

O USO DE TÉCNICAS DE RECOMENDAÇÃO EM UM SISTEMA PARA APOIO À APRENDIZAGEM COLABORATIVA

Daniel Lichtnow

Ramiro Saldaña Garin

Luiz A. Moro Palazzo

Universidade Católica de Pelotas
lichtnow@ucpel.tche.br
rsaldana@ucpel.tche.br
palazzo@ucpel.tche.br

Stanley Loh

Universidade Católica de Pelotas - PPGINF
Universidade Luterana do Brasil
sloh@terra.com.br

Adriana Justin Cerqueira Kampff

Universidade Luterana do Brasil
Universidade Federal do Rio Grande do Sul
adriana@maristas.org.br

Tiago Primo

Pontifícia Universidade Católica do
Rio Grande do Sul
tiagoprino@gmail.com

José Palazzo Moreira de Oliveira

José Valdeni de Lima

Universidade Federal do Rio Grande do Sul
palazzo@inf.ufrgs.br
valdeni@inf.ufrgs.br

Resumo: Este artigo apresenta o SisRecCol – Sistema de Recomendação para Apoio à Colaboração. O sistema oferece ferramentas que tem por finalidade apoiar o processo de aprendizagem colaborativa. O sistema segue a perspectiva sócio-interacionista de Vygotsky, procurando a partir da identificação dos perfis interagentes, realizar recomendações que possam auxiliar no processo de aprendizado. As recomendações são feitas a partir da análise do comportamento dos usuários em um web chat e do acesso a uma biblioteca digital.

Palavras-Chave: aprendizagem Colaborativa, Groupware, Sistemas de Recomendação

Abstract: This paper presents SisRecCol system. The system offers tools which aims to support collaborative learning processes. The system follows Vygotsky's socio-interactionist theory, identifying user's profiles to produce recommendations that can assist the learning process. Recommendations are produced from the users behavioral analysis on a web chat and from a digital library access.

Keywords: collaborative Learning, Groupware, Recommender Systems

1. INTRODUÇÃO

Embora não existindo consenso quanto a sua definição, a Aprendizagem Colaborativa pode ser definida como “uma situação em que duas ou mais pessoas aprendem ou tentam aprender algo juntas” Dillenbourg *apud* [3]. Diversos trabalhos ressaltam o fato de que a elaboração do conhecimento é construção coletiva. Nonaka & Takeuchi [15], por exemplo, destacam que o conhecimento dentro de uma organização é criado a partir da interação entre os seus membros. Alguns autores, considerando o ambiente empresarial, salientam que devem ser proporcionados

momentos para que esta interação ocorra, citando inclusive situações que não estão relacionadas a atividades produtivas, mas que assumem importância por permitirem a interação entre os membros da organização [4].

Vygotsky [27], a partir de sua teoria sócio-histórica, destaca que a interação social é a base da aprendizagem. É através da interação, em um dado contexto social e cultural, que os membros de um grupo se apropriam, internalizam e geram novos conhecimentos. O desenvolvimento intelectual aparece primeiro no nível interpessoal e depois no nível intrapessoal. Assim, o processo de comunicação

assume fundamental importância no processo de aprendizagem, sendo a linguagem um signo mediador.

Através da linguagem ocorrem as trocas entre os membros de um grupo e desses com indivíduos mais experientes. Neste sentido, Vygotsky [27] destaca que cada sujeito apresenta um nível de conhecimento real (aquilo que é capaz de fazer sozinho, que já domina) e um nível de conhecimento potencial (aquilo que é capaz de fazer com a ajuda de outros elementos do grupo). A diferença entre esses níveis é chamada de zona de desenvolvimento proximal.

A interação entre indivíduos deve ser facilitada de forma que a probabilidade de ocorrência da aprendizagem colaborativa aumente [3]. Os recursos computacionais podem, em muitos casos, favorecer estas interações, permitindo, em um primeiro momento, que o contato entre pessoas ocorra mediante o uso de alguma ferramenta de *Groupware*.

Porém, a simples utilização de ferramentas que permitam o contato entre pessoas, através da troca de mensagens, por exemplo, pode não ser suficiente. Um dos problemas que podem advir do uso dos recursos computacionais está na sobrecarga de informações. Além disto, existem no contexto de uma organização, problemas recorrentes, para os quais já foram discutidas prováveis soluções havendo por vezes material relacionado que nem sempre é acessado facilmente. Para muitos destes problemas a solução pode ser indicada por membros mais experientes, que nem sempre são identificados com facilidade.

O sistema SisRecCol busca exercer o papel de mediador entre os sujeitos, considerando as características individuais, e sugerindo materiais complementares para esclarecimento ou para o aprofundamento dos assuntos de interesse dos usuários do sistema, buscando desta maneira ampliar as zonas de desenvolvimento dos aprendizes e, portanto, permitir que novas aprendizagens aconteçam.

Para tentar atingir estes objetivos, o SisRecCol faz uso de técnicas que são utilizadas nos chamados Sistemas de Recomendação. A idéia básica de um Sistema de Recomendação é que sejam reconhecidas as preferências e/ou necessidades dos seus usuários sendo recomendado de forma automática aquilo que pode ser de seu interesse. Sistemas de Recomendação vêm sendo utilizados em diversos contextos, como comércio eletrônico, por exemplo. No SisRecCol, as recomendações podem ser feitas durante uma seção de *chat* ou quando o usuário está acessando a Biblioteca Digital. O apoio à aprendizagem coletiva se dá durante a sessão de *chat* e também através da troca indireta de fontes de informações entre os usuários cadastrados; este último caso acontece quando um usuário incluiu na Biblioteca Digital um documento que pode ser relevante a outros usuários e o próprio sistema de recomendação identifica isto e recomenda tal documento a estes usuários.

A seção 2 deste artigo trata de temas relacionados aos Sistemas de Recomendação, sendo descritas as principais abordagens utilizadas no seu desenvolvimento.

A seção 3 descreve a arquitetura e as técnicas que foram utilizadas na implementação do SisRecCol. Já na seção 4 é feito um relato sobre os experimentos realizados com o sistema. Finalmente, a seção 5 apresenta as considerações finais do trabalho.

2. SISTEMAS DE RECOMENDAÇÃO

O objetivo dos Sistemas de Recomendação é o de que os usuários não apenas recebam o retorno de itens a partir da formulação de consultas, mas que o possível interesse do usuário por um determinado item possa ser previsto. Esta tentativa de prever o que é mais adequado para um usuário é uma forma de tentar evitar a sobrecarga de informações.

Os Sistemas de Recomendação vêm sendo implementados mediante o uso de diversas técnicas. As técnicas variam em função de aspectos como o objetivo a ser atingido e o tipo de item a ser recomendado. Um item a ser recomendado pode ser qualquer coisa tal como um artigo, um DVD, um filme, um CD ou uma peça de teatro.

