

Gerador Automático de Arquivos HTML de Ajuda para Aplicação em Educação a Distância (GAAHA)

PAULO SÉRGIO SALLA SÁ
PROF. DR. DILVAN DE ABREU MOREIRA

Universidade de São Paulo – USP
Instituto de Ciências Matemáticas e de Computação – ICMC
Departamento de Ciência de Computação e Estatística
Av. Dr. Carlos Botelho, 1465 – Caixa Postal –668
CEP: 13560-970 – São Carlos - SP
{paulo,dilvan}@icmc.sc.usp.br

Resumo. Este trabalho propõe o desenvolvimento de um software: o Gerador Automático de Arquivo Html de Ajuda (GAAHA) para a geração de arquivos de indexação e ajuda em HTML baseado na análise automática de documentos. Esses arquivos de indexação e ajuda têm uma dinâmica muito rápida de atualizações de seu conteúdo. Este trabalho focaliza principalmente: a manutenção de arquivos de ajuda para documentos de cursos para Educação a Distância (via WWW) e a criação de um formato de arquivos de ajuda (Help files) eficiente, para facilitar e agilizar a pesquisa de tópicos ou palavras nesses cursos. O aluno, que participa de um treinamento ou curso de educação a distância, poderá aproveitar o sistema de ajuda para pesquisa sobre tópicos relacionados ao curso ou treinamento em questão.

Abstract. This whork proposes the development of a computer program: the Automatic Generator of HTML Help Files (in Portuguese GAAHA) for the generation of index and help files in HTML based in the automatic analyses of documents. Index and help files have a very fast rate of changes in their contents. This work is main focus is in the maintenance of help files for distant learning courses documents and the creation of an efficient format for help files to ease topics and words searches in those courses. The student taking part in a distant learning course can take advantage of this help system to research about topics related to the course.

Keywords: Ensino a Distancia e Search Engines.

1. Introdução

A área de informática tem sido caracterizada como uma daquelas em que os profissionais mais têm buscado aprimoramento constante e dinâmico, como forma de resposta à competitividade e dinamismo do mercado. Em especial cada vez mais tem sido requerida uma visão abrangente da informática em geral e dos processos de gestão de recursos de informática em particular, tanto para profissionais acadêmicos como aqueles do mercado não acadêmico.

Com o avanço das tecnologias de comunicação e com a disseminação da Internet, cada vez mais se torna imprescindível a elaboração de projetos que possam se beneficiar deste potencial. Nesta linha de raciocínio, tem-se desenvolvido muitas aplicações na área comercial, de pesquisa, de entretenimento e de Educação a Distância (EAD), que se beneficiam do padrão *HyperText Mark-up Language* (HTML) como meio de documentação e treinamento. Este padrão oferece características atraentes, tais como a possibilidade de se escrever um material uma vez e usá-lo em qualquer plataforma. Ele também é mais fácil de trabalhar, possibilitando adicionar às páginas detalhes sutis, como aplicações Java e cliques de som.

Esses documentos, escritos em HTML, tem uma evolução de conteúdo muito dinâmica, *softwares* tem novas versões e cursos (de educação a distância ou treinamento) mudam seus conteúdos frequentemente. Em adição, os programas de EAD vem se intensificando e se tornando uma tendência marcante nos últimos tempos, inovando a forma de aprendizagem e possibilitando vários tipos de interação entre alunos e professores. Este tipo de aplicação se apresenta como uma das maneiras para se resolver o problema da concentração de potencialidades em alguns centros de excelência e da difusão dessas potencialidades a locais que não possuem tanta tecnologia e estejam geograficamente distantes.

A dinâmica rápida de mudança de conteúdo da documentação HTML, cria o problema de como manter esses documentos indexados corretamente e como manter arquivos de ajuda (*Help Files*) eficientes para os usuários deste material.

Visando resolver o problema exposto acima, focalizando principalmente o campo da educação e treinamento a distância, foi desenvolvido um *software*: o Gerador Automático de Arquivo Html de Ajuda (GAAHA) para a geração de arquivos de indexação e ajuda em HTML, oferecendo assim, uma maior rapidez, eficiência e elegância na procura e visualização de palavras e tópicos desejados.

Este sistema tem uma forte integração Web com banco de dados, já que todo o conteúdo de um determinado curso é indexado em um banco dados para posterior pesquisa via Web. Uma técnica de seleção de

palavras significantes é empregada no momento da indexação, com o objetivo de armazenar apenas palavras que realmente tenham um valor significativo.

2. Educação a Distância (EAD)

Serão abordados aqui alguns pontos referentes a EAD que julgamos importantes serem expostos, já que, a proposta principal deste projeto é voltada para a EAD.

2.1. Caracterização

A Educação a Distância é mais antiga do que parece, pois já contabiliza mais de um século de existência. Seus primórdios remontam ao ano de 1881 quando William Rainey Harper, primeiro reitor e fundador da Universidade de Chicago, ofereceu, com absoluto sucesso, um curso de Hebreu por correspondência. Em 1889 o Queen's College do Canadá deu início a uma série de cursos a distância, registrando sempre grande procura pelos mesmos motivos, principalmente, seu baixo custo e às grandes distâncias que separam os centros urbanos daquele país (Loyolla e Prates, 1999).

