

Cyberdocs Documentation

Université Lyon 2 - 2009

Table des matières

1. Introduction.....	3
2. Communauté.....	4
3. Documentation.....	4
3.1. Aide à l'installation.....	4
3.1.1. Les pré-requis matériels.....	4
3.1.2. Les pré-requis logiciels.....	5
3.1.3. Les composants sous-jacents.....	6
3.1.4. Télécharger Cyberdocs.....	7
3.1.5. Installation étape par étape.....	7
3.1.5.1. Le fichier de propriétés pcd.properties.....	8
3.1.5.2. Compilation.....	11
3.1.5.3. Installation des modules.....	12
3.1.5.3.1. Installation du module de conversion.....	12
3.1.5.3.2. Installation du module de gestion.....	12
3.1.5.3.3. Installation du module de publication.....	14
3.1.6. Personnaliser cyberdocs.....	15
3.1.6.1. Créer une nouvelle institution.....	15
3.1.6.2. Ajout d'un nouveau code de styles.....	16
3.1.6.3. Conversion dans une autre langue.....	17
3.1.6.4. Les feuilles de styles CSS.....	17
3.1.6.5. Publier d'autres documents.....	17
3.1.6.6. Modifier les pages statiques.....	18
3.1.6.7. Modifier l'application Web.....	18
3.1.6.8. Personnaliser le module de gestion.....	19
3.1.6.9. Lier Apache et Tomcat : installer le module JK.....	19
3.1.6.10. Utiliser une base MySQL.....	21
3.1.6.11. Intégrer le module de publication à votre site web.....	23
3.1.7. Outils.....	23
3.1.7.1. Réindexer tous vos documents.....	23
3.1.7.2. Modifier vos métadonnées en masse.....	25
3.2. Aide aux utilisateurs.....	25
3.2.1. Approche technologique.....	25
3.2.2. Convertir un document via le module de gestion.....	26
3.2.2.1. S'identifier sur Cyberdocs.....	27
3.2.2.2. Créer un espace pour une thèse.....	27
3.2.2.3. Procédure de conversion.....	28
3.2.2.3.1. Transfert des fichiers stylés et des fichiers associés (onglet Gestion de fichiers).....	28
3.2.2.3.2. Saisie des métadonnées (onglet Métadonnées).....	30
3.2.2.3.3. Conversion (onglet Traitements).....	32
3.2.2.3.4. Vérification du résultat de la conversion.....	36
3.2.3. Convertir un document sans le module de gestion : en ligne de commandes.....	39
3.2.3.1. Opérations et étapes de conversion.....	39
3.2.3.2. Préparer un espace de travail.....	41
3.2.3.3. Préparer un document.....	42
3.2.3.4. Effectuer une conversion.....	44
3.2.3.5. Résultats de la conversion.....	46
3.2.4. Module de publication.....	46
3.2.4.1. La gestion et l'identification des utilisateurs.....	47

3.2.4.2. La gestion des documents.....	47
3.2.4.3. L'utilisation des métadonnées du document.....	49
3.2.4.4. La consultation des documents.....	51
3.2.4.5. La recherche.....	53
3.2.4.6. Les résultats de recherche.....	53
3.2.4.7. Les versions pour impression.....	54
3.3. Aide aux développeurs.....	54
3.3.1. Signaler des bogues.....	54
3.3.2. Contribuer à Cyberdocs.....	54
3.4. Téléchargements de Cyberdocs.....	54
4. Téléchargement.....	54
4.1. Distributions de Cyberdocs.....	55
4.2. Récupérer les sources avec Subversion.....	55
5. Listes.....	56
6. FAQ.....	58
7. Contacts.....	58

1. Introduction

La plate-forme Cyberdocs offre des outils, mais également un *modèle*, permettant à une institution de **publier** de la **documentation scientifique**, par exemple des thèses, des rapports, des revues, etc.

