema de forma proativa (automática e *online*) durante as interações. Para tomar esta decisão, o sistema analisa as mensagens enviadas pelos participantes da sessão e identifica o contexto ou tema da discussão. Os itens recomendados podem ser documentos, páginas Web, discussões anteriores sobre temas semelhantes ou autoridades no assunto (pessoas com uma certa competência).

Na seção 2 deste artigo, são aprofundados alguns aspectos relacionados à Colaboração e aos Sistemas de Recomendação. Na seção 3, é apresentada a arquitetura do sistema proposto, e as sub-seções detalham os componentes do sistema, seu papel e funcionamento dentro do sistema. Já a seção 4 apresenta um exemplo de interação auxiliada pelo sistema de recomendação proposto. Por fim, as conclusões discutem contribuições e limitações da proposta.

2 Colaboração e Sistemas de Recomendação

Existem diferentes tipos de conhecimento dentro de uma organização. Aquele conhecimento que não se encontra documentado e está relacionado às experiências adquiridas pelos indivíduos é denominado de *conhecimento tácito*. Por outro lado, o *conhecimento explícito* refere-se ao que é codificado de forma concreta, por exemplo, em bancos de dados e documentos (NONAKA & TAKEUCHI, 1997).

Para que exista a criação do conhecimento organizacional é necessário que seja feita a identificação dos tipos de conhecimento e sua conversão de um modo para outro (STEWART, 1998). O processo de conversão é um processo social podendo ser um de quatro modos (NONAKA & TAKEUCHI, 1997):

1) **Socialização: do conhecimento tácito em conhecimento tácito.** É caracterizado pelo ato de compartilhar experiências. Refere-se àqueles indivíduos que aprendem observando

outros trabalharem ou aprendem através da imitação. Neste processo é necessário que mestre e aprendiz compartilhem as experiências.

2) **Externalização: do conhecimento tácito em conhecimento explícito.** O conhecimento tácito torna-se explícito, expresso na forma de metáforas, analogias, conceitos ou modelos. A escrita é uma forma de externalização.

3) **Combinação: do conhecimento explícito em conhecimento explícito.** Os processos de categorização do conhecimento explícito podem levar a um novo conhecimento.

4) **Internalização: do conhecimento explícito em conhecimento tácito.** O conhecimento explícito torna-se tácito, isto é, os indivíduos incorporam um determinado conhecimento.

Boa parte da geração de novos conhecimentos se dá pela colaboração entre pessoas. A colaboração acontece quando pessoas interagem trocando conhecimento e experiências, para realizarem algum trabalho em conjunto ou simplesmente para uma ajudar outra.

Um tipo de sistema automatizado que pode auxiliar na colaboração são os sistemas de recomendação. Estes sistemas procuram encaminhar para uma pessoa soluções que foram úteis para outras pessoas. Este tipo de serviço é especialmente útil quando não se sabe o que pode ajudar na solução de um problema ou não se sabe onde esta solução se encontra.

Segundo TERVEEN & HILL (2001), há quatro tipos principais de sistemas de recomendação.

Os *sistemas baseados em conteúdo (content-based systems)* utilizam apenas as preferências dos usuários. Estes se concentram em recomendar itens semelhantes aos que o usuário obteve no passado. Seu foco está em algoritmos para aprender sobre as preferências dos usuários e na filtragem de um conjunto de itens que mais se assemelhem a essas preferências.

Os *sistemas de apoio a recomendações (recommender support systems)* não automatizam o processo de recomendação. Servem como ferramentas para auxiliar pessoas a compartilhar recomendações, ajudando tanto aqueles que produzem as recomendações quanto aos que as procuram.

Os *sistemas de mineração social de dados (social data mining systems)* mineram preferências implícitas em registros computacionais de atividades sociais, como mensagens trocadas através do Usenet, troca de documentos, histórico de uso de sistemas, citações ou hiperlinks.

Os *sistemas de filtragem colaborativa* requerem que as pessoas que procuram por recomendações expressem suas preferências. Estes sistemas trabalham com algoritmos para agrupar pessoas através de suas preferências e para pesar o interesse de pessoas com os mesmos gostos para produzir recomendações.