A EAD é caracterizada:

- Pela separação do professor e do aluno no espaço e/ou tempo (Keegan, 1986), (Perraton, 1988), (Mcisaac e Gunawardena, 1997).
- O controle da aprendizagem é realizado mais intensamente pelo aluno do que pelo professor instrutor (Jonassem, 1992).
- A interação entre alunos e professores é mediada por alguma forma de tecnologia (Keegan, 1986). (Garrison e Shale, 1990), (Kearsley, 1995).

“A Educação a Distância é uma aprendizagem planejada que normalmente ocorre em um local diferente do tradicional e como resultado requer projeto de curso e técnicas instrucionais especiais, métodos especiais de comunicação eletrônica ou outra tecnologia, bem como sistemas organizacionais e administrativos especiais” (Moore e. Kearsley, 1996).

As tecnologias utilizadas para fornecer informação na EAD, tem o objetivo de promover interação entre professor/aluno e fornecer o *feedback* para o aluno a distância (Mcisaac e Gunawardena, 1997).

O uso do ferramental pedagógico atualmente disponível pela EAD, permite o oferecimento de condições assíncronas de aprendizado, que podem, e devem ser combinadas com o ferramental do sistema convencional, permitindo uma combinação estreita de grande flexibilidade e alta eficiência no aprendizado final. Uma outra particularidade da EAD é que as modernas tecnologias, atualmente disponíveis, permitem o oferecimento de múltiplas combinações de ferramentas pedagógicas, modernas e tradicionais, com inegável e significativo melhoramento da relação custo-benefício de implantação e manutenção dos programas

de pós-graduação nestes moldes (Loyolla e Prates, 1999).

2.2. História da Educação a Distância

O primeiro curso por correspondência, nasceu na Inglaterra no final do século XIX sendo Sir Issac Pitman, da empresa *Correspondence Colleges*, o principal responsável por esta forma de transmissão de conhecimentos. Este primeiro curso por correspondência, tal como os que lhe seguiram, foi criado com o intuito de poder dar formação a um grupo de pessoas que, por motivos geográficos, de comodidade, económicos ou sociais, não podiam deslocar-se aos tradicionais locais de ensino (Matos, 1998).

Com o decorrer dos anos, a evolução tecnológica foi também pondo a descoberto a dificuldade que os técnicos, das mais diversas áreas laborais, tinham em se adequar às novas tecnologias adotadas e a educação a distância foi a solução mais rentável e eficaz para resolver esse grande problema (Matos, 1998).

O emprego da informática em um sistema de EAD pode viabilizar melhorias na qualidade de ensino em larga escala e a custos reduzidos. A interatividade pode ser incorporada neste processo através da hipermídia interativa, a qual integra textos, imagem, vídeos, fotos, som e animação, proporcionando um ambiente de ensino atraente, envolvente e multissensorial (Salvador, 1995).

Em finais dos anos 60 eis que surge a Internet, e tudo se torna ainda mais simples, pois as distâncias são substancialmente encurtadas, até que em 1993, com o *World Wide Web*, se atinge o ponto mais alto neste desenvolvimento. Este ambiente visual com amplas potencialidades, permite que se utilizem inúmeras tecnologias de apoio a educação a distância e o crescimento espetacular existente nos últimos anos é um reflexo disso.

2.3. Porque Ensinar a Distância

Segundo Matos (Matos, 1998), as principais características e vantagens da educação a distância podem resumir-se da seguinte forma:

- É abrangida uma maior diversidade de alunos, que de outra forma seria impossível.
- O estudante e o professor normalmente não se encontram no mesmo espaço físico, o que é vantajoso para casos onde seria muito dispendioso que isso acontecesse.
- O estudante não se desloca aos locais tradicionais de ensino, exceto nos casos em que seja necessário algum apoio de material existente em laboratórios ou em oficinas.

- Os horários praticados pelo aluno não são rígidos, muito pelo contrário, são bastante flexíveis permitindo o aproveitamento do tempo livre.
- O ritmo adotado pelo aluno é de sua exclusiva responsabilidade.
- Os temas de aprendizagem são mais vastos do que nos tradicionais locais de ensino.
- É possível receber contribuições de pessoas que, por razões de disponibilidade, não o poderiam fazer em um sistema tradicional.

Todas estas vantagens podem ainda ser acrescidas se houver uma “humanização” do espaço ou meio onde são lecionados os cursos, tornando-os o mais interativos possível e aperfeiçoando-os com constantes remodelações e atualizações de forma a que o ambiente seja dinâmico e não um espaço onde simplesmente é publicada informação. Deve também haver o cuidado de expor claramente os conteúdos para não suscitem dúvidas e, no caso de estas existirem, devem ser prontamente retiradas e esclarecidas.

2.4. Educação a Distância na Internet

Em primeiro lugar, vale realçar que o aparecimento de novas tecnologias para a Internet (especialmente para WWW), assim como o aumento constante da largura de banda disponível, permitem utilizar novas ferramentas mais poderosas, que não existiam há alguns anos atrás. Com este panorama favorável, a Internet vai certamente tornar-se no principal veículo utilizado para a educação a distância.

As vantagens inerentes a este ambiente cada vez mais familiar são muitas, sendo as mais importantes:

- Ensino centralizado no aluno.

