

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

Última alteração: 5 de junho de 2017

1º semestre de 2017

Professores: Berthier Ribeiro-Neto e Nivio Ziviani

Monitor: Marlon da Silva Dias <mailto:marlonsdias@gmail.com>

Trabalho Prático 3:

22/05/2017

Data de Entrega do Resultado de Rankings das Consultas: 23/06/2017

Data de Entrega do TP3: 02/07/2017

Penalização por Atrazo: 1 ponto até 05/07/2017 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 2.

Cada aluno deverá:

1. Apresentar um sistema capaz de realizar ordenação (*ranking*) usando:
 - Modelo de espaço vetorial.
 - Pagerank.
 - Anchor text.
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 o conjunto de consultas a ser disponibilizado, utilizando:
 - O modelo de espaço vetorial.
 - Uma combinação do modelo de espaço vetorial com Pagerank.
 - Uma combinação do modelo de espaço vetorial com anchor text.

O conjunto de relevantes para cada uma das consultas vai estar disponível em endereço a ser fornecido.

4. Descrever cuidadosamente os resultados observados para as diversas consultas com relação a qualidade do ranking em cada situação apresentada no item anterior.
5. O conjunto de relevantes para cada consulta será determinado com a participação dos alunos.

Referências

- [1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, 2011 second edition).
- [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.