TERVEEN & 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 de forma explícita, ou seja, as pessoas decidem realizar o processo e tomam a iniciativa. Neste caso, o sistema automatizado funciona de forma **reativa**, esperando um estímulo da pessoa para realizar as suas ações.

O ideal é que o sistema de recomendação funcione de forma proativa, ou seja, decidindo por conta própria o que, para quem e quando recomendar. Tais sistemas analisam o comportamento e as preferências das pessoas (por exemplo, produtos comprados e sites visitados), sem que as mesmas precisem determinar o que armazenar e em que momento. Entretanto, são sistemas normalmente voltados a recomendações no comércio eletrônico (SRIVASTAVA et al., 2000).

Um exemplo de sistema proativo 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.

Neste trabalho, é apresentada a arquitetura de um sistema de recomendação que procura auxiliar as pessoas no momento de discussões *online*. A base do sistema é a identificação de assuntos sendo discutidos e a recomendação de itens relacionados aos assuntos. Para tanto, o sistema utiliza técnicas de *Text Mining* para análise e classificação de mensagens contendo textos livres. As recomendações podem ser de documentos (bibliografias, artigos, arquivos eletrônicos e páginas Web), de discussões anteriores sobre o mesmo tema e de pessoas reconhecidas como autoridades no assunto. As recomendações baseiam-se no conteúdo das mensagens e, portanto, nos assuntos sendo discutidos. Uma estrutura do tipo *thesaurus* auxilia na identificação de contextos e na resolução de ambigüidades.

O sistema proposto incorpora características dos quatro tipos de sistemas de recomendação, conforme a classificação de TERVEEN & HILL (2001) descrita nesta seção.

3 Arquitetura e Funcionamento do Sistema Proposto

O objetivo do sistema proposto é auxiliar processos de colaboração realizados através de ferramentas de interação síncrona.

O auxílio se dá mediante a recomendação de algum material que possa auxiliar na discussão. Porém, o sistema também pretende auxiliar o registro, a recuperação e a classificação de discussões, para que estas também possam ser recomendadas em discussões futuras.

A base para o funcionamento do sistema proposto é um módulo de análise textual das mensagens trocadas durante as sessões de discussão. Para tanto, serão utilizadas ferramentas de *Text Mining*, associadas a um *thesaurus*, que permite a identificação de temas ou assuntos presentes nos textos. O sistema poderá ser utilizado em domínios diferentes, desde que um *thesaurus* específico do domínio seja inserido no sistema.

O conhecimento será explicitado na forma de:

- documentos: arquivos eletrônicos ou referências a artigos, livros e páginas Web;
- base de perfis: contendo a identificação e características das pessoas do grupo; e
- histórico das discussões: registrando as mensagens trocadas durante as sessões.

A figura 1 apresenta a arquitetura geral do sistema proposto, destacando os seus componentes e as trocas entre estes.

A figura 2 detalha os sub-módulos de análise *offline* do log das sessões.

No contexto deste artigo, uma sessão é uma troca de mensagens eletrônicas e textuais entre pessoas, realizada durante um período com início e fim determinados. A princípio, uma sessão deve ter um único objetivo, assumindo-se, portanto, que se trata de uma única discussão e vice-versa (uma discussão somente é feita durante uma sessão). Entretanto, é possível que uma sessão trate de mais de um tema ou assunto, fazendo com que seja classificada por mais de uma entrada do *thesaurus*. Neste trabalho, serão consideradas

somente sessões síncronas, que ocorrem quando todos os participantes estão conectados ao mesmo tempo e as mensagens são lidas em tempo real.

Geralmente, em discussões de colaboração, reúnem-se pessoas com interesses específicos, um objetivo comum e certos conhecimentos prévios relacionados. Este fato garante antecipadamente que o diálogo estabelecido entre os participantes não irá, pelo menos na grande maioria das vezes, abranger assuntos que estejam fora do contexto no qual estão inseridos os participantes.

