

TD – Recherche d'information – Recherche et évaluation

Exercice 1 – Recherche dans le modèle vectoriel

Fournir les résultats des requêtes suivantes pour le corpus de l'exercice 3 des TDs précédents :

Q1 : pomme de terre

Q2 : recherche d'information textuelle

Q3 : domaine du modèle vectoriel

Question 1

Construire les vecteurs des requêtes. Commencer par analyser les requêtes comme les documents (anti-dictionnaire), et utiliser une pondération des requêtes par le tf.idf (similaire aux documents). Calculer les normes de chaque vecteur requête.

Question 2

Calculer les similarités $RSV(d,q)$ en reprenant la formule du cours (transparent 10) sans utiliser l'index inversé, avec chacun des documents. En déduire la réponse du système pour chaque requête.

Question 3

Refaire les calculs de la requête Q3 avec l'utilisation de l'index inversé, en utilisant la description du transparent 12.

Exercice 2 – Evaluation de SRI

Nous réalisons ici une évaluation d'un système de recherche d'information.

Question 1

Supposons que pour une requête Q1 le système de recherche d'information testé renvoie les réponses suivantes:

rang	n° doc	pertinent	rappel	précision
1	588	1		
2	589	1		
3	576	0		
4	590	1		
5	986	0		
6	592	1		
7	884	0		
8	988	0		
9	578	0		
10	985	0		
11	103	0		
12	591	0		
13	572	1		
14	990	0		

Les documents pertinents pour Q1 sont : 572, 588, 589, 590, 592.

Calculer les taux de précision et de rappel du système à chaque réponse et remplir le tableau ci-dessus.

Donner le tableau de résultats normalisé pour cette requête, et en déduire la courbe de rappel/précision.

Question 2

Réaliser le même travail pour la requête Q2, avec les réponses suivantes :

Rang	n° doc	pertinent	Rappel	précision
1	324	X		
2	589	X		
3	528	X		
4	590	X		
5	986	X		
6	592	X		
7	899	X		
8	988	X		
9	578			
10	985			
11	537	X		
12	591	X		
13	772	X		
14	990			

La liste des tous les documents pertinents pour la requête Q2 est : 324, 528, 537, 589, 590, 591, 592, 772, 899, 986, 988.

Question 3

En regardant les courbes, que pouvez-vous déduire de la qualité relative du système pour ces deux requêtes?

Question 4

Donner le tableau global des résultats du système pour les deux requêtes et dessiner le schéma résultant.

Exercice 3 – Comparaison de SRI

Nous voulons comparer deux systèmes de recherche d'information.

Le premier système S1 est celui de l'exercice 9. Le second système, S2, a pour tableau de rappel/précision pour les deux requêtes Q1 et Q2:

Rappel	Précision
0	0.92
0.1	0.88
0.2	0.86
0.3	0.84
0.4	0.80
0.5	0.75
0.6	0.72
0.7	0.70
0.8	0.65
0.9	0.63
1.0	0.61

Tracer les courbes de S1 et S2 sur la même figure.

Analyser les courbes pour en déduire lequel des deux systèmes semble le meilleur.