Tél : +352 27 39 351
Fax : +352 27 39 52 68
Email : formation@oxiane.lu

Apache Solr

Implémenter un moteur de recherche « scalable » avec SOlr

La Recherche d'Information (RI) consiste à trouver des ressources (habituellement des documents) de nature non structurée (habituellement des textes) qui répondent à un besoin d'information parmi une large collection stockée (sur ordinateur).
La recherche d'information comporte deux phases : l'indexation et la recherche.
L'indexation est le processus qui consiste, à partir des données sources, à construire la structure de données (index inversé) qui va faciliter l'accès à l'information.
La phase de recherche consiste à trouver les documents pertinents à partir de l'index inversé.
Pour mettre en place un moteur de recherche, il est capital de comprendre les fondamentaux de cette filière.
Bien choisir les librairies et solutions de développement est tout aussi primordial.
A l'issue de cette formation, vous serez au fait des fondamentaux de la recherche d'information.
Vous comprendrez les principaux composants du module d'indexation d'un moteur de recherche Web ou d'entreprise.
Nous vous montrons les structures de données utilisées, les infrastructures nécessaires.
Nous mettons un accent sur les techniques pour améliorer la pertinence des résultats et l'expérience de l'utilisateur.
Nous vous donnons les clés pour aller au delà de la simple recherche par mots clés en capitalisant sur les connaissances de votre domaine d'activité.

Détails

- Code : DB-SOLR
- Durée : 3 jours (21 heures)

Public
- Architectes
- Developpeurs

Pré-requis
- Expérience de développement

Objectifs
- Démarrer un projet de recherche d'information
- Modéliser les unités d'indexation
- Développer les services de recherche
- Analyser les performances de votre moteur de recherche
- Déployer le moteur de recherche suivant différentes topologies

Programme

Fondamentaux de la Recherche d'Information (RI)
- Définitions
- RI Web vs RI Entreprise
- concepts de base
- Structure et construction de l'index
- Modèle booléenne de recherche d'information
- Recherche ordonnée
- Modèle vectoriel de recherche d'information

Indexation du Web: un état de l'art
- Historique de l'innovation des principaux moteurs de recherche
- Organisation des documents du Web
- Construction du dictionnaire des termes
- Stockage de l'index
- Répondre à une requête de l'utilisateur
- Mise à l'échelle du moteur de recherche
- Le cas Google Search Engine

Solutions open source Lucene/Solr
- présentation de la librairie Apache Lucene
- présentation du serveur Apache Solr

Indexation avec Solr
- Mise en place d'un projet Solr avec Apache Maven
- Structure du répertoire d'installation de Solr
- Comprendre le concept Solr Core
- Les fichiers de configuration
- Schéma des documents et analyse des textes
- Les modes de communication avec Solr
- Le framework Data Import Handler (DIH) de Solr
- Indexer les fichiers avec Solr Cell

Recherche avec Solr
- Les paramètres de recherche
- La syntaxe des requêtes
- Parseur de requête Lucene vs Parseur de requête Dismax
- Recherche Géospatiale
- Influencer la pertinence des résultats
- Recherche par facettes pour une meilleure expérience de l'utilisateur
- Les composants Highlight et MoreLikeThis
- Aller au delà de la recherche par mots clés

Tél : +352 27 39 351
Fax : +352 27 39 52 68
Email : formation@oxiane.lu

Mise à l'échelle de Solr

- Evaluer les performances de Solr avec SolrMeter
- Optimiser une instance unique de Solr
- Passer à plusieurs serveurs avec Solr Replication et/ou Solr Cloud

Modalités

- Type d'action :Acquisition des connaissances
- Moyens de la formation :Formation présentielle – 1 poste par stagiaire – 1 vidéo projecteur – Support de cours fourni à chaque stagiaire
- Modalités pédagogiques :Exposés – Cas pratiques – Synthèse
- Validation :Exercices de validation – Attestation de stages