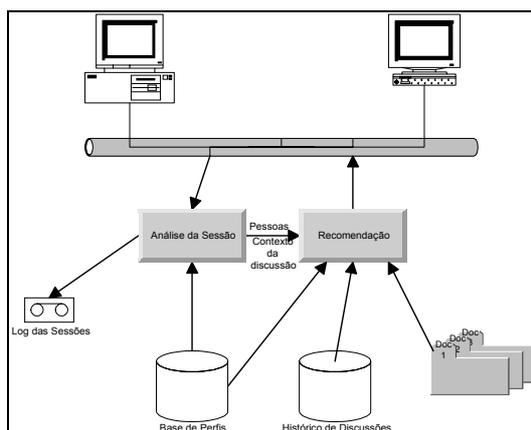


Figura 1: Arquitetura do sistema proposto

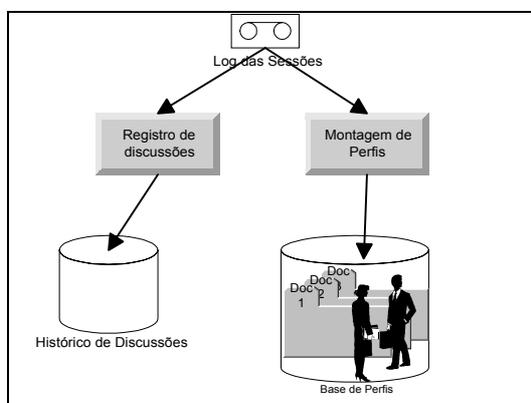


Figura 2: Sub-módulos de análise *offline*

Assim, os termos utilizados na interação tendem a ser termos intimamente relacionados ao assunto tratado e, portanto, que estão definidos no thesaurus do domínio. É possível que ocorram diálogos com uma abrangência maior do que a dos assuntos sendo tratados. Neste caso, seria necessário ter disponíveis vários *thesauri*.

Este fato gera uma maior complexidade no que se refere a elaboração de recomendações, pois uma abrangência maior fará com que o número de recomendações cresça e acabe perdendo a precisão necessária, podendo ser produzida a chamada sobrecarga de informações.

Nas próximas seções, cada componente do sistema é descrito em detalhes.

3.1 Módulo de análise da sessão

Toda mensagem enviada por um participante da sessão será analisada em busca de termos-chave. Termos muito gerais, denominados *stopwords*, são eliminados. Os termos restantes são utilizados num processo de classificação, cujo objetivo é determinar o assunto da mensagem. Os assuntos que podem ser identificados encontram-se no *thesaurus* do domínio, bem como os termos que identificam cada assunto.

O método de classificação utilizado é baseado nos algoritmos de Rocchio e Bayes e foi apresentado em LOH et al. (2000). Este método analisa o contexto geral do texto, ou seja, todos os termos significativos presentes e não somente a presença de determinadas palavras-chave. Isto evita, por exemplo, que a presença do termo ‘*inteligência*’ gere uma ambigüidade por se referir tanto ao tema “inteligência artificial” quanto ao tema “inteligência competitiva”.

Mesmo que vários assuntos tenham sido identificados numa mensagem, apenas um assunto é identificado como o foco da mensagem (aquele com maior presença ou que ainda não foi citado na sessão).

Durante uma sessão, o assunto sendo tratado pode mudar, já que cada mensagem pode tratar de um assunto específico diferente. Em geral, os assuntos devem estar relacionados, ou seja, próximos na hierarquia definida no *thesaurus* (irmãos, pais ou filhos).

O sistema deverá ter um **apontador de assunto** para saber qual é o assunto corrente. Este apontador se deslocará pela hierarquia de assuntos definidos no *thesaurus*. Quando um assunto “pai” e um assunto “filho” estiverem presentes na mesma mensagem, o assunto “filho” (mais específico) será o assunto apontado (corrente). Não é necessário que o apontador se desloque somente por assuntos vizinhos ou próximos na hierarquia, pois a discussão pode mudar para um assunto mais distante. Entretanto, supõe-se que isto ocorra momentaneamente e que a discussão retorne para o assunto principal.

O assunto corrente (onde o apontador está posicionado) será repassado ao módulo de recomendação, para que sejam selecionados itens neste assunto para serem recomendados. Pode haver mais de um assunto discutido numa mesma sessão. Neste caso, o sistema deverá manter a relação de assuntos mencionados na sessão (histórico).