Para implementação dos Sistemas de Recomendação existem duas abordagens principais: *Content-based Filtering* – Filtragem Baseada em Conteúdo e *Collaborative Filtering* – Filtragem Colaborativa [22,24,17]. Eventualmente, as duas abordagens podem ser aplicadas simultaneamente.

A Filtragem Baseada em Conteúdo parte do princípio de que os usuários tendem a interessar-se por itens similares aos que demonstraram interesse no passado. Assim é definida a similaridade entre os itens. Porém estabelecer esta similaridade entre itens pode não ser fácil em algumas situações [24]. Para que seja estabelecida, por exemplo, a similaridade entre itens como roupas e brinquedos seria necessário que fossem identificados nos itens atributos a serem comparados (cor, preço, marca, por exemplo).

Já quando os itens correspondem a artigos (ou *sites*), este processo de comparação pode ser facilitado, pois documentos podem ser considerados similares se compartilharem termos em comum. Assim esta abordagem guarda forte relação com a área de *Information Retrieval* – Recuperação da Informação [23].

Além de salientarem que a comparação entre itens nem sempre é facilmente automatizada, Shardanand & Maes [24] destacam o fato de que não é feita, no caso da abordagem Baseada em Conteúdo, a avaliação qualitativa dos itens recomendados. No caso de documentos textuais, por exemplo, basta que o conteúdo seja similar (presença de termos comuns) para que dois textos sejam considerados similares.

Em função destas limitações, foram propostas outras técnicas de recomendação. Dentre estas destacam-se as técnicas de Filtragem Colaborativa. O termo Filtragem Colaborativa está originalmente ligado a um sistema denominado *Tapestry* [7]. Esta abordagem procura verificar

usuários que tem gostos semelhantes. Assim, no caso de um sistema que recomende filmes, a recomendação de um filme será feita a um usuário se este filme tiver sido bem avaliado por algum usuário com perfil semelhante ao seu. Usuários com perfil semelhante, neste caso, são usuários que no passado avaliaram os mesmos filmes de maneira similar.

As técnicas de Filtragem Colaborativa levam em conta que no dia-a-dia, é comum pessoas recomendarem ou pedirem recomendações quanto a filmes, CDs, livros, software e outros produtos. Assim, um sistema de recomendação concebido segundo esta abordagem terá por objetivo a automação do processo “boca-a-boca” – *Word of Mouth* [24].

Balabanovic & Shoham [2], destacam que no caso da Filtragem Colaborativa, o principal problema reside no fato de que os novos itens que surgem somente são recomendados após terem sido avaliados por algum usuário. Deve ser considerado que os sistemas de recomendação que utilizam a Filtragem Colaborativa apresentarão melhores resultados a medida que um número considerável de usuários avaliarem os mesmos itens. A aplicação desta abordagem pode portanto ser um problema para os casos em que existem usuários com preferências bastante distintas.

Em função das limitações de cada uma destas abordagens existem vários exemplos de Sistemas de Recomendação que utilizam as duas abordagens em conjunto. Em [2] é apresentado o sistema de recomendação *Fab* que procura trabalhar com as duas abordagens: Baseada em Conteúdo e Colaborativa.

Sistemas de Recomendação podem ser usados para facilitar o processo de aprendizado. Em [19] é apresentado um assistente virtual que além de interagir com os usuários usando linguagem natural, procura determinar conteúdo a ser recomendado aos usuários a partir do acompanhamento das ações destes usuários, bem como da comparação entre os usuários. Assim, o sistema constata que para resolver determinado exercício, disponibilizado no ambiente da ferramenta, a maioria dos alunos acessa determinados conteúdos.

3. DESCRIÇÃO DO SISTEMA SISRECOL

O objetivo principal do sistema SisRecCol é apoiar o processo de construção do conhecimento dos usuários. Este apoio é fornecido a partir das recomendações feitas aos usuários. Inicialmente o SisRecCol foi concebido como um sistema acessível via *Web* onde as recomendações ocorreriam sempre dentro de uma sessão de *chat*.

Na arquitetura do sistema mostrada na Figura 1, pode-se notar o módulo de *text mining* que é responsável por acompanhar a troca de mensagens no *chat* e identificar os assuntos tratados a partir de consultas feitas à ontologia. Identificados os assuntos tratados, esta informação é passada para o módulo de recomendação, que consultando

a base de perfis gera as recomendações para cada um dos usuários participantes da sessão.

Posteriormente, novas funcionalidades foram implementadas e as recomendações passaram a ser feitas em outros contextos, não ficando limitadas as sessões de *chat*. Atualmente, as recomendações feitas pelo SisRecCol ocorrem em dois contextos:

- em um *web chat*, as recomendações são feitas durante a discussão a partir da identificação dos assuntos tratados e da identificação de algum recurso que esteja relacionado a este assunto. Estes recursos consistem de artigos e *sites* previamente indexados e armazenados em uma biblioteca digital privada, assim como de indicações de outros usuários, que supostamente têm maior experiência na área e de sessões anteriores relacionadas ao mesmo assunto;
- fora do *web chat*, o usuário pode recuperar as recomendações feitas durante a sessão do *chat*, sendo que novas recomendações podem ser acessadas, recomendações estas feitas a partir da análise das ações dos usuários (documentos acessados e/ou avaliados por outros usuários, por exemplo).

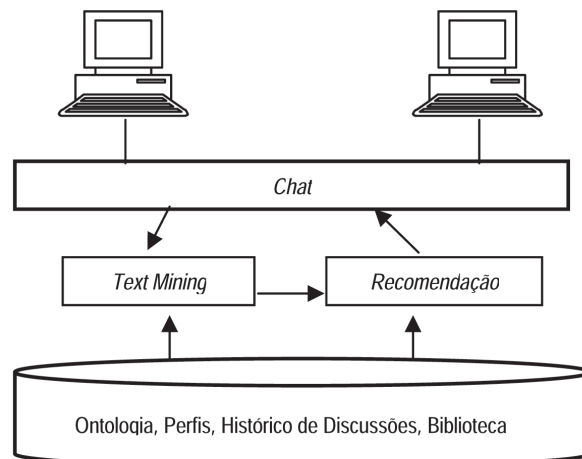


Figura 1: Arquitetura do SisRecCol.

O sistema foi implementado utilizando as linguagens de programação *PHP* e *Javascript*, o banco de dados *MySQL*, servidor *web* *Apache* e sistema operacional *Linux*. Um protótipo do sistema encontra-se disponível no endereço <http://gpsu.ucpel.tche.br/sisrec>. As próximas seções descrevem sucintamente o funcionamento de cada um dos módulos do sistema. Algumas funcionalidades do sistema foram apresentadas em trabalhos anteriores [9,11]. Em relação aos trabalhos anteriores, este artigo descreve algumas novas técnicas de recomendação que foram incorporadas ao SisRecCol no ambiente do *chat* e especialmente fora do *chat*.

