

Utiliser ORI-OAI-indexing (Indexation)

Utilisation du module ORI-OAI-indexing

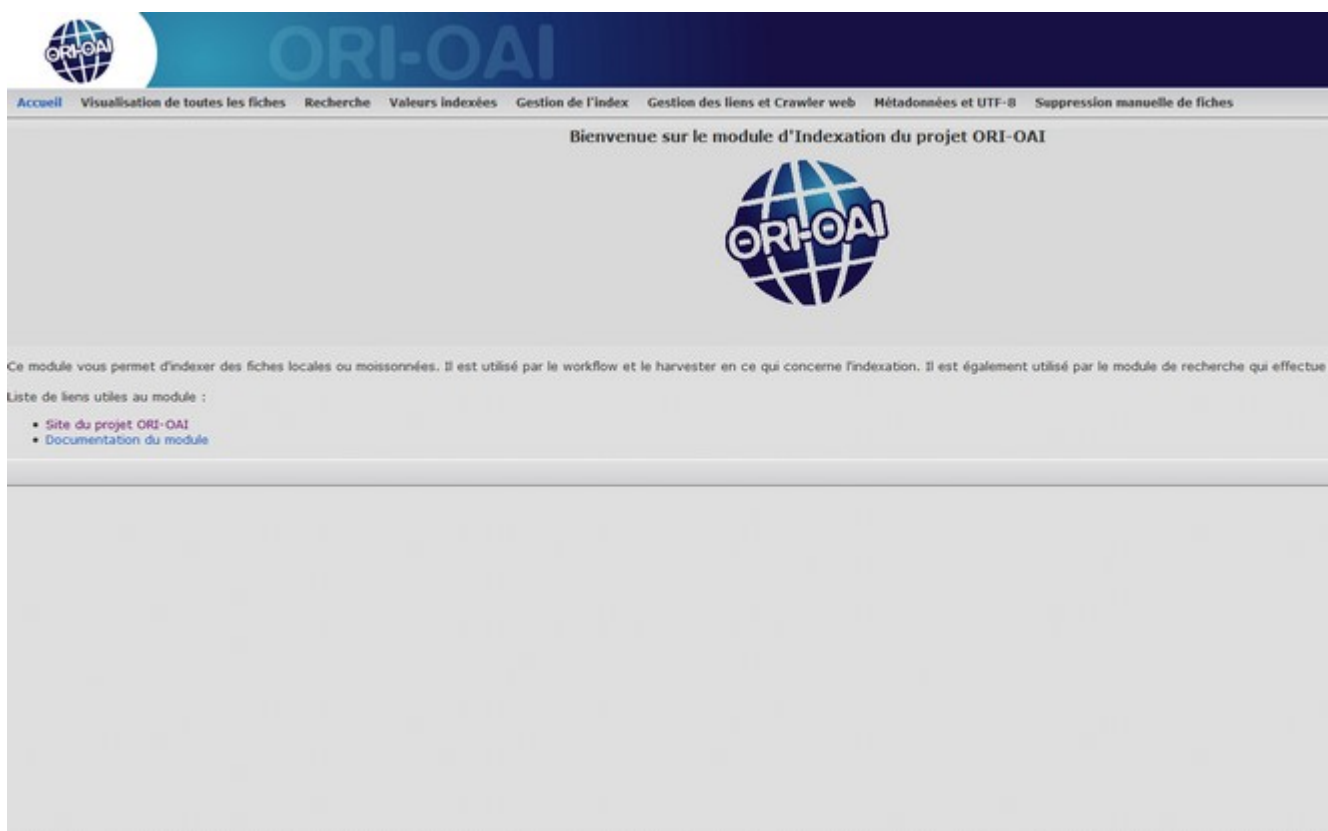
Ouvrir un navigateur web et, suivant l'installation effectuée, aller à l'adresse du module.
Si le module a été installé selon la documentation proposée avec un Tomcat par module, cette adresse devrait ressembler à :

```
http://NOM_SERVEUR_ORI:8182/ori-oai-indexing
```

Quelques fonctionnalités permettent de consulter l'index depuis le navigateur.

Accueil

La page se présente de la manière suivante :



Cette page indique que le module a été correctement déployé et qu'il est prêt à être utilisé. Cette page donne également le lien vers le site du projet ORI-OAI ainsi que le lien vers la documentation du module.

