

## 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
  - Développeurs

- 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

##### Gestion des cores

- Core schemaless
- Options de démarrage
- Chemin personnalisé pour la localisation des cores
- Configurations prédéfinies configSets
  - Création core
- Utilisation de l'API REST (endpoint \_introspect)
- Création core avec des chemins personnalisés

##### Configuration d'un core

- Configuration en mode schemaless et autoguessing
- Création d'un schéma personnalisé

##### Indexation

- API - Update Request Handler
  - json
  - xml
- Utilitaire post
  - xml
- Utilitaire *tika*

##### Data Manipulation Language

- Recherches structurées - Les filtres
  - Recherches basiques de type filtre - 1
  - Recherches basiques de type filtre - 2

- Recherches *full-text*
  - Recherches sur un champ unique
  - Recherches multichamps - un texte spécifique à chaque champ
  - Recherches multichamps - un texte commun à tous les champs
- Proximité
  - Phrase match
- Partial matching (Part One)
  - Chargement du jeu d'essai
  - Partial matching sans modification de l'indexation
  - Partial matching avec le type natif search\_as\_you\_type
- Spécificités du langage
  - Préambule
  - Mapping inféré
  - Mapping personnalisé
- Spellcheckers
  - Chargement des données
  - Configuration du handler /select et requêtes
  - Configuration du handler /spellcheck et requêtes
- Suggesters
  - Chargement des données
  - Implémentation individuelle
  - Implémentation groupée
  - Combinaison d'un search et d'une suggestion

- Agrégations
  - Chargement des données
  - Requêtes
- Cas concret : Oxiane Trainings

### Solr Cloud

- Zookeeper embarqué
- Zookeeper standalone

### Test de charge

- Installation
- Interprétation des résultats
  - View Results Tree
  - Aggregate Graph
  - Summary report

### Modalité

- Stage pratique en présentiel
- Stage pratique en distanciel
- Nombre de stagiaires minimum : 4
- Nombre de stagiaires maximum : 10

### Méthodes pédagogiques

- Exposés
- Cas pratiques
- Echanges d'expérience

### Profils des intervenants

- Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

### Modalités d'évaluation

- Evaluation des acquis de la formation par le biais de cas pratiques et/ou mises en situation.
- Attestation de formation remise à chaque participant.

### Démarche qualité

- Questionnaire d'évaluation de satisfaction à chaud complété par chaque participant à l'issue de la formation.

### Moyens pédagogiques

- Salle équipée de PC (1 poste par stagiaire), vidéo-projecteur.
- Espace de pause.

Dernière mise à jour le 27/05/2021