

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

Última alteração: 3 de junho de 2013

1º semestre de 2013

Professores: Berthier Ribeiro-Neto e Nivio Ziviani

Monitor: Aline Bessa

Trabalho Prático 3:

22/05/2013

Data de Entrega: 17/06/2013

Penalização por Atrazo: 1 ponto até 24/06/2012 mais 1 ponto por dia útil a seguir

---

## Processamento de Consultas

O objetivo deste trabalho é implementar dois mecanismos para melhorar a ordenação de respostas (*ranking*) do sistema de processamento de consultas desenvolvido no Trabalho Prático 2. A ordenação de respostas no TP2 envolveu:

1. O modelo de espaço vetorial.
2. O modelo BM25.
3. Uma combinação dos dois modelos anteriores.

## O Que Fazer

1. Realizar ordenação (*ranking*) usando uma combinação do sistema criado no TP2 com:
  - (a) Análise de *links* usando PageRank.
  - (b) Texto âncora (*anchor text*).
2. Apresentar as curvas de precisão-revocação para as mesmas 34 consultas utilizadas no TP2, disponíveis no link <http://garnize.latin.dcc.ufmg.br/colecaoRI/consultas-34>, considerando:
  - (a) O processador de consultas do TP2 envolvendo o modelo de espaço vetorial e o BM25.
  - (b) As melhores combinações entre o processador de consultas do TP2 com novas evidências usando PageRank e texto âncora.

O conjunto de relevantes para cada uma das 34 consultas está disponível no endereço:  
<http://garnize.latin.dcc.ufmg.br/colecaoRI/relevantes-34.tar.gz>