Deux principes essentiels sont associés à la plate-forme :

1. Les logiciels libres : **Cyberdocs est une plate-forme libre** ([licence GPL](#)) et utilise elle-même des logiciels libres. Cette approche permet à tous de l'utiliser sans contrainte, de l'améliorer, de contribuer à son rayonnement, de partager son expérience, de la traduire, de la documenter, etc. Il s'agit donc d'un véritable projet de coopération entre différents partenaires, financeurs, développeurs, utilisateurs.
2. Les documents structurés : le modèle de publication de Cyberdocs est basé sur des **documents structurés avec la norme XML**, ce qui comporte plusieurs avantages, notamment d'intéressantes possibilités de consultation, un archivage pérenne, des échanges facilités, etc.

Le modèle de publication

Le modèle de publication de Cyberdocs consiste à produire un document structuré en format XML (DTD TEILite), document qui devient ainsi le document de référence qui servira à l'archivage et à la publication.

L'objectif est d'obtenir ce document de référence avec le moins d'efforts possibles, en partant des documents tels qu'ils sont produits par les auteurs, soit le plus souvent avec un traitement de texte.

Le processus de publication

Le processus de publication de la chaîne Cyberdocs est constitué de ces étapes :

1. Préparation des documents

Un opérateur (ça pourrait être l'auteur, un documentaliste, etc.) va prendre un document traitement de texte et le préparer adéquatement. Cette préparation consiste à utiliser certains styles pour identifier des parties du contenu, s'assurer que les styles *Titre 1*, *Titre 2* sont utilisés pour la table des matières, préparer les dessins vectoriels pour les pixelliser, etc. Mais de plus en plus souvent, ce travail est réalisé directement par l'auteur qui utilise les modèles de documents proposés par le logiciel de traitement de texte ou par l'institution scientifique au sein de laquelle il travaille.

Exemple : le modèle de documents [Cyberthèses](#).

2. Préparation des métadonnées

Les métadonnées associées au document doivent être préparées, ce qui consiste en la création d'un document XML contenant ces métadonnées.

3. Conversion vers XML

Le **module de conversion** de la plate-forme Cyberdocs permet de convertir le document traitement de texte préalablement préparé en un document XML respectant la DTD TEILite. Le **module de gestion** de la plate-forme Cyberdocs permet de piloter la plupart de ces opérations à l'aide d'une interface Web très facile à utiliser.

4. Production des fichiers pour la publication statique

Si nécessaire, des versions HTML, [XHTML](#) et [PDF](#) sont produites (depuis le format XML) afin de permettre une diffusion statique des documents, sur un réseau ou sur un cédérom par exemple.

5. Publication

Le **module de publication** permet de rendre disponible les documents, avec une consultation aisée et exploitant la structure des documents, de même que des recherches

documentaires précises à l'intérieur d'un document ou dans une collection de documents publiés.

2. Communauté

Cyberdocs a pu voir le jour grâce au soutien de l'Agence Intergouvernementale de la Francophonie et à son programme Cyberthèses, par l'intermédiaire du Fonds Francophone des Inforoutes.

Cette plate forme de conversion, placée sous licence GPL, a été réalisée avec l'aide de la société AJLSM, et est diffusée en Afrique, dans l'Amérique du Sud, en Suisse et en France.

Aujourd'hui, la plateforme est principalement maintenue à travers le projet [Cyberthèses](#), dont les acteurs majeurs sont notamment [l'université Lumière Lyon 2](#), [l'université de Santiago du Chili](#), [les universités du Pérou](#) et la société [Silecs](#).

3. Documentation

3.1. Aide à l'installation

La plate-forme Cyberdocs permet de diffuser des documents numériques dans un environnement Internet. Elle offre des services de conversion et de publication de ces documents numériques.

