TD 2 - Modèles d'ordonnancement

Exercice 1 – Modèles d'ordonnancement

On considère trois documents d_1, d_2 et d_3 ainsi qu'un vocabulaire $\{t_1, ...t_10\}$. La répartition des termes du vocabulaire dans les documents d_1, $_2$ et d_3 est la suivante :

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
d_1	0	1	0	2	0	0	0	1	2	0
d_2	1	1	0	0	0	0	1	2	0	3
d_3	1	0	0	0	1	0	1	2	0	2

Q 1.1 Calculer le score de chaque document en utilisant le modèle booléen pour les requêtes suivantes :

- $q_1 = t_2 \wedge t_8 \wedge t_7$
- $q_2 = t_8 \wedge (t_2 \vee t_9)$
- $q_3 = t_{10} \wedge ((t_2 \vee t_9) \vee (t_3))$

Q 1.2 Calculer le score de chaque document en utilisant le modèle vectoriel (similarité cosinus) sur une pondération tf pour les requêtes suivantes :

- $q_1 = t_2, t_8$
- $q_2 = t_8, t_2, t_2$
- $q_3 = t_{10}, t_2, t_7$

Q 1.3 On souhaite utiliser le modèle probabiliste BIM. On sait dans les jugements de pertinence que les documents d_1 et d_3 sont pertinents. Calculer le score de chaque document en utilisant le modèle probabiliste BIM pour la requête $q_1 = t_2, t_8$. Pour cela, il vous faut en amont compléter le tableau d'occurrences qui permet de calculer la probabilité des termes grâce au maximum de vraisemblance.

Exercice 2 – Ré-injection de pertinence

On considère les documents de l'Exercice 2 et la requête $q_1 = t_2, t_8$. On pose également l'hypothèse que les documents d1 et d3 sont pertinents et que le document d2 est non pertinent pour la requête q_1.

Q 2.1 Utiliser la formule de Rocchio pour reformuler la requête, avec alpha=1, beta=0.5 et gamma=0.3.

Q 2.2 Calculer le score des documents pour cette nouvelle requête avec le modèle BIM.