



Utilisation de *Lucene*

MC AP8 - RI et Web

Antoine.Doucet@iut3.unicaen.fr
Semestre 4, printemps 2014

Ce document résume les commandes principales dont vous aurez besoin pour installer et utiliser *Lucene*.

Installation :

- Sous linux, récupérez le code source depuis la page de *Lucene*.
<http://lucene.apache.org/>
Le nom du fichier est lucene-3.0.0.tar.gz, pour la version 3.0.0.
- Décompressez le fichier:

```
tar -xvzf lucene-3.0.0.tar.gz
```
- Mettez à jour votre *CLASSPATH* en ajoutant le chemin des fichiers .jar utiles. Si vous aviez placés le fichier lucene-3.0.0.tar.gz dans */home/doucet/riweb/*, vous devez taper la commande suivante (sur une ligne):

```
export CLASSPATH=  
$CLASSPATH:/home/doucet/riweb/lucene- 3.0.0/lucene-  
core-3.0.0.jar:/home/doucet/riweb/lucene-  
3.0.0/lucene-demos-3.0.0.jar
```
- Vous pouvez vérifier que tout est en ordre en tapant:

```
echo $CLASSPATH
```

Construction de l'index :

- En illustration, nous pouvons indexer le code source de lucene, soit les documents qui se trouvent dans */home/doucet/riweb/lucene-3.0.0/src/*, et en supposant que le répertoire actuel est */home/doucet/riweb/*:

```
java org.apache.lucene.demo.IndexFiles index/src/
```
- Cela a pour effet de créer un répertoire 'index/'.
- Dans le cadre du projet, vous placerez tous vos documents dans un même répertoire d'indexation
- N'hésitez pas à faire appel à la documentation de *Lucene* !

Requêtes :

- Rappel : n'hésitez pas à faire appel à la documentation de *Lucene* ! La syntaxe y est décrite en détail.
- Pour pouvoir soumettre des requêtes à cet index, tapez :

```
java org.apache.lucene.demo.SearchFiles
```
- Par exemple :

```
doucet$ java org.apache.lucene.demo.SearchFiles
Enter query:
new
Searching for: new
17 total matching documents
1. src/demo/org/apache/lucene/demo/HTMLDocument.java
2. src/demo/org/apache/lucene/demo/FileDocument.java
3. src/demo/org/apache/lucene/demo/html/Test.java
4. src/demo/org/apache/lucene/demo/IndexHTML.java
5. src/demo/org/apache/lucene/demo/IndexFiles.java
```

```
6. src/demo/org/apache/lucene/demo/html/SimpleCharStream.java
7. src/demo/org/apache/lucene/demo/html/HTMLParser.java
8. src/demo/org/apache/lucene/demo/SearchFiles.java
9. src/demo/org/apache/lucene/demo/html/Token.java
10. src/demo/org/apache/lucene/demo/html/HTMLParser.jj
Press (n)ext page, (q)uit or enter number to jump to a page.
n
11. src/demo/org/apache/lucene/demo/DeleteFiles.java
12. src/jsp/results.jsp
13. src/demo/org/apache/lucene/demo/html/Entities.java
14. src/demo/org/apache/lucene/demo/html/ParseException.java
15. src/demo/org/apache/lucene/demo/html/Tags.java
16. src/demo/org/apache/lucene/demo/html/TokenMgrError.java
17. src/demo/org/apache/lucene/demo/html/HTMLParserTokenManager.java
Press (p)revious page, (q)uit or enter number to jump to a page.
q
Enter query:
new AND delete
Searching for: +new +delete
3 total matching documents
1. src/demo/org/apache/lucene/demo/DeleteFiles.java
2. src/demo/org/apache/lucene/demo/IndexHTML.java
3. src/demo/org/apache/lucene/demo/IndexFiles.java
Press (q)uit or enter number to jump to a page.
```