TD 1

**Techniques d'indexation et recherche multimédia **

Groupe : Licence Science Informatique LSI A2 (LSI A2)

Enseignant Nefzaoui Nabil

Voici les exercices que je recommande de faire :

Exercice 1: 17qsts

Qu'est-ce que l'indexation de documents ?

L'**indexation** est une procédure consistant à associer ou à marquer des **documents** ou d'autres contenus avec des informations permettant par la suite de rechercher et récupérer ces **documents**

L'**indexation** est le processus selon lequel le contenu d'un **document** est analysé pour être ensuite reformulé dans une forme permettant d'accéder au contenu et de le manipuler.

Pourquoi indexer un document ?

L'**indexation** a pour principal objectif de rendre accessibles des informations, que l'on repère pour cela au moyen d'**index**. + def

L'**indexation** est le processus selon lequel le contenu d'un **document** est analysé pour être ensuite reformulé dans une forme permettant d'accéder au contenu et de le manipuler.

Comment indexer des documents ?

Dans votre **document** Word, sélectionnez l'un des mots à ajouter à l'**index**, puis tapez Alt + Maj + X. La fenêtre de dialogue Marquer les entrées d'**index** s'ouvre alors.

Quel est le principe de l'indexation ?

L'**indexation** correspond à la représentation d'un texte ou d'un document par un indice ou un mot clé, avec l'aide ou non d'un langage documentaire, en vue d'en faciliter le repérage et la consultation. ...

Il s'appuie le plus généralement sur un répertoire de mots normalisés

Qu'est-ce que l'indexation d'un moteur de recherche ?

L'indexation d'un site ou d'une page web désigne le processus par lequel un robot **d'**

un moteurs de recherche (Google ou autre) va parcourir, analyser, répertorier et classer les pages **d'un** site avant de les afficher dans une page de résultat

Quelles sont les tâches principales réalisées par un moteur de recherche ?

Sur internet, un **moteur de recherche** réalise 3 **principales tâches** : Exploration du web par ses robots (spider, crawler) Indexation de contenus. **Recherche** dans **son** index.

Quels sont les principaux critères d'indexation de Google ?

Voici comment optimiser votre site Web pour les recherches Google

- URLs de mots-clés (noms de domaine à mots-clés) Si une URL comporte un mot-clé, les moteurs de recherche vont l'associer aux recherches correspondantes par pertinence. ...
- Contenus d'actualité ...
- Contenu multimédia.

Quels sont les trois types d'informations conservées par un moteur de recherche ?

Quelles informations sont conservées par les moteurs de recherche...

- les dates de vos visites.
- les mots-clés que vous entrez.
- les liens sur lesquels vous cliquez

Comment indexer des pages sur Google ?

Dans un premier temps, assurez-vous que la **page** en question n'est pas déjà connue de **Google**.

...

Faire indexer votre page par Google

1. Aller dans le menu Exploration / Explorer comme **Google**.
2. Dans le champ du milieu, saisissez l'URL de la **page** à faire **indexer** (sans le nom de domaine)
3. Cliquez sur le bouton Explorer et afficher.

Comment s'appelle le processus qu'utilisent les moteurs de recherche ?

Le **processus s'appelle** le "Crawling", on parle également de "collecte" ou d'"indexation" PageRank est un algorithme utilisé par Google Search pour classer les sites Web dans les résultats de leurs **moteurs de recherche**. PageRank est un moyen de mesurer l'importance des pages de site Web

C'est quoi une page indexée ?

Une **page indexée est** une **page** enregistrée dans les bases de données du moteur de recherche. ... Pour **qu'une page** soit visible, il faut **qu**'elle réussisse l'indexation par les robots des moteurs de recherche.

La structure de la **page** Web ainsi **que les** mots clés sont scrutés par les robots d'indexation

Comment savoir si un site est indexé par Google ?

Pour le **savoir**, inspecter une URL. **Si** ,vous lisez Explorée avec Googlebot pour smartphone, c'**est qu**'elle a été **indexée** à partir de sa version mobile.

Vérifier avec **Google**Search Console avec quelle version de Googlebot votre **site est** analysé.

Pourquoi mon site n'est pas référencé par Google ?

1. Votre **site** web n'est **pas** indexé dans **Google**.
   Si votre **site** web **n'apparait pas sur Google**, c'est peut-être parce que

Si votre **site** web **n'apparait pas sur Google** via cette requête, c'est qu'il n'est **pas** indexé.

Comment savoir quelles urls de mon site sont réellement indexées par Google ?
Traditionnellement, la mesure de l'indexation de ses contenus sur **Google** est faite grâce à la commande «**site**:**monsite**.com» dans **Google**.
Cette requête renvoie l'ensemble des pages **indexées** de ton domaine par le moteur de recherche mais aussi, et surtout le nombre de pages que le moteur associe à ton nom de domaine

Quelle technique Peut-on utiliser pour éviter de solliciter inutilement un moteur de recherche ?
Mots clés « exclus » : lors d'une **recherche** on **peut** également vouloir exclure un mot qui pourrait induire le robot en erreur et apporter des résultats non pertinents. **Pour** cela, il suffit de faire précéder le mot d'un signe – (moins)

Comment ecrire une expression de recherche ?
Chercher une **expression** exacte. « **Recherche** » –Si vous souhaitez faire une **recherche** précise comprenant tous les mots de votre **phrase**, utilisez les guillemets. Attention
, Google va chercher uniquement les pages où votre **expression** apparaît telle que vous l'avez écrite entre guillemets.