Différents publics sont concernés par la plate-forme et donc par cette documentation. On distingue notamment :

- les **utilisateurs**, pour qui le module de publication, avec ses fonctionnalités d'affichage et de recherche constitue l'élément le plus important
- les responsables de la **préparation** et de la **publication** des documents, qui piloteront la conversion puis la publication des documents numériques
- les **implémenteurs**, c'est-à-dire les personnes qui seront amenées à installer, déployer, exploiter une plate-forme Cyberdocs, mais également la personnaliser, voire l'adapter
- les **développeurs**, qui ajouteront de nouvelles fonctionnalités à la plate-forme ou qui corrigeront les erreurs

La plate-forme Cyberdocs se découpe en trois grands modules :

- **le module de conversion** : il permet de transformer des documents traitement de texte en format XML selon la DTD TEILite.
- **le module de gestion** : il permet de piloter, via une interface web, les conversions et la publication des documents.
- **le module de publication** : il permet de diffuser, dans une architecture Web dynamique, des documents, avec des fonctions de recherche et de consultation.

L'installation du **module de gestion** requière au moins l'installation du **module de conversion**. Mais vous n'êtes pas obligé d'utiliser le module de gestion pour faire fonctionner le module de conversion en ligne de commande.

En ce qui concerne les modules de publication et de conversion, ils sont **indépendants**.

3.1.1. Les pré-requis matériels

Les informations données dans cette page concernent d'abord et avant tout les outils **serveurs** de la plate-forme, c'est-à-dire ce qui est nécessaire pour exécuter les conversions, installer l'outil de recherche, etc. Les **postes clients** n'ont besoin que d'un navigateur Web récent, qui supporte les normes HTML, CSS et Javascript, tel Internet Explorer, Mozilla, Opera, etc.

Le matériel informatique peut être très varié, ses caractéristiques précises dépendent en grande partie de la charge de travail que vous prévoyez ainsi que des modules que vous souhaitez installer.

L'**espace disque** requis sera d'environ 200 Mo pour une installation complète, incluant PHP, SDX et Java. Les outils Cyberdocs eux-mêmes prennent relativement peu de place, environ 20 Mo. Bien entendu, l'espace disque pour stocker les documents convertis n'est pas compté ici, et à ce sujet il faut prévoir environ 10 fois la taille du document original, même si ce chiffre peut varier en fonction du nombre de ressources externes telles que les images. Cet espace disque est nécessaire parce que les outils de conversion produisent beaucoup de documents :

- Une version XML / TEI Lite du document
- Deux versions XHTML statiques du document : une pour la consultation statique hors de tout contexte, et une autre pour la fonction d'impression dans l'application Web dynamique
- Deux versions HTML statiques du document : une pour la consultation statique hors de tout contexte, et une autre pour la fonction d'impression dans l'application Web dynamique, ces versions HTML étant destinées principalement au navigateur Web Internet Explorer.
- Une version PDF du document, avec un fichier PDF pour le document au complet et un fichier par partie de document
- Trois versions XML intermédiaires du document, produites lors de l'exécution des outils de conversion

Seul le document XML de référence en format TEI Lite est réellement nécessaire, tous les autres sont des produits dérivés qui peuvent être conservés ou non ; à noter qu'il est même possible de ne pas les faire générer par les outils. Si vous ne conservez pas tous les documents, l'espace disque requis diminue également.

On doit également mentionner que si vous utilisez l'environnement Web dynamique basé sur SDX pour diffuser les documents, vous devez compter environ 2,5 fois la taille des documents dans leur version TEI Lite en terme d'espace disque. En effet, SDX doit conserver une copie du document en entier, ainsi que chacune des parties. La taille des index pour la recherche est variable, mais on peut l'estimer à environ 50% de la taille des documents en format XML TEI Lite.

