PRÉ-PROCESSAMENTO

ARQUIVO . db SQLITE 3 com texto em ascií

#include < sqlite 3, h>

Importar texto p/programa

Ler do arquivo reparadores.txt

e unar on reparadores p/converter o

texto em vetor solução: Atrtok

Parecido com , split (...)

Let do arquivo "stopwords.txt" e remover as stopwords.

Filtrá-las do vetor anterior.

Hinclude < libstenener. h>

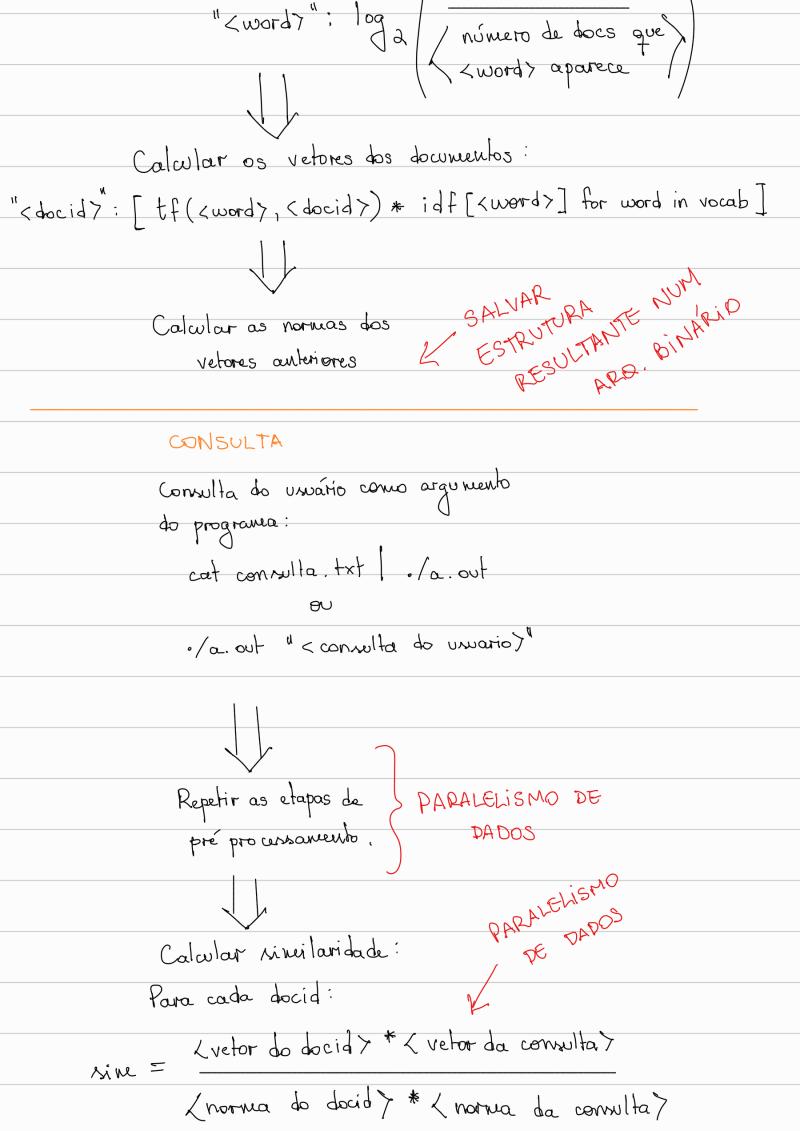
Aplicar a função sb_stemmer_stem () do Snowball Stemmer em cada palavra do vetor anterior.

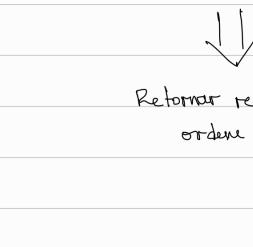
BARREIRA ->

Montar a matriz de frequência ou indice de arquivo invertido.

Calcular o idf:

(< número de docs>





Retornar resultados en ordens decrescente de sincilaridade.