

Recherche de documents

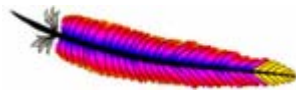
Laboratoire N° 4

A. Objectifs

Ce laboratoire a pour but de découvrir l'outil d'indexation et de recherche Open-Source *Apache Lucene* et de l'intégrer dans le framework de RIM.

Les points étudiés dans ce laboratoire seront :

1. Découverte des caractéristiques de *Lucene*
2. Indexation et recherche dans la collection CACM avec *Lucene*



B. Références

Cours «Recherche d'Information Multimédia » de Nastaran Fatemi.

Site Web du projet *Lucene* : <http://lucene.apache.org/>

C. Rapport

A remettre : Vos sources et un rapport comportant les objectifs, la description des démarches adoptées, les réponses aux questions, l'analyse des résultats obtenus et une conclusion personnelle.

A remettre : Au plus tard au début de la séance du 12 Décembre 2007.

D. Donnée

1. Découverte de Lucene

Pour commencer, rendez-vous sur le site (<http://lucene.apache.org/>) et prenez connaissance des descriptions et de la documentation de *Lucene*, de manière à pouvoir répondre aux questions suivantes :

- Qu'est-ce que *Lucene* ? Est-ce destiné à *Java* uniquement ?
- Dans *Lucene*, à quoi sert un *Analyzer* ? Peut-on écrire son propre *Analyzer* ?
- Comment un document est-il représenté dans *Lucene* ?
- Quelles sont les possibilités offertes par *Lucene* pour la formulation d'une requête ?
- Expliquez brièvement comment *Lucene* gère ses index (en termes de fichiers) et expliquez l'intérêt de cette façon de procéder.

2. Indexation et recherche avec Lucene

Dans ce laboratoire vous devrez écrire une nouvelle version de votre moteur de recherche, à l'aide de *Lucene*. Pour ce faire, vous devrez reprendre les classes de votre laboratoire n°3 et **remplacer votre précédent code par des appels aux classes de *Lucene***.

Pour vous simplifier la tâche, vous êtes fortement encouragés à vous inspirer du code source donné en exemple dans l'API de *Lucene*, sur la page Web suivante :

http://lucene.zones.apache.org:8080/hudson/job/Lucene-Nightly/javadoc/overview-summary.html#overview_description

Votre travail consistera essentiellement à comprendre cet exemple et à en intégrer les différentes parties au bon endroit dans le Framework de RIM.

E. Indications

- Vous n'êtes pas obligés de fournir d'implémentation pour les méthodes *searchTerm()* et *searchDocument()* du *Retriever*. Retournez simplement la valeur *null*.
- Faites attention à bien placer les opérations d'indexation dans l'*Indexer* et les opérations de recherche dans le *Retriever*.
- En cas de doute, n'hésitez pas à poser des questions à l'assistant ou au professeur.