

Exercice 1 (8pts) Répondez brièvement aux questions suivantes

- 1- Peut-on utiliser un **SGBD** pour rechercher des attributs dans un corpus de documents ? justifiez votre réponse.
- 2- **a-** Définissez **Précision** et **Rappel**, les 2 mesures utilisées pour évaluer la performance d'un SRI en précisant chaque terme dans les formules.
b- Pourquoi ne peut on pas se limiter à la mesure **Précision** ?
- 3- Dans la phase indexation, comment sont déterminés les termes représentants au mieux le contenu d'un document en utilisant la loi de **ZIPF**?
- 4- Dans la phase indexation et dans l'approche **tf** (term frequency) **idf** (inverted term frequency) , expliquez le rôle de chacune de ces valeurs (**tf** et **idf**) dans un corpus de documents.
- 5- Quelle est la difficulté majeure pour un utilisateur qui utilise un SRI basé sur l'**appariement exact** par rapport un SRI basé sur l'appariement rapproché ?.

Exercice 2 (8pts)

Soit le corpus de documents suivant représentant la fréquence de chaque terme (**t1, t2, t3**) dans le document. Quel est le terme le plus discriminant selon l'approche basée sur la **valeur de discrimination**.

t1, t2, t3

D1 (9, 1, 3)

D2 (8, 0, 0)

D3 (8, 4, 7)

Exercice 3 (4 pts) Soit la représentation suivante :

	T1	T2	T3	T4
D1	1	0.9	1	1
D2	1	0.9	0	0.7
D3	0.8	0	1	0.9

Soit la requête **Q= T1 et (T2 ou T4)**

Donnez le résultat de Q par ordre de pertinence (en classant les documents) en utilisant le modèle booléen classique.

Exercice (Contrôle continu)

1-Soit la représentation suivante

D1 : (1, 0, 1, 0)

D2 : (3, 0, 2, 1)

D3 : (1, 2, 3, 0)

Q : (2, 0, 2, 0)

En appliquant le modèle vectoriel, quel est le document qui sera classé en premier ?

(Utilisez la mesure de similarité de votre choix)