O ensino é centralizado no aluno e não no professor, o que fomenta a colaboração entre os estudantes e proporciona um método menos rígido de aprendizagem, tornando-a mais interativa e mais interessante.

- Conveniência.

Os horários de trabalho e o local de ensino são bastante flexíveis permitindo o desenvolvimento de atividades paralelas por parte do aluno, como por exemplo, exercer uma profissão e, nas horas vagas, fazer o seu curso sem ter problemas de incompatibilidades.

- Ferramentas muito fáceis de utilizar.

As ferramentas (computador, clientes dos diversos serviços Internet utilizados, etc.), normalmente utilizadas para o apoio do aluno, são extremamente simples de utilizar e permitem uma rápida ambientação às mesmas.

- Desenvolvimento dos ambientes virtuais rápido e fácil.

As empresas e universidades que disponibilizam cursos *on-line* também têm a sua tarefa facilitada devido a existência de ferramentas de composição de criação das páginas *Web*. Hoje em dia, a construção de um *site* na Internet, ainda que com as devidas exceções, demora muito pouco tempo a projetar e ainda menos a concretizar.

- Aproveitamento dos recursos já existentes.

A grande quantidade de recursos atualmente disponível na Internet permite que o estudo de um determinado tema se possa reduzir a uma série de *links* para material já existentes, o que poupa imenso tempo. A facilidade de encontrar os mesmos, também aumenta de dia para dia com a implementação de poderosos motores de procura gratuitos, destinados a todos os usuários em nível mundial.

- Fácil alteração dos conteúdos temáticos.

O formato eletrônico dos documentos permite alterações simples e sem grandes custos de tempo, o que é uma grande vantagem quando há falta de recursos humanos disponíveis para o fazerem.

- Informação e tecnologias em formato padrão.

Neste ambiente virtual, tudo o que é publicado e todas as tecnologias utilizadas são neste momento acessíveis em qualquer parte do globo e num formato uniformizado, para que não haja incompatibilidades impositivas do bom funcionamento deste sistema.

Com todas estas vantagens, percebe-se que o futuro da educação a distância, passa pela Internet. A maior tendência visível é a de um crescimento exponencial para os anos que se avizinham.

3. Agente Automático de Análise de Texto.

A área de Recuperação de Informações (IR -*Information Retrieval*) procura coletar documentos e artigos com texto relevante (Cowie e Lehnert, 1996). As ferramentas de Recuperação de Informações, geralmente, trabalham com técnicas de indexação, capazes de indicar e acessar mais rapidamente documentos de um Banco de Dados textual (Yates, 1996).

Existem três tipos de indexação (Yates, 1996):

- Indexação tradicional: é aquela onde uma pessoa determina os termos descritivos ou caracterizadores dos documentos, os quais farão parte do índice de busca.

- indexação *full-text* (ou indexação do texto todo): onde todos os termos que compõem o documento fazem parte do índice.

- Indexação por *tags* (por partes do texto): onde apenas algumas partes do texto são escolhidas automaticamente, para gerar as entradas no índice (somente aquelas consideradas mais importantes ou mais caracterizadas).

No desenvolvimento deste sistema (GAAHA), a técnica utilizada como base para a seleção de palavras significativas de um texto é a técnica baseada na frequência de ocorrências de palavras. A ênfase neste método é mais da área estatística do que de uma aproximação lingüística para análise de texto automatizado (Luhn, 1958), (Chen et al., 1996). As razões para isto são várias. A primeira é que a análise lingüística é muito dispendiosa para ser implementada (Rijsbergen, 1970). Parte do problema é devido ao pouco progresso que tem sido realizado na teoria da semântica formal. Sem dúvida alguma, a teoria da linguagem será de extrema importância para o desenvolvimento de sistemas IR, mas atualmente, tais teorias não tem sido suficientemente desenvolvidas para serem aplicadas eficazmente em sistemas IR. A segunda razão é que a aproximação estatística tem sido examinada e testada, com resultados satisfatórios (Montgomery, 1972).

Luhn (Luhn, 1958) em um dos seus primeiros artigos publicados, propõe que a frequência da ocorrência de uma palavra em um texto forneça uma medida útil para a significância da palavra. Isto favorece que a posição relativa das palavras dentro de uma sentença tenha valores significativos, fornecendo assim, uma medida útil para determinar sentenças significativas. O fator significativo de uma sentença, consequentemente, será baseado na combinação destas duas medidas. Luhn assume que a frequência do dado pode ser usada para extrair palavras e sentenças para representar um documento.

Seja f a frequência de ocorrência de várias palavras em um texto e r sua ordem de classificação (*rank order*), isto é, a ordem de sua frequência de ocorrência como demonstrado na tabela 1. O resultado da interseção do relacionamento f e r , gera uma curva similar a uma hiperbólica, como exemplificado na figura 1. Esta curva demonstra a lei de Zipf (Zipf, 1949), ou seja, o produto da frequência do uso das palavras com a classificação das classes (*rank order*) é aproximadamente uma constante. Zipf testou sua lei analisando as palavras de um jornal (*American Newspaper English*).

Palavra do texto	Frequência que ocorrem (f)	Rank Order (r)	(Zipf) Peso da Palavra
E	73	1	73
No	56	2	112
Matemática	13	3	33
Scanner	5	4	20
Oráculo	1	5	5

Tabela 1 - Representação das palavras de um texto.

