

2º TP – Modelo Vetorial

Alunos:

- João Guilherme Araújo Viana – 12311BSI238
- Matheus Gualter Silva Resende – 12311BSI310

4) Responda:

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

Segundo os documentos selecionados e a coleção montada, o vocabulário possui 1125 termos.

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

*Após um trecho de script utilizado no código, consta-se que o termo com maior peso TF-IDF é **'help'**, no documento **'04 – I Had Some Help'**, como peso TF-IDF equivalente a **0.7690**.*

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

Na primeira execução, demorou aproximadamente 18 segundos. Nas demais, demora aproximadamente 4 segundos.

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

Na primeira execução, levou aproximadamente 21 segundos. Nas demais, após a primeira execução, levou aproximadamente 3 segundos.