

Tarefa de Programação 2

João Lucas Pontes Freitas

Respostas questão 4:

- a) 1300 termos de indexação.
- b) A palavra que teve maior peso TF-IDF foi 'luck' no documento 'GoodLuckBaby.txt'.
O valor de TF-IDF foi de: 24.13780634075915
- c) Tempo gasto na leitura, normalização dos documentos e criação do vocabulário:
0.010229349136352539 segundos
- d) Tempo gasto no processamento das matrizes TF e IDF: 0.13655805587768555 segundos

Print com os resultados:

```
Tamanho do vocabulário: 1300
Tempo gasto na leitura, normalização dos documentos e criação do vocabulário: 0.010229349136352539 segundos

Matriz TF-IDF:
[[0.37647521 0.30857687 0.22200174 ... 0.19128873 0.07400058 0.28174648]
 [0.         6.         0.         ... 0.         0.         0.         ]
 [0.         0.         0.         ... 0.         0.         0.         ]
 ...
 [0.         0.         0.         ... 0.         3.32192809 0.         ]
 [0.         0.         0.         ... 0.         0.         0.         ]
 [0.         4.32192809 0.         ... 0.         0.         0.         ]]

A palavra que teve maior peso TF-IDF foi 'luck' no documento 'GoodLuckBaby.txt'
O valor de TF-IDF foi de: 24.13780634075915
Tempo gasto no processamento das matrizes TF e IDF: 0.13655805587768555 segundos

Digite a consulta: you

Tempo gasto na consulta: 0.0012731552124023438 segundos

Similaridades da consulta com os documentos:

BirdsOfAFeather.txt: (cosseno 0.0173) (sem divisao: 0.0026)
Snooze.txt: (cosseno 0.0171) (sem divisao: 0.0026)
LoseControl.txt: (cosseno 0.0169) (sem divisao: 0.0026)
BeautifulThings.txt: (cosseno 0.0166) (sem divisao: 0.0025)
IHadSomeHelp.txt: (cosseno 0.0161) (sem divisao: 0.0025)
AgoraHills.txt: (cosseno 0.0160) (sem divisao: 0.0024)
Greedy.txt: (cosseno 0.0159) (sem divisao: 0.0024)
LovinOnMe.txt: (cosseno 0.0141) (sem divisao: 0.0021)
GoodLuckBaby.txt: (cosseno 0.0138) (sem divisao: 0.0021)
TooSweet.txt: (cosseno 0.0135) (sem divisao: 0.0021)
CruelSummer.txt: (cosseno 0.0119) (sem divisao: 0.0018)
PleasePleasePlease.txt: (cosseno 0.0118) (sem divisao: 0.0018)
IRememberEverything.txt: (cosseno 0.0112) (sem divisao: 0.0017)
MillionDollarBaby.txt: (cosseno 0.0109) (sem divisao: 0.0017)
LikeThat.txt: (cosseno 0.0109) (sem divisao: 0.0017)
StickSeason.txt: (cosseno 0.0108) (sem divisao: 0.0016)
Espresso.txt: (cosseno 0.0108) (sem divisao: 0.0016)
NotLikeUs.txt: (cosseno 0.0096) (sem divisao: 0.0015)
ABarSong.txt: (cosseno 0.0000) (sem divisao: 0.0000)
Saturn.txt: (cosseno 0.0000) (sem divisao: 0.0000)
```