Luhn usou esta lei como uma hipótese nula, definindo assim os limites de *cut-offs*, um superior (*upper*) e um inferior (*lower*), excluindo assim as palavras não significantes, ou seja, as palavras que estão fora destes limites (Montgomery, 1972). As palavras que excedem os limites impostos pelo *upper cut-off* são consideradas comuns e aquelas abaixo do limites do *lower cut-off* são consideradas raras, e portanto, não contribuem significativamente no conteúdo de um texto. Com isso, Luhn chegou a uma técnica capaz de encontrar palavras

significativas em um texto, e também assumiu que o poder de resolução das palavras significantes, no que diz respeito a habilidade de distinguir conteúdos de palavras, atinge um pico no meio do caminho entre os dois extremos do *cut-off*, caindo em direção aos extremos. Uma certa arbitrariedade é envolvida na determinação do *cut-off*. Não existe um oráculo que dê os valores do *cut-off*, o estabelecimento deles se dá via um julgamento de tentativas e erros.

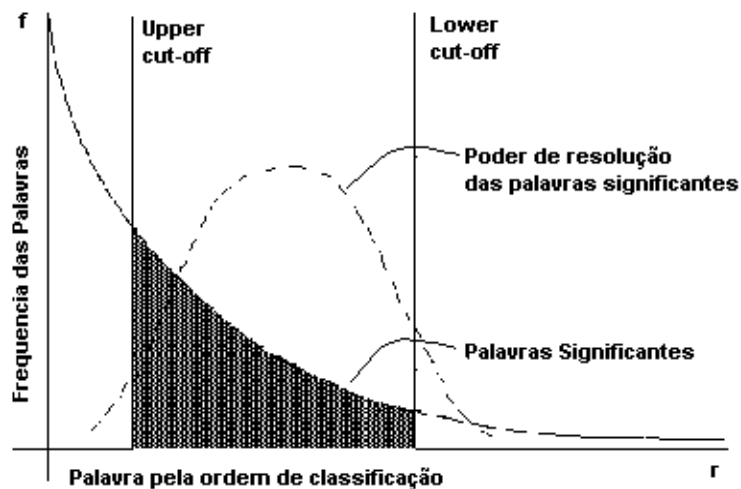


Figura 1 - Gráfico relacionando a frequência de ocorrência f e a sequência de posição r .

As Figuras 2 e 3 representam os resultados da aplicação da teoria de Luhn e Zipf em um texto

técnico em Inglês e em uma obra da literatura Brasileira.

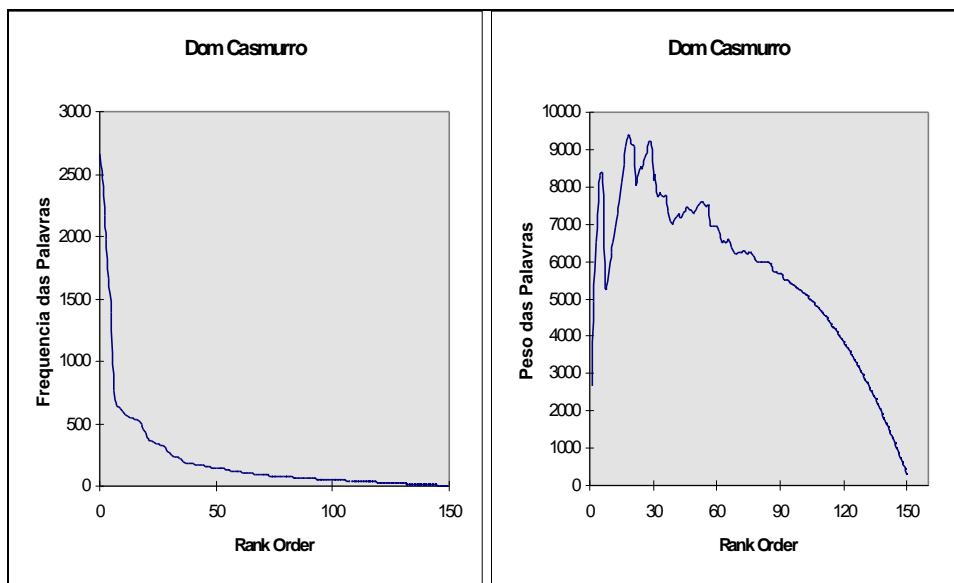


Figura 2 - Gráfico da Obra de Machado de Assis - Dom Casmurro.