3.1. O CHAT

O *chat* desenvolvido para a ferramenta assemelhasse a um *chat* convencional, onde usuários trocam mensagens textuais. Uma limitação do protótipo reside no fato de que não é permitida a existência de diversos canais de discussão. Apenas usuários cadastrados podem usar o sistema.

3.2. A ONTOLOGIA DE DOMÍNIO

Uma ontologia de domínio (*domain ontology*) é uma descrição de “coisas” que existem ou podem existir em um domínio [16,26] e contém o vocabulário relacionado ao domínio em questão [8].

No sistema proposto, a ontologia é implementada como um conjunto de conceitos em uma estrutura hierárquica (um nó raiz, e nós pais e filhos). Cada conceito, que representa um assunto, tem associado a si uma lista de termos e seus respectivos pesos, que ajudam a identificar os conceitos nos textos. O SisRecCol utiliza a ontologia de domínio para identificar os assuntos presentes nas mensagens do *chat*, nos itens da biblioteca digital e para auxiliar na definição do perfil dos usuários. No momento, há somente uma ontologia no sistema e esta define a área de Ciência da Computação.

A ontologia foi criada de forma semi-automática. A seleção de conceitos (áreas e subáreas da hierarquia) foi feita manualmente por especialistas seguindo a classificação da ACM (*Association for Computing Machinery* – www.acm.org) para a área de Computação. Estes especialistas selecionaram documentos (aproximadamente 100 documentos para cada conceito) sendo então utilizada uma ferramenta de software para identificar os termos mais relevantes e determinar o seu peso, isto é, o quanto um termo identifica um assunto. Este peso nada mais é do que um valor numérico que foi calculado com base na frequência do termo dentro dos documentos. Posteriormente, uma revisão dos termos e pesos foi feita pelos especialistas. Na ontologia existem termos em português e inglês, sendo que foi necessário incluir os termos nas duas línguas quando o processo automático não conseguiu identificá-los. O método não utiliza *stemming* [1], razão pela qual foi necessário gerar manualmente as variações lingüísticas.

3.3. O MÓDULO DE TEXT MINING

Este módulo identifica os assuntos tratados a partir da análise dos termos presentes nas mensagens e da comparação destes com conceitos presentes em uma ontologia de domínio. O método de classificação de textos que foi utilizado foi apresentado em [12]. Este método de classificação de textos está baseado em técnicas probabilísticas e, portanto, não utiliza técnicas de PLN (Processamento de Linguagem Natural) para analisar a sintaxe de cada mensagem. O algoritmo é baseado nos algoritmos Rocchio e Bayes [21,18,10].

Basicamente, o método avalia a similaridade entre o texto e um conceito, usando uma função de similaridade que calcula a distância entre os dois vetores, um relacionado às mensagens do *chat* e outro relacionado aos conceitos da ontologia. Os vetores que representam as mensagens e os conceitos são compostos por uma coleção de termos, existindo um peso (valor numérico) associado a cada termo. No caso das mensagens do *chat*, o peso de cada termo é dado pela frequência relativa de cada termo no texto. Na montagem dos vetores, relacionados aos assuntos presentes na ontologia e as mensagens do *chat*, são ignoradas as *stopwords*, isto é, termos como preposições que aparecem com muita frequência.

Aplicando a função de similaridade relacionada ao método é feita a multiplicação dos pesos dos termos que estão presentes nos dois vetores (vetor construído a partir das mensagens e cada vetor relacionado a cada um dos assuntos presentes na ontologia), sendo que a soma destes produtos, limitada a 1, é o grau de similaridade existente entre a mensagem e o assunto.

Cabe ressaltar que o método produz uma classificação graduada, isto é vários assuntos podem ser identificados em uma mensagem e para cada um dos assuntos, pode estar associado um grau distinto. Este grau indica qual a probabilidade do assunto estar presente na mensagem. Quando este grau está abaixo de um valor mínimo, um limiar, então os assuntos identificados são desconsiderados. Este limiar foi definido a partir de alguns experimentos, sendo ainda objeto de avaliação. Conforme indicado na Figura 1, o módulo de *Text Mining* faz a identificação dos assuntos e envia esta informação para o módulo de recomendação.

3.4. A BIBLIOTECA DIGITAL

Todo material a ser recomendado aos usuários do SisRecCol está armazenado em uma Biblioteca Digital. A Biblioteca Digital consiste de uma coleção de documentos no formato *pdf*, *doc* e *rtf*, sendo que são também armazenados os dados e o conteúdo (o texto) de *web sites*.

A indexação de todo o material disponível na Biblioteca Digital é feita de forma automática, usando o método descrito na seção 3.3. Assim, todo documento é associado a um ou diversos conceitos, sendo que para cada conceito é atribuído um valor numérico que indica a probabilidade deste conceito estar presente em um documento. A indexação prévia do material permite maior agilidade ao processo de recomendação.

O cadastro é permitido a qualquer usuário, mas uma vez cadastrado alguns usuários com privilégios específicos deverão liberar o acesso do material cadastrado aos outros usuários da ferramenta. A idéia é manter a qualidade do material cadastrado e incentivar o processo de colaboração entre os usuários da ferramenta.

Mesmo usuários que não participem de uma sessão de *chat*, podem acessar o conteúdo da Biblioteca Digital. Para tanto, estão disponíveis aos usuários recursos que permitem a elaboração de consultas. Nestas consultas, cada usuário informa como parâmetros o conceito e/ou alguma palavra-chave a ser procurada no título do documento e/ou dentro do seu conteúdo. De forma a evitar o retorno de um grande número de documentos, o usuário deve informar ainda um limiar (*threshold*), que corresponde a um valor numérico que está relacionado ao grau atribuído a relação documento-conceito quando da indexação. Assim, caso o usuário pesquise por documentos relacionados à área de Inteligência Artificial, somente serão retornados documentos cujo grau da relação esteja acima do valor do limiar indicado pelo usuário. Ao realizar consultas, abrir e avaliar documentos o perfil de cada usuário vai sendo construído e atualizado.

3.5. A BASE DE PERFIS

A base de perfis mantém informações sobre os usuários do SisRecCol. Inicialmente eram armazenadas informações cadastrais como nome, endereço e *e-mail*. Atualmente o grau de interesse dos usuários em relação a determinado assunto (conceito da ontologia) passou a ser também armazenado.

