

Sistema de Recomendação de Documentos na Internet Utilizando Meta-Busca para o Moodle

Francisco Adelson Alves Ribeiro¹, Luis Carlos Costa Fonseca², Miguel de Sousa Freitas³

¹ Instituto Federal de Educação Ciência e Tecnologia do Maranhão (IFMA) – Codó, MA - Brasil

² Universidade Estadual do Maranhão (UEMA) – São Luis, MA – Brasil

³ Secretaria de Educação do Maranhão (SEDUC) – Caxias, MA – Brasil

adelton@ifma.edu.br, lccf@engcomp.uema.br,
qdmisofre@gmail.com

Abstract: *This research aims to improve the Learning Object Recommendation System developed for the Virtual Learning Environment (VLE) Moodle "Modular Object-Oriented Dynamic Learning Environment", which retrieves information just from the platform document base. After some tests of system usage, regarding educational needs and its architecture analysis, show the possibility of expanding the site search document repository for Web, integrating into the search engines of the Internet.*

Resumo: *A pesquisa tem como objetivo aprimorar o sistema de recomendação de Objeto de Aprendizagem – OA desenvolvido para o Ambiente Virtual de Aprendizagem – AVA Moodle “Modular Object-Oriented Dynamic Learning Environment”, o qual buscava informações na base de dados restrita da plataforma. O uso frequente do sistema, em função das necessidades pedagógicas e a análise de sua arquitetura, demonstrou a possibilidade de ampliação da busca do repositório local de documentos para a Web integrando-o ao motores de buscas da Internet.*

1. Introdução

Vivemos em um mundo no qual a sociedade atingiu o mais alto grau de evolução da técnica, sendo mais imperioso continuar evoluindo em função da tecnologia. Nessa perspectiva e corroborando com o pensamento de Freire [1996, p.25] “ensinar não é transferir conhecimento, mas criar as possibilidades para a produção ou a sua construção“. A informática na Educação insere-se nesse contexto, por ser uma ferramenta capaz de possibilitar ao processo ensino-aprendizagem, uma educação mais dinâmica e moderna.

Segundo, Flores [1996] a informática deve habilitar e dar oportunidade ao aluno de adquirir novos conhecimentos, facilitar o processo ensino-aprendizagem, enfim ser um complemento de conhecimento de conteúdos curriculares visando o desenvolvimento integral do indivíduo.

As pesquisas do IBOPE mostram uma tendência crescente do número total de pessoas com acesso a internet no Brasil de 105,1 milhões no segundo trimestre de 2013. O novo número é 3% maior que os 102,3 milhões registrados no trimestre anterior. Apesar do grande volume de dados disponíveis na internet e de sites especializados em busca, tais informações necessitam de uma curadoria para serem analisadas, organizadas e transformadas em conhecimento sistematizado.

O estudo de Cazella, Drumm, Barbosa [2010] aborda este problema, mencionando que o usuário não é capaz de selecionar as informações mais adequadas à sua necessidade, seja por desconhecimento do tema, ou pela sobrecarga de informação provocada pelas ferramentas de busca utilizadas (exemplo, google). Atendendo aos novos modelos educacionais em que novos paradigmas surgem como constantes desafios das necessidades dos professores e alunos, na objetivação e delimitação de consultas através de grupo com características específicas em plataformas de ensino. Tais necessidades vêm aos poucos sendo incorporadas através de plataformas educacionais que especificam resultados de buscas personalizadas.

Nesta perspectiva, os usuários dos Ambientes Virtuais de Aprendizagem – AVA, necessitam de uma ferramenta para coletar e sugerir conteúdos relevantes e relacionados ao tema em estudo. Autores como Burke [2007], O’Donovan, Smyth [2005] definem Sistema de Recomendação - SR como sendo ferramentas de software personalizadas que fornecem recomendações e sugestões de itens que possam ser de utilidade para o usuário. Tais recomendações e sugestões devem satisfazer as necessidades constantes dos novos modelos educacionais. Surge um novo cenário educacional no qual alunos e professores tem autonomia, confiança, inovação, sociabilidade, solidariedade e criatividade, possibilitando o uso desta ferramenta em qualquer lugar e a qualquer hora de forma personalizada.