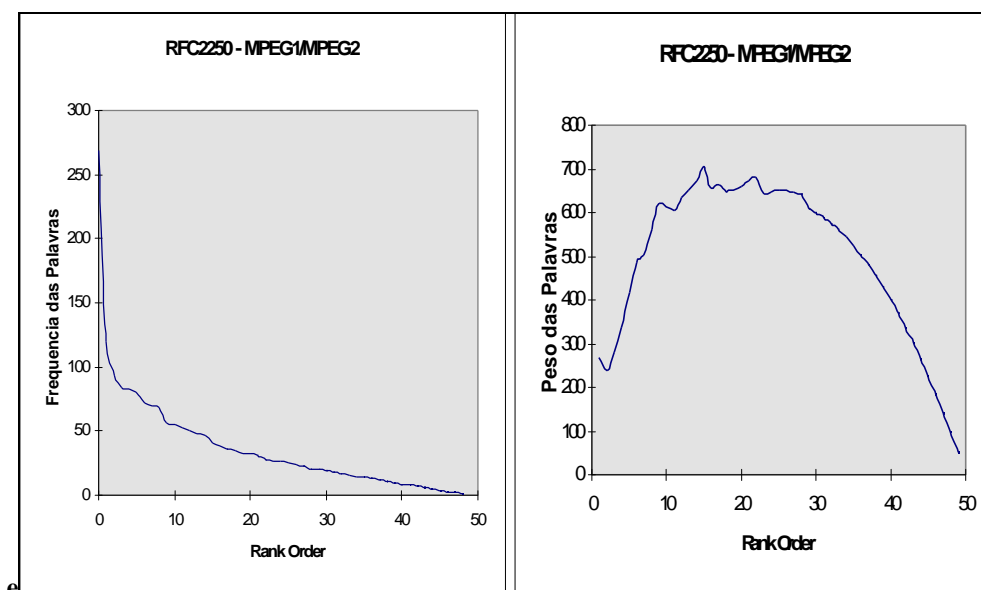


Figura 3 - Gráfico da RFC 2250 MPEG1/MPGE2.

Rijsbergen (Rijsbergen, 1970) resumiu a proposta de Luhn sobre a análise automática de um texto, dizendo que a frequência dos dados pode ser usada para extrair palavras e sentenças que representam um documento.

Não existe nenhuma razão para que tais análises fiquem restritas somente a palavras. Elas podem ser igualmente aplicadas para frases e de fato isto tem ocorrido (Rijsbergen, 1970).

4. O sistema GAAHA

Existe uma grande frustração dos usuários com o problema da "sobrecarga de informações" disponíveis para consulta hoje na Internet. Ela ocorre quando o usuário tem muita informação ao seu alcance, mas não tem condições de tratá-la ou de encontrar o que realmente deseja ou lhe interessa.

Visando a solução para o problema descrito acima, foi desenvolvido um sistema de ajuda chamado de GAAHA. O GAAHA é uma solução para a análise do conteúdo de um curso residente em um *Web Site* e posterior criação de um mapa gráfico preciso, intuitivo e interativo de posicionamento em tópicos do curso. O GAAHA tem um robô *parser* responsável pela geração de uma base de dados indexada que é utilizada para pesquisas de palavras-

chave entradas pelo usuário e também para geração de uma vista hierárquica de todos os *links* de um curso, aqui chamado de mapa do curso.

Este sistema é composto por três subsistemas: um agente *crawler* para a indexação das informações, figura 4; um agente *mentor* que é uma rotina de pesquisa executada no servidor, responsável pela pesquisa no banco de dados, figura 5; e um cliente WWW para a entrada dos dados de pesquisa.

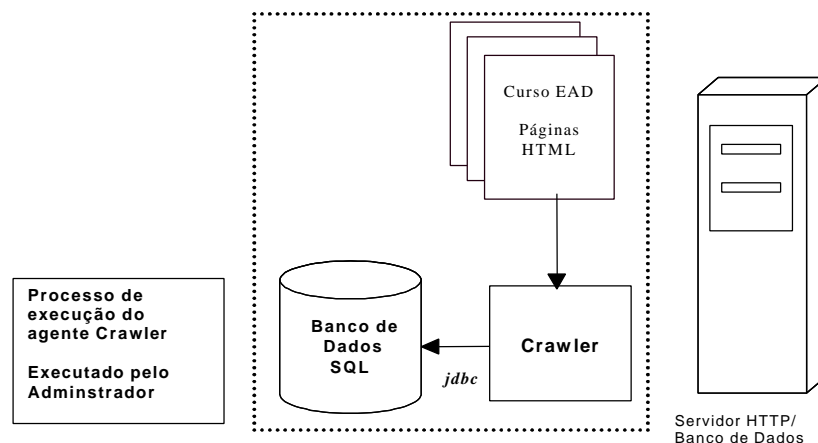


Figura 4 – Estrutura do agente crawler

O agente *crawler* e o agente *mentor* são os dois módulos principais desenvolvidos neste projeto. O cliente consiste de um *Web Browser* (tal como o Netscape ou Explorer) e um documento HTML criado dinamicamente pelo agente *mentor*.

Os módulos *crawler* e *mentor* são caracterizados como agentes devido a algumas características próprias de funcionamento que os integram, tais como, a manifestação de características de inteligência distribuída, por serem capazes de desempenhar o papel de mediador entre pessoas e

programas, e também por desempenharem o papel de um "assistente inteligente".

O agente *crawler* se caracteriza principalmente pela autonomia do seu funcionamento, ou seja, através de um pedido de indexação de um curso de EAD, todos os documentos que compõem este curso são indexados sem qualquer interação externa do administrador. Quanto ao agente *mentor*, sua característica é mais voltada para um mordomo que fica a disposição do usuário para responder as dúvidas solicitadas.

