

Uso de recursos educacionais para melhorar a eficiência das buscas na Internet

João C. Prates, Sean W. M. Siqueira

Departamento de Informática Aplicada
Universidade Federal do Estado do Rio de Janeiro (UNIRIO) – Rio de Janeiro, RJ -
Brasil

{joao.prates, sean}@uniriotec.br

***Resumo.** A internet é uma fonte incalculável de informações, que pode e deve ser utilizada para auxiliar o processo educacional. Uma das possibilidades do uso da Internet na educação é a busca por recursos educacionais que possam complementar o processo de aprendizagem, normalmente realizadas com o apoio de ferramentas de busca. Esta busca, em geral, é realizada a partir de palavras-chave, o que implica em resultados pouco contextualizados. Este trabalho utiliza técnicas de extração de informações aplicadas a recursos educacionais para expandir as consultas feitas pelos estudantes, introduzindo informações do contexto educacional na busca e com isso recuperando recursos educacionais mais adequados.*

1. Introdução

A expansão no uso da Internet e o crescente volume de informações disponíveis causam impacto e provocam mudanças em diversos setores da sociedade, dentre eles a educação. Como exemplos destas mudanças estão o uso da Internet pelo corpo docente para encontrar recursos educacionais para incorporar em suas aulas e a orientação aos estudantes para que pesquisem e utilizem recursos por conta própria [Ryan, Scott, Freeman, e Patel 2000]. Uma pesquisa realizada por Hargis (2001) com estudantes de nível superior mostrou que é possível aprender conceitos técnicos com a utilização da Internet, apoiando a sua utilização como uma ferramenta complementar para a educação. Ainda segundo Hargis (2001), não é mais suficiente ter o professor como principal fonte de informação em um mundo onde o conhecimento dobra a cada sete anos e em que 10.000 artigos científicos são publicados a cada ano.

Porém, a utilização correta deste ambiente tecnológico e das incontáveis fontes de informações disponíveis para apoiar o ensino e a aprendizagem envolve uma série de desafios a serem superados. Segundo Janette e Hannafin (2001), os recursos tecnológicos se mostram promissores para o ensino e aprendizagem, mas as práticas atuais de ensino podem se mostrar insuficientes para otimizar o uso dos recursos disponíveis e para a preparação de indivíduos para aprender em um ambiente rico em recursos. O aumento do acesso e da disponibilidade de recursos é necessário, porém não suficiente para promover um aprendizado efetivo.

Alguns desafios são apontados por Janette e Hannafin (2001), dentre os quais podemos destacar a necessidade de mecanismos flexíveis para a recuperação e uso de informações; a utilização de informações do contexto em que os recursos foram

produzidos ou serão utilizados; e a dificuldade em realizar a identificação, seleção e avaliação dos recursos adequados.

A recuperação de informações educacionais normalmente é realizada de duas formas: a partir de consulta a objetos de aprendizagem em sistemas gerenciadores de aprendizagem ou em repositórios de objetos de aprendizagem; e a partir de ferramentas de buscas disponíveis na Internet, tais como Google e Yahoo.

Um objeto de aprendizagem [IEEE LTSC 2002] deve ser projetado e desenvolvido com o preenchimento de metadados padronizados que auxiliem a classificação e recuperação destes recursos. A consulta aos objetos de aprendizagem é realizada com base nestes metadados cadastrados, permitindo a especificação do contexto de aprendizagem em que um determinado recurso será utilizado. Porém, a tarefa de preenchimento destes metadados acarreta um aumento de esforço para o autor (ou equipe para descrição dos objetos), dificultando a adoção de conjuntos de metadados, como o LOM [IEEE LTSC 2002], em ambientes como a web. Portanto, novos mecanismos para facilitar a recuperação de conteúdos adicionais são necessários.