La quantité de **mémoire vive (RAM)** sur le serveur est un facteur critique. Le module de conversion utilise de la mémoire de manière proportionnelle à la taille des documents qu'il doit traiter, de même qu'au nombre d'images comprises dans ces documents. Une règle simple pour évaluer le besoin en mémoire est de prévoir 50Mo au départ, et ajouter 1,25Mo par page du document. Ainsi, pour un document de 500 pages, il peut être nécessaire d'avoir près de 700Mo de mémoire pour le traiter, même si très souvent une quantité moins grande sera suffisante.

De son côté, la mémoire utilisée par le module de publication dépend largement du nombre d'utilisateurs simultanés que l'on souhaite servir. Pour une application telle Cyberdocs, une mémoire de 2Go à 4Go sera en général confortable.

La puissance de l'**unité de traitement (CPU)** n'est pas un facteur critique, sachant que plus le processeur est rapide, plus rapides seront les conversions, mais également les recherches et l'affichage dans le module de publication.

3.1.2. Les pré-requis logiciels

Les différents modules de la plate-forme Cyberdocs ont été conçus de manière à être le plus portables possibles. Toutefois, ils font appel à des technologies différentes, d'où une certaine variation dans les spécifications des systèmes où on peut installer chaque module.

Il est important de mentionner que dans cette page et toute la documentation en général, le nom Windows fait référence aux versions Serveur, XP ou Vista de Windows. En particulier, cela exclut les versions 95, 98 ou Millenium ; certains outils pourront y être installés, mais ces

systèmes d'exploitation ne sont pas conçus pour ce genre d'opérations et il peut être difficile de s'en sortir. Une chose est certaine, les procédures ne sont pas documentées ici.

1) Le module de conversion

Le module de conversion fait appel à différentes technologies :

- Java pour piloter les conversions qui sont pour la plupart effectuées à l'aide du langage XSLT

2) Le module de gestion

Le module de gestion est une interface Web développée avec le langage PHP 5. On peut donc l'installer sur toute plate-forme supportant un tel environnement, ce qui ouvre beaucoup de possibilités. A noter un excellent support PHP sur les systèmes Linux et UNIX, mais aussi sous Windows.

3) Le module de publication

Le module de publication est une application Web dynamique développée à l'aide de la plate-forme SDX, elle-même conçue sur l'infrastructure Cocoon. Il s'agit en quelque sorte d'une application Web dynamique selon l'API des Java Servlets.

Pour l'héberger, il faut donc un système qui peut accueillir une machine virtuelle Java, et une telle machine virtuelle existe pour tous les systèmes d'exploitation récents, de Mac OS X à Windows en passant par Linux et de nombreux UNIX.

3.1.3. Les composants sous-jacents

En fonction des modules Cyberdocs à installer, vous devrez d'abord prévoir l'installation de composants sous-jacents, nécessaires pour utiliser les fonctions Cyberdocs. Ces composants sont : Java, un serveur Web, PHP, un moteur de servlets et SDX.

1) Java

Une machine virtuelle Java est nécessaire pour les modules de conversion et de publication. Elle est même nécessaire pour procéder à l'installation de ces modules, c'est pourquoi il s'agit en général de la première installation à faire. Les modules Cyberdocs fonctionnent avec Java 1.4, mais si vous procédez à une nouvelle installation, nous vous conseillons fortement d'utiliser la dernière version de Java, la 1.6.

Ci-dessous, vous trouverez des liens vers les distributions Java pour les plates-formes les plus populaires. A noter que vous devriez utiliser un **Java Development Kit (JDK)**, et non un **Java Runtime Environment (JRE)**.

[Distribution de référence de Sun](#) (pour Windows, Linux et Solaris)

[HP-UX](#) (UNIX pour les stations Hewlett-Packard)

[IBM](#) (pour AIX, Windows, Linux, OS/2, z/OS)

[Apple](#) (pour Mac OS X)

Ces distributions sont en général très faciles à installer, vous n'avez qu'à suivre les instructions fournies ou à consulter leur documentation. L'installation de Java devrait prendre environ 40Mo.

