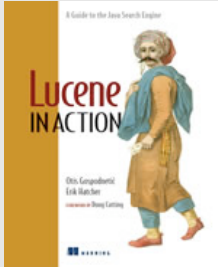




zenika
ARCHITECTURE INFORMATIQUE

www.zenika.com

Repartez avec le livre



Durée
2 jours

Répartition
50% théorie
50% pratique

Pré-requis
Connaissance de Java

Public
Architecte
Développeur
Chef de projet

Tarif (déjeuners inclus)
1200 € (HT)
Tarif (1 mois avant)
1100 € (HT)

Lieu
Paris 9ème

Sessions 2009
29 au 30 janvier
5 au 6 mars
18 au 19 mai
2 au 3 juillet
10 au 11 septembre
15 au 16 octobre

Intra-entreprise sur
demande

Inscription et
renseignements
+33(0)1.45.26.19.15
training@zenika.com
www.zenika.com

Lucene

Optimiser les recherches avec Lucene

Objectifs

- Apprendre à indexer et à chercher des documents à l'aide de Lucene
- Intégrer un moteur de recherche à une application profitant de Lucene pour accéder rapidement à un contenu
- Améliorer les temps de recherche grâce à une API performante

Contenu

Introduction

- Lucene : l'API de recherche plein-texte et d'indexation des informations
- Accéder rapidement à l'information désirée
- Ajouter un Google-like dans votre application sur vos documents !

Les bases

- Notion de token, index, parser
- Les packages et classes clés de l'API
- Une utilisation basique : Recherche simple dans une chaîne de caractère

Les packages

- Analysis : Conversion de texte en tokens
- Document : Modèle de représentation d'un document
- Index : Gestion des index
- Search : Effectuer la recherche dans l'index
- Store : Stockage de l'index, en mémoire ou sur disque

Utilisation de Lucene

- Définir la base des documents à indexer
- Choisir les « champs » accessibles des documents pour la recherche
- Stocker les index
- Générer la requête de recherche à partir d'une expression plein-texte
- Obtenir les résultats à partir de l'index

Indexer vos documents

- Support de tous les documents dont le texte est lisible : HTML, XML, PDF, Word, Excel, Visio...
- Indexer un site web en HTML pur et effectuer une recherche sur son contenu

Séparation des fonctionnalités

- Parsing - Indexation - Recherche
- Toutes les fonctionnalités sont séparées afin de permettre une utilisation cross-platform de Lucene

Intégration

- Lucene existe dans plusieurs autres langages : Delphi, Perl, C#, C++, Python, Ruby et PHP

Les fonctionnalités intéressantes

- La puissance de la recherche plein-texte : wildcards, opérateurs...
- La recherche dans la recherche : utilisation de filtres pour raffiner une recherche
- Optimisation : l'usage des index comme facteur de réduction du temps de recherche
- Créer un index en mémoire et un autre en système de fichier
 - Comparer les temps de recherche
 - Utiliser des wildcards pour profiter de la recherche plein-texte

Fonctionnalités avancées

- Recherche dans plusieurs index (parallèle ou séquentielle)
- Retrouver la position du terme recherché dans le document cible (pour une précision absolue)
- Caching des recherches pour une optimisation des recherches fréquentes

Les projets périphériques à Lucene et leur

possibilités

- Nutch : moteur de recherche complet utilisant Lucene comme API d'indexation et de recherche
- Solr : Un serveur de recherche complet s'appuyant sur Lucene et Tomcat
- Compass : mapping des objets pour les rendre « cherchables » à l'aide de Lucene