Essa pesquisa resultou em dissertação de Mestrado, que rendeu amplos debates, discussões e contribuições para o II Congresso Brasileiro de Informática na Educação (CBIE 2013) através da publicação DOI: 10.5753/CBIE.SBIE.2013.82 no XXIV Simpósio Brasileiro de Informática na Educação (SBIE 2013). O uso frequente do sistema evidenciou a necessidade de ampliação da busca para Web integrando-o ao motor de busca do google. Durante a apresentação da pesquisa no SBIE 2013 foi grande o questionamento em relação a limitação de busca realizada pelo sistema de recomendação no repositório local de arquivos do AVA, tais discussões motivaram a produção deste artigo que contribuirá como aprimoramento do sistema de recomendação.

2. Arquitetura Inicial

Nesta arquitetura os usuários do AVA Moodle, atribuem hashtags às suas postagens nos fóruns, com a finalidade de criar um agrupamento textual em torno da marcação, possibilitando ao sistema classificar e recomendar objetos de aprendizagem aos usuários. As consultas são feitas no repositório local do Moodle, conforme figura a baixo:

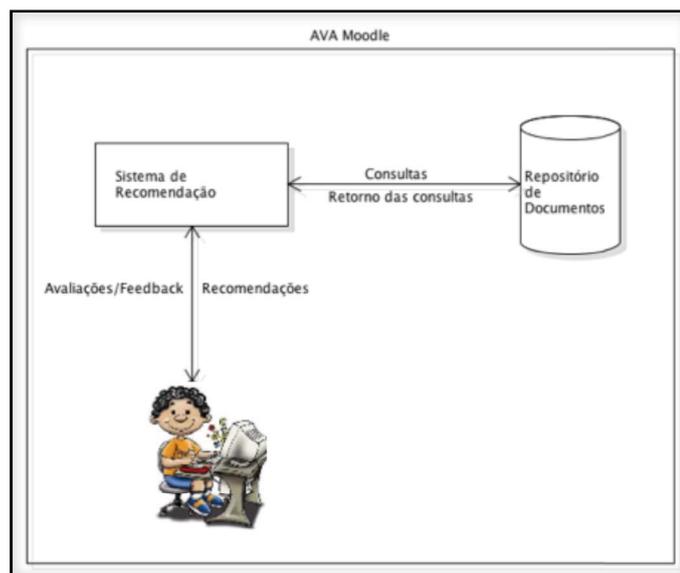


Figura 1 - Visão Geral do Sistema e Repositório Local Moodle.
Fonte: [RIBEIRO, 2013].

Nesta arquitetura o sistema apresentava pequenas restrições como: a necessidade de constantes atualizações da base de dados, realizada pelos professores, tutores ou administradores; necessidade de gerenciamento da base de dados e manutenção corretiva, preventiva e evolutiva do AVA, podendo ocasionar problemas relacionados com o desempenho computacional da plataforma.

A extensão da pesquisa, contribuiu para uma reflexão crítica, objetivando aprimorar o sistema, sendo necessário a criação de uma nova arquitetura que será descrita na figura 2.

