

## Exercices

### Exercice 1

#### Énoncé

Collection de 3 documents :

- D1 : « Shipment of gold arrived in a truck »
- D2 : « Shipment of gold damaged in a fire »
- D3 : « Delivery of silver arrived in a silver truck »

Requête : « gold silver truck »

#### Question

Quel est le classement des documents par le modèle vectoriel ?

#### Réponse

TFs :

Document	Shipment	of	gold	arrived	in	a	truck	damaged	fire	delivery	silver
D1	1	1	1	1	1	1	1				
D2	1	1	1		1	1		1	1		
D3		1		1	1	1	1			1	2
Q			1				1				1

IDFs :

Shipment	of	gold	arrived	in	a	truck	damaged	fire	delivery	silver
0.18	0	0.18	0.18	0	0	0.18	0.48	0.48	0.48	0.48

TFIDFs :

Document	Shipment	of	gold	arrived	in	a	truck	damaged	fire	delivery	silver
D1	0.18	0	0.18	0.18	0	0	0.18	0	0	0	0
D2	0.18	0	0.18	0	0	0	0	0.48	0.48	0	0
D3	0	0	0	0.18	0	0	0.18	0	0	0.48	0.96
Q	0	0	0.18	0	0	0	0.18	0	0	0	0.48

Dernière ligne contestable, on pourrait mettre des 1 là où il y a des valeurs différentes de 0.

Similarités :

$$- Sim(Q, D_1) = \frac{\sum_{i=1}^{11} W_{i,q} \cdot W_{i,d_1}}{|\vec{Q}| \cdot |\vec{D_1}|} = \frac{2 \times 0.18^2}{0.54 \times 0.36} = 0.33$$

$$- Sim(Q, D_2) = \frac{2 \times 0.18^2}{0.54 \times 0.72} = 0.08$$

$$- Sim(Q, D_3) = \frac{0.48 \times 0.96 + 0.18^2}{0.54 \times 0.36} = 0.83$$

Tri :  $D_3 > D_1 > D_2$