

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

Última alteração: 16 de abril de 2012

1º semestre de 2012

Professores: Berthier Ribeiro-Neto e Nivio Ziviani

Monitor: Marco Túlio Ribeiro

Trabalho Prático 2:

18/04/2012

Data de Entrega: 16/05/2012

Penalização por Atrazo: 1 ponto até 21/05/2012 mais 1 ponto por dia útil a seguir

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 1.

Cada aluno deverá:

1. Apresentar um sistema capaz de realizar ordenação (*ranking*) usando:
 - o modelo de espaço vetorial.
 - o modelo BM25.
2. Desenvolver uma interface HTML com uma caixa de consultas e colocar a interface disponível para consultas.
3. Apresentar as curvas de precisão versus revocação para as 30 consultas disponíveis no endereço <http://garnize.latin.dcc.ufmg.br/colecaoRI/consultas-30>, utilizando:
 - O modelo de espaço vetorial.
 - O modelo BM25.
 - Uma combinação dos dois modelos anteriores.

O conjunto de relevantes para cada uma das 30 consultas está disponível no endereço:
<http://garnize.latin.dcc.ufmg.br/colecaoRI/relevantes-30.tar.gz>

Referências

- [1] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. Addison-Wesley, 2011 (second edition).
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Ian H. Witten, Alistair Moffat e Timothy C. Bell: *Managing Gigabytes: Compressing and Indexing Documents and Images*, Morgan Kaufmann Publishers, 1999, second edition.