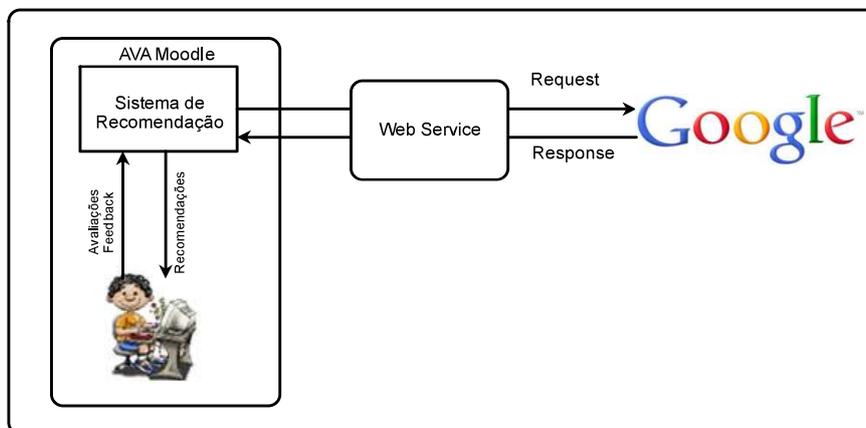


Figura 2 - Visão Geral do Sistema e Web Service Google.
Fonte: Elaborado pelos autores (2014).

Essa nova arquitetura permite que as pesquisas sejam feitas diretamente na Web, para tanto é necessária a integração do sistema desenvolvido para o AVA Moodle com motores de busca da Internet (como o Google), através do uso de Web Services que é uma tecnologia utilizada na integração de sistemas e na comunicação entre aplicações diferentes, permitindo a compatibilidade de sistemas desenvolvidos em diferentes plataformas. Para enviar e receber mensagens padronizadas os Web Services utilizam a linguagem XML¹, que permite a universalização das diferentes linguagens utilizadas por aplicações em um único formato.

3. Arquitetura conceitual de Web Services

A arquitetura padrão de Web Services faz uso das tecnologias universais (TCP/IP², HTTP³, Java⁴, XML), podendo ser descrita em diversas camadas, sendo que cada uma delas ficam responsável por diferentes operações. Kreger [2001] denomina “Pilha de Web Services” essas diversas camadas, conforme ilustra a figura 3.

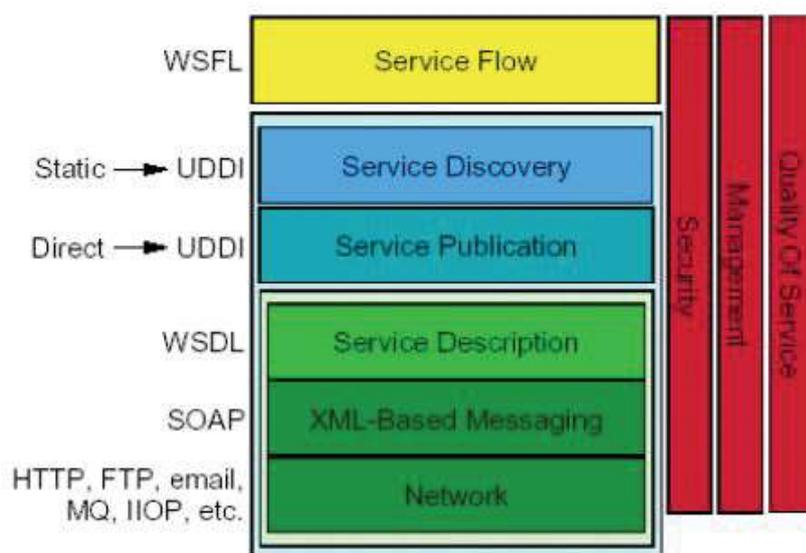


Figura 3 – Pilha de protocolos da arquitetura Web Services
Fonte: [KREGER, 2001].

A camada Network é a base da pilha responsável por permitir o acesso ao cliente através de seus protocolos. Na segunda camada, a linguagem XML é responsável pela troca de mensagens. Na terceira camada encontram-se o Service Description que é o padrão utilizado para interoperar o Web Services. As camadas Service Discovery e Service Flow são responsáveis pela publicação, descrição, colaboração, fluxo de dados e comunicação entre serviços, podendo ser implementadas de forma diferente para uma variedade de soluções.

¹ Extensible Markup Language - Linguagem de marcação recomendada pela W3C para a criação de documentos com dados organizados hierarquicamente.

² Transmission Control Protocol/Internet Protocol - Conjunto de protocolos de comunicação entre computadores em rede.