Na recuperação de informações educacionais em ferramentas de busca disponíveis na Internet, o usuário informa a necessidade de informação digitando palavras-chave ou expressões de busca. Embora esta abordagem apresente como grande vantagem para os usuários a facilidade em expressar a necessidade de informação, ela possui alguns problemas, tais como: (i) o usuário precisa conhecer as palavras-chave que devem estar presentes no conteúdo dos recursos a serem recuperados; (ii) palavras-chave podem ter diferentes significados e a desambiguação destes conceitos deve ser feita pelo próprio usuário através de realização de novas consultas com a inclusão ou alteração de palavras-chave; e (iii) não são considerados na pesquisa aspectos pessoais do usuário e do ambiente em que esta consulta está sendo executada.

Estes problemas podem ser tratados através do contexto da consulta ou do usuário. Contexto é qualquer informação que possa ser utilizada para caracterizar a situação de uma entidade, sendo esta entidade uma pessoa, lugar ou objeto que possa ser considerada relevante para a interação entre o usuário e a aplicação, incluindo o próprio usuário e a aplicação [Dey e Abowd 2000]. Segundo [Shambaugh e Magliaro 1997], o contexto exerce um importante papel na implementação de qualquer evento de aprendizagem.

Outra questão importante a ser observada é o comportamento do usuário ao realizar buscas na Internet. Usuários acessam poucas páginas de resultado, frequentemente se limitando aos 5 primeiros resultados [Spink e Jansen 2004], mas as estatísticas mostram que as expressões de busca normalmente são compostas de poucos termos. 65% das buscas realizadas na Internet contêm de 1 a 3 termos [Experian 2010]. Para tornar os resultados de busca mais relevantes considerando este comportamento dos usuários ao elaborar uma expressão, é utilizada uma técnica de recuperação de informações conhecida como expansão de consultas, que consiste em adicionar termos na consulta original feita pelo usuário em uma tentativa de prover uma maior contextualização, recuperando documentos mais úteis [Yates e Neto 1999].

Considerando a grande utilização de ferramentas de busca para a recuperação de recursos para o ensino, e a utilização de poucos termos para a construção de expressões

de busca, e considerando também a necessidade de estudantes encontrarem recursos que possam complementar a aprendizagem em uma determinada aula, o objetivo deste trabalho é expandir as consultas realizadas pelos estudantes, incluindo termos extraídos de recursos educacionais utilizados nesta aula, adicionando à consulta original informações do contexto de domínio, e melhorando a eficiência em seus resultados, tratando assim alguns dos desafios apontados para a utilização da Internet na educação. Um cenário típico de aplicação deste trabalho é o seguinte: um professor de uma universidade seleciona recursos educacionais para suas aulas em uma determinada disciplina, e os cadastra em uma ferramenta de gerenciamento de cursos. Os alunos, ao acessarem a página de uma aula nesta ferramenta, podem realizar pesquisas na internet por materiais adicionais/complementares à aula, utilizando as expansões de consulta com termos do contexto desta aula. Ao acessar a página de outra aula, outro contexto é utilizado como fonte para as expansões de consulta.

A estrutura deste trabalho é a seguinte: na seção 2 será detalhado o uso de informações de contexto para a recuperação de informações; na seção 3 será explicado como as atividades de extração de informações podem ser utilizadas para a captura de contexto; na seção 4 detalharemos o experimento; e na seção 5 é feita a conclusão e apontados os trabalhos futuros.

2. Informação de contexto para recuperação de informações

O uso de informações do contexto para a recuperação de informações é alvo de diversas pesquisas. De forma geral, o contexto modelado é utilizado para alterar o processo normal de recuperação dos documentos indexados, alterando o conjunto de documentos retornados e também a sua ordem de relevância. A criação destes contextos pode ser realizada de forma manual ou automática.

Na criação manual de contextos, algumas pesquisas definem ontologias (ou utilizam ontologias pré-existentes) para representar o conhecimento em algum domínio [Pan, Wang e Gu 2004] e [Cesarano, Acierno e Picariello 2003]. Outros exemplos envolvem o preenchimento de formulário de preferências pelos usuários ou a atribuição de classificação de contexto para cada documento [Chanana, Ginige e Murugesan 2004]. A maior desvantagem no uso de métodos manuais para a criação de contexto é o custo de tempo e esforço envolvidos, de usuários ou de especialistas.