Comment savoir le nombre de pages indexées d'un site ?
On peut facilement obtenir un aperçu des **pages indexées** du domaine via la Google Search Console.
Si tu n'utilises pas encore la Google Search Console, tu peux l'activer pour ton **site web** gratuitement (www.google.com/webmasters/tools/? hl=en) en seulement quelques minutes

Exercice 2:

Q1- quelle est la différence entre: «compression avec perte» et «compression sans perte». Donner un exemple de chaque technique pour la compression d'image.

REP

Le taux de compression est plus important si on utilise une compression avec perte mais l'image décompressée n'est pas égal à l'image compressée.

Compression avec perte: jpeg
compression sans perte: codage des répétitions, codage entropique, LZW

Q2- quel est le résultat de l'opération d'indexation dans les deux cas suivants: indexation des textes et indexation des images.

REP

Textes: des termes d'indexation (mots, phrase,..) 1pts
Images: des descripteurs d'images: couleurs, texture, formes

Q3 –«on doit stocker les images sous forme d'un BLOB dans la base de données pour faire une recherche par contenu». Est-ce que cette phrase et correcte. Justifiez

REP

NON.0.5 pts L'approche de recherche dépend de l'indexation et ne dépend pas de la technique dé~~-stockage~~

Exercice 3:

On reçoit souvent des demandeurs pour enseigner des heures supplémentaires en tant que vacataire, on veut réaliser un carnet de contacts qui contient les CV, et les photos des demandeurs. On veut concevoir la base de données selon le cahier de charge suivant:

- Stocker les photos.

- Recherche des noms approximatifs,

- Une recherche avancée dans les CV.

Questions:

1. Donner la requête de création de la table.

2. Donner la requête qui donne les personnes qui ont des compétences en Réseaux sociaux, et pas les réseaux électriques.

3. Donner la requête qui donne les personnes qui ont des compétences en bases de données «MySQL, SQL server, oracle»

4. Donner la requête qui donne les personnes dont leurs noms ressemblent à «Rezki».

5. Soit le code PHP suivant: Compléter le code pour qu'il affiche les images correctement.

```
$cnx = new PDO("mysql:host=localhost;dbname=bdtp",
"user", "p@ssW0");

$req = "SELECT img_id, img_type, img_blob FROM images
WHERE img_id = 1";

$result = $cnx->query ($req);

if($result)

{ $col = $result->fetch();

 if ($col[0])

 {
.....................................................

echo $col[2];

 }

}
```

Réponses:

1- Donner la requête de création de la table.
La réponse utilise
- Blob pour les images
- Fulltext pour les CV
Create table cv
(nom varchar(20),
Prenom varchar(20),
photo blob,
competance text,
fulltext(competance))

2- Donner la requête qui donne les personnes qui ont des compétences en Réseaux sociaux, et pas les réseaux électriques.
La réponse utilise le Boolean Mode
SELECT * FROM cv WHERE MATCH (competance) AGAINST ('Réseaux sociaux - réseaux électriques' in BOOLEAN MODE);

3- Donner la requête qui donne les personnes qui ont des compétences en bases de données «mysql, sql server, oracle»
La réponse utilise le Natural Langauge Mode
SELECT * FROM cv WHERE MATCH (competance) AGAINST ('mysqlsql server oracle' in NATURAL LANGUAGE MODE);

4- Donner la requête qui donne les personnes dont leurs noms ressemblent à «Rezki».
   La réponse utilise le soundex
   Select * from cv where soundex(nom) = soundex('rezki')
5- Soit le code PHP suivant: Compléter le code pour qu'il affiche les images correctement.
   header ("Content-type: " . $col[1]);

Exercice 4:

Quelle est la différence entre **cotation** et **indexation**

1. Définitions **L'indexation** correspond à l'analyse, à l'identification et au renseignement du contenu d'un document dans sa notice bibliographique. Il s'agit d'en identifier les thèmes et postulats principaux afin de rendre le document visible au sein du catalogue de la bibliothèque et d'en faciliter ainsi l'accès au public. L'indexation associe une approche sémantique et sémiologique puisqu'il s'agit de refléter le sens des documents et de les signaler ; ce n'est qu'une fois ce travail effectué qu'un document pourra être repérable d'après son contenu dans le catalogue de la bibliothèque.

La cotation est un processus qui consiste à attribuer une cote à un document. Une cote est constituée de symboles alphanumériques et constitue « l'adresse » du document au sein des rayonnages de la bibliothèque. Outil de localisation, de rangement et de classement, la cote relève de la signalétique documentaire. Elle doit donc être reportée sur le document ainsi que dans sa notice exemplaire. L'attribution de la cote dépend notamment de la nature du document et du plan de classement de la bibliothèque.
L'indexation (indiquer des contenus) traduit et signale le contenu d'un document, tandis que la cotation (localiser des documents) attribue à ce document une adresse physique, la plus judicieuse possible

Exercice 5:

Question de cours

1. Est ce qu'il y a une différence entre un système de gestion de base de données et un système de recherche d'information ? Expliquez

OUI.

2- Un terme qui apparaît dans un seul documents d'un corpus est-il discriminant ou non ? Justifier votre réponse. oui

Un terme qui apparaît dans un document est discriminant car ce terme distingue ou faire la différence entre ce document des autres documents.

3. Dans le modèle vectoriel, à quoi correspondent les axes de l'espace vectoriel

Les axes de l'espace vectoriel correspondent aux termes de la collection de documents.