A base de perfis procura atender a dois objetivos principais. O primeiro diz respeito à qualidade da recomendação, pois sabe-se que quanto melhor for o conhecimento dos interesses de um usuário tanto melhor será a qualidade das recomendações feitas. Já o segundo objetivo é o de que esta base de perfis seja usada como Páginas Amarelas, indicando quem dentro de uma comunidade de usuários possui determinados conhecimentos ou habilidades (ou pelo menos interesses). As Páginas Amarelas não armazenam a solução de problemas, mas indicam que certas pessoas possuem determinado tipo de conhecimento [4].

No SisRecCol estão sendo feitas algumas tentativas de manter atualizado o perfil do usuário. Esta atualização é feita mediante o acompanhamento das ações do usuário tais como:

- cadastrar documentos na Biblioteca Digital;
- cadastrar *web sites* na Biblioteca Digital;
- participar de sessões do *chat*;
- avaliar documentos recomendados.

Para cada uma destas ações foi atribuída uma pontuação. Para ilustrar, no caso da avaliação dos documentos, a pontuação é dada de acordo com o conceito atribuído pelo usuário. Quando um usuário considera o documento nada relevante é atribuído peso -1 (menos um) a esta ação, 1 (um) quando é pouco relevante e 2 (dois) quando o usuário considera o documento muito relevante.

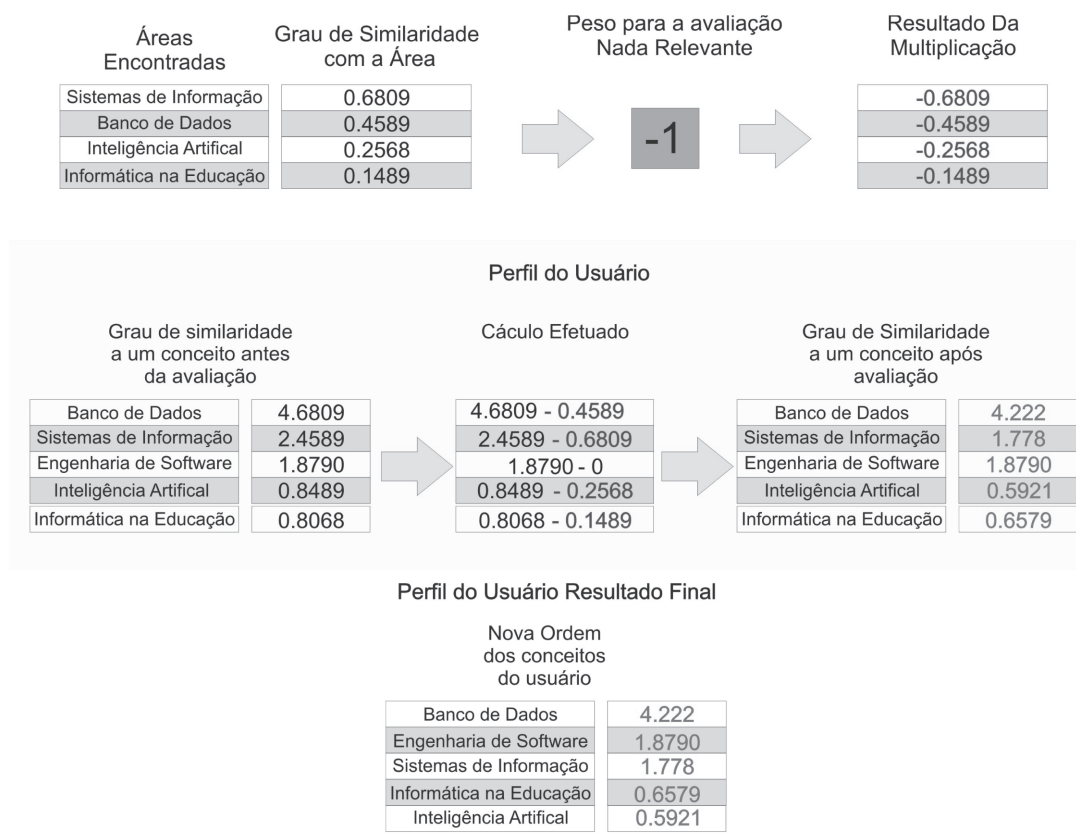


Figura 2: Representação do processo de atualização do perfil.

Estes valores (-1, 1 e 2) são multiplicados pelos valores dos graus de afinidade encontrados no documento avaliado. O resultado da multiplicação é então somado ou subtraído do peso associado ao usuário. Na Figura 2 é mostrado o processo. Como o usuário considerou que um documento relacionado ao conceito de Sistemas de Informação era “Nada Relevante” o seu grau de afinidade a esta área e as outras presentes no documento sofreram um decréscimo.

Cabe ressaltar que a avaliação dos documentos pode ser feita pelos usuários posteriormente, ou seja, em outro momento que o usuário entrar no SisRecCol e não apenas no momento em que ele recebe a recomendação. O usuário pode alterar as avaliações feitas anteriormente, neste caso quando um usuário reavaliar um documento, o seu perfil é atualizado.

A avaliação por um usuário de um documento recomendado pode ser bastante subjetiva, sendo dependente do seu nível de conhecimento e de seus interesses momentâneos. Neste sentido, existem trabalhos que consideram o peso da opinião de um usuário com base em sua reputação [25]. Ainda cabe destacar que usuários com um mesmo nível de conhecimento irão provavelmente avaliar um documento de forma similar, razão pela qual mesmo a opinião de usuários com menor conhecimento pode ser considerada.

Além disto, embora problemas possam ser gerados pela avaliação feita pelos usuários, esta avaliação fornece indicativos que não são obtidos quando todo processo de recomendação têm por base o conteúdo dos documentos. Neste caso, a simples presença de termos dentro de um documento (de boa ou má

qualidade) poderia fazer com que este fosse recomendado.

Um problema relacionado aos Sistemas de Recomendação reside no fato de que no início do uso, o perfil do usuário não está definido. Este problema conhecido por *cold start* pode ser minimizado mediante a construção de um perfil inicial. Isto pode ser feito solicitando que o usuário informe suas preferências ou que ele avalie alguns itens que podem ser recomendados [14]. No SisRecCol, uma proposta foi construir o perfil inicial do usuário a partir das informações do seu currículo Lattes. [20].

3.6. A BASE DE DISCUSSÕES ANTERIORES

As sessões anteriores são armazenadas de forma a permitir que estas possam vir a ser lidas em um momento posterior por algum dos participantes ou por alguém que não tenha participado da sessão, mas que tenha interesse nos assuntos discutidos. No SisRecCol, toda sessão que é armazenada é recomendada como um recurso a ser consultado. Espera-se assim aproveitar de forma mais efetiva o conhecimento gerado pela troca de mensagens entre os diversos usuários em alguma sessão anterior.