A criação automática de contexto pode ser realizada a partir do comportamento do usuário, analisando o histórico de consultas realizadas [Shen et al 2005], as informações recentemente visualizadas [Dumais et al 2003] ou o histórico de navegação do usuário na internet. O uso de métodos automáticos para a criação de contextos elimina o problema do custo envolvido na abordagem manual, porém pode apresentar resultados piores na aplicação deste contexto na recuperação de informações. Por exemplo, seja um contexto modelado com base nos últimos documentos visualizados por um usuário. O uso deste contexto para melhorar a recuperação de informações supõe que o usuário deseja pesquisar informações relacionadas com os documentos recentemente visualizados. Caso o usuário visualize documentos de assuntos totalmente divergentes, o contexto modelado tende a levar a resultados incorretos.

Outro problema importante tratado em diversos trabalhos é a evolução destes contextos modelados, que deve ocorrer quando o usuário possui novas necessidades de

informações diferentes das anteriores. Caso o contexto não acompanhe as mudanças, corre-se o risco de uma perda de eficiência na recuperação de informações.

No ambiente educacional, alguns trabalhos abordam o uso do contexto na recuperação de informações. Zhuhadar and Nasraoui (2008) capturam o contexto com a utilização de taxonomias de conceitos do domínio e do perfil do usuário, e reclassificam os resultados de pesquisa de acordo com a similaridade entre os termos presentes nos resultados da busca e nas taxonomias. A taxonomia de conceitos é definida manualmente enquanto a taxonomia que representa o perfil do usuário é gerada automaticamente com base na visitação do usuário a documentos.

Khribi et al (2009) recomenda recursos educacionais baseado em técnicas de mineração de uso da web. A partir da leitura de arquivos de *log* (registros de endereços visitados), sessões de usuários são agrupadas de acordo com similaridade e regras de associação são criadas entre estas sessões e elementos visitados. Recursos são recomendados a partir da identificação da sessão do usuário atual e dos elementos que estão associados a esta sessão.

Em [Baghi, Biletskiy e Li 2008] é apresentada uma abordagem para a recuperação de objetos de aprendizagem cuja classificação de relevância é calculada com base em uma combinação de três medidas de similaridade: entre palavras-chave, conceitos (extraídos com o suporte de uma ontologia) e o perfil do usuário (informações de preferência, experiência e interesses), obtidas através de preenchimento de formulário. Silva (2009) utiliza apresentações selecionadas por professores para, com base em técnicas de mineração de texto, recomendar documentos armazenados em um repositório e sugerir expansões de consultas a serem realizadas na web.

Este trabalho diferencia-se dos trabalhos anteriores nos seguintes aspectos: (i) a criação do contexto é feita de forma semi-automática, com a seleção de recursos sendo manual e a extração de termos por assuntos automática; (ii) a seleção do contexto a ser utilizado em cada busca é feita pelo usuário, simplesmente selecionando a aula desejada, evitando o problema da evolução dos contextos e (iii) os diferentes assuntos lidos em um contexto são considerados, sendo realizada uma expansão de consulta para cada assunto identificado.

3. Extração de informações em recursos educacionais para captura de contexto

A solução proposta neste trabalho utiliza atividades de extração de informações em recursos educacionais para capturar o contexto de uma aula e o utilizar para melhorar a eficiência nos resultados de buscas na web. A captura deste contexto é feita através da extração dos termos que aparecem com maior frequência dentre os assuntos identificados nos recursos educacionais disponibilizados para uma determinada aula e a aplicação deste contexto na recuperação de informações é feita através da expansão de consultas realizadas pelos usuários em um buscador web com o uso dos termos extraídos para cada assunto. Para detalhar melhor as atividades envolvidas nesta abordagem, podemos dividi-la em três módulos: Seleção de recursos, Modelagem do contexto e Recuperação de informações.



Figura 1. Módulos da solução proposta

3.1. Seleção de recursos

O contexto a ser utilizado na expansão de consultas é definido a partir de extração dos termos mais frequentes dos recursos educacionais selecionados pelo professor da disciplina. Logo, a atividade de seleção destes recursos é uma tarefa essencial para o correto funcionamento da aplicação. Caso não sejam bons “representantes” do contexto, este acabará sendo modelado com informações incorretas e as consultas expandidas acabarão não retornando os resultados desejados.

