Université Mohamed Khider - Biskra
Département d'Informatique
Module: BDI & BigData

2017/2018
Niveau: Master 1
Option IA

TP 3

- Nous allons étudier l'utilisation de la bibliothèque open source Lucene, afin d'analyser et indexer les bases de données documentaires.
- Créez un nouveau projet java sous Eclipse ou Netbeans.
- Créer les deux classes: indexer.java et recherche.java dans votre projet
- Indexez tous les documents textuels qui se trouvent dans votre pc (Doc, Docx, Pdf, txt, Xml,…). Analyser ces documents afin d'éliminer les mots vides, et lemmatiser le texte.
- Créez une classe : Termes_Importants, qui permet de récupérer à partir de l'index les 10 meilleurs termes pour chaque document, en calculant leur TF.IDF

- Indexez un site web (https://fr.wikipedia.org/wiki/Algérie) et tous les liens sortants (liens href) de chaque page en restant dans le même domaine (wikipedia.org). d'une manière récursive. (>1000 liens)
- Créez la classe : Termes_Importants pour ce site
- Développer les interfaces nécessaires