

---

### ÍNDICE INVERTIDO

Considere os documentos abaixo sobre a *Copa do Mundo de 2010* para resolver os exercícios a seguir.

- DOC 01** Fabricante se diz surpresa com críticas à bola da Copa
- DOC 02** Felipe Melo põe Argentina atrás do Brasil na Copa e vira espelho de Dunga
- DOC 03** Seleção completa quatro jogos em três treinos e mostra Dunga “carrasco”
- DOC 04** Seleção fica no 0 a 0 em primeiro coletivo na África; gols só na bola parada

01. Mostre a *tokenização* para os documentos apresentados.
02. Mostre a eliminação das *stop words* (termos discriminates) para os documentos apresentados.
03. Cite 5 exemplos de normalização que poderiam ser feitos para a construção do índice invertido (em documentos em geral e não somente com base nesse exemplo).
04. Faça a análise de frequência para os 6 primeiros termos.
05. Desenhe o índice invertido para a coleção de documentos apresentada (utilize somente os 6 primeiros termos).
06. Desenhe o dicionário de termos para a coleção de documentos apresentada (utilize somente os 5 primeiros termos).
07. Considerando o seu índice invertido, quais documentos retornariam as seguintes consultas? Qual o caminho percorrido no índice para responder as essas consultas?
  - a. Copa
  - b. Brasil
  - c. Copa AND Brasil
  - d. Gols OR Dunga
  - e. Brasil AND Argentina
08. Qual a importância da ordenação dos termos do índice? Dê um exemplo.
09. Qual a importância da ordenação dos documentos na lista invertida? Dê um exemplo.
10. Faça uma análise comparativa entre armazenar a frequência dos termos na lista de termos ou armazenar a frequência de termos para cada documento? Como essas informações poderiam ser utilizadas?
11. Qual a vantagem (utilidade) de se armazenar a posição em que cada termo ocorre dentro dos documentos? Como isso poderia ser implementado (Estrutura de dados)?
12. Quais as vantagens e desvantagens do uso de índice invertido?