Visualisation des fiches

Cet onglet montre toutes les fiches présentes dans l'index. Voici une copie d'écran de cette page :



Le champ crawled indique l'état de crawling de la fiche. Différentes valeurs sont possibles pour ce champ :

- **no** : Aucune tentative de crawling n'a été faite sur cette fiche
- **failed_n** : Le crawling a échoué sur cette fiche pour la n^{ème} fois
- **unreachable** : Le lien à crawler est injoignable
- **forbidden** : Il n'est pas possible de crawler le lien
- **some_unreachables** : La fiche a été crawlée avec succès mais certains liens n'ont pas pu être joints
- **yes** : La fiche a été crawlée avec succès

Pour de plus amples informations, veuillez consulter l'aide du module.

Fiches de l'index - 1 à 50 sur 536

Identifiant	Deadlink	Crawled	Namespaces des formats indexés
oai@orwww.unit.eu/unit-ori-wf-1-5	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-25	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-27	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-33	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-35	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-37	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-39	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-29	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-31	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-47	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-49	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-59	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-77	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-79	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-81	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-103	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-109	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-193	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-209	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-211	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-213	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-215	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-227	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-239	unknown	no	http://ftsc.ieee.org/xsd/LOM
oai@orwww.unit.eu/unit-ori-wf-1-243	unknown	no	http://ftsc.ieee.org/xsd/LOM

Il s'agit d'un tableau renseignant sur l'identifiant de chaque fiche indexée, l'état de ses liens, son état de crawling, ainsi que les formats dans lesquels l'identifiant a été indexé. Il faut noter que les identifiants ne sont pas triés dans l'ordre alphabétique mais dans l'ordre dans lequel ils ont été indexés, c'est-à-dire dans l'ordre croissant de leur position dans l'index Lucene.

Au dessus du tableau, les différentes valeurs de la colonne "Crawled" sont expliquées. [A noter : *forbidden*, pour ressources d'URL, de types MIME ou d'entrepôts dont le crawling n'est pas souhaité (configurable) ; *unreachable*, pour ressources dont l'URL est inaccessible au moment du crawling ou plus valide (typiquement erreur 404) ; *some_unreachables*, pour ressources dont l'URL est accessible, mais contenant elles-mêmes des liens qui n'ont pu être crawlés]

Pour augmenter les performances de l'affichage de cette page, seules 50 fiches sont présentées par page. Vous trouverez en bas de la page des liens "page suivante" et/ou "page précédente" pour naviguer entre les différentes pages.

Par ailleurs les identifiants des fiches sont des liens cliquables. En cliquant sur l'un de ces identifiants, on peut en visualiser toutes ses données. Pour retourner à la page de visualisation de toutes les fiches, il suffit de cliquer sur l'onglet "Visualisation de toutes les fiches" ou de cliquer sur le bouton "Page précédente" de votre navigateur. La page se présente comme suit :



Visualisation de la fiche

oai@oriwww.unit.eu@unit-ori-wf-1-5	
//lom:general/lom:title /lom:string[starts-with{@language,'fr'}]	Quels supports de diffusion utiliser pour quels contenus de télévision numérique ? (TIM région 2 Q9)
//lom:general/lom:language	fr
//lom:general/lom:description /lom:string[starts-with{@language,'fr'}]	Bien choisir le support de diffusion en fonction du contenu à véhiculer.
//lom:general/lom:keyword /lom:string[starts-with{@language,'fr'}]	support de diffusion
//lom:general/lom:keyword /lom:string[starts-with{@language,'fr'}]	compression vidéo
//lom:general/lom:keyword /lom:string[starts-with{@language,'fr'}]	codage
//lom:general/lom:keyword /lom:string[starts-with{@language,'fr'}]	compression d'images
//lom:general/lom:keyword /lom:string[starts-with{@language,'fr'}]	télévision numérique
//lom:general/lom:keyword /lom:string[starts-with{@language,'fr'}]	image numérique
//lom:general/lom:keyword /lom:string[starts-with{@language,'fr'}]	traitement d'images
//lom:general/lom:keyword /lom:string[starts-with{@language,'fr'}]	gestion mémoire
//lom:general/lom:keyword /lom:string[starts-with{@language,'fr'}]	télécommunications
//lom:general/lom:keyword	...