Além disto, os registros das sessões anteriores podem ser utilizados em análises posteriores, permitindo descobrir informações relevantes relacionadas ao comportamento dos usuários e a eventuais desvios que ocorrerem ao longo da discussão. Cabe ressaltar que as recomendações feitas ao longo da sessão também ficam registradas.

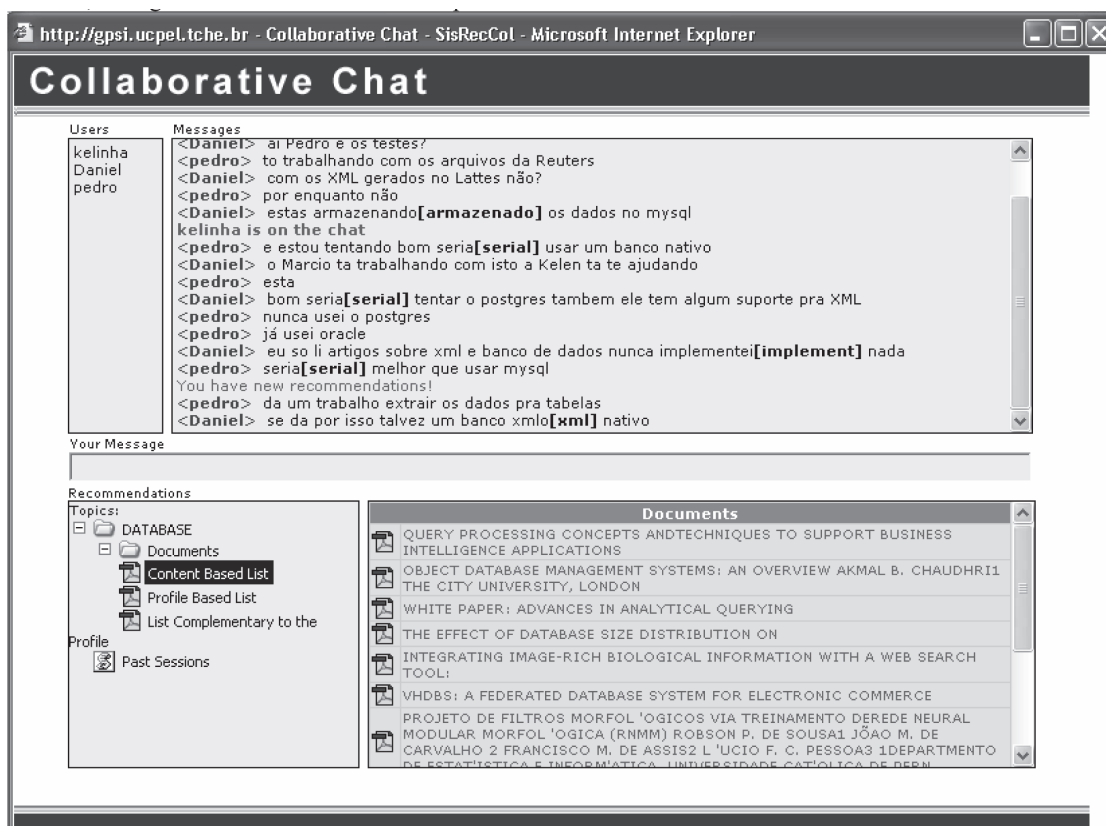


Figura 3: Recomendação na sessão de chat.

3.7. RECOMENDAÇÃO NO SISRECOL

Conforme já destacado, a recomendação no SisRecCol é feita dentro de uma sessão de *chat* e fora dela a partir do momento que o usuário acessa o sistema. São inicialmente descritas as recomendações feitas ao longo de uma sessão de *chat*, sendo posteriormente descritas as recomendações feitas fora do *chat*.

Para que algumas das recomendações fossem feitas, foram implementadas funções que permitem calcular a similaridade existente entre o perfil do usuário e os documentos existentes na Biblioteca Digital. Para isto foi usada uma variação do método de *Jaccard* [28].

3.7.1. RECOMENDAÇÕES FEITAS DURANTE A SESSÃO DO CHAT

Em uma sessão de *chat*, o módulo de recomendação recebe um conceito (assunto) identificado pelo módulo de *text mining* e oferece conteúdos relacionados ao tema discutido, durante uma sessão na sala de discussões. A Figura 3 apresenta a interface do *chat* do SisRecCol com a simulação de uma sessão de *chat*. Cada conceito é identificado a partir do momento em que os termos presentes nas mensagens permitem esta identificação.

O diálogo mostrado na Figura 3, está relacionado a discussão sobre o desenvolvimento de uma aplicação que irá armazenar informações em um banco de dados. O sistema analisa cada mensagem e identifica os conceitos usando o método descrito na seção 3.3. Maiores informações sobre a identificação de assuntos no *chat* podem ser obtidos em [13].

Na Figura 3 são apresentadas as recomendações feitas durante a sessão. Estas recomendações estão agrupadas por tipos. A *Content Based List*, apresenta a recomendação de documentos de acordo com o conceito identificado na discussão corrente do *chat*. Para gerar a *Profile Based List* é percorrida a Biblioteca Digital, sendo recomendados ao usuário os documentos que estão relacionados ao seu perfil. Esta lista tem como intenção recuperar os documentos considerando as áreas de maior interesse (ou afinidade) do usuário. A *List Complementary to the Profile* tem como finalidade complementar o conhecimento de um usuário, com os documentos relacionados às áreas menos acessadas pelo usuário no sistema, isto é, aquelas áreas com as quais o usuário tem provavelmente menor afinidade. O objetivo deste tipo de recomendação é auxiliar o usuário a se fortalecer em áreas onde ele tenha algum déficit, por exemplo, áreas onde ele esteja começando ou temas periféricos ao seu interesse, mas que podem contribuir com sua formação. Isto pode ser útil em um contexto restrito, como alguma disciplina de um curso de graduação, servindo para auxiliar alunos que estão apresentando algumas dificuldades em tópicos específicos.

A lista de *Past Sessions* apresenta as sessões que trataram do assunto que foi identificado através do processo de *text mining*. Podem ainda ser recomendados

aos usuários uma lista de *Websites* e uma lista dos usuários que possuem, em princípio, maior afinidade em relação ao conceito da ontologia que foi identificado ao longo da sessão. No caso dos usuários recomendados é levado em conta o perfil que é construído a partir das ações destes usuários.

As recomendações não devem interromper a discussão. Em função disto, as mesmas aparecem em um quadro separado. As recomendações são individualizadas, isto é, cada usuário recebe uma lista diferente de recomendações. Além disto, o sistema não recomenda o mesmo item mais de uma vez na mesma sessão, nem itens já acessados pelo usuário.