3.2. Modelagem de contexto

O objetivo da modelagem de contexto é identificar os principais termos de cada recurso educacional selecionado na etapa anterior e, a partir destes termos, realizar a expansão de consulta. Para isso, é calculado o peso de cada termo através da métrica TF X IDF, muito utilizada na área de recuperação de informações [Manning, Raghavan e Schütze 2008].

Porém, um documento pode tratar de mais de um assunto em sua extensão. A simples extração dos termos mais frequentes poderia combinar termos referentes a assuntos diferentes, o que acabaria comprometendo a qualidade das consultas expandidas a serem geradas nas etapas posteriores. A solução encontrada para este problema foi considerar a segmentação de textos e a *clusterização*.

A segmentação de texto é uma atividade de processamento de linguagem natural que objetiva identificar subtópicos dentro de um documento, delimitando os seus limites.

Cada subtópico identificado pode ser tratado em nosso experimento como um documento independente. Mas como vários documentos da coleção podem tratar do mesmo assunto ou de assuntos semelhantes, optou-se então por agrupá-los de alguma forma para a modelagem do contexto.

Para isto utilizamos a *clusterização*, técnica estatística que permite uma geração automática de agrupamento de dados (documentos). Nesta técnica, os vetores de termos de cada documento são agrupados de acordo com o cálculo de funções de similaridade e distância aplicadas a cada par de vetores.

3.3. Recuperação de informações

O subconjunto de termos obtido na execução da atividade de extração de informações, executada em cada agrupamento de assunto, é a entrada para a expansão de consultas. A expressão de busca digitada pelo usuário será expandida n vezes, sendo n o número total

de agrupamentos obtidos, representando todos os assuntos tratados na coleção de documentos selecionada. Estes termos obtidos também são expandidos com a utilização de um dicionário, sendo incluídos termos sinônimos em cláusulas OR na consulta.

Cada consulta expandida é executada em um motor de busca, podendo ser este um buscador web comercial (Google, Yahoo) ou alguma ferramenta interna. Seus resultados são exibidos em regiões distintas, uma para cada agrupamento de assunto.

4. Experimento

Para validar a proposta, foi implementada em Java uma aplicação Web para capturar e modelar um contexto representando o domínio de conhecimento de um curso. No módulo de seleção de recursos foi implementada uma integração com o sistema educacional Moodle. No sistema, o professor de um determinado curso pode selecionar, dentre os recursos disponibilizados para uma aula, quais devem ser utilizados para representar o seu contexto.

Para realizar a segmentação de texto, utilizamos o software MorphAdorner¹, que implementa o método de segmentação linear Texttiling [Hearst 1997]. Para o agrupamento de segmentos, utilizamos uma implementação do algoritmo de *clusterização* k-means [Manning, Raghavan e Schütze 2008] disponível na suíte de ferramentas de mineração de texto chamada Textgarden². O dicionário utilizado para a expansão de termos foi o Wordnet³, uma vez que os textos selecionados seriam em inglês.

Utilizamos este sistema para realizar um estudo de caso em uma disciplina de mestrado de uma universidade federal. Um professor selecionou, dentre os recursos disponíveis na plataforma Moodle para uma determinada aula, quais seriam boas fontes para a construção do contexto. Como nos artigos selecionados era possível distinguir dois grandes assuntos, a aplicação foi configurada para a geração de 2 agrupamentos para este contexto. A aplicação também foi configurada para executar as consultas realizadas no motor de busca Google e exibir seus resultados da seguinte forma: um painel exibindo os resultados da consulta original e outros dois painéis exibindo os resultados das consultas expandidas, um para cada agrupamento (cada agrupamento contribuindo com termos diferentes para a expansão).

Os estudantes matriculados na disciplina, que já tinham conhecimento sobre os assuntos tratados na aula selecionada, foram divididos em 2 grupos e utilizaram a ferramenta em duas situações: na primeira parte deveriam executar uma consulta previamente elaborada por um especialista. Na segunda parte teriam a liberdade de elaborar a sua própria consulta, com o objetivo de pesquisar recursos que pudessem complementar o aprendizado em algum assunto tratado na aula. Nas duas etapas, as consultas foram executadas duas vezes, com configurações diferentes para o número de termos utilizados nas expansões: na primeira vez com 4 termos e na segunda com 7.