La partie gauche présente les métadonnées indexées et la partie droite montre leur contenu. Ceci permet de vérifier notamment que la fiche est bien indexée et que toutes les métadonnées sont dans l'index.

Recherche

Cette page permet de lancer une requête Lucene. Pour se faire il suffit d'entrer votre requête dans le formulaire et de cliquer sur "Lancer la recherche". Voici une copie d'écran de cette page :



Recherche dans l'index

Cette page vous permet d'effectuer une recherche dans l'index. Pour cela il vous suffit d'entrer une **requête Lucene**. Une requête se construit de cette manière : "Champ_Lucene:(Valeur_a_recherche)".

Dans Ori-Oai-Indexing on distingue deux types de champs Lucene :

- **Les Métadonnées Annexes :**
 - md-ori-oai-id représente l'identifiant de la fiche,
 - md-ori-oai-namespace correspond au format,
 - md-ori-oai-repository indique l'entrepôt,
 - md-ori-oai-datestamp renseigne sur la date.
- **Les Métadonnées de la fiche :** Il s'agit ici du xpath qui a été encodé UTF-8. Pour encoder un xpath, il suffit d'utiliser `java.net.URLEncoder.encode(xpath_a_encoder, "UTF-8");`

Entrez votre requête ci-dessous :

%2F%2F%3Atitle:(test)

Lancer la recherche

Il faut noter que le nom des métadonnées doit être encodé en UTF-8. Ainsi "//dc:title" devient "%2F%2Fdc%3Atitle". Ceci a pour objectif de ne pas utiliser le caractère ":" qui est déjà sollicité dans une requête Lucene pour faire la séparation entre la métadonnée et le contenu recherché.

La copie d'écran suivante montre les résultats d'une recherche :

Identifiant	Namespaces des formats indexés
oa@orwww.unit.eu/unit-ori-wf-1-1633	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-1635	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-1845	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-247	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-273	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-3049	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-3053	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-3061	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-3083	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-341	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-3445	http://ltsc.ieee.org/xsd/LOM
oa@orwww.unit.eu/unit-ori-wf-1-547	http://ltsc.ieee.org/xsd/LOM

Les résultats se présentent sous la forme d'un tableau contenant l'identifiant de la fiche ainsi que les formats indexés. Là encore l'identifiant est un lien cliquable qui aboutit à la visualisation de la fiche.

Valeurs indexées

Cette page permet de retrouver toutes les valeurs indexées d'un champ de l'index. Pour connaître les valeurs indexées il suffit d'entrer le nom du champ dans la zone de saisie puis de cliquer sur le bouton "Lancer la recherche".



Valeurs indexées d'un champ

Cette page vous permet de retrouver toutes les valeurs indexées en fonction d'un champ donné. Il peut s'agir d'un champ provenant de la fiche descriptive ou d'un autre champ tel que l'identifiant de la fiche triée par ordre alphabétique.

Entrez le xpath désiré ci-dessous :

//dc:title

Lancer la recherche

La copie d'écran suivante montre le résultat d'une recherche pour le titre du format LOM :