Termos novos, que não aparecem no *thesaurus* como *stopword* ou indicando um assunto, deverão ser apresentados aos participantes da discussão para definição.

3.2 Thesaurus

Segundo FOSKET (1997), *thesaurus* é um dispositivo de controle de termos usado na representação de documentos. Os *thesauri* provêm mapas de conhecimento, apresentando conceitos ou idéias do domínio de aplicação e indicando relações entre eles. Estes conceitos aparecem representados através de termos, os quais indicam quando um determinado conceito está sendo tratado. As relações geralmente indicam hierarquias entre conceitos (divisão de conceitos em menores). No *thesaurus*, também deverão constar as chamadas *stopwords*, que são termos muito freqüentes ou genéricos e sem significado específico, tais como preposições, artigos e conjunções. Alguns termos, mesmo que não tão freqüentes, podem ser definidos como *stopwords* por não representarem um conceito do domínio (vide exemplo na seção 4).

A figura 3 apresenta um exemplo de *thesaurus* para a área de Computação (somente a parte de conceitos e hierarquia). Neste exemplo, o conceito “Redes Neurais” pode ser representado pelos termos ‘rede neural’, ‘redes neurais’, ‘redes neuronais’, ‘*neural network*’ e ‘redes de neurônios’.

No sistema proposto, deverá ser definido um *thesaurus* para o domínio do grupo de trabalho. Este *thesaurus* será utilizado para indexar o conhecimento explícito (documentos, sessões e competências ou áreas de interesse de pessoas do grupo). O *thesaurus* pode ser

montado manualmente ou extraído de um já existente, por exemplo, a classificação da ACM para Informática e a Classificação Internacional de Doenças (CID) para a Medicina.

Um item interessante no sistema proposto é que o *thesaurus* pode crescer à medida que as interações vão realizando-se. Como dito anteriormente, quando um novo termo (ainda não presente no *thesaurus*) é citado numa mensagem, será solicitado que o participante defina este termo no *thesaurus*. Esta definição significa relacionar o termo a um conceito já existente ou resulta na criação de um novo conceito (uma nova entrada no *thesaurus*). Neste último caso, deve ser fornecida uma explicação (como em um dicionário) e o posicionamento do conceito na hierarquia (indicar o conceito “pai”).

O refinamento do *thesaurus* pode ocorrer durante a sessão ou após. Na interface proposta, aparece no menu principal uma lista de termos novos, os quais devem ser definidos.

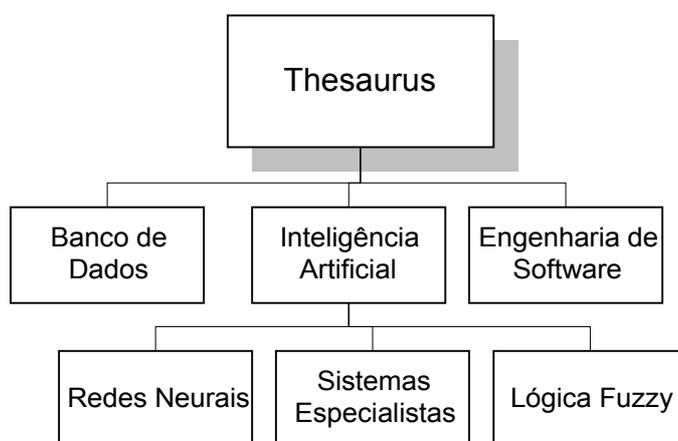


Figura 3: Exemplo de parte de um *thesaurus* para a área de Computação

3.3 Arquivo de Log das sessões

Este arquivo registra tudo o que ocorre durante uma sessão, tais como os participantes, dia e horário, as mensagens enviadas pelas pessoas, as recomendações feitas, os documentos lidos ou recuperados durante a sessão e os temas identificados durante a sessão.

Os temas identificados durante a sessão devem ser listados na ordem em que foram citados, formando assim um caminho que corresponde aos conceitos percorridos pelo apontador de assunto dentro do *thesaurus*. Isto permite analisar os rumos que a sessão seguiu através dos conceitos citados. Por exemplo, pode-se observar se a discussão passou de um tema mais genérico para um mais específico, qual foi a profundidade e a abrangência da discussão, se foi citado algum conceito sem relação hierárquica com os demais (desvio de assunto) e se houve algum conceito central, para o qual convergiam as mensagens.

