

Examen Recherche d'information et Recommandation

Partie Recherche d'information (6 pts)

(4 pts) Question 1 – Indexation et modèle vectoriel

Le modèle vectoriel de recherche d'information représente documents et requêtes par des vecteurs.

- (a) (1 pt) Expliquer le processus général de recherche d'information dans l'environnement Elastic Search utilisé en TP (quels sont les modules, les paramètres, l'enchaînement des étapes).
- (b) (1 pt) Que signifie « indexer » un document ? Décrire la structure de données obtenue et les informations stockées dans l'index (plusieurs solutions sont possibles selon la nature et le format des documents, selon le paramètre du logiciel d'indexation).
- (c) (1 pt) Comment peut-on estimer l'importance d'un mot dans un document ?
- (d) (1 pt) La similarité entre un document et une requête peut être calculée par un produit scalaire ou un cosinus. Quelles sont les conséquences du choix de l'un ou de l'autre ?

(2 pts) Question 2 — Documents structurés et méta-données

Les documents peuvent être encodés en XML ou en JSON.

- (a) Quels sont les avantages et inconvénients du format PDF par rapport à XML ou JSON ?
- (b) Soit l'extrait de document suivant, ici en XML. Expliquez quelles sont les méta-données présentes, leur intérêt et comment elles pourraient être exploitées dans un système de recherche d'information.

```
<title-group>~
  <article-title>Atypical sleep architecture and the autism phenotype</article-title>
</title-group>~
<contrib-group>~
  <contrib contrib-type="author">~
    <name>~
      <surname>Limoges</surname>~
      <given-names>E&#769;lyse</given-names>~
    </name>~
    <xref rid="AFF1">1</xref>~
  </contrib>~
  <contrib contrib-type="author">~
    <name>~
      <surname>Motttron</surname>~
      <given-names>Laurent</given-names>~
    </name>~
  </contrib>~
</contrib-group>~
```

Partie Recommandation (14 pts)

(2pt) Question 1 : Qu'est-ce que la longue traine et le cold start dans les systèmes de recommandation

(2pt) Question 2 : Sur quel principe est basé les systèmes de recommandation de type collaborative filtering et de content-based.

(2pt) Question 3 : Définissez le terme Item et le terme User dans le cadre de la recommandation

(2pt) Question 4 : Quels sont les inconvénients des systèmes de recommandation de type collaborative filtering et ceux de type content based

(3pts) Question 5 : Dans le cas d'un système de recommandation basé sur le collaborative filtering, calculez les ratings manquant pour l'utilisateur 4. Vous utiliserez la mesure de similarité pearson. U correspond à un utilisateur. I correspond à un Item. Expliquez votre démarche

	I1	I2	I3	I4	I5
U1	9	5	7	2	7
U2	1	3	2	?	2
U3	8	?	4	5	?
U4	?	6	?	7	1
U5	6	5	?	?	5

(3pts) Question 6 : Dans le cas d'un système de recommandation basé sur le contenu, indiquer si l'utilisateur aime ou n'aime pas les Items n'ayant pas de like ou dislike. I sont les items et C correspondent aux caractéristiques des items. 1 présence de la caractéristique, 0 absence de la caractéristique. Expliquez votre démarche.

	C1	C2	C3	C4	Like ?
I1	0	1	1	0	dislike
I2	1	0	0	1	like
I3	0	1	1	1	dislike
I4	1	1	1	0	like
I5	1	1	0	1	?