3.7.2. RECOMENDAÇÕES FEITAS FORA DA SESSÃO DO CHAT

Conforme já destacado, o SisRecCol, deixou de ser apenas um *chat* com recomendações e passou a disponibilizar alguns módulos que fazem recomendações aos usuários fora de uma sessão do *chat*. Assim, fora da sessão de *chat* são feitas as seguintes recomendações:

- **Documentos mais acessados no sistema.** Leva em consideração todos os documentos abertos pela comunidade que usa o SisRecCol, sendo limitada a 10 documentos. O objetivo deste tipo de recomendação é apresentar uma lista de quais seriam os documentos que no histórico do sistema obtiveram o maior número de acessos. Pode ser útil para identificar quais documentos possuem, aparentemente, maior aceitação dentro da comunidade de usuários do SisRecCol.
- **Documentos mais acessados pelos usuários que mais acessaram/leram documentos.** O objetivo deste tipo de recomendação é apresentar uma lista de documentos que são, provavelmente, documentos importantes por terem sido lidos por um grupo seletivo (a princípio) de usuários.
- **Documentos acessados pelos usuários que mais incluíram documentos na Biblioteca Digital.** Novamente está se supondo que usuários que atualizam a Biblioteca Digital costumam ter bons conhecimentos relacionados às áreas de interesse da comunidade de usuários da biblioteca, assim os documentos por eles acessados podem ser documentos importantes.
- **Documentos lidos pelos usuários mais ativos em um determinado conceito.** A suposição é que estes usuários podem fornecer pistas acerca de que documentos de uma área poderão ser importantes.
- **Documentos lidos nos últimos 30 dias pelos usuários mais ativos em um determinado conceito.** Semelhante a lista de usuários mais ativos, com a diferença que nesta recomendação apenas os documentos mais recentemente acessados são considerados.

- **Quem Leu Também Leu.** Esta técnica é similar aquelas presentes em *sites* como o *Amazon.com*. Assim, quando um usuário abre um documento são apresentados documentos que outros usuários que acessaram aquele documento também acessaram. Este tipo de recomendação é feito também no contexto da sessão de *chat*.

4. AVALIAÇÃO PRELIMINAR DO SISRECCOL

O sistema aqui descrito pode ser avaliado sob diferentes aspectos como: a qualidade da recomendação, a satisfação dos usuários com o sistema, os benefícios gerados pelo uso do sistema, a identificação dos conceitos presentes nas mensagens na sessão de *chat*, por exemplo. Para avaliar o uso do sistema, foram realizados experimentos com estudantes de diferentes disciplinas de cursos de Computação.

Alguns experimentos consistiram em utilizar o sistema em aulas de disciplinas da área de Computação (Banco de Dados, Sistemas de Informação, Linguagens Formais e Autômatos e Mineração de Dados). Não foi definida nenhuma medida de avaliação. O objetivo era simplesmente avaliar o comportamento dos alunos nesta situação. Os alunos poderiam estar em salas diferentes e até mesmo fora da universidade. O professor estimulava os alunos postando mensagens no *chat* com perguntas ou explicações sobre o assunto. Notou-se que a participação dos alunos no *chat* era bastante variada. Alguns alunos contribuíam seriamente com o tema e outros dispersavam com outros assuntos.

No tocante às recomendações, poucos alunos abriram os documentos recomendados durante o *chat*. Destes, alguns tiveram dificuldade em ler os documentos, pois estavam em inglês e não tinham muita fluência nesta língua. Entretanto, os poucos que conseguiram acessar os documentos e compreender o que estava escrito, conseguiram ter uma idéia melhor do tema e isto ajudou-os a entender melhor o assunto sendo discutido. Cabe salientar que, conforme esperado, nenhum aluno conseguiu ler um documento inteiro durante o *chat*. Foi possível apenas ler o resumo ou acessar algumas partes mais específicas. Disto conclui-se que o sistema não deve ser utilizado com o intuito de que os alunos leiam os documentos durante a sessão, mas apenas que acessem algumas partes ou que verifiquem a relevância do documento durante a sessão e o marquem para uma leitura detalhada posteriormente.

No entanto, manter a recomendação de documentos durante uma sessão de *chat* pode ser útil uma vez que o estímulo a leitura, mesmo posterior é dado em um momento em que existe normalmente a necessidade e o desejo de buscar maiores informações sobre os assuntos tratados. Deve ainda ser considerado que é comum que os participantes de uma sessão de *chat* indiquem algum material (livro, site, artigo, etc) que possa vir a ser consultado posteriormente. Assim, o sistema procura assumir e colaborar com esta tarefa.

Um outro experimento foi realizado de forma mais direcionada. O professor entregou aos alunos (total de 15) uma lista de exercícios com questões a serem respondidas (5 perguntas sobre temas discutidos na disciplina), sabendo que as respostas encontravam-se em documentos que estavam armazenados na Biblioteca Digital do sistema. Foi feito um treinamento básico com os alunos para eles entenderem as funções do sistema. Depois, os alunos em duplas começaram a pesquisar documentos no sistema. Com isto, o sistema pôde gerar perfis para cada usuário e assim iniciar as recomendações. Foi dada uma semana para os alunos utilizarem o sistema e conseguirem as respostas às questões. Apenas alguns alunos conseguiram completar a tarefa (apenas 2 duplas) e ainda assim não obtiveram respostas para todas as questões.

Após o experimento, foi feita uma reunião com os alunos para uma avaliação qualitativa do processo. Grande parte dos alunos teve dificuldades em utilizar ou entender as funções do sistema; ficou evidente que o treinamento foi muito rápido. Os que não conseguiram responder nenhuma questão indicaram como causa principal a dificuldade em entender a lista de documentos recomendada pelas várias técnicas disponíveis, cada qual apresentando uma ordem diferente para os documentos (*ranking*), de acordo com sua importância na lista. Isto ocorreu principalmente porque a explicação das listas é muito rápida e há trechos em inglês. Outro fator que influenciou negativamente a procura de material relevante é que a configuração do limiar mínimo de importância deveria ser feita pelo próprio usuário (ver seção 3.4). Uma vez que este não sabia como fazer esta configuração, as listas estavam muito grandes, gerando sobrecarga.

Os alunos que conseguiram responder alguma questão relataram que foi necessário acessar vários documentos e ler cada documento inteiro; isto porque o sistema não colocava no topo os documentos com a resposta, nem indicava o trecho do documento com a informação desejada. Esta pode ser uma melhoria futura no sistema: separar o documento em partes. Estamos trabalhando em pesquisas que consigam separar automaticamente um texto em partes de acordo com a similaridade entre as partes, de forma a poder mais tarde recomendar apenas parte de um documento.

