

Professor: Gustavo Guedes
Aluno: Rodrigo Tavares de Souza

Trabalho 2

Parte 2 - Long

Tempo de processamento: 66282 (ms) / 66,282 (s).

Total de documentos processados: 485
10 documentos mais relevantes para a query: kids dog

Documento	Dist. Coseno	Similaridade
cv119_9867.txt	00,12	03,19
cv390_11345.txt	00,08	02,52
cv304_28706.txt	00,08	04,14
cv327_20292.txt	00,07	04,56
cv153_10779.txt	00,07	02,90
cv344_5312.txt	00,07	02,32
cv298_23111.txt	00,06	02,93
cv112_11193.txt	00,06	02,80
cv136_11505.txt	00,05	02,80
cv380_7574.txt	00,05	01,49
cv269_21732.txt	00,05	02,90

Tempo de escrita (console / PDF): 1016 (ms) / 01,016 (s).

Existem 10 documentos relevantes que possuem as palavras kids e dog, com as seguintes frequências, respectivamente:

cv327_20292.txt (4 e 10)
cv304_28706.txt (7 e 4)
cv119_9867.txt (2 e 2)
cv298_23111.txt (1 e 2)
cv284_19119.txt (1 e 1)
cv153_10779.txt (0 e 7)
cv136_11505.txt (0 e 6)
cv112_11193.txt (0 e 6)
cv256_14740.txt (7 e 0)
cv380_7574.txt (5 e 0)

O sistema retornou 8 desses documentos. Kids aparece 84 vezes em 54 arquivos e dog 69 vezes em 26 arquivos. Dessa forma, dog tem peso maior do que kids. O sistema TF e DF não considera um peso maior na busca para documentos que apresentem simultaneamente os dois termos, por isso o documento cv284_19119.txt não foi retornado. O documento apresenta apenas 1 ocorrência de cada termo, o que resulta em um peso final baixo. O outro documento não retornado - cv256_14740.txt, possui baixa frequência dos demais termos além de kids. A distância por cosseno utiliza a raiz do somatório dos quadrados das frequências desses termos excedentes, o que fez com que o mesmo tivesse baixo peso e portanto não fosse retornado. Para efeito de cálculo, são considerados então: 10 documentos retornados, 8 relevantes de um total de 10 relevantes esperados. Nessa configuração temos a seguinte precisão, abrangência e F-Measure:

Precisão: 00,800

Abrangência: 00,800

F-Measure: 00,800