Este arquivo servirá de base para geração do histórico de discussões e para a montagem dos perfis das pessoas.

3.4 Histórico das Discussões

O histórico das discussões contém as sessões de discussões anteriores e suas características gerais (participantes, data, mensagens e o caminho percorrido pelo apontador de assunto). Cada sessão deve estar armazenada separadamente. Além disto, a cada sessão deverá ser associado um ou mais assuntos, aqueles identificados nas mensagens enviadas durante a sessão.

Como uma sessão pode ter associado a si mais de um assunto, é necessário associar a cada contexto (assunto tratado) os documentos lidos ou recuperados e as recomendações

feitas durante a sessão, mas dentro do contexto específico. Isto permitirá aos participantes reverem as recomendações feitas durante a sessão mas relativas a determinado tema.

3.5 Base de documentos

Um dos objetivos do sistema proposto é recomendar documentos (artigos, bibliografias e páginas Web). Para tanto, deve existir uma base ou biblioteca de documentos. A inclusão de documentos deve ser feita pelos próprios membros do grupo e não precisa ser feita durante uma sessão (pode ser *offline*).

Estes documentos devem ser previamente indexados por assuntos, de acordo com o thesaurus criado no sistema. O processo de indexação é feito de forma automática por ferramentas automatizadas que utilizam algoritmos de classificação (ver LOH et al., 2000).

No momento em que um documento é associado a um tema ou conceito do *thesaurus*, é também determinado o seu grau de afinidade com aquele tema. Um documento pode ser classificado em mais de um tema.

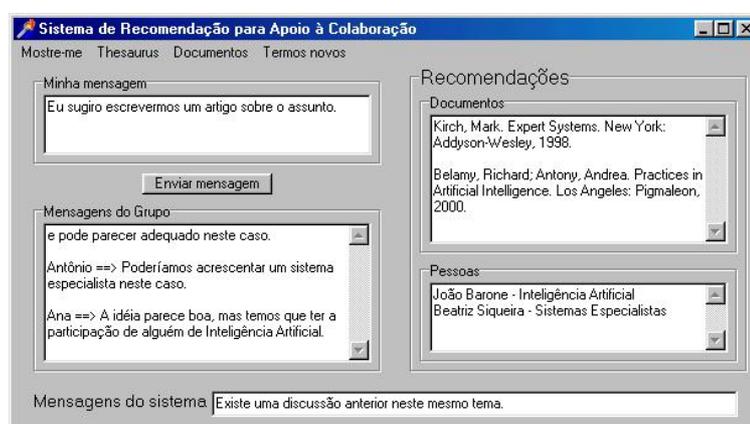


Figura 4: Interface do sistema de recomendação

3.6 Base de Perfis

A base de perfis contém a identificação das pessoas do grupo 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 no *thesaurus* do grupo) e um grau associado a cada conceito. Este grau determina o quanto uma pessoa está relacionada a 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.

Os conceitos associados às pessoas na base de perfis não necessariamente precisam ser disciplinas previamente estabelecidas por métodos tradicionais (cursos e diplomas). Segundo o conceito de Árvores de Conhecimento de LÉVY & AUTHIER (1995), novas competências e saberes podem surgir da coletividade, e cada indivíduo tem uma posição única nesta coletividade. As competências têm valores diferentes dependendo do contexto em que se inserem. Assim, a pessoa deixa de ser identificada apenas por um rótulo (nome de família, número da carteira de identidade, cargo ou diploma) e passa à condição de portador de um perfil de conhecimentos ou competências.

Desta forma, cada conceito e o grau correspondente podem ser estabelecidos previamente na inclusão da pessoa no grupo (na base de perfis), analisando-se seu *curriculum vitae*. Depois disto, novas áreas podem ser acrescentadas ou o grau pode ser alterado à medida que a pessoa participa em sessões de discussão ou lê documentos. O grau de competência será usado para recomendar pessoas e funciona como uma pontuação.