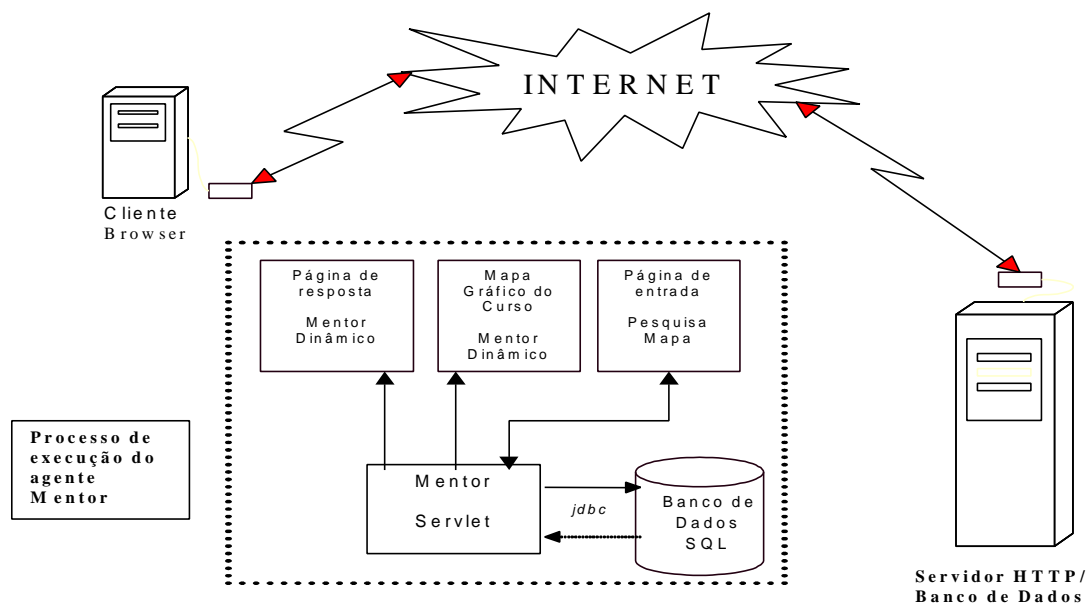


Figura 5 – Estrutura do agente mentor

O GAAHA usa duas técnicas principais:

- Análise de texto usando ferramenta estatística.
- Programação distribuída usando agentes de software.

Basicamente o programa gera um banco de dados relacional com as palavras mais significativas, os *links* de referências, os *headers* e os títulos do texto. Cada informação útil gerada, é relacionada com os locais onde elas aparecem. Obviamente o texto analisado deverá estar formatado no padrão HTML.

O tipo de indexação aplicada neste projeto para a análise da estrutura de um texto, é a indexação automática conhecida como *full-text* (Yates, 1996). Como descrito no item 3, este tipo de indexação, indexa automaticamente todas as palavras do documento. O grande problema deste tipo de indexação é o tamanho dos índices gerados, pois, como todas as palavras são armazenadas, os índices ficam volumosos e com muitas palavras insignificantes.

Para diminuir o problema do tamanho dos índices e também para armazenar apenas palavras significativas, um filtro foi implementado. Para a implementação deste filtro, foi utilizada uma técnica estatística de indexação automática, baseando-se na frequência de ocorrência de palavras nos documentos (Luhn, 1958), (Chen et al., 1996), ou seja, a palavra só é armazenada no banco de dados se estiver dentro de uma faixa de limite mínimo (*lower cut-off*) ou

limite máximo (*upper cut-off*) de frequência de ocorrência.

A escolha do tipo de indexação (*full text*) e do filtro (por frequência) foi pelo fato de que os mesmos não exigem qualquer tipo de formatação especial ou *tags* de reconhecimento em um texto, facilitando e até incentivando a utilização deste sistema (GAAHA).

Quando existem exigências quanto a inserções de *tags* ou formatações especiais, o desenvolvedor tem que conhecê-las para então aplicá-las no curso a ser desenvolvido. Como atualmente o desenvolvedor não tem que se preocupar com os comandos HTML para o desenvolvimento de um curso, já que, existem diversas ferramentas que facilitam o desenvolvimento para este padrão, seria muito inconveniente e até mesmo desestimulante a adoção de uma técnica que exigisse o uso de *tags* ou formatações especiais nos textos que compõem um curso de EAD.

Outro grande motivo da adoção destas técnicas é a possibilidade imediata de indexação de cursos já existentes, utilizando este sistema (GAAHA).

Depois do agente *crawler* gerar um banco de dados, aplicando os mecanismos de indexação e filtragem das palavras correspondente a um determinado curso, como mostrado na figura 6, é possível via uma URL, executar um servlet Java passando como parâmetro o nome do curso a ser pesquisado.

