

## Techniques d'indexation

Durée : 1h00 || Calculatrice autorisée || Documents interdits

---

### Questions de cours (6 points)

1. Expliquez brièvement la raison pour laquelle la **distance Euclidienne** est inappropriée pour le calcul de la similarité entre deux documents (vous pouvez, si vous le souhaitez, donner un exemple simple).
2. Expliquez brièvement ce qu'est un **copyField** dans SolR ainsi que son intérêt. Dites également si un champ copyField doit être indexé et stocké ou bien seulement indexé.
3. Expliquez brièvement la différence entre le type **TEXT** et le type **STRING** dans SolR.
4. Un utilisateur fait une recherche dont il sait qu'elle devrait retourner 10 documents. Le système retourne 20 documents : 5 documents parmi ceux attendus par l'utilisateur, en plus de 15 autres non pertinents pour la recherche. Donnez la **précision** et le **rappel** du système pour cette recherche.

### Exercice 1 (6 points)

Nous sup posons que l'on ne connaît que 3 pages du Web :  $A$ ,  $B$  et  $C$ .  $A$  référence  $B$  et  $C$ .  $B$  référence  $A$ .  $C$  référence  $B$ . Trouvez le pageRank de ces pages après 3 itérations.

### Exercice 2 (4 points)

Soit un graphe contenant 3 pages  $A_1$ ,  $A_2$  et  $A_3$  et  $n$  pages  $\{B_1, B_2, \dots, B_n\}$ .  $A_1$  et  $A_2$  se référencent l'une l'autre.  $A_1$  et  $A_3$  se référencent l'une l'autre.  $A_2$  et  $A_3$  se référencent l'une l'autre. Chaque page  $B_i$  référence  $A_1$ . Trouvez le **pageRank** de  $A_1$ ,  $A_2$ ,  $A_3$  et  $B_i$ .

### Exercice 3 (4 points)

La table ci-contre montre une matrice d'incidence avec une ligne par terme, une colonne par document et le  $tf$  (*Term Frequency*) dans chaque cellule. Considérez la requête  $q$  : "SolR, Cosinus". En supposant que le vocabulaire considéré soit  $V = \{\text{SolR}, \text{Index}, \text{Cosinus}\}$ , donnez la **similarité cosinus** entre les documents et la requête  $q$  (sans tenir compte de l' $idf$ , donc seul le  $tf$  entre en compte).

	$d_1$	$d_2$	$d_3$
SolR	0	1	0
Index	2	1	0
Cosinus	0	0	2