3.6.1 Esquema de pontuação nos perfis

A pontuação da pessoa em um determinado assunto depende das atividades desta pessoa. Cada sessão de discussão ou documento lido/recuperado aumenta uma certa quantidade de pontos no assunto em questão (referente à sessão ou ao documento). A razão disto é que a pessoa começa a se especializar e a ganhar competência interagindo com outras ou lendo documentos. Se a interação for com uma autoridade na área, a pontuação deve aumentar mais ainda.

Autoridade em um assunto é uma pessoa com pontuação acima de um limite pré-estabelecido para aquele assunto, significando que esta pessoa possui um elevado grau de competência. A pontuação deve diminuir a cada ano, se a pessoa não tiver nenhum evento (sessão ou documento lido) nesta área. Além disto, caso a pessoa venha a colocar um documento na base, ela terá sua pontuação incrementada.

Isto é semelhante à proposta do sistema ReferralWeb (avaliado em TERVEEN & HILL, 2001), que permite identificar pessoas que tenham determinado conhecimento (*expertise*), analisando documentos da Web para achar co-ocorrências de nomes de pessoas com o intuito de criar uma rede social.

Na base de perfis, ainda deverão existir, associadas a cada pessoa, uma lista de documentos lidos ou recuperados e uma lista de recomendações feitas. Isto evitará que o sistema recomende algo que a pessoa já conheça ou que o sistema repita uma mesma recomendação.

3.7 Módulo de Recomendação

O módulo de recomendação recebe como estímulo o contexto da sessão. Este contexto é a posição do apontador de assunto, dentro da estrutura do *thesaurus*, conforme identificado anteriormente pelo módulo de análise. Quando mais de um assunto é tratado na mesma sessão de discussão, o sistema faz as recomendações no assunto corrente.

Como a discussão é síncrona, o sistema não deve interromper a discussão para que seja analisada a recomendação. Entretanto, acredita-se ser melhor fazer a recomendação durante a sessão e não após, para aproveitar o momento. Para tanto, o sistema proposto utiliza janelas específicas em sua interface para tais sugestões (ver figura 4). As recomendações são individuais. Isto quer dizer que diferentes sugestões podem ser apresentadas na mesma sessão para pessoas diferentes. Isto dependerá do perfil do participante, conforme definido na base de perfis (evita que a mesma recomendação seja feita mais de uma vez).

As recomendações podem ser de documentos (extraídos da base de documentos), de discussões anteriores (recuperadas do histórico de discussões) e de autoridades no assunto corrente (identificadas na base de perfis).

De forma básica, são feitas recomendações de itens que estão classificados no assunto tido como corrente (conforme o apontador de assunto).

Entretanto, recomendações mais complexas podem ser feitas. Por exemplo, podem ser recomendados documentos semelhantes aos documentos vistos (lidos ou recuperados) durante a sessão, mesmo que não estejam classificados no mesmo assunto.

Outro exemplo é a recomendação de documentos vistos em sessões semelhantes. Neste caso seria feita a procura de interações semelhantes, sendo recuperados documentos sugeridos pelo sistema ou lidos pelas pessoas durante estas sessões.

4 Exemplo hipotético de uma sessão com apoio do sistema

A seguir, é apresentado um exemplo de funcionamento do sistema proposto. A sessão corresponde a uma conversa online (síncrona) entre um professor e um aluno sobre o trabalho final da disciplina. São apresentadas as mensagens dos participantes e o funcionamento interno do sistema é descrito logo após o sinal *.

Aluno: *Estou pensando em implementar uma rede neural.*

* Após esta mensagem, o sistema reconhece as *stopwords* ‘estou’, ‘em’, ‘uma’ e ‘implementar’. O termo ‘pensando’ é novo (não aparece no *thesaurus*). Será solicitado que as pessoas participantes desta sessão definam o termo no *thesaurus* (o qual será definido também como *stopword*). Neste momento, o sistema reconhece que o assunto da sessão é “redes neurais” e posiciona o apontador de assunto para este conceito no *thesaurus*. O termo ‘rede’ aparece no *thesaurus* tanto em “redes de computadores” quanto em “redes neurais”. Mas a presença de “neural” elimina a ambigüidade. Não é feita nenhuma recomendação, pois os documentos armazenados na base e relacionados ao assunto em questão já são conhecidos dos participantes.