Valeurs indexées d'un champ

Valeurs

comment faire confiance à un compilateur ? (serie : colloquium jacques morgenstern)
optique des lasers et faisceaux gaussiens (optique pour l'ingenieur : lasers et optique non lineaire)
"parlez-vous chinois ?" modules d'apprentissage du chinois en ligne
"parlez-vous chinois ?" pour iphone
accidents nucleaires : exemples de three mile island et tchernobyl (filere linguistique preparatoire aux etudes en france - flépe)
acquérir et pretraiter des images (le traitement d'image en questions, tm region 1)
advection d'un scalaire et caracteristiques (ondes de surface et ressauts)
aide-memoire de probabîtes (serie : calcul des probabîtes)
algebre lineaire
algorithmes et programmation
algorithmes pour le calcul scientifique (serie : cours polytech'ille)
algorithmique
algorithmique et techniques de base des systemes repartis (serie : enseignements en sar)
algorithmique et techniques de base des systemes repartis : introduction (serie : enseignements en sar)
analyse d'images (serie : vision par ordinateur)
analyse de donnees : aide-memoire
analyse de risque incendie sur un etablissement recevant du public (cyberisques : cyber ingenierie des risques en genie civil)
analyse de risque quantitative pour les structures petrolieres - projet tuteur (cyberisques : cyber ingenierie des risques en genie civil)
analyse de risques : identification et estimation, partie 1 (cyberisques : cyber ingenierie des risques en genie civil)
analyse de risques : identification et estimation, partie 2 (cyberisques : cyber ingenierie des risques en genie civil)
analyse et programmation langage ada
analyse mathematique, modeles et simulations numeriques (serie : colloquium jacques morgenstern)
annales d'algebre et analyse des donnees (serie : enseignements de statistique en biologie)
annales d'analyse de donnees (serie : enseignements de statistique en biologie)
annales de biologie et modelisabon (serie : enseignements de statistique en biologie)
annales de biostatistique (serie : enseignements de statistique en biologie)
annales de stabsitique (serie : enseignements de stabsitique en biologie)
annales de sujets d'examen (serie : enseignements en sar)
annales des epreuves de la section applications des mathematiques et de l'informatique a la genomique (serie : enseignements de statistique en biologie)
appel de procedure a distance : corba (serie : enseignements en sar)
applications a la mesure sans contact et aux cnd optiques (optique pour l'ingenieur : interferences et diffraction)

Gestion de l'index

Cette page est utile lorsque vous souhaitez réinitialiser l'index ou lorsque vous faites des recherches. Les caches permettent de ne pas trop solliciter l'index. Si une requête est lancée plusieurs fois, la première servira à mettre en cache les résultats de la requête. Les suivantes ne feront que consulter le cache. Trois boutons sont présents :

- Réinitialiser l'index : cela vous permettra de supprimer l'index et de le recréer à vide. Le cache sera lui aussi vidé. **[Attention !** Le

bouton "Réinitialiser l'index" recrée l'index vide, mais ne réindexe pas les fiches. Pour réindexer les fiches reportez-vous à la page consacrée à la gestion de l'index.]

- Optimiser l'index : ce bouton lancera l'optimisation de l'index. Depuis la version 1.6, l'index s'optimise automatiquement grâce à un processus qui se lance régulièrement. Un log en INFO indique le moment où est lancée l'optimisation automatique. Mais pour devancer cette phase vous pouvez utiliser ce bouton.
- Vider le cache : cela supprimera toutes les entrées du cache. Cette fonctionnalité est utile notamment si l'index a été modifié après avoir effectué des recherches. Vider le cache permettra de prendre en compte ces modifications dans les résultats des recherches.

Un message de confirmation apparaîtra alors et ce sur chacun des deux boutons.



Depuis la version 1.4, arrêter et relancer le Tomcat du module d'indexation ne permet plus de vider le cache.

La page de gestion de l'index se présente comme suit :

Accueil Visualisation de toutes les fiches Recherche Valeurs indexées **Gestion de l'index** Gestion des liens et Crawler web Métadonnées et UTF-8 Suppression manuelle de fiches

Gestion de l'index

Cette page vous permet de vider l'index. Ce dernier sera supprimé puis sera recréé à vide. Pour reconstituer votre index, il vous faudra ensuite lancer les procédures de réindexation depuis ori-oai-wordpress. Cette opération videra également les caches.

Pour vider l'index, cliquez sur le bouton ci-dessous :

Vous pouvez également optimiser votre index dans le but d'améliorer les performances à la recherche. Cette fonctionnalité est surtout utile après avoir terminé le crawling.

Pour optimiser l'index, cliquez sur le bouton ci-dessous :

Vous pouvez aussi simplement vider le cache contenant le nombre de résultats. La durée de vie des éléments du cache peut être modifiée en éditant le fichier ehcache.xml situé dans le dossier properties.

Pour vider le cache, cliquez sur le bouton ci-dessous :

Gestion des liens et crawler web

Cet onglet permet de lancer le gestionnaire de liens ainsi que le crawler web. La copie d'écran suivante montre la page du gestionnaire :



Gestion des liens et crawler web

Cet outil vous permettra de vérifier si le lien inclus dans la fiche descriptive est valide ou non. Tant que le lien n'a pas été vérifié une métadonnée annexe appelée md-ori-oai-state-link contiendra la v

Il est possible de définir dans le fichier configIndexing.xml le nombre de tentatives avant de considérer que le lien est mort. A chaque lancement du gestionnaire on comptabilise une seule tentative. L'attribut deadlink_x où x représente la x-ième tentative. Si à la dernière tentative le lien n'est toujours pas actif, la valeur de la métadonnée devient deadlink.

