

Greenstone Travail Pratique 2

Greenstone et le protocole OAI

Le logiciel *Greenstone*, de même que d'autres outils de gestion de bibliothèques numériques, permet l'exploitation des données diverses grâce à l'interopérabilité offerte par le protocole OAI-PMH. *Greenstone* permet d'exporter, mais aussi d'effectuer la moisson de métadonnées provenant de ce protocole.

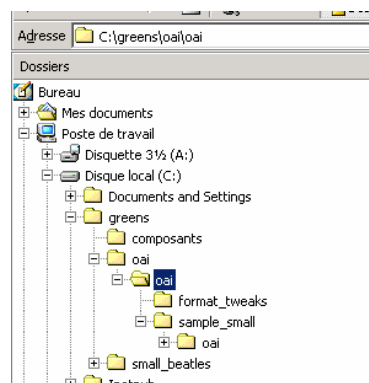
Notre exercice consistera en

- I) Exploitation des données OAI pour reconstruire une collection (les métadonnées ont déjà été recueillies).
- II) Étape plus délicate, se connecter à un serveur OAI pour télécharger directement des données. Ici on devra travailler en ligne de commande. Il est nécessaire en principe de posséder une connexion directe à Internet, car les firewall et proxys peuvent faire échouer la connexion.

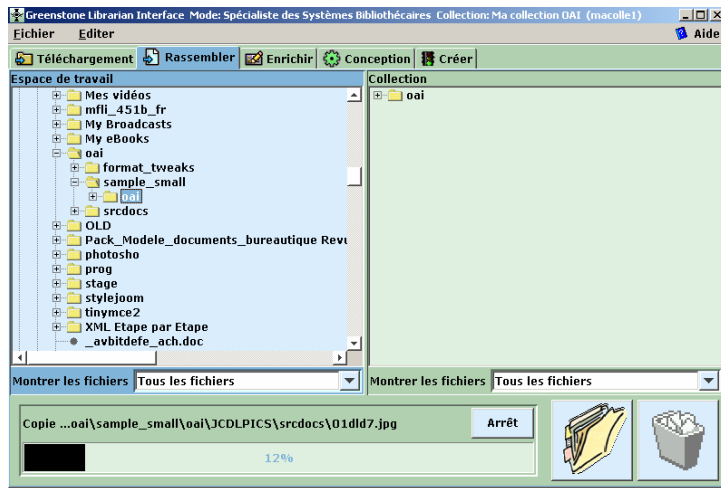
Première partie. Exploitation des données OAI

Créer la collection

- 1) Dans l'interface du bibliothécaire de *Greenstone*, créez une nouvelle collection appelée « monoai », « mon_cher_oai » ou un autre nom de votre cru. Suivez la procédure de l'exercice précédent : basez votre collection sur « Nouvelle collection » et acceptez les options proposées par défaut.
- 2) Allez dans **Poste de travail->C->greens**. Ouvrez le dossier oai.
- 3) Depuis le volet « Rassembler », naviguez jusqu'au dossier **oai** contenu dans **sample_small**, lui-même dans le dossier **oai** de niveau supérieur (voir capture d'écran). Chargez-le dans votre collection (à droite) par glisser-déposer.



- 4) Durant le transfert vers la collection, on vous demandera s'il faut ajouter le plugin *OAIplug* pour traiter les fichiers avec l'extension .oai. Vous l'ajoutez. Vous continuez en n'ajoutant pas d'autre plugin pour le moment.

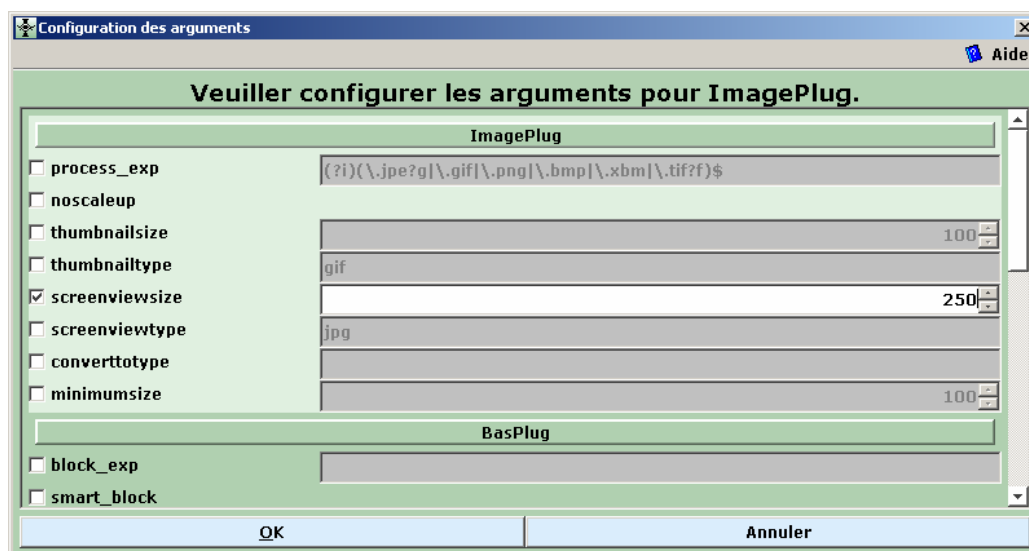


Configuration du plugin Image.