Professor: *Que tipo de rede tu vais usar ?*

Aluno: *Uma rede back propagation.*

* Os termos ‘back’ e ‘propagation’ já aparecem no *thesaurus* formando o conceito “back propagation”, o qual encontra-se abaixo do conceito “redes neurais” (pois é um tipo de rede neural). Neste momento, o apontador de assunto posiciona-se no conceito mais específico “back propagation”. Como há um documento sobre este assunto que ainda não foi lido pelo aluno nem recomendado ao mesmo, o sistema recomenda este documento na janela apropriada, somente na interface do aluno.

Professor: *Por que não implementar uma rede CNM ?*

* Neste momento, o sistema identifica um termo novo ‘CNM’, o qual deverá ser definido no *thesaurus*. Como CNM é um tipo de rede neural, este conceito deve aparecer logo abaixo de “redes neurais” na hierarquia do *thesaurus*.

Aluno: *Eu ouvi falar que a rede CNM é adequada somente para implementar sistemas especialistas.*

* Neste momento, o sistema reconhece que o assunto da sessão moveu-se para o conceito “sistemas especialistas” e recomenda alguns documentos novos sobre este tema.

5 Conclusões e Discussão

Este trabalho descreveu a arquitetura de 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. Sua aplicação é útil, sobretudo, em organizações onde os membros estiverem geograficamente dispersos.

Além de possibilitar a comunicação entre os membros de uma organização, o sistema permite que parte do conhecimento tácito possa ser explicitado e recuperado na forma de registros das discussões, contendo as mensagens trocadas, as recomendações feitas e o rumo da discussão em termos de assuntos analisados.

Uma contribuição importante da proposta está na construção de uma base de perfis dos usuários. Com esta base, existe a possibilidade de se descobrir que pessoas possuem determinado conhecimento e qual o nível de atividade de cada pessoa numa determinada área. Isto permite que a organização possa avaliar que conhecimentos ela possui ou estão sendo

trabalhados e ainda avaliar os seus próprios membros. Como consequência, pode-se determinar a necessidade (ou não) de investimento na aquisição de determinados conhecimentos, mediante treinamento ou contratação de pessoal.

Embora haja 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.

Outra barreira está na construção do *thesaurus*. Esta construção é normalmente complexa e requer um grande investimento em termos de tempo. Sugere-se o uso de técnicas automatizadas ou, conforme já destacado na seção 3.2, a adoção de alguma classificação pré-existente. Com o sistema proposta, tal barreira pode ser minimizada pelo fato de que o *thesaurus* pode ser incrementado e refinado através da colaboração entre os usuários ao longo do uso da ferramenta.

Alguns pontos ainda merecem maiores estudos. Um deles é a identificação do assunto tratado durante uma sessão. Este é identificado mediante a análise das mensagens trocadas. Neste sentido existe a preocupação em relação ao tempo gasto para que seja determinado qual é o assunto discutido. Nem sempre isto ficará claro a partir das primeiras mensagens e além disto discussões anteriores poderão vir a ser retomadas.

Outro ponto crítico está no tempo gasto pelo sistema para realizar a recomendação de documentos. Esta análise deverá ser feita de tal forma que não venha a comprometer a interação, e de forma que a recomendação se dê a partir da análise das mensagens atuais, caso contrário, as recomendações poderão parecer estranhas e fora do contexto da discussão.

A construção e uso de diferentes *thesauri* também merece estudos mais detalhados. No momento, o sistema está planejado para trabalhar com somente um *thesaurus*. Entretanto, há a possibilidade de se utilizar mais de um *thesaurus* para suportar a discussão entre pessoas de áreas diferentes (por exemplo, uma da área de Computação e outra de Educação). Uma solução futura deverá integrar vários *thesauri* de forma a garantir a correta identificação do assunto tratado, assim como a recomendação adequada de recursos.

Ainda em relação ao *thesaurus*, é necessário analisar o caso de um mesmo termo possuir diferentes significados para usuários distintos. Esta situação pode ser contornada permitindo que haja definições diferentes para o mesmo termo ou conceito e que no *thesaurus* seja indicado o autor de cada definição.