Important : Si le lien est actif, vous pouvez lancer la crawler web sur la fiche ce qui indexera le plein texte associé à la fiche descriptive. Pour se faire il suffit de mettre la paramètre crawlingAuthorized à true. Le crawling de la fiche se lancera automatiquement.

Pour lancer le gestionnaire de liens, cliquez sur le bouton ci-dessous.

Lancer le gestionnaire de liens

Pour lancer le gestionnaire de liens manuellement il suffit de cliquer sur le bouton "Lancer le gestionnaire de liens".

Au bout de quelques secondes une nouvelle page indiquera la progression du gestionnaire. Cette page se rafraichit toutes les 10 secondes mettant ainsi à jour l'indicateur d'avancement du gestionnaire.



Pour que le crawler web se lance automatiquement après le passage du gestionnaire, il faut positionner la valeur INDEXING_CRAWLER_FULL_TEXT du fichier commons-parameters.properties du quick-install à true. Si vous avez choisi une installation manuelle, il faut remplir la balise "crawlingAuthorized" du fichier configIndexing.xml à true.



- 1 - Quitter cette page ne stoppe pas le gestionnaire.
- 2 - Un bouton vous permet de stopper le gestionnaire de liens. Une attente de quelques secondes est alors nécessaire pour que les liens en cours d'analyse arrivent à terme correctement.

Métadonnées et UTF-8

Cette page vous permet d'encoder ou de décoder facilement vos xpaths en UTF-8. Ceci est utile lorsque vous souhaitez ajouter un nouveau xpath dans le fichier liusConfig.xml. La page se présente comme suit :



Métadonnées et UTF-8

Cette page permet de vous faciliter le remplissage du fichier `liusConfig.xml`. En effet il vous suffit de donner le xpath que vous souhaitez ajouter et de cliquer sur le bouton "Lancer l'opération" et vous obtenez la valeur correspondante.
Cette donnée correspond au nom du champ utilisé dans le fichier de configuration de LIUS.
Il est également possible de réaliser l'opération inverse en sélectionnant l'option "decode" de la liste déroulante.

Veuillez renseigner les champs suivants : encode ▾

Lancer l'opération

Pour encoder un xpath en UTF-8, entrez sa valeur dans le formulaire (exemple : `//dc:title`) et cliquez sur le bouton "Lancer l'opération". La copie d'écran suivante présente le résultat :



Métadonnées et UTF-8

La valeur correspondante est : `%2F%2Fdc%3Atitle`

[Retour](#)