Outra melhoria futura é rever o modo como o *ranking* dos documentos é criado. Atualmente, a maioria das técnicas é baseada no perfil do usuário e não nas consultas feitas por ele. Isto pode ter influenciado o resultado do experimento já que o perfil é algo mais perene, enquanto que as consultas refletem necessidades momentâneas (*ad hoc*). Uma das pesquisas em andamento é justamente estudar a evolução do perfil do usuário de acordo com períodos de tempo, permitindo avaliar a persistência das áreas de interesse e assim avaliar a relevância dos documentos de acordo com áreas mais constantes ou momentâneas.

Pode-se concluir destes experimentos que o sistema oferece melhores resultados quando utilizado como indicativo

da relevância de documentos para interesses mais persistentes, não sendo tão útil quando o usuário possui uma necessidade muito específica e passageira. Neste caso, um bom sistema de busca por palavras-chave poderia ser mais útil.

Ainda no que se refere à qualidade da recomendação feita durante a sessão de *chat* constatou-se que a qualidade das recomendações (na visão dos usuários) está mais associada à qualidade dos itens presentes na Biblioteca Digital do que propriamente a acertos no processo. Assim, o emprego de técnicas relacionadas à abordagem Colaborativa (ver seção 2) que é utilizada em alguns sistemas de recomendação, poderia aprimorar a qualidade do material recomendado. Cabe destacar que a recomendação feita durante a sessão de *chat* no SisRecCol é tipicamente uma recomendação feita seguindo a abordagem Baseada em Conteúdo.

Algumas das recomendações feitas fora da sessão de *chat* são um passo nesta direção, uma vez que, em muitos casos, estas recomendações são feitas considerando ações feitas por outros usuários (como no caso dos documentos mais lidos por outros usuários). No SisRecCol foi levado em conta que as ações de outros usuários, representam avaliações implícitas, que podem indicar em muitos casos, a qualidade dos recursos.

No futuro, além de ações como o acesso de usuários a documentos, poderão ser consideradas também as avaliações explícitas feitas pelos usuários, que atualmente são utilizadas basicamente para realizar a atualização do perfil, conforme mostrado na seção 3.5. Desta forma a abordagem Colaborativa estaria sendo utilizada de forma mais efetiva. Aqui cabe considerar que os usuários deverão estar motivados para contribuir com as suas avaliações. Propostas para motivar os usuários a contribuírem com avaliações criteriosas são discutidas por exemplo em [6].

No que se refere à avaliação da correção na identificação dos conceitos (assuntos) das mensagens durante a sessão, esta representa sem dúvida um desafio. Elnahrawy [5] destaca que esta identificação *on-line* é bastante difícil, pelo fato de que vários assuntos podem ser tratados simultaneamente em uma sessão e pelo fato de que mensagens de diferentes assuntos são intercaladas. A avaliação foi feita pelas próprias pessoas que participaram das sessões, sendo que foi avaliado o grau de acerto na identificação dos assuntos relacionados às mensagens, assim como as ocasiões em que os assuntos deveriam ter sido identificados e não o foram. Na avaliação foram considerados dois métodos. No primeiro, cada mensagem era avaliada individualmente, isto é as mensagens anteriores não tinham qualquer influência na identificação do assunto. No segundo, procurou-se considerar o contexto de cada mensagem, levando em conta as mensagens anteriores. Neste caso, o assunto só era identificado quando estivesse presente em várias mensagens ou fortemente presente em uma mensagem.

Na avaliação, foram consideradas a precisão e a abrangência. A precisão foi medida pela razão entre o número de mensagens onde o assunto foi corretamente

identificado e o número de mensagens onde o assunto foi identificado. A abrangência foi medida pela razão entre número de mensagens onde o assunto foi corretamente identificado e o número de mensagens onde deveria ter sido identificado um assunto.

Três sessões foram avaliadas. O melhor resultado obtido foi mediante o uso do segundo método, que avalia um conjunto de mensagens, sendo obtida precisão de 85,7% e abrangência de 85,7%. Já o pior resultado foi obtido com o método que avaliava cada mensagem individualmente, tendo sido obtida a precisão de 37,5 % e abrangência de 66,7 %. Embora novas e mais consistentes avaliações devam ser feitas, fica claro, como já esperado, que análise de um conjunto de mensagens favorece o método. Outras considerações podem ser vistas em [13].

Importante ressaltar que a análise posterior dos assuntos identificados nas mensagens permitiu verificar que usuários costumavam “fugir” do assunto tratado, ou situações onde o assunto foi aprofundado, o que foi indicado pela identificação inicial de um conceito pai passando nas mensagens seguintes à identificação de assuntos “filhos”.

5. CONSIDERAÇÕES FINAIS

Levando em conta que a aprendizagem colaborativa ocorre a partir do momento em que a interação e troca de informações é facilitada, este trabalho apresentou o SisRecCol que é um sistema que procura auxiliar na construção coletiva e individual do conhecimento mediante o uso de técnicas de recomendação que procuram indicar aos seus usuários material que possa ser útil, que incida na zona de desenvolvimento proximal de cada usuário [27].

No sistema apresentado, a recomendação de itens é feita no contexto de uma sessão de *chat* sendo que os usuários podem recuperar as recomendações feitas durante a sessão e ainda acessar outras recomendações fora do contexto do *chat*. Dentro de uma sessão de *chat*, para que as recomendações sejam feitas, são considerados o perfil do usuário e o conteúdo dos documentos. Durante as sessões, a recomendação utiliza basicamente a chamada abordagem Baseada em Conteúdo. Já no caso das recomendações feitas fora do contexto do *chat* é considerado o uso dos documentos por outros usuários, assim pode-se dizer que a neste contexto o processo de recomendação está mais fortemente relacionado à Abordagem Colaborativa. Foi possível constatar que se pode conhecer melhor o perfil e os interesses de uma comunidade e de seus membros individualmente mediante a análise das mensagens trocadas por eles e acompanhamento das suas ações dentro de um ambiente, como o oferecido pelo SisRecCol.

A forma de garantir que o perfil do usuário esteja atualizado, isto é a estratégia de pontuação ainda está sendo avaliada. O que o sistema permite indicar com maior

probabilidade de acerto é o grau interesse de um usuário por determinados assuntos.

Algumas das limitações do sistema estão relacionadas à construção de novas ontologias e à atualização da Biblioteca Digital, de forma a atender comunidades com interesses diversos. O processo de construção da ontologia é demorado e demanda muito esforço por parte de especialistas. Ainda que ferramentas de software estejam sendo usadas, há o trabalho árduo de selecionar documentos de treino para representar cada conceito. Já no caso da Biblioteca Digital, a ausência de material relevante pode causar desinteresse pelas recomendações feitas e conseqüentemente pelo uso do sistema. É preciso ressaltar que a idéia é que o SisRecCol seja utilizado por uma comunidade de usuários por um período considerável de tempo, assim o esforço na construção da ontologia e na atualização da Biblioteca Digital podem ser recompensados. O desenvolvimento de ferramentas que auxiliem tanto na construção da ontologia quanto na inserção de documentos na Biblioteca Digital vêm sendo realizado.