- 5) Allez au volet *Conception*.

a) Sélectionnez **Plugins des documents**, puis la ligne **pluginImagePlug**

b) Cliquez sur *Configurer plugin* (ou double cliquez sur la ligne sélectionnée). Une fenêtre popup s'ouvre. Allez à l'option **screenviewsize**, tapez 250 comme valeur. Cela créera des vignettes de 250pixels. Faites OK.



- 6- Allez au volet *Création*. Construisez la collection et effectuez une prévisualisation.

Amélioration des options de recherche

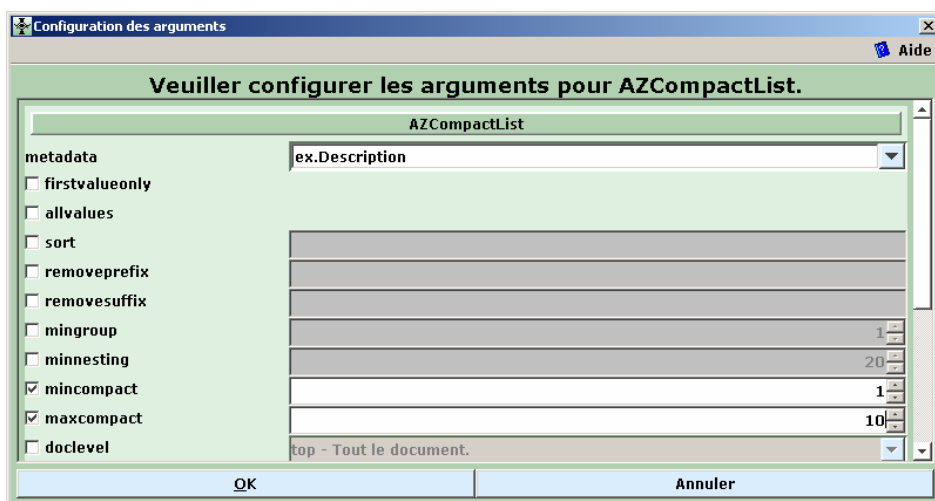
Nous allons peaufiner la collection en ajoutant des options de recherche basées sur les métadonnées récoltées par le protocole OAI-PMH (dans les fichiers .oai).

7) Nous travaillerons dans le volet *Conception*

a) Dans la section *Classificateurs de navigation* ajoutez un classificateur **AZCompactList** basé sur les métadonnées **ex.Subject**.

b) Ajoutez un autre classificateur **AZCompactList** basé sur les métadonnées **ex.Description**. Configurez-le de la manière suivante : mincompact(nombre minimum de documents à afficher par page)=1, maxcompact=10. Nommez le **buttonname** en *Description images*

c) Eliminez les deux classificateurs **AZList (ex.Title et ex.Source)**.



d) Dans la section *Indexes de recherche* procédez comme ceci : effacez tous les index et ajoutez-en un nommé « description » basé sur les métadonnées **ex.Description**

e) Dans *Créer*, construisez à nouveau votre collection et pré visualisez

Création de formats personnalisés pour une meilleure présentation

8) Nous restons dans le volet *Conception*

a) Sélectionnez *Caractéristiques de format*. Remplacez la définition de format **VList** de la manière suivante :

```
<td>
  {f}{[numleafdocs],[link][icon][link],[link][thumbicon][link]}
</td>
<td valign=middle>
  {f}{[numleafdocs],[Title],<i>[Description]</i>}
</td>
```

Cette définition personnalise l’affichage de listes verticales ainsi que les résultats de la recherche et les listes comprenant une vignette et des métadonnées de description. Greenstone extrait automatiquement les données des champs, ici[Description] équivaut à ex.Description

b) Dans *Caractéristiques de format* sélectionnez **DocumentHeading** dans la liste déroulante **Choisir caractéristique** ou en cliquant sur l’élément dans la liste.

c) Changez la définition à :

```
<h3>[Subject]</h3>
```

et cliquez sur « Remplacer le format »

Cela produit l’effet suivant : l’entête du document apparaît au-dessus. Par défaut, DocumentHeading affiche les métadonnées de ex.Title.

9) A côté des images, il y a un icône de document. Si l’on clique dessus, il apparaît « ce document n’a pas de texte ». Pour y remédier

a) Dans Conception -> *Caractéristiques de format* -> Choisissez **DocumentText** dans la liste déroulante. Dans « Chaîne au format HTML tapez le text suivant ou bien copiez-le à partir du fichier *doctxt_tweak.txt* qui se trouve dans le dossier *format_tweaks* lui même dans le dossier oai., tout en modifiant les intitulés en français comme ci-dessous (en gras) :

```
<center><table width=_pagewidth_ border=1>
<tr><td colspan=2 align=center>
<a href=[OrigURL]>[screenicon]</a></td></tr>
<tr><td>Description de image:</td><td> <i>[Description]</i> <br>
(<a href=[OrigURL]>original [ImageWidth]x[ImageHeight] [ImageType] available</a>)
</td></tr>
<tr><td>Sujet:</td><td> [Subject]</td></tr>
<tr><td>Editeur:</td><td> [Publisher]</td></tr>
<tr><td>Droits:<td> [Rights]</td></tr>
</table></center>
```

b) Cliquez sur *Remplacer le format*.

