

Aluno: Pedro Henrique Silva Santana

Matrícula: 12011BSI218

## ES3- Modelo Vetorial

D1 – {logan e ororo são x-men}

D2 – {stark, parker e logan já foram vingadores parker gostaria de ser novamente}

D3 – {ororo e stark não são guardiões como groot e rocket}

D4 – {eu sou groot, logan, todos somos groot e groot precisa de ajuda}

D5 - {logan e rocket formariam uma dupla melhor do que logan e groot}

Assuma que o vocabulário dessa coleção seja formado pelos seguintes termos de indexação: V = {logan, ororo, stark, parker, groot, rocket}.

a) Calcule os pesos TF-IDF para a coleção.

TFi1	TFi2	TFi3	TFi4	TFi5		IDF = log(N/ni)	d1	d2	d3	d4	d5
1	1	0	1	2		0,321928	0,321928	0,321928	0	0,321928	0,643856
1	0	1	0	0		1,321928	1,321928	0	1,321928	0	0
0	1	1	0	0		1,321928	0	1,321928	1,321928	0	0
0	2	0	0	0		2,321928	0	4,643856	0	0	0
0	0	1	2,584963	1		0,736966	0	0	0,736966	1,905028	0,736966
0	0	1	0	1		1,321928	0	0	1,321928	0	1,321928
						Tamanho	1,360563	4,839063	2,405327	1,932038	1,644738

b) Considere a seguinte consulta q1 = {logan parker groot}. Calcule o grau de similaridade dos documentos da coleção para a consulta q1. Não se esqueça de calcular as normas dos documentos. Explique os resultados obtidos.

q1 = {logan parker groot}

Wi,q		0,321928	0	0	2,321928	0,736966	0	tamanho	2,457256	
Produto interno							Total	Score	ordenando	
d1	0,103638	0	0	0	0	0	0,103638	0,030999	d2	0,915525
d2	0,103638	0	0	10,7827	0	0	10,88634	0,915525	d4	0,317551
d3	0	0	0	0	0,543118	0	0,543118	0,09189	d5	0,18567
d4	0,103638	0	0	0	1,40394	0	1,507578	0,317551	d3	0,09189
d5	0.207275	0	0	0	0.543118	0	0.750394	0.18567	d1	0.030999

c) É possível usar o modelo vetorial para comparar documentos? Se sim, explique qual seria a similaridade entre os documentos D1 e D5.

Sim, é possível utilizar o modelo vetorial para comparar documentos.

Produto interno							Tamanho	
d1	0,321928	1,321928	0	0	0	0	1,360563	
d5	0,643856	0	0	0	0,736966	1,321928	1,644738	Score
Total	0,207275	0	0	0	0	0	0,207275	0,250568