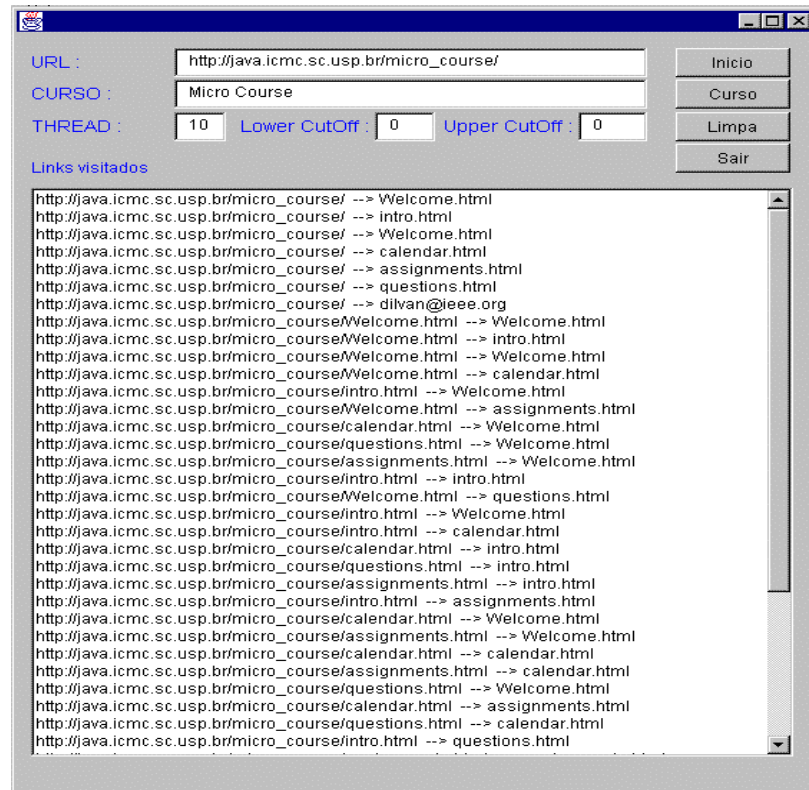


Figura 6 – Execução do agente crawler

O resultado desta execução, é a construção dinâmica pelo agente *mentor* de uma página HTML de ajuda, correspondente ao curso em questão, como mostrado na figura 7. Esta página contém basicamente um botão que possibilita a construção dinamicamente

do mapa do curso, um campo para digitação das palavras-chave de pesquisa ao banco de dados e um *link* para uma página informativa contendo informações de utilização da máquina de pesquisa.



Figura 7 – Execução do agente mentor.

O mapa do curso é uma estrutura em árvore que representa a hierarquia de todos os documentos e seus respectivos *links* contidos, possibilitando ao usuário uma visão mais ampla de todo o curso e um rápido posicionamento do tópico desejado, como mostrado na figura 8.

Após a interpretação das palavras-chave, o agente *mentor* constrói uma página HTML com os resultados da pesquisa, retornando-a ao usuário.

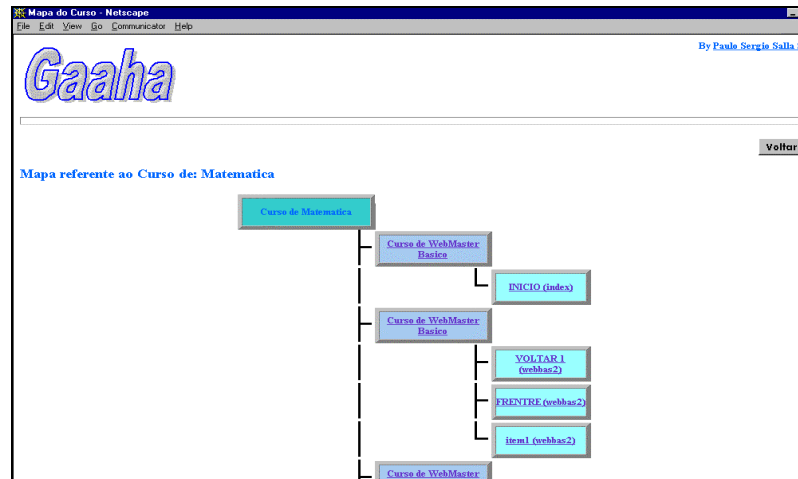


Figura 8 – Mapa gráfico do curso.

Quanto ao agente *mentor*, cuja responsabilidade é a interpretação das palavras-chave entradas pelo usuário e geração das páginas dinâmicas de resposta, é utilizado um Servlet Java para pesquisa no banco de dados. Também o agente *mentor* é responsável

pela construção dinamicamente de um mapa gráfico hierárquico do curso .

A figura 9 mostra o fluxo de dados na máquina de pesquisa quando solicitado pelo cliente (aluno).

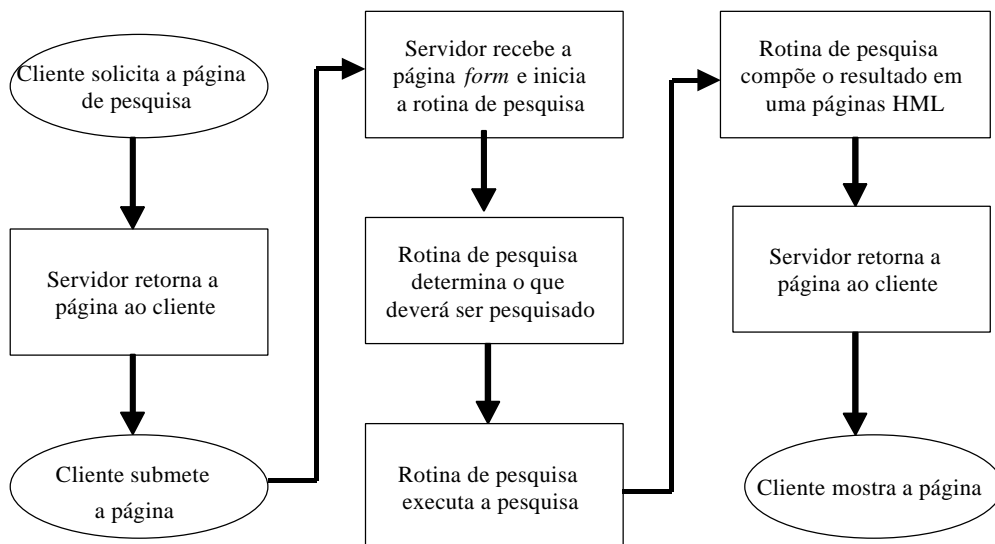


Figura 9 – Fluxo de dados no sistema de pesquisa.