A diferença entre os experimentos realizados pelos dois grupos foi a forma com que os termos utilizados para a expansão foram selecionados. No primeiro grupo, os termos eram selecionados automaticamente, pelo critério de relevância. Já no segundo,

¹ <http://morphadorner.northwestern.edu/>

² <http://textgarden.org/>

³ <http://wordnet.princeton.edu/>

os estudantes tinham acesso a uma lista de termos sugeridos e poderiam selecionar quais termos seriam utilizados na expansão.

Os resultados obtidos foram avaliados de acordo com métricas consolidadas por [Tang e Sun 2003] para a avaliação de buscadores web: Precisão total dos 10 primeiros resultados, profundidade da busca e correlação de *ranking*. A primeira medida é uma adaptação da precisão total aplicada na pesquisa original, aplicada aos 20 primeiros resultados. A alteração do número de resultados avaliados se justifica por estudos que mostram o comportamento ao realizar buscas na internet, em que a maioria acessa somente a primeira página de resultados e em geral se limitando aos 5 primeiros resultados [Spink e Jansen 2004] e [Jansen et al 1998]. Como a ferramenta de busca utilizada em nossa aplicação tem como configuração padrão a exibição de 10 resultados por página, foram avaliados somente este número de resultados.

Os estudantes julgaram cada resultado obtido, dentre os 10 primeiros, de acordo com sua relevância em uma escala de 5 pontos, usando “0” para indicar sem relevância e “4” indicando alta relevância. Estes julgamentos foram utilizados para gerar a medida de precisão total dos 10 primeiros resultados, calculada com a aplicação da seguinte fórmula:

$$\text{Precisão Total dos 10 primeiros resultados} = \frac{\sum_{i=1}^{10} \text{ju\textit{lgamento}}_i}{10 \times 4}$$

Onde $\text{ju\textit{lgamento}}_i$ corresponde ao julgamento de relevância atribuído ao i -ésimo resultado, 10 é o número de resultados considerados e 4 é o valor máximo de relevância atribuído para cada resultado.

A segunda métrica calculada foi a profundidade da busca, que reflete o número de documentos não relevantes que o usuário deve avaliar até que encontre x documentos relevantes. Logo, quanto menor o valor, menor é o esforço do usuário para encontrar resultados relevantes. No experimento, os estudantes avaliaram, para os resultados obtidos com a consulta original e com as consultas expandidas, a partir de que posição na lista de resultados foram obtidos 2 resultados consecutivos com relevância atribuída maior ou igual a 3.

A terceira métrica, correlação de *ranking*, foi realizada através da comparação entre a posição de cada resultado obtido com a atribuição de relevância feita pelos estudantes. Espera-se que os resultados obtidos sejam ordenados em ordem decrescente de relevância, e o objetivo desta métrica é verificar a correlação entre a priorização ideal e a obtida nas buscas. Dado que foram avaliados os 10 primeiros resultados, e a escala de avaliação é de 5 posições, a priorização ideal a ser considerada possui nas primeiras posições dois documentos com avaliação 4, dois com avaliação 3, e assim sucessivamente até os dois últimos resultados com avaliação 0. Para o cálculo da correlação foi utilizado o coeficiente de correlação de Pearson [Kendall e Stuart 1973].

4.1. Resultados obtidos e discussões

Os resultados obtidos na métrica de precisão total, detalhados na tabela 1, mostram uma melhora nas expansões com 4 termos nos dois assuntos, e na expansão com 7 termos em um dos assuntos. No grupo em que os termos foram selecionados automaticamente a

melhora só pode ser identificada em um dos assuntos, enquanto no grupo em que os termos foram selecionados manualmente a partir de uma lista de sugestões a melhora foi nos dois assuntos e nas duas configurações de números de termos.