Important

Une fois Java installé, vous devez manuellement définir une variable d'environnement nommée **JAVA_HOME**, et lui donner comme valeur le chemin où vous avez installé Java. Sur UNIX ou Linux, vous pouvez définir cette valeur dans l'un des scripts de démarrage situés dans le répertoire maison de l'utilisateur utilisé par Java (par exemple dans le fichier `.bashrc`). Sur Windows, vous pouvez le faire en ouvrant le panneau de contrôle **Système** et en choisissant l'onglet *Avancée*, où vous trouverez un bouton qui vous permet de définir des variables d'environnement.

2) Serveur Web

Un serveur Web est nécessaire pour utiliser les modules de publication et de gestion de Cyberdocs. A noter que si vous souhaitez seulement utiliser le module de publication pour des démonstrations ou des tests, vous pouvez éviter cette installation et passer directement à l'installation du moteur de servlets.

Mais pour un site de production, ou encore pour installer le module de gestion, nous vous encourageons fortement à installer le [serveur Web Apache](#). Sur certains systèmes (Linux, UNIX) ce serveur est probablement déjà installé ; sinon aller au [site de téléchargement](#) et prendre la version qui convient à votre environnement. La [documentation](#) d'Apache est très complète, incluant une section sur la [compilation et l'installation](#), mais aussi sur la [configuration](#).

Si vous souhaitez utiliser le module de gestion, le « module Rewrite » (ou équivalent) doit être installé et activé pour votre serveur web.

3) PHP

L'environnement de script PHP doit être installé pour utiliser le module de gestion de Cyberdocs. PHP s'appuie sur un serveur Web, le plus souvent Apache, qui doit donc aussi être installé. A noter que très souvent le support PHP est déjà prévu par Apache, et que la version 5.x de PHP est requise.

Le [site principal de PHP](#) vous fournira toute l'information nécessaire sur cet outil. Il existe une page spécifique pour [les téléchargements](#).

Important :

Pour une question de gestion de processus, le module de gestion nécessite que PHP soit installé en tant que **module** du serveur web et non en CGI ([Common Gateway Interface](#)). Reportez-vous à la [documentation d'installation de php](#).

4) Moteur de servlets

Le moteur de servlets est un environnement d'exécution de sites Web dynamiques écrits en langage Java. Le module de publication a besoin d'un tel environnement pour fonctionner.

[Tomcat](#) de la fondation Apache est le moteur de servlets le plus utilisé. Vous pouvez le télécharger et l'installer aisément, si vous avez préalablement installé une machine virtuelle Java. Tomcat inclut son propre serveur Web, qui est en général suffisant pour des tests ou des démonstrations, ce qui évite l'installation du serveur Web. Si toutefois vous souhaitez utiliser Tomcat avec le serveur Web Apache, vous devrez installer et configurer le module [JK](#) pour faire le lien entre les deux outils. Afin de simplifier son installation, nous vous proposons [un rapide guide d'installation](#) pour Linux.

5) SDX

La plate-forme SDX est un outil de recherche et de consultation de documents XML. Elle est utilisée comme outil de base pour le module de publication Cyberdocs. **La version 2.3 ou supérieure est nécessaire.** Vous pouvez [télécharger SDX](#) et ses sources, mais si vous cherchez simplement à l'installer rapidement nous vous conseillons d'utiliser le fichier `sdx.war` et de le placer dans le dossier `webapps` du moteur de servlets. Une fois votre moteur redémarré, le fichier WAR sera décompressé et SDX sera disponible.

3.1.4. Télécharger Cyberdocs

Retrouvez toutes les instructions à la page « Téléchargements ».

3.1.5. Installation étape par étape

Lorsqu'on [télécharge](#) la plate-forme Cyberdocs, il est nécessaire de faire un certain nombre d'opérations avant de pouvoir l'utiliser ; dans cette documentation, nous allons appeler ces opérations le **processus d'installation** de la plate-forme.