³ HyperText Transfer Protocol - protocolo utilizado para transferência de páginas HTML do computador para a Internet.

⁴ Linguagem de programação orientada a objeto.

A tecnologia de Web Services foi considerada apropriada para integrar o sistema de recomendação de documentos desenvolvido para o AVA Moodle ao sistema de busca do google pelas possibilidades de reutilização do código, disponibilidade de APIs⁵ para download, e interoperabilidade entre sistemas.

4. Integração do sistema com o Google

Devido constante atualização da base de dados do google, iniciou-se uma pesquisa que teve como objetivo integrar o sistema de recomendação desenvolvido para o AVA Moodle com seu motor de busca.

O Moodle foi desenvolvido pelo australiano Martin Dougiamas, objetivando ofertar cursos não presenciais auxiliados pela rede mundial de computadores. Em virtude da facilidade de manuseio e controle das ações desenvolvidas pelos seus usuários, o Moodle vem sendo utilizado por inúmeros países, sobretudo, os subdesenvolvidos. Segundo dados extraídos do site <https://moodle.org/stats/>, até abril de 2014, 240 países e 85,389 sites usavam a plataforma. Esses dados confirmam a preferência pelo uso do Moodle e sua constante atualização.

A plataforma foi escolhida como o ambiente hospedeiro para o desenvolvimento do sistema por ser utilizado em várias instituições de ensino e por ser um ambiente com código aberto (open source) disponível para a comunidade de programadores. Ademais, é baseado na Web e foi desenvolvido fundamentado em princípios pedagógicos, possibilitando a participação de uma grande comunidade de usuários que contribuem para a melhoria do ambiente, além de ser acessível em diversos idiomas.

Para a integração do sistema com o google, elaborou-se uma pesquisa que resultou lista de repositórios de documentos, nos quais foram levantados dados relativos a disponibilização de código para integração com outros sistemas e como esses repositórios realizam a busca por conteúdos armazenados em cada um deles. O google foi escolhido como o site padrão para o sistema, por ser um motor de busca “robô” especializado em vasculhar a internet em intervalos regulares visitando sites, lendo seu conteúdo, sugerindo links para páginas e os documentos encontrados pelos robôs são encaminhados aos indexadores (inclusão da palavra no índice), que extraem a informação e as armazenam em sua base de dados, excluindo a necessidade de atualizações manuais do repositório local da plataforma, além de fornecer aos desenvolvedores um conjunto de rotinas e padrões estabelecidos por um software para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do software, mas apenas usar seus serviços, chamado de API.

5. Funcionamento do sistema

Partindo-se do principio de que a finalidade básica das hashtags, seja classificar, a atribuição de hashtags às postagens de um usuário cria um agrupamento contextual em torno da marcação, ou seja, as postagens fornecem contexto às hashtags. Essa observação ganha relevância quando se visualiza cada hashtag como um agregado de

⁵ Application Programming Interface ou Interface de Programação de Aplicativos.

postagens, formando assim, um documento do usuário composto de recortes desconexos em torno de um mesmo assunto.

O problema desta abordagem é definir como as informações serão tratadas e comparadas com os documentos contidos em repositório. Para tanto, será criado para cada hashtag h_j um vetor de termos e pesos, contendo todos os termos das postagens de um determinado usuário que esteja marcado com h_j . Para esta formulação, utilizar-se á uma técnica de atribuição de pesos a documentos conhecida na área de RI que é o esquema a frequência de ocorrência do termo no documento (TF) e o inverso da frequência do termo entre documentos da coleção (IDF). Desta forma, para cada termo (palavra) das postagens associadas a uma hashtag h_j , teremos o seguinte esquema de cálculo dos pesos W_{ij} :

$$w_{i,j} = \begin{cases} (1 + \log f_{i,j}) \times \log \frac{N}{n_i} & \text{se } f_{i,j} > 0 \\ 0 & \text{se } f_{i,j} \leq 0 \end{cases}$$

Fonte: [SALTON & BUCKEY, 1998].