Tabela 1. Precisão total dos 10 primeiros resultados

	Consulta original	1° assunto (4 termos)	2° assunto (4 termos)	1° assunto (7 termos)	2° assunto (7 termos)
Grupo 1	0,74	0,72	0,79	0,38	0,80
Grupo 2	0,46	0,73	0,62	0,65	0,63
Valor global	0,63	0,73	0,72	0,49	0,73

Na profundidade da busca (tabela 2), que representou o número de resultados não relevantes que o usuário teve que avaliar até encontrar dois resultados consecutivos com relevância maior ou igual a 3, a tendência de melhora se repetiu, porém com os resultados das expansões no grupo 1 sendo sempre melhores, e o grupo 2 apresentando o único resultado de aumento na profundidade na expansão realizada com 7 termos no primeiro assunto.

Tabela 2. Profundidade da busca

	Consulta original	1° assunto (4 termos)	2° assunto (4 termos)	1° assunto (7 termos)	2° assunto (7 termos)
Grupo 1	4,16	3,5	3,5	4	3,5
Grupo 2	4,75	2,25	2,5	6	3,5
Valor global	4,4	3	3,1	4,8	3,5

De forma geral, a métrica correlação de ranking demonstrou que as expansões, embora possam retornar resultados mais precisos e com menor profundidade de navegação, retornam resultados com ordenação de prioridade menos correlacionada com a priorização ideal. Somente em um assunto no grupo 1, com expansão de 7 termos, foi observada uma melhoria no coeficiente de correlação.

Tabela 3. Correlação de ranking

	Consulta original	1° assunto (4 termos)	2° assunto (4 termos)	1° assunto (7 termos)	2° assunto (7 termos)
Grupo 1	0,29	0,22	0,32	0,40	0,17
Grupo 2	0,17	0,18	-0,14	-0,22	-0,16
Valor global	0,24	0,20	0,13	0,15	0,03

4.2. Questionário

Após a execução do experimento, os estudantes responderam a um questionário contendo perguntas fechadas sobre o uso da ferramenta, com escala de 5 pontos, variando de “0” indicando discordância total até “4” indicando concordância total.

Todos os estudantes concordaram que os resultados de busca apresentados nos painéis dos dois assuntos representavam diferentes visões sobre o tema pesquisado. A maioria dos estudantes (80%) também concordaram que estas visões eram complementares. Também houve unanimidade em concordar que os resultados obtidos com a consulta original foram mais genéricos dos que os obtidos com as expansões.

A expansão de consulta trouxe como consequência a obtenção de mais resultados contendo referências acadêmicas, tais como livros, apresentações, artigos, etc. na avaliação de todos os participantes do experimento. A maioria deles (80%)

também considerou que o aumento no número de termos contribuiu com a obtenção destes resultados.

O aumento do número de termos teve influência direta na queda da qualidade dos resultados para 40% dos estudantes. 20% consideram que não houve piora, enquanto 40% se mostraram neutros.

5. Conclusões e trabalhos futuros

O uso de atividades de extração de informações pode ser considerado para aumentar a eficiência em recuperação de informações em um ambiente educacional. Os experimentos realizados indicaram que recursos educacionais utilizados em uma aula e selecionados por um professor podem ser utilizados, após a aplicação de técnicas de segmentação de texto, *clusterização* e extração de termos, como fonte para a expansão de consultas, trazendo para os estudantes resultados que indicam novos recursos educacionais que podem complementar o seu aprendizado.

Como trabalhos futuros, podemos destacar: Aplicação desta abordagem para modelar outros tipos de contextos, como por exemplo as preferências dos usuários ou o contexto didático/pedagógico; Aplicação do experimento em diferentes cenários de aprendizagem, tais como baseados em problema, projeto ou discussão, com colaboração entre os alunos; Aplicação da avaliação em diferentes situações, domínios e grupos de usuário; Aplicação em recursos de diferentes idiomas; Desenvolvimento de uma interface em que seja permitida uma melhor visualização dos resultados obtidos em cada agrupamento, facilitando a identificação dos resultados mais relevantes.

Agradecimentos

Os autores agradecem aos alunos que participaram do experimento, ao Departamento de Informática Aplicada/UNIRIO, bem como FAPERJ (através do auxílio E-26/101.509/2010 - BOLSA/BBP Representação e recuperação contextualizada de conteúdos de aprendizagem).

