
TP-1 : ANALYSE ET RECHERCHE DES DONNEES

Objectif : être capable de analyser et chercher de l'information contenue dans un jeu de données.

Matériel : Java

ETAPE 1 – RECHERCHE VECTORIELLE

Téléchargez dans Moodle le jeu de données relatif aux avis d'hôtel *TripAdvisor*. Prenez connaissance de la structure du programme Java donné et mettez en place une fonction pour effectuer des recherches vectorielles. Votre programme JAVA doit accepter une requête textuelle, et produire une liste ordonnée des documents comme réponse.

La requête est entrée sous une forme de texte libre. Votre programme doit indexer cette requête comme pour l'indexation de documents, et évaluer cette requête en utilisant le fichier inversé fournit sur Moodle.

Input : Document et Keyword (class Java disponibles sur Moodle) + données déjà sérialisées. (lien sur Moodle)

Output : HashMap<Integer,Double>, ou l'Integer sera l'id du document correspondant à la requête et le double la valeur de classement.

ETAPE 2 – BINARY INDEPENDENCE RETRIEVAL ET RELEVANCE FEEDBACK IMPLICITE

L'objectif du projet lié à ce sujet est d'implémenter un système de recherche basé sur le modèle probabiliste présenté.

Le moteur devra, comme pour la recherche vectorielle, fournir un classement des documents (qui sera basé sur le cosinus).

Input : Document et Keyword (class Java disponibles sur Moodle)+ données déjà sérialisées. (lien sur Moodle)

Output : HashMap<Integer,Double>, ou l'Integer sera l'id du document correspondant à la requête et le double la valeur de classement.