

Licence Fondamentale Sciences de l'informatique	Classes : LGLSI-2
TP n°3 Techniques d'indexation et recherche multimédia	
Objectifs	Recherche textuelle

TP3: Recherche TEXTUELLE

L'objectif de ce TP est de réaliser le module de recherche textuelle pour le système de recherche d'information que nous sommes entrain de développer. Notons que ce TP est la suite du TP2.

A la fin de ce TP, l'étudiant sera évalué sur la totalité du travail (TP2 + TP3). La note obtenue après cette évaluation sera considérée comme sa note de TP.

Partie 1: indexation textuelle

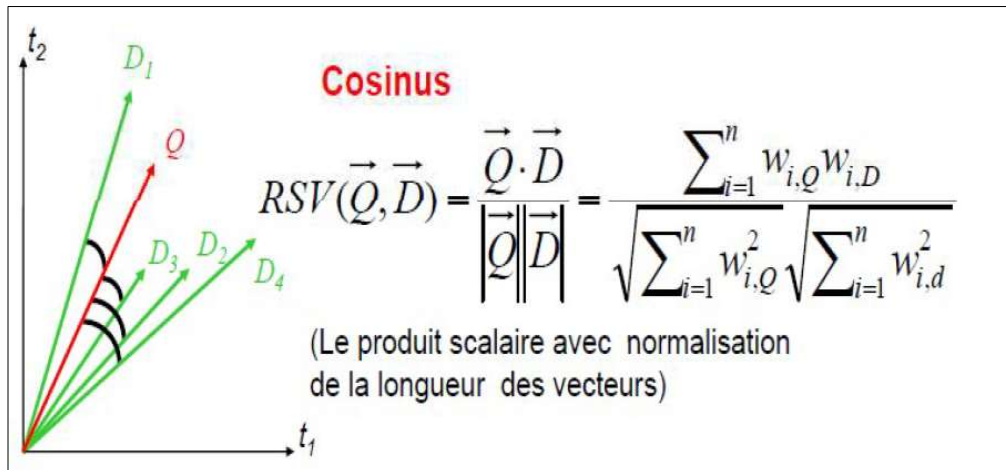
- Cette partie est réalisée pendant les séances du TP2.

Partie 2: Recherche textuelle

Afin de finaliser le système de recherche d'information, ajouter à votre code les fonctions suivantes :

- Fonction de lecture de la requête utilisateur saisie au clavier
- Fonction d'indexation de la requête utilisateur (il faut appliquer le même algorithme d'indexation qui a été utilisé pour l'indexation du corpus qui est constitué de 10 documents).
- Fonction de recherche des documents qui sont jugés pertinents par rapport à la requête de l'utilisateur. Les documents retournés par cette fonction doivent être triés par ordre de pertinence. Pour ce faire, il faut utiliser une fonction de similarité pour calculer le score de chaque document par rapport à la requête utilisateur que nous notons $S(Q,D)$. $S(Q,D)$ sera calculée selon la fonction de similarité $RSV(Q, D)$ ci-dessous (voir chapitre 4 pour plus de détails). Il faut donc utiliser ce score pour afficher la liste des documents pertinents pour la requête utilisateur (c.a.d dont le score $S(Q,D) \neq 0$) triés par ordre de pertinence. Notons que le document le plus pertinent est celui dont le score est le plus élevé.

Licence Fondamentale Sciences de l'informatique	Classes : LGLSI-2
TP n°3 Techniques d'indexation et recherche multimédia	
Objectifs	Recherche textuelle



avec:

n : nombre de termes du fichier inversé

$w_{i,Q}$: poids du terme i du fichier inversé dans la requête Q (=1 si présent, 0 sion)

$w_{i,D}$: poids du terme i dans le document D

Vous pouvez éventuellement améliorer votre code en implémentant d'autres fonctionnalités, par exemple, vous pouvez implementer d'autre fonctions de similarité (présentées dans le chapitre 4), etc.