

UFU/FACOM  
Bacharelado em Sistemas de Informação  
GSI024 - Organização e Recuperação de  
Informação

18 de Julho de 2013

## Prática 4 - Modelo vetorial e ranking

Nesta prática deverão ser implementadas estruturas e operações para prover recuperação de informação baseada no modelo vetorial e ranking.

- O sistema deverá retornar documentos ordenados de mais relevante para menos relevante para as consultas dos usuários.
- Para isso, para cada termo no dicionário, deverá ser calculado e armazenado seu IDF.
- No índice invertido, para cada referência de um termo a um documento, deverá ser calculado e armazenado seu TF na forma logaritmica.
- Os valores  $TF \times IDF$  deverão ser normalizados para cada documento utilizando a magnitude do vetor que representa o documento  $||d||$ .
- A consulta também deverá ser normalizada com sua magnitude.
- A pontuação de cada par consulta-documento utilizará o cosseno entre o vetor-consulta e o vetor-documento.
- A ordenação dos resultados será em ordem decrescente.