Il est important de réaliser que ce processus d'installation ne concerne que les opérations spécifiques à la plate-forme Cyberdocs. Pour l'installation des composantes sous-jacentes (serveur Web, PHP, moteur de servlets, SDX, Java), vous devez suivre les instructions fournies avec ces outils. Dans la page sur le [téléchargement des composantes](#), vous trouverez des liens vers les sites Web des outils qui vous permettront de réaliser votre installation.

Le processus d'installation comporte essentiellement deux types d'opération : la modification de fichiers de propriétés et l'exécution de différents scripts, en fonctions des modules à installer.

Afin de ne rien oublier, veuillez respecter l'ordre suivant :

- Modification du fichier pcd-properties
- Compilation
- Installation des modules

3.1.5.1. Le fichier de propriétés pcd.properties

Dans le dossier racine des sources Cyberdocs se trouve un fichier nommé pcd.properties. Ce fichier contient quelques informations qui permettent de déterminer les paramètres d'installation de la plate-forme. Il est essentiel de comprendre ces paramètres et de leur donner des valeurs appropriées pour votre environnement.

Le format de ce fichier est le format de propriétés standard de l'environnement Java. Il s'agit d'un fichier texte, et pour être certain qu'il soit correctement interprété dans tous les environnements Java, il ne contient que des caractères de l'ASCII. Les lignes qui commencent par le caractère # sont des commentaires, elles sont donc ignorées ; vous pouvez en ajouter si vous le jugez nécessaire. Idem pour les lignes blanches, elles n'ont aucune signification sinon que de rendre le fichier plus lisible. Enfin, les lignes significatives sont celles où l'on retrouve le nom d'une propriété, suivi d'un caractère =, suivi de sa valeur.

Il est important de préciser que vous pouvez définir ces propriétés ailleurs que dans le fichier pcd.properties. En effet, afin de faciliter les mises à jour, il est conseillé de ne pas toucher ce fichier, mais plutôt placer les propriétés dans l'un ou l'autre de ces fichiers :

- Un fichier nommé pcd.properties situé dans le répertoire « maison » de l'utilisateur qui effectue l'installation.
- Un fichier nommé .ant.properties situé dans le répertoire « maison » de l'utilisateur qui effectue l'installation.
- Un fichier nommé .ant.properties situé dans le répertoire racine des sources Cyberdocs qui serviront à l'installation.

Ces différents fichiers de propriétés sont lus dans cet ordre, et la première fois qu'une propriété est définie, elle conserve sa valeur jusqu'à la fin du processus. **Le dernier fichier lu est le fichier pcd.properties situé à la racine du répertoire des sources Cyberdocs, et les valeurs propriétés qu'on y trouve constituent donc des valeurs par défaut.**

Après installation, le fichier pcd.properties se trouve à l'emplacement suivant : outils/bin/pcd.properties

1. Le dossier d'installation des outils de conversion

La propriété dossier.installation.up permet de définir l'emplacement du répertoire d'installation du module de conversion. A noter que ce répertoire devient également le répertoire de travail où les documents sont convertis.

La valeur de cette propriété est un chemin qui peut être soit absolu (par exemple /usr/local/cyberdocs/pcd-install ou C:/programmes/cyberdocs/pcd-install) ou relatif (par exemple

../.. /pcd-install). S'il est relatif, la résolution du chemin se fait par rapport à la racine des sources Cyberdocs. Précisons également que même sous Windows, il est conseillé d'utiliser le caractère / pour séparer les répertoires dans un chemin ; sinon, vous devez doubler les séparateurs : \\.

La valeur par défaut est ../pcd-install, le répertoire de travail sera donc le dossier pcd-install situé dans la hiérarchie de votre poste de travail au même niveau que le répertoire contenant l'ensemble des sources de la plate-forme.

