

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

Última alteração: 17 de junho de 2009

1º semestre de 2009

Professores: Berthier Ribeiro-Neto e Nivio Ziviani

Monitor: Anísio Mendes Lacerda

Trabalho Prático 3:

27/05/2009

Data de Entrega: 18/06/2009

Penalização por Atraso: 1 ponto até 22/06/2009 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.

## O Que Fazer

1. Realizar ordenação (*ranking*) usando uma combinação do sistema criado no TP2 com análise de *links* usando PageRank.
2. Participar do processo de avaliação de documentos relevantes utilizando o *pool* de respostas disponível no endereço:  
<http://garnize.latin.dcc.ufmg.br/colecaoRI/pool>.  
para as 50 consultas disponíveis no endereço:  
<http://garnize.latin.dcc.ufmg.br/colecaoRI/consultas-50>  
Para tal vamos utilizar o sistema EESIR disponível em  
<http://garnize.latin.dcc.ufmg.br/ri-evaluation>,  
cujo uso foi explicado em sala de aula.
3. Apresentar as curvas de precisão-revocação para as 50 consultas mencionadas acima, 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 (*anchor text*).