As variáveis da fórmula são detalhadas abaixo:

- W_{ij} é o peso de cada termo k_i associado à uma hashtag h_j ;
- f_{ij} é a frequência de ocorrência de um termo k_i nas postagens de h_j ;
- N é o número de documentos armazenados na base de dados do Moodle;
- n_i é o número de documentos com nomes relacionados as hashtag postadas nos fóruns de uma determinada sala de aula, disciplina ou curso (esta terminologia varia conforme a estrutura de cursos adotada);

Convém salientar que este estudo faz adequação do tradicional esquema de ranking denominado TF-IDF [Baeza-Yates e Ribeiro-Neto, 1999; Manning et al., 2009], à estrutura dos cursos do Moodle. No esquema original, a variável N é o número de documentos de uma coleção e n_i é a quantidade de documentos nos quais o termo k_i ocorre.

Desta forma, é possível comparar a estrutura de dados das postagens (hashtags) com uma coleção de documentos em repositório. A princípio, as hashtags de melhor ranqueamento, serão aferidas com os documentos existentes no repositório local do MOODLE, de modo a calcular suas similaridades. Para realizar o cálculo fazemos uso do Modelo Vetorial, conforme a fórmula abaixo:

$$\text{sim}(H_j, D) = \frac{\sum_{t=1}^t W_{i,j} \times W_{i,d}}{\sqrt{\sum_{t=1}^t W_{i,j}^2} \times \sqrt{\sum_{j=1}^t W_{i,d}^2}}$$

Nessa fórmula, H_j é o vetor de termos das postagens associadas à hashtag H_j , D é o vetor de termos do documento que está sendo comparado, $W_{i,j}$ são os pesos dos termos das postagens associadas à hashtag H_j e $W_{i,d}$ são os pesos dos termos do

DesafIE - III Workshop de Desafios da Computação Aplicada à Educação documento.

Para a fórmula acima se observa uma nova variável D , que representa os itens (documentos) que potencialmente podem ser recomendados a um usuário. Estes itens são ranqueados de acordo com o seu grau de similaridade com alguma hashtag H_j . Portanto, torna-se possível recomendar os mais variados documentos de um repositório, sejam planilhas eletrônicas, arquivos de textos, apresentações, arquivos PDF, etc, bastando apenas que eles contenham alguma informação textual que seja indexável. A figura 4, apresenta o diagrama e fluxo de atividade que o novo sistema realiza.