A partir dos experimentos, foi constatada a necessidade de que sejam revistas algumas das interfaces oferecidas pelo sistema. Dentre as mudanças necessárias está o oferecimento de uma interface multilíngüe que possa ser configurada pelos usuários da ferramenta.

Finalmente, o sistema desenvolvido deverá ainda ser aplicado e avaliado em situações nas quais possam ser verificados, de forma mais concreta, os eventuais ganhos obtidos pela adoção das ferramentas disponíveis.

6. AGRADECIMENTOS

O presente trabalho foi realizado com o apoio do CNPq, uma entidade do Governo Brasileiro voltada ao desenvolvimento científico e tecnológico, e FAPERGS (Fundação de Apoio à Pesquisa do Estado do Rio Grande do Sul).

REFERÊNCIAS

- [1] R. Baeza-Yates & B. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, Harlow, England, 1999.
- [2] M. Balabanovic, Y. Shoham. *Fab: content-based, collaborative recommendation*. In: *Commun. ACM*, New York, 40(3), páginas 66–72, 1997.
- [3] L. M. Cunha, H. Fuks, C. J. P. Lucena. *A Adaptação do Ambiente AulaNet para dar Suporte a Grupos de Aprendizagem e sua Formação Utilizando os Conceitos de Agentes de Software*. *Revista Brasileira de Informática na Educação -SBC*, 11(1):26-46, Abril 2003.
- [4] T. H. Davenport, L. Pruzac. *Working knowledge – How organizations manage what they know*, Harvard Business School Press, Harvard, 1997.
- [5] E. Elnahrawy. *Log-Based Chat Room Monitoring Using Text Categorization: A Comparative Study*. In *Proc. of IKS 2002 St. Thomas, US Virgin Islands*, 2002.
- [6] A. C. C. Garcia, M. Ekstrom, H. Björnsson. *HYRIWYG: leveraging personalization to elicit honest recommendations*. In *Proceedings of the 5th ACM conference on Electronic commerce New York, NY, USA*. páginas 232 - 233, 2004
- [7] D. Goldberg, D., D. Nichols, B. M. Oki, D. Terry. *Using collaborative filtering to weave an information tapestry*. In *Commun. ACM*, New York, NY, USA, 35(12), páginas 61–70, 1992.
- [8] N. Guarino. *Formal Ontology and Information Systems*. In: *International Conference on Formal Ontologies*. In *Information Systems - FOIS'98*, Trento, Itália. páginas 3-15, Junho de 1998.
- [9] A. Kampff, et.al. *Sistema de Recomendações em Discussões Eletrônicas - Apoio à Construção do Conhecimento em Comunidades Virtuais de Aprendizagem*. In *XXV Congresso da SBC- Workshop de Informática na Escola, São Leopoldo*. páginas 2504-2512, 2005.
- [10] D. D. Lewis. *Naive (bayes) at forty: the independence assumption in information Retrieval*, in: *Proc. European Conference on Machine Learning, Lecture Notes in Computer Science*, v.1398, Springer, Berlin, páginas 4-15, 1998.
- [11] D. Lichtnow, et. al. *Análise Automática de Textos em uma Comunidade Virtual*. In: *XXXI SEMISH - XXIV Congresso da SBC*, Salvador, páginas 202-216, 2004.
- [12] S. Loh, L. K. Wives, J. P. M. Oliveira. *Concept-based knowledge discovery in texts extracted from the Web*. *ACM SIGKDD Explorations* 2 (1), páginas 29-39, 2000.
- [13] S. Loh, et. al. *Investigação sobre a Identificação de Assuntos em Mensagens de Chat*. In: *TIL - XXIV Congresso da SBC*, Salvador, páginas 187-195, 2004.
- [14] MovieLens, MovieLens - *Movie Recommendations*, <http://movielens.umn.edu>, 20 de setembro de 2006.
- [15] I. Nonaka, T. Takeuchi. *The knowledge-creating company: how Japanese companies create the dynamics of innovation*, Oxford University Press, Cambridge, 1995.
- [16] N. F. Noy, D. L. McGuinness. *Ontology Development 101: a guide to creating your first ontology*, <http://protege.stanford.edu/publications15> de dezembro de 2003.
- [17] J. M. Pazzani. *A Framework for Collaborative, Content-Based and Demographic Filtering*. *Artificial Intelligence Review*, 13(5-6). páginas 393–408, 1999.
- [18] H. Ragas, C. H. A. Koster. *Four text classification algorithms compared on a Dutch corpus*, in: *Proc. SIGIR'98 International ACM-SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, Washington, páginas 369-370, 1998.
- [19] E. Reategui, A. Lorenzatti. *Um Assistente Virtual para Resolução de Dúvidas e Recomendação de Conteúdo*. In: *ENIA 2005, São Leopoldo, RS*. páginas 821-830, 2005.

- [20] L. C. Ribeiro Junior, et. al. Identificação de áreas de interesse a partir da extração de informações de currículos Lattes/XML. In: I Escola Regional de Banco de Dados. Porto Alegre: Ufrgs. páginas 67-72, 2005.
- [21] J. J. Rocchio. Document retrieval systems - optimization and evaluation, Ph.D. Thesis, Harvard Computation Laboratory, Harvard University, Report ISR-10 to National Science Foundation, 1996.
- [22] P. Resnick, N. Iacovou, M. Suchak, P. Bergtorm, P. J. Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In: Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, 1994, Chapel Hill, North Carolina. páginas 175–186, 1994.
- [23] G. Salton, J. M. McGill. Introduction to Modern Information Retrieval. McGraw Hill, New York, 1983.
- [24] U. Shardanand, P. Maes. Social Information Filtering: Algorithms for Automating “Word of Mouth”. In: ACM CHI’95 Conference on Human Factor in Computing Systems, Proceedings, (1), páginas 210–217, 1995.
- [25] W. D. Silva Filho, S. C. Cazella. STAR: Um Framework para recomendação de artigos científicos baseado na relevância da opinião dos usuários e em filtragem colaborativa. In: ENIA 2005. páginas 1042 – 1051, 2005.
- [26] J. F. Sowa. Building, sharing, and merging ontologies, AAAI Press / MIT press. páginas 3-41, 2002.
- [27] L. S. Vygotsky. A formação social da mente. Martins Fontes, São Paulo, 1984.
- [28] P. Willet. Recent trends in hierarchic document clustering: A critical review. Inf. Process. Manage., 24(5). páginas 577–597, 1988.