Si vous ne prévoyez pas d'utiliser ni le module de conversion ni le module de gestion de Cyberdocs, cette propriété est bien entendue inutile.

2. Le dossier d'installation et la version d'OpenOffice

Note importante : contrairement aux anciennes versions de Cyberdocs, OpenOffice n'est plus utilisé par défaut pour convertir les fichiers DOC et ceci pour des raisons de stabilité.

Vous pouvez ignorer ce paramètre.

3. Le style utilisé par défaut lors la conversion

Il s'agit de l'identifiant des styles utilisés par défaut lors de la conversion. La liste des styles et des codes se trouvent dans le fichier : src/oo-vers-tei/xslt/utiles/styles.xml. Au moment de l'installation, ce fichier est disponible ici : outils/xslt/utiles/styles.xml.

Reportez-vous à la partie « liste des styles supportés par Cyberdocs » pour plus d'information.

La valeur par défaut « lyon2 » est fonctionnelle pour effectuer des tests.

4. Le chemin du module de publication sous SDX

Le module de publication de Cyberdocs utilise SDX comme infrastructure de site Web dynamique ; une fois installé, ce module devient donc une *application SDX*. Chaque application SDX doit être installée dans un dossier situé immédiatement dans le dossier d'installation de SDX lui-même.

La propriété sdx.application.path indique le nom de ce dossier. Par défaut, c'est la valeur pcd qui est utilisée.

Cette propriété ne concerne que les installations où il est prévu d'utiliser le module de publication.

5. Informations complémentaires pour SDX

Pour les propriétés sdx.logicsheet.path, cocoon.class.include et cocoon.class.wildcard, reportez-vous aux commentaires du fichier pcd.properties fourni par défaut, car ces propriétés sont amenées à évoluer très rapidement.

6. L'habillage de l'interface de consultation

Le module de publication peut avoir différents habillages pour la consultation des documents sur le Web. La propriété habillage permet de spécifier lequel choisir. La valeur donnée à cette propriété doit correspondre au nom d'un dossier qui se trouve dans le dossier src/web/habillages à partir de la racine des sources servant à l'installation.

Dans la distribution normale de Cyberdocs, seul l'habillage *pcd* est proposé. Si vous téléchargez les sources, vous aurez également l'occasion d'utiliser l'habillage *cybertheses*, mais celui-ci devrait être réservé à un usage spécifique au projet Cyberthèses.

Cette propriété ne concerne que les installations où il est prévu d'utiliser le module de publication.

7. L'identifiant du module de publication sous SDX

La plate-forme SDX offre un éventail de fonctionnalités dont certaines peuvent être distribuées sur différents serveurs ou différentes applications. C'est pourquoi toutes les applications sous SDX doivent avoir un identifiant, et les responsables des applications doivent faire en sorte que cet identifiant soit unique parmi toutes les applications SDX existantes.

Un bon moyen d'y arriver est d'utiliser le principe des *noms de domaines inversés*. Par exemple, si vous pensez installer le module de publication à une adresse de type `http://www.monsite.org/sdx/pcd/`, vous êtes normalement responsable du nom de domaine *monsite.org*. Dans ce cas, vous devriez choisir un identifiant qui commence à *org.monsite* et qui est complété de différentes parties séparées par un « . », et vous êtes responsables de faire en sorte que ce complément soit unique parmi les applications SDX qui sont sous la responsabilité de *monsite.org*.

La propriété `application.sdx.identifiant` contient cette valeur, et il est **nécessaire de la modifier** afin d'éviter que tous utilisent la même valeur.

Cette propriété ne concerne que les installations où il est prévu d'utiliser le module de publication.

8. Le dossier d'installation de SDX

Pour installer le module de publication, il est nécessaire de connaître le dossier où la plate-forme SDX est installée. La propriété `dossier.installation.consultation` doit contenir le nom de ce dossier.

Cette propriété ne concerne que les installations où il est prévu d'utiliser le module de publication.

