

Recuperação de Informação - Máquinas de Busca na Web

DCC – UFMG

1º. Sem/07

Plano de Curso

Professores: Berthier Ribeiro-Neto e Nivio Ziviani, Depto de Ciência da Computação, UFMG

Dados: 1º. Semestre de 2006, 07:30-09:10 horas, 2ª. e 4ª. Feiras

1. Objetivos

A Web conta hoje com centenas de milhões de usuários em todo o mundo, tendo se tornado num novo meio de comunicação de massa. Como resultado, as empresas e corporações têm intensificado o uso da Web como veículo de divulgação de seus produtos e serviços. Isso é feito através da *propaganda digital*, um mercado que cresceu abruptamente. De fato, estima-se que o mercado de propaganda digital faturou algo em torno de 25 bilhões de dólares em 2006.

Para veicular sua propaganda na Web, o anunciante associa seus produtos e serviços ao interesse momentâneo do usuário expresso, por exemplo, através de uma consulta em uma máquina de busca. Como consequência, máquinas de busca se tornaram um elemento central da chamada “economia da Web”.

Como *máquinas de busca* são sistemas de Recuperação de Informação, o conhecimento e domínio das técnicas, processos e métricas da área de Recuperação de Infomação tornaram-se requisitos básicos para uma melhor compreensão das questões centrais na Web nos dias de hoje.

Recuperação de Informação lida com a representação, armazenamento, organização e acesso a itens de informação (tais como livros, documentos diversos, páginas Web, etc), e tem por objetivo maior prover ao usuário acesso facilitado à informação de seu interesse.

Neste curso, o objetivo maior é introduzir o aluno aos conceitos fundamentais relacionados com a área, discutindo-os no contexto da Web. Os tópicos centrais a serem abordados incluem processamento de consultas, modelos para ordenação da respostas, avaliação da qualidade das respostas, indexação, interfaces, novos paradigmas para pesquisa por informação.

2. Calendário e Avaliações

Fevereiro: 28 (1 aula)

Março: 5-7, 12-14, 19-21, 26-28 (8 aulas)

Abril: 2-4, 9-11, 16-18, 23-25, 30 (9 aulas)

Maiο: 7-9, 14-16, 21-23, 28-30 (8 aulas)

Junho: 4-6, 11-13, 18-20, 25-27 (8 aulas)

Avaliação

2 provas: 40 pontos

- Prova 1: ~ aula 13 (9 de abril)
- Prova 2: ~ aula 22 (21 de maio)

Trabalho Prático: 45 pontos

Seminário em Sala: 10 pontos

Presença em Aula: 5 pontos

3. Trabalho Prático

O trabalho prático consiste em desenhar, implementar e validar uma máquina de busca para a Web brasileira. A máquina de busca deverá atender a requisitos mínimos de qualidade e desempenho, conforme explicado durante o curso. A implementação deverá ser feita em C++, sem a utilização de módulos externos por ventura disponíveis.

4. Bibliografia de Referência

Ricardo Baeza-Yates & Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, 1999.

Ian H. Witten, Alistair Moffat & Timothy C. Bell. *Managing Gigabytes*. Morgan Kaufmann, 1999, second edition.