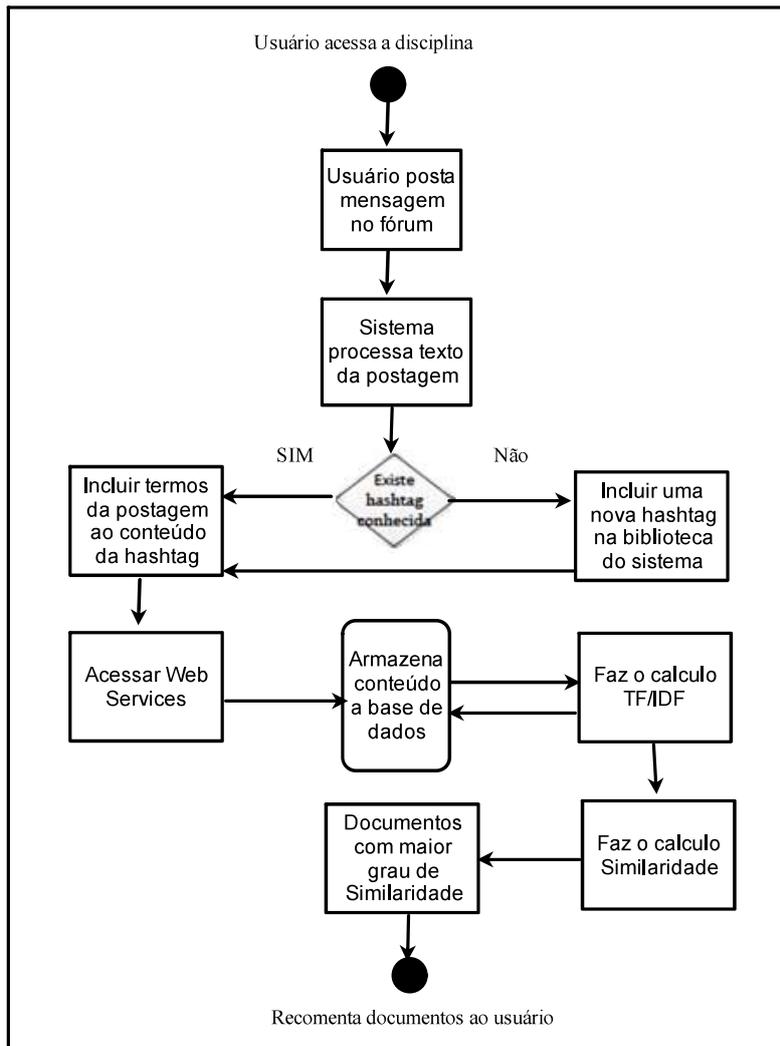


Figura 4 - Diagrama e Fluxo de Atividades.
Fonte: Elaborado pelos autores (2014).

O sistema proposto é baseado na técnica de recomendação de conteúdos, no qual o algoritmo recebe consultas textuais, para extrair todas as hashtags presentes nos fóruns do AVA Moodle de forma idêntica ao processo de indexação, retornando, então, a lista de documentos do Google que possuem nomes relacionados com as hashtags. Os documentos retornados são armazenados em uma base de dados, possibilitando ao sistema a execução do cálculo TF-IDF (Term Frequency – Inverse Document Frequency) utilizado para avaliar o quanto uma palavra “hashtag” é importante para um documento em relação a uma coleção “base de dados”, permitindo a operação de um algoritmo para classificação e ranqueamento de

conteúdos a partir dos termos mais utilizados nos fóruns. Finalizando o processo, o sistema realiza o cálculo de similaridade para determinar o quando as hashtags estão relacionadas com o conteúdo resultantes da busca realizada pelo google.

6. Resultados

Para realizar os primeiros testes com o sistema de recomendação, foi criado um curso no ambiente Moodle intitulado “Arquitetura de Computadores” para 25 alunos do segundo ano do curso Técnico de Informática do Instituto Federal de Educação, Ciência e Tecnologia do Maranhão IFMA – Câmpus Codó, ofertado na modalidade de distância, durante 30 dias, com dois encontros presenciais e carga horaria de 15 horas. Durante os encontros presenciais, alguns alunos relataram, que já haviam participado de cursos na modalidade a distância. Dos alunos inscritos apenas 15 frequentaram regularmente o curso, resultando numa evasão de aproximadamente 42%.

Ao finalizar o curso verificamos que o sistema gerou 46 recomendações, o que representa uma média pouco superior a três recomendações por aluno. O feedback das avaliações feitas pelos alunos revelou que 30 recomendações foram avaliadas, indicando que nem todas foram visualizadas, ou que algumas não despertaram o interesse dos alunos.

De acordo com as 30 avaliações realizadas pelos alunos, (com os conceitos bom, rum ou péssimo), foi analisado que mais de 65% foram consideradas boas, número satisfatório devido a exiguidade temporal da disponibilização do curso. Esta limitação não permitiu avaliar de forma considerável a evolução das recomendações em decorrência da limitação do tempo de uso. Porém, a fim de avaliar o funcionamento do sistema, relegando o aspecto pedagógico, as avaliações foram eficientes.

As recomendações ocorreram a contento, pois os resultados revelaram o bom funcionamento do sistema. A arquitetura utilizada mostrou-se eficiente, apresentando recomendações padronizadas em caráter diversificado, segundo interesse do usuário.

