

Questions de cours :

1. donner une définition d'un système de recherche d'information.
2. Donner le schéma général d'un système de recherche d'information.
3. expliquer en quelques lignes le modèle vectoriel.
4. quelles sont les mesures de performance pour évaluer un système de recherche d'information SRI ? donner les formules et l'explication de chaque indicateur.
5. donner une explication de loi de Zipf.
6. est ce que la phase de reformulation de la requête est obligatoire pour un SRI ? argumenter votre réponse.
7. Donner le principe d'algorithme PageRank.
8. Quelle est la différence entre le modèle booléen strict et le modèle booléen étendu ?

Exercice 01 :

On vous propose une collection de trois (3) documents :

D1 : plus un mot est fréquent dans un document, plus est non pertinent.

D1 : plus un mot est rare dans la collection, plus est non pertinent.

D3 : les mots pertinents sont les mots qui sont trop rares et pas trop fréquents dans un document.

Le vocabulaire proposé pour cette collection est : {'fréquent', 'document', 'pertinent', 'rare', 'mot', 'être'}.

1. Donner la représentation Tf des documents précédents.
2. Donner la représentation Tf normalisé des documents précédents.
3. Donner la représentation Idf des termes.
4. calculer Tf.Idf. des termes.

Exercice 02 :

La liste suivante de R et N {pertinent, non pertinent} des documents retournés dans une liste de classement des 20 documents récupérés en réponse à une requête.

N N N N N N N N R N R N N N R N N N N N.

La collection contient 10.000 documents. Le document le plus à gauche de la liste de classement est considéré le plus pertinent. Cette liste montre 3 documents pertinents. Supposons qu'il y a 6 documents pertinents au total dans cette collection.

1. Quelle est la précision du système sur le top 20?
2. La moyenne harmonique F1 du rappel et précision est présenté par la formule suivante :

$$F1 = 1 / ((1/p) + (1/r))$$
, où r : rappel & p : précision.
 Question : calculer la moyenne harmonique F1 sur le top 20.
3. Calculer est la précision du système sur le top 10.
4. Calculer la moyenne harmonique F1 sur le top 10.

Bon courage.