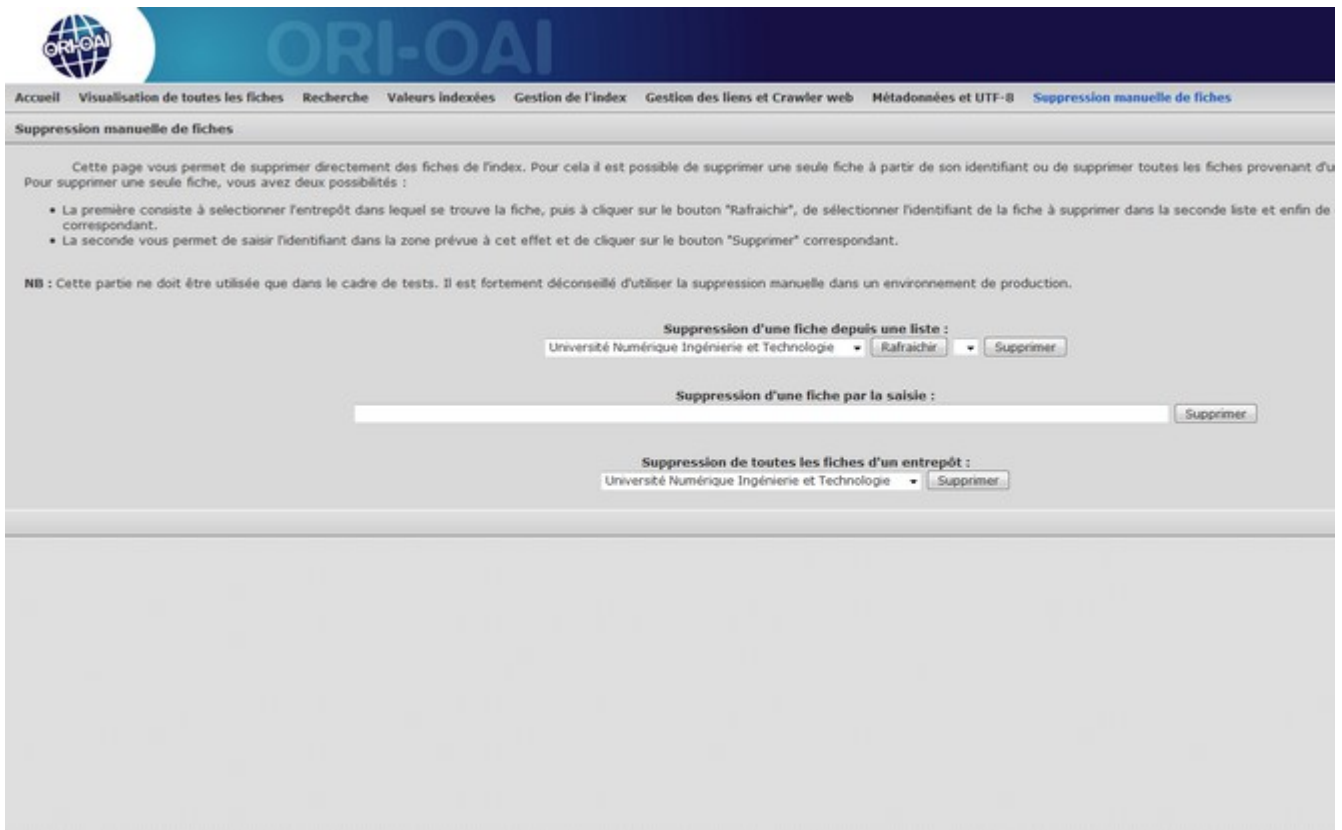


Il est possible d'effectuer l'opération inverse en sélectionnant "decode" et en donnant le xpath encodé.

Suppression manuelle de fiches

Cette page permet de supprimer manuellement des fiches de l'index. Il est possible de supprimer toutes les fiches d'un entrepôt ou de

supprimer une seule fiche, soit en saisissant directement son identifiant ou en sélectionnant son entrepôt puis en cliquant sur "Rafraichir" et enfin en sélectionnant son identifiant. Voici une copie d'écran de cette page :



The screenshot shows the 'Suppression manuelle de fiches' page of the ORI-OAI system. At the top, there is a navigation bar with links: Accueil, Visualisation de toutes les fiches, Recherche, Valeurs indexées, Gestion de l'index, Gestion des liens et Crawler web, Métadonnées et UTF-8, and Suppression manuelle de fiches. The main content area is titled 'Suppression manuelle de fiches' and contains the following text:

Cette page vous permet de supprimer directement des fiches de l'index. Pour cela il est possible de supprimer une seule fiche à partir de son identifiant ou de supprimer toutes les fiches provenant d'un entrepôt. Pour supprimer une seule fiche, vous avez deux possibilités :

- La première consiste à sélectionner l'entrepôt dans lequel se trouve la fiche, puis à cliquer sur le bouton "Rafraichir", de sélectionner l'identifiant de la fiche à supprimer dans la seconde liste et enfin de cliquer sur le bouton "Supprimer".
- La seconde vous permet de saisir l'identifiant dans la zone prévue à cet effet et de cliquer sur le bouton "Supprimer" correspondant.

NB : Cette partie ne doit être utilisée que dans le cadre de tests. Il est fortement déconseillé d'utiliser la suppression manuelle dans un environnement de production.

The interface includes three main sections for deletion:

- Suppression d'une fiche depuis une liste :** A dropdown menu showing 'Université Numérique Ingénierie et Technologie', a 'Rafraichir' button, and a 'Supprimer' button.
- Suppression d'une fiche par la saisie :** A text input field and a 'Supprimer' button.
- Suppression de toutes les fiches d'un entrepôt :** A dropdown menu showing 'Université Numérique Ingénierie et Technologie' and a 'Supprimer' button.



Il est déconseillé d'utiliser la suppression manuelle dans un environnement de production.