5. Conclusão

O resultado da execução do sistema GAAHA é realmente a geração de uma base de dados indexada contendo as palavras que compõem um curso, sendo que, a totalidade de armazenamento das palavras só é alcançada quando os valores de *lower cut-off* e *upper cut-off* são iguais a 100 %. Quanto maior for o valor de *lower cut-off* e *upper cut-off*, maior a chance de ser armazenada palavra sem muita significancia, segundo a teoria de Luhn (Luhn, 1958). Caso estes valores não sejam bem definidos, corre-se o risco de, ou armazenar palavras insignificantes, ou não armazenar palavras significantes.

A conseqüência do não armazenamento de todas as palavras que compõem um curso é a construção de um banco de dados mais enxuto e eficiente. Portanto, a construção dinâmica do arquivo de ajuda contendo as respostas a uma consulta feita por um usuário, torna-se mais “limpa” e rápida.

A execução do sistema GAAHA, também tem como resultado a geração dinâmica de um mecanismo de posicionamento gráfico hierárquico em árvore, contento todos os *links* de referencias de um curso de EAD.

Apesar de existir uma semelhança deste programa com os *searchs engines* disponíveis na Internet, como o Alta Vista (<http://www.altavista.com>), o GAAHA tem uma interface com o usuário mais trabalhada. Além disso, esse programa é capaz de usar técnicas e heurísticas muito mais particulares ao problema em questão (geração de arquivos de ajuda), pois o seu universo de aplicação é muito mais restrito que o dos *searchs engines*, tanto quanto ao seu escopo de aplicação, quanto a quantidade de documentos a serem pesquisados.

Os resultados desse trabalho demonstram que os objetivos do projeto foram alcançados e que o GAAHA é uma aplicação inovadora que deverá se mostrar muito útil para professores e alunos de cursos de EAD.

6. Referências.

- H. Chen et all. A concept space approach to addressing the vocabulary problem in scientific information retrieval: na experiment on the worm community system. MIS Department, University of Arizona, 2 de Julho de 1996. (<http://ai.bpa.arizona.edu/papers/>, 1998).
- J. Cowie., W. Lehnert. Information extraction. *Communications of the ACM*, v.39, n.1, p. 80-91, jan 1996.
- D. R. Garrison, D. Shale. An analysis and evaluation of audio teleconferencing to facilitate education at a distance. *International Journal of Distance Education*, v.1, n.1, p. 7-13, 1990.
- D. H. Jonassen. Applications and limitations of hypertext technology for distance learning. In: *Distance Learning Workshop*. San Antonio, Texas: Armstrong Laboratory, 1992.
- G. Kearsley. *The nature and value of interaction in distance learning*. In: DISTANCE EDUCATION RESEARCH SYMPOSIUM, 3., Washington: George Washington University, May p.18-21, 1995.
- D. Keegan. *The foundations of distance education*. London: Croom Helm, 1986.
- W. P. D. C. Loyolla, M. Prates. *Educação à Distância Mediada por Computador (EAD): Uma Proposta Pedagógica*. (<http://www.puccamp.br/~prates/edmc.html>, 1999).
- H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, v.2, p.159-165, 1958.
- H. A. B. Matos. *Sistemas de Formação*. Universidade de Coimbra. Coimbra, Portugal. (<http://student.dei.uc.pt/~kikas/DLIndex.html>, 1998).
- M. S. Mcisaac, C. N. Gunawardena. Distance education. In: JONASSEN, D. (ed.) *Handbook for Research on Educational Communications and Technology*. New York: MacMillan, 1997.
- C. A. Montgomery. Linguistics and Information Science. *Journal of the American Society for Information Science*, v.23, p.195-219, 1972.
- M. G. Moore, G. Kearsley. *Distance education: a systems view*. Belmont, California: Wadsworth, 1996.
- H. Perraton. A theory for distance education. In: SEWART, D., KEEGAN, D., HOLMBERG, B. (ed.) *Distance education: International perspectives*. New York: Routledge, 1988. p.34-35.

B. S. Rijsbergen. *Information Retrieval*. London: Butterworths, 1970.

V. L. G. Salvador,. Hipermídia interativa – a educação do futuro, no presente. *Tecnologia Educacional*. v.22, n.123/124, p.22-23, 1995.

R. B. Yates. An extended model for full text databases. *Journal of the Brazilian Computer Society*, v.2, n.3, abril, 1996.

H. P. Zipf. *Human Behavior and the Principle of Least Effort*. Cambridge, Massachusetts: Addison-Wesley, 1949.