

O que é um Índice Invertido

Documento 1:

Espaço: a fronteira final. Estas são as viagens ...

Documento 2:

Ele é mau, ele é o número um. Ele é o cowboy do espaço com a arma laser!

Índice Invertido

espaço:	1, 2
a:	1, 2
final:	1
fronteira :	1
ele:	2
mau:	2
...	

Claro que
não é tão
simples

- TF-IDF: Term Frequency * Inverse Document Frequency
- Term Frequency: frequência com que uma palavra aparece em um documento
- Document Frequency é a frequência com que um termo aparece em todos os documentos
- Relevância: Term Frequency / Document Frequency, medida em um documento