

CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
RECUPERAÇÃO DE INFORMAÇÃO - MÁQUINAS DE BUSCA NA WEB

Última alteração: 20 de junho de 2007

1º semestre de 2007

Professores: Berthier Ribeiro-Neto e Nivio Ziviani

Monitor: Anísio Mendes Lacerda

Trabalho Prático 3:

30/05/2007

Data de Entrega: 29/06/2007 (até as 10:00 horas) (até as 10:00 horas) (até as 10:00 horas) (até as 10:00 horas) (até as 10:00 horas) (até as 10:00 horas) (até as 10:00 horas) (até as 10:00 horas) (até as 10:00 horas)

Processamento de Consultas

O objetivo deste trabalho é projetar e implementar um sistema de programas para responder eficiente e eficazmente consultas realizadas sobre o índice gerado a partir dos programas do Trabalho Prático 2.

Características do Sistema de Consultas

O sistema deverá:

1. Realizar ordenação (*ranking*) usando o modelo de espaço vetorial.
2. Desenvolver uma interface HTML com uma caixa de consultas e colocar a interface disponível para consultas.
3. No documento final, o aluno deverá comentar sobre a qualidade das respostas por meio da inspeção de algumas consultas por ele/ela selecionados.
4. Nota: se o aluno implementar análise de links com "page rank" e mostrar que os resultados melhoram, ele ganhará pontos adicionais que serão utilizados para compensar pontos perdidos nas etapas anteriores do TP. Tais pontos adicionais estarão restritos às seguintes condições:
 - Somente podem ser utilizados para melhorar a nota final do TP;
 - O número máximo de pontos adicionais está limitado a 10.

Referências

- [1] B. A. V. Pôssas. Set-Based Vector Space Model: A new Approach for Correlation Based Model. Tese de Doutorado, Departamento de Ciência da Computação, UFMG, 22 de agosto de 2005.
- [2] S. E. Robertson and K. S. Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3):129–146, May-June 1976.
- [3] S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. Okapi at trec-3. In *Proceedings of the Third Text Retrieval Conference (TREC-3)*, pages 109–126. NIST Special Publication 500-225, April 1995.
- [4] Ian H. Witten, Alistair Moffat e Timothy C. Bell: *Managing Gigabytes: Compressing and Indexing Documents and Images*, Morgan Kaufmann Publishers, 1999, second edition.