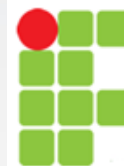




INSTITUTO FEDERAL
MINAS GERAIS

RECUPERAÇÃO DE INFORMAÇÃO

Profa. Mcs. Patrícia Proença
patricia.proenca@ifmg.edu.br



ATENÇÃO!!!

- ⤵ O material a seguir é uma videoaula apresentada pela professora PATRÍCIA APARECIDA PROENÇA AVILA, como material pedagógico do IFMG, dentro de suas atividades curriculares ofertadas em ambiente virtual de aprendizagem. Seu uso, cópia e ou divulgação em parte ou no todo, por quaisquer meios existentes ou que vierem a ser desenvolvidos, somente poderá ser feito, mediante autorização expressa deste docente e do IFMG. Caso contrário, estarão sujeitos às penalidades legais vigentes”.
- ⤵ Conforme Art. 2º§1º da Nota Técnica nº 1/2020/PROEN/Reitoria/IFMG (SEI 0605498, Processo nº 23208.002340/2020-04



INSTITUTO FEDERAL
MINAS GERAIS

Exemplo Modelo Vetorial

Exemplo

- Considere uma coleção de quatro documentos.
- Assuma que todos os documentos e consultas passam por um pré-processamento, e que somente os termos presentes na tabela abaixo são incluídos no índice.

	doc 1	doc 2	doc 3	doc 4
recuperação	4		2	1
informação	6		4	1
IFMG	1	2		
Ciência		2	3	3
Computação		4		2

Exemplo

- Calculando o TF-IDF para os termos i nos documentos j :

$$W_{i,j} = (1 + \log f_{i,j}) \times \log (N/n_i)$$

Exemplo

- Calculando o TF-IDF para os termos i nos documentos j:

$$W_{i,j} = (1 + \log f_{i,j}) \times \log (N/n_i)$$

	TF 1	TF 2	TF 3	TF 4			IDF 1	IDF 2	IDF 3	IDF 4
recuperação					recuperação					
o	3	0	2	1	o	0,415037	0,415037	0,415037	0,415037	
informação	3,584963	0	3	1	informação	0,415037	0,415037	0,415037	0,415037	
IFMG	1	2	0	0	IFMG	1	1	1	1	
Ciência	0	2	2,584963	2,584963	Ciência	0,415037	0,415037	0,415037	0,415037	
Computação					Computação					
o	0	3	0	2	o	1	1	1	1	

Exemplo

- Calculando o TF-IDF para os termos i nos documentos j:

$$W_{i,j} = (1 + \log f_{i,j}) \times \log (N/n_i)$$

	Vetor de Pesos com tamanho 5			
	w1	w2	w3	w4
recuperação	1,245112	0	0,830075	0,415037
informação	1,487894	0	1,245112	0,415037
IFMG	1	2	0	0
Ciência	0	0,830075	1,072856	1,072856
Computação	0	3	0	2

Exemplo

- Calculando os TF-IDF para a consulta:
 - “Recuperação de Informação”

$$W_{i,j} = (1 + \log f_{i,j}) \times \log (N/n_i)$$

Exemplo

- Calculando os TF-IDF para a consulta:
 - “Recuperação de Informação”

	Vetor consulta
	q
recuperação	0,415037499
informação	0,415037499
IFMG	0
Ciência	0
Computação	0

$$W_{i,j} = (1 + \log f_{i,j}) \times \log (N/n_i)$$

Exemplo

- Calculando as similaridades:

$$\frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}}$$

Exemplo

- Calculando as similaridades:
- Doc 1 e consulta:
- Sim =
$$(1,2451 * 0,4150 + 1,4879 * 0,4150 + 1 * 0 + 0 * 0 + 0 * 0) / (\text{raiz}(1,2451^2 + 1,4879^2 + 1^2) * \text{raiz}(0,4150^2 + 0,4150^2)) = 0,8853$$

$$\frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}}$$

Exemplo

	SIMILARIDADE
doc 1	0,885388126
doc 2	0
doc 3	0,796929768
doc 4	0,250378725



MUNDO1
MENSAGENS

COMEÇA A SEMANA, RENOVAM-SE
AS EXPECTATIVAS, E SOMOS INVADIDOS
POR UMA GRANDE VONTADE
DE CONQUISTAR ALGO GRANDIOSO.