7. Conclusão

De acordo com os resultado obtidos, o sistema de recomendação de documentos para os usuário do AVA Moodle apresentou-se satisfatoriamente nos testes iniciais. A nova arquitetura, em geral, permitiu ao sistema buscar informações relevantes para os usuários e recomendar documentos relacionados ao tema de interesse. É importante citar também que agora o sistema utiliza-se de um mecanismo de meta-busca, pois acessa os motores de busca por meio de WebServices.

Um dos pontos mais desafiadores desta pesquisa, foi criar o algoritmo gerador das recomendações, adequando-o ao esquema de atribuição de pesos TF/IDF e ao cálculo de similaridade. O sistema faz uso da técnica de filtragem baseada em conteúdo, porém antes de executar essa tarefa, foi necessário analisar as diferentes abordagens de recomendações e criar a arquitetura geral do sistema. Outro desafio foi estudar o núcleo do AVA Moodle, para conhecer suas 285 tabelas e entender seus relacionamentos, permitindo criar novas tabelas e seus relacionamentos, sem alterar a performance da plataforma.

Durante o desenvolvimento e o período de teste do sistema, foi possível vislumbrar a ampliação de suas funções através de trabalhos futuros. Dentre estas aspirações, está o estudo de outros algoritmos para o cálculo de similaridade entre os documentos dos repositórios e as postagens, levando-se em consideração o custo

DesafIE - III Workshop de Desafios da Computação Aplicada à Educação computacional. Faz parte do escopo desta pesquisa também, a vinculação das hashtags ao usuário, tornando possível ao avaliador de um curso verificar especificamente, para cada aluno o que foi dito nos fóruns, configurando-se uma ferramenta interessante de avaliação. Entretanto, é importante ressaltar que a cultura de uso das hashtags deve estar devidamente difundida aos usuários da plataforma Moodle.

Referências

- Baeza-Yates, R. A.; Ribeiro-Neto, B. Modern information retrieval. Addison-Wesley Longman Publishing Co., Inc., 1999.
- Burke, R. Hybrid Web Recommender Systems. In Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.). *The Adaptive Web: Methods and Strategies of Web Personalization.*, Lecture Notes in Computer Science, Vol. 4321, pp. 377-408, Springer, Berlin-Heidelberg, 2007.
- Cazella, S. C.; Bhear, P.; Schneider, D.; Silva, K. K.; Freitas, R. (2012), “Desenvolvendo um Sistema de Recomendação de Objetos de Aprendizagem baseado em Competências para a Educação: relato de experiências”. In *Anais do XXIII Simpósio Brasileiro de Informática na Educação, SBIE*, Rio de Janeiro, RJ.
- Estatística Moodle. Disponível em: <http://MOODLE.org/stats>. Acesso em: 04/01/2014.
- Flores, A. M.; *A Informática na Educação: Uma Perspectiva Pedagógica – monografia- Universidade do Sul de Santa Catarina 1996 -* <http://www.hipernet.ufsc.br/foruns/aprender/docs/monogr.htm> (nov/2002).
- Freire, P. *Pedagogia da autonomia : saberes necessários à prática educativa*. 8. ed. São Paulo : Paz e Terra, 1996.
- IBOPE – Instituto Brasileiro de Opinião Pública e Estatística. Disponível em: http://www.appbrasil.org.br/ibope/free/internet_brasil_chega_105_milhoes.pdf. Acesso em: 24/03/2014.
- O'Donovan, J.; Smyth, B. Trust in recommender systems. In: *Proceedings of the international conference on intelligent user interfaces (IUI'05)*. 10, São Diego, Estados Unidos. Nova Iorque: ACM Press, 2005.
- Ribeiro, F.; Fonseca, L. & Freitas, M. Recomendando Objetos de Aprendizagem a partir das hashtags postadas no Moodle. In: *Anais Simpósio Brasileiro de Informática na Educação, SBIE*, Campinas, SP, 2013.
- Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. In: *Information processing & management*, vol. 24, pages 513-523, 1988.