



Projet ORI-OAI

Outil de **R**éférencement et d'**I**ndexation
Réseau de portails **OAI**

Moteur d'indexation ORI-OAI-indexing

Formation technique niveau II :
Configuration avancée (comment naviguer dans les fichiers de configuration)





Sommaire

Points d'entrée principaux

Gestionnaire de liens / crawler web

Points d'entrée principaux

- Voir doc « Navigation dans ... configuration »
- properties/
 - ../configIndexing.xml : configuration générale
 - ../liusConfig.xml : déclaration des MD à indexer
- **+ d'infos sur wiki.ori-oai.org**
 - <http://wiki.ori-oai.org/pages/viewpage.action?pageId=54821207>

Gestionnaire de liens / crawler web

- Idée :
 - renseigner l'administrateur sur la disponibilité du document distant (lien brisé ou non)
- via l'IHM de l'indexing
 - onglet "Gestion des liens et crawler web" : lancement manuel
- via properties/configIndexing.xml
 - **<scheduleCrawler>** : planification du gestionnaire
 - **<sendMail>** : envoi d'un mail à l'administrateur
 - ...

■ Idée :

- Permettre d'enrichir la fiche de MD avec une indexation plain texte
 - ❶ contenu textuel dans MD "full Text"
 - ❷ indexation de cette MD
- ➔ Requête possible sur cette MD depuis le moteur de recherche
- **Dépendant** du gestionnaire de liens
 - lancement par le gestionnaire de liens

- via properties/configIndexing.xml
- **<!-- Gestionnaire de liens ... Crawler web -->**
 - <crawlingAuthorized> : activation du crawler
 - <repository ... > : profondeur, types Mime
 - <doNotCrawl> : entrepôts à ne pas crawler
- **+ d'infos sur ori-oai.org**
 - <http://wiki.ori-oai.org/pages/viewpage.action?pageId=54821209>

- Copyright © 2012 Consortium ORI-OAI
 - Ce document peut être copié et distribué dans son intégralité, sans modification, retrait ou ajout.
 - Tout usage commercial est interdit.
 - L'utilisation de ce document dans un cadre de formation collective est soumise à l'approbation explicite et préalable de ses auteurs.