

Fiche TD

EX 0 :

Considérons la collection faite des 3 documents suivants : t_i des mots

$D1=(t1,t3,t4)$

$D2=(t1,t2,t4,t6)$

$D3=(t1,t2,t3,t4,t5)$

En utilisant le modèle booléen quel est le résultat de la requête suivante : $Q = t1 \wedge (t3 \vee t4)$.

Quel est l'inconvénient de ce modèle.

EX1 :

Considérons les documents textuels suivants :

- **A:** Le loup est dans la bergerie.
- **B:** Les moutons sont dans la bergerie.
- **C:** Un loup a mangé un mouton, les autres loups sont restés dans la bergerie.
- **D:** Il y a trois moutons dans le pré, et un mouton dans la gueule du loup.

Prenons le vocabulaire suivant : {"loup", "mouton", "bergerie", "pré", "gueule"}.

- 1- Donnez la matrice de correspondance terme document de ces documents (matrice binaire)
- 2- Donnez le résultat des requêtes suivantes en calculant leur score de correspondances (Cosinus)

- q_1 "loup et pré"
- q_2 "loup et mouton"
- q_3 "bergerie"
- q_4 "gueule du loup"

EX2 :

Considérons un corpus de test pour un SRI qui contient 150 documents et 1 requête avec le jugement suivant :

Q_1 : d12, d141, d04, d15, d91, d87, d149

Le résultat de Q_1 avec ce SRI a donné les fichiers suivants :

D12	D61	D13	D111
D141	D72	D15	D107
D01	D87	D26	D91
D04	D101	D27	D98
D32	D100	D34	D150

- 1- Donnez le rappel et précision et f-mesure sur les 10 premiers documents retournés puis sur l'ensemble du corpus
- 2- Tracer la courbe R-P sur les 20 premiers documents retournés en utilisant 11 pnts de rappels.

Si on ajoute une 2 eme requêtes avec le même SRI et donne le résultat et jugement suivants
 Q2 : d26, d31, d7, d3, d71, d87, d131

D26	D19	D14	D04
D45	D111	D87	D99
D31	D120	D62	D131
D10	D3	D18	D148
D7	D71	D21	D2

3- Tracez la courbe R-P de ce SRI.

-EX3 :

Soit la matrice de correspondance termes-documents M suivantes : (4 documents, 10 termes)

Et le vecteur de fréquence de termes dans une requête $Q = \{0,1,1,1,0,0,0,0,2\}$

- 1- Donner la matrice TF*IDF correspondante.
- 2- Donner le résultat de la requête en utilisant la fonction Cosinus.

$$3- M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 2 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$