Pour rendre effectives les modifications de ce genre, il n'est pas nécessaire de Reconstruire la collection. Allez seulement à *Créer* et pré visualisez la collection. Cliquez sur un des icônes de texte à côté de l'image pour vérifier. Le document ainsi modifié inclut une version redimensionnée de l'image, reliée à une version haute définition accessible via le web. On y trouve aussi des informations sur le format de l'image.

Deuxième partie : télécharger grâce au protocole OAI

Moyennant une connexion à Internet et quelques instructions en ligne de commande, nous allons pouvoir obtenir les données OAI-PMH directement du serveur.

Avant de commencer

1-Sauvegarder votre collection (Fichier->Sauvegarder).

a) Notez le nom de répertoire (il se trouve dans la barre supérieure de la fenêtre *Interface bibliothécaire*.... Le nom sera **monoai** ou autre selon le nom donné à la collection.

b) Quittez *l'interface du bibliothécaire*

Préparation des fichiers pour la connexion

2-Dans un éditeur de texte (de préférence le *notepad*) ouvrez le fichier de configuration de la collection. Celui-ci se trouve dans :

C:\ProgramsFiles\Greenstone\collect\monoai\etc\collect.cfg.

a) Ajoutez la ligne suivante au-dessous de *public true* (ce sera en principe pour vous la 4^{ème} ligne) :

acquire OAI -src rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI1.1/jcdlpix.pl -getdoc

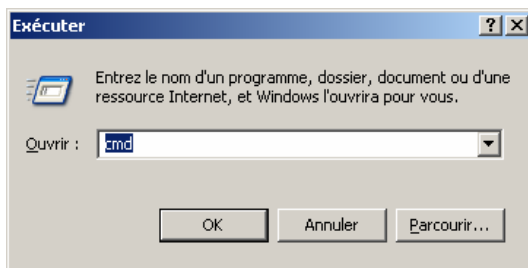
Ici nous avons donné l'adresse du serveur, l'action à entreprendre ainsi que le type de protocole (OAI 1.1 en l'occurrence).

3-Effacez le **contenu** du dossier *import* de la collection. Il contient les données de l'ancienne collection. Nous allons les recevoir maintenant directement du serveur distant (Nous pourrons toujours remettre les images jpg depuis le dossier oai placé dans **greens**. Ce qui nous intéresse maintenant c'est la récolte des métadonnées).

Travail en ligne de commande

4-Il faut ouvrir une fenêtre DOS pour accéder à la ligne de commande

a)Faites **Démarrer->Exécuter**. Saisissez, s'il n'est pas déjà, **cmd** dans la fenêtre qui apparaît. Faites **OK**.



b) Dans la fenêtre DOS, déplacez-vous au dossier racine de Greenstone. Pour cela il faut taper :

cd C:\Program Files\Greenstone

Le chemin dépend de l'endroit où se trouve installé *Greenstone*. Ici en l'occurrence c'est dans C.

c) Saisissez ensuite :

setup.bat

Pour activer la possibilité d'employer les commandes de Greenstone en ligne

5-Changement de répertoire afin de se rendre à celui contenant la collection OAI telle qu'a vu dans première partie.

a)Tapez :

cd collect\monoai (monoai correspond au nom de votre collection tel qu'il apparaît dans le dossier de collection).

b) Lancez :

perl -S importfrom.pl monoai

Greenstone se connectera au serveur et lancera une série de messages de diagnostic. Le programme `importform.pl` se connecte au fournisseur de données OAI spécifié dans le dossier de configuration de la collection. Il fait ceci ligne par ligne, trouve les données et les exporte vers notre collection.

(Attention à bien respecter les minuscules /majuscules et les espaces)

6- Vérification du chargement des données (une fois le téléchargement fini)

a) Allez dans le dossier *import* de votre collection. Vérifiez que les données ont bien été chargées.

b) Vous pouvez si vous le désirez refaire l'exercice de la première partie, cette fois-ci avec les données « vraiment » téléchargées.

Note : si par hasard il y avait un message d'erreur au moment de recréer la collection, il faudra effacer ou renommer un fichier nommé *Program* qui aurait été généré dans votre C (ou dans la partition correspondante selon la configuration de l'ordinateur et le système d'exploitation). Vérifiez toutefois la date de création

Adapté des *Tutorials Greenstone* dans <http://greenstone.sourceforge.net/wiki/index.php>



Cette création est mise à disposition selon le Contrat Paternité-NonCommercial-NoDerivs 2.5 disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/2.5/> ou par courrier postal à Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.