Em algumas situações o sistema deverá analisar o perfil dos usuários visando determinar que recomendações deverão ser feitas aos usuários. Quando dois usuários comentarem a respeito de um determinado assunto, mas não existindo no perfil de ambos a indicação deste conhecimento, pode-se considerar que a discussão sofreu um desvio de tema ou ainda que há a necessidade de atualizar as áreas de interesse dos usuários.

Quanto à recomendação, é necessário definir com mais precisão se uma recomendação deverá ou não ser refeita. Por vezes o fato de um assunto estar sendo rediscutido pode indicar que o usuário não utilizou de fato a recomendação. Isto pode ter ocorrido pela baixa qualidade do material recomendado ou pode revelar a necessidade de que os usuários devam ser lembrados de antigas recomendações. Neste caso, podem ser enquadradas aquelas recomendações feitas há algum tempo ou aquelas feitas em contextos diferentes, já que um mesmo material poderá tratar de diversos assuntos devendo ser portanto recomendado a cada vez que possa ser útil.

Futuramente, será necessário levar em conta o nível de conhecimento de cada usuário. Usuários com um nível elevado de conhecimento em uma área não terão interesse em receber

recomendações relacionadas a algum material que fornece apenas os conceitos iniciais relacionados à área.

A partir destas observações finais, pode-se verificar a necessidade de uma contínua avaliação e aperfeiçoamento da arquitetura aqui proposta de tal forma que o sistema venha de fato a fornecer um auxílio relevante.

6 Referências Bibliográficas

- DAVENPORT, T. H & PRUZAC, L. (1997) **Working knowledge – how organizations manage what they know**. Harvard Business School Press, 224 p.
- FOSKET, D. J. (1997) Theory of clumps. In: SPARCK-JONES, Karen & WILLET, Peter (eds). **Readings in Information Retrieval**. San Francisco: Morgan Kaufmann, 1997.
- GREENBERG, S. (1997) Computer supported cooperative work and groupware: an introduction to the special edition. **International Journal of Man Machine Studies**. [s.l.]: February, v34 n.2, p. 133-143.
- KOMOSINSKI, L. J. et al. (2000) Uso de Agentes para Apoio à Mediação de Diálogos Entre Estudantes via Internet. In: Congresso Iberoamericano de Informática na Educação (RIBIE 2000), V, Dez, 2000. **Proceedings...** Viña del Mar (publicado em CD).
- LAWRENCE, R. D. et al. (2001) Personalization of supermarket product recommendations. **Journal of Data Mining and Knowledge Discovery**, v.5, n.1/2, January, p.11-32.
- LÉVY, Pierre & AUTHIER, Michel. (1995) **As árvores do conhecimento**. São Paulo: Ed. Escuta, 192 p.
- LOH, S. et al. (2000) Concept-based knowledge discovery in texts extracted from the Web. **ACM SIGKDD Explorations**, v.2, n.1, Julho, p.29-39.
- NONAKA, I. & TAKEUCHI, T. (1997) **Criação de conhecimento na empresa**. 2ª Edição. Rio de Janeiro:Campus, 358p.
- RESNICK, P. Varian, H. (1997) Recommender systems. **Communications of the ACM**, v.40 n.3, Março, p.56-58.
- SCHAFER, J. Ben et al. (2001) E-commerce recommendation applications. **Journal of Data Mining and Knowledge Discovery**, v.5, n.1/2, Janeiro, p.115-153.
- SENGE, Peter M. (2001) **A quinta disciplina: arte e prática da organização que aprende**. 9. ed. Best Seller, 444p.
- SRIVASTAVA, J. et al. (2000) Web usage mining: discovery and applications of usage patterns from web data. **ACM SIGKDD Explorations**, v.1, n.2, Janeiro, p.12-23.
- STEWART, T. A. (1998) **Capital intelectual: a nova vantagem competitiva das empresas**. 4. Ed. Rio de Janeiro: Campus, 239p.
- TERVEEN, L. & HILL, W. (2001) Beyond recommender systems: helping people help each other. In: CARROLL, Jack (ed.). **Human computer interaction in the new millennium**, Addison-Wesley.