Referências

- Baghi, H. Biletskiy, Y. Li, H. (2008) "A Blended Approach for Search of Learning Objects" In Conference on Electrical and Computer Engineering
- Cesarano C., Acierno A., Picariello A. (2003) "An Intelligent Search Agent System for Semantic Information Retrieval on the Internet" In International workshop on Web information and data management
- Chanana V., Ginige A., Murugesan S. (2004) "Improving information retrieval effectiveness by assigning context to documents" In international symposium on Information and communication technologies
- Dey, A. K., Abowd, G. D. (2000) "Towards a better understanding of context and contextawareness" in Conference on Human Factors in Computing Systems
- Dumais S., Cutrell E., Cadiz J., Jancke G., Sarin R e Robbins D (2003) "Stuff I've Seen: A System for Personal Information Retrieval and Re-Use" In ACM SIGIR conference
- Experian Hitwise (2010) Searches statistics. Disponível em <http://www.hitwise.com/us/press-center/press-releases/google-searches-feb-10/>
- Hargis, J. (2001) "Can Students Learn Science Using the Internet?" in Journal of Research

- on Technology in Education, v33 n4.
- Hearst, M. A. (1997) "TextTiling: segmenting text into multi-paragraph subtopic passages", Computational Linguistics, Volume 23, P 33 - 64
- IEEE LTSC (2002) – IEEE – Learning Object Metadata 2002. Disponível em: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
- Janette R. e Hannafin, M. J. (2001) "Teaching and Learning in Digital Environments: The Resurgence of Resource-based Learning" in Educational Technology Research and Development, v49 n3 p37-52
- Jansen, B. J., Spink, A., Bateman, J., e Saracevic, T. (1998) "Real life information retrieval: A study of user queries on the Web" in *SIGIR Forum*, 32(1), 5-17.
- Kendall, M.G., Stuart, A. (1973) The Advanced Theory of Statistics, Volume 2: Inference and Relationship, Griffin
- Khribi, M. K., Jemni, M. e Nasraoui, O. (2009). "Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval". In IEEE International Conference on Advanced Learning Technologies
- Manning C., Raghavan P., Schütze H. (2008) Introduction to Information Retrieval, Cambridge University Press.
- Moodle, Sistema de Gerenciamento de Cursos, <http://moodle.org/>
- MorphAdorner, <http://morphadorner.northwestern.edu/>
- Pan X., Wang Z. e Gu X. (2004) "Context-Based Adaptive Personalized Web Search for Improving Information Retrieval Effectiveness" In IEEE Transactions on Knowledge and Data Engineering
- Ryan, S., Scott, B., Freeman, H. and Patel, D. (2000) "The Virtual University: The Internet and Resource-Based Learning. Kogan Page Open and Distance Learning Series.", Stylus Publishing Inc.
- Shambaugh, N. e Magliaro S. (1997) "Mastering the Possibilities: A Process Approach to Instructional Design Book Description, Allyn & Bacon
- Shen X., Tan B., Zhai C. X. (2005) "Context-sensitive information retrieval using implicit feedback". In ACM SIGIR conference
- Silva, L. O. (2009) "BOOKISH - Uma Ferramenta para Contextualização de Documentos Utilizando Mineração de Textos e Expansão de Consulta", Dissertação de Mestrado, Universidade Federal de Goiás
- Spink, A., Jansen, B.J. (2004) "A study of Web search trends". Webology, 1(2), Artigo 4, disponível em <http://www.webology.ir/2004/v1n2/a4.html>
- Tang M.C. e Sun Y. (2003) "Evaluation of Web-Based Search Engines Using User-Effort Measures", LIBRES, Volume 13, Issue 2
- Text-Garden, Text-Mining Software Tools, <http://kt.ijs.si/software/TextGarden/>
- Yates, R. B. e Neto, B. R. (1999) "Modern Information Retrieval", Addison Wesley, 1a ed.
- Zuhadar L., Nasraoui O. (2008) "Semantic Information Retrieval for Personalized -E-learning" In 20th IEEE International Conference on Tools with Artificial Intelligence