

Sistemas de Informação - UFU
Organização e Recuperação da Informação
TP02

Aluna: Beatriz Ribeiro Borges

Matrícula: 12021BSI231

Exercício 3

2) Use o programa criado no Exercício 2 para responder as seguintes perguntas:

a) Qual o tamanho do vocabulário da coleção (quantos termos)?

O vocabulário da coleção tem 751 termos.

```
742  years
743  yes
744  you
745  youll
746  young
747  your
748  youre
749  youve
750  zombie
751  zone
752
```

b) Qual termo, em qual documento, possui o maior peso TF-IDF?

```
Termo: eileen
TF-IDF: 23.8724376017338
Documento: come on eileen - dexys midnight runners.txt
> 
```

c) Quanto tempo o programa demorou para criar o vocabulário da coleção de músicas?

0:00:00.004027

```
0:00:00.004027
> 
```

d) Quanto tempo o programa demorou para calcular o TF-IDF da coleção inteira?

0:00:11.851704

```
0:00:11.851704
> 
```