9. L'adresse du serveur SDX

Spécifiez ici l'URL pour accéder à SDX. Par exemple : <http://www.monserveur.fr:8080/sdx>.

Le port 8080 est nécessaire seulement si vous utilisez Tomcat dans sa configuration par défaut. Pour une utilisation de Cyberdocs en production, nous vous conseillons d'utiliser [module JK](#) permettant de lier Apache à Tomcat.

Cette propriété servira de base au URL des documents disponibles via le module de publication. Modifiez cette valeur par l'adresse externe de votre serveur, ne laissez pas « localhost » sauf pour une utilisation dans le cadre d'un test.

Cette propriété ne concerne que les installations où il est prévu d'utiliser le module de publication.

10. L'ouverture de l'application de consultation

Dans une installation SDX, on peut retrouver un dossier contenant les sources d'une application sans que celle-ci soit réellement visible publiquement. Pour qu'elle le soit, elle doit être *ouverte* dans SDX, c'est-à-dire apte à être utilisée.

L'installation de la plate-forme Cyberdocs peut effectuer automatiquement cette ouverture, si la propriété `sdx.application.open` a une valeur *1*, ce qui est la valeur par défaut que nous vous conseillons de garder.

Cette propriété ne concerne que les installations où il est prévu d'utiliser le module de publication.

11. Identification des erreurs de conversion

Lorsque le module de conversion traite un document traitement de texte, il peut signaler certaines erreurs (style manquant, etc.) propre à un certain contexte d'utilisation. Pour l'instant, le seul contexte où de telles erreurs sont définies est le projet CyberThèses. Ces

erreurs sont signalées par des attributs spécifiques dans le document XML en format TEILite, qui ne seront donc pas valide et devra être modifié.

Pour activer l'identification des erreurs, il faut donner la valeur *1* au paramètre `erreurs.cyberdocs`. Cette fonctionnalité est désactivée pour toute autre valeur.

12. Les paramètres OAI

La plate-forme Cyberdocs permet de rendre disponibles des métadonnées associées aux documents via le protocole [OAI-PMH](#). Pour que cette fonctionnalité fonctionne, vous devez utiliser la version 2.2 ou plus récente de SDX. Pour activer cette fonctionnalité, vous devez modifier trois propriétés dans le fichier `pcd.properties` :

- `entrepot.oai` : donner la valeur *true* pour activer la fonctionnalité OAI.
- `entrepot.oai.nom` : le nom de votre entrepôt OAI, requis par le protocole.
- `admin.oai.email` : l'adresse de courrier électronique de l'administrateur de cet entrepôt OAI, requis par le protocole.

13. Location du binaire php

Cette propriété ne concerne que les installations où il est prévu d'utiliser le module de gestion.

Il s'agit du répertoire où se trouve le binaire de php (l'interface en ligne de commande). Le chemin doit être absolu.

Par exemple : `/usr/local/bin`.

Si vous ne savez pas où trouver le fichier, aidez-vous de la commande : `which php`

Si PHP est actif et que vous ne trouvez pas l'exécutable php, essayez d'installer « `php-cli` » (« `cli` » pour « `Command Line Interface` ») avec votre gestionnaire d'installation habituel.

Par exemple pour une Debian ou une Ubuntu : `apt-get install php-cli`.

Avertissement

- Vérifiez que le binaire porte bien le nom `php` ou `php.exe` à l'intérieur du répertoire spécifié. Créez un lien symbolique dans le cas contraire.
- Votre version de php doit être supérieure ou égale à 4.3 (version 5 conseillée). Vous pouvez vérifier à l'aide la commande : `php --version`.

14. Et en suite ?

Vérifiez une dernière fois vos propriétés, notamment qu'il n'y a pas d'espace dans les chemins d'accès, etc...

Votre fichier de configuration est maintenant prêt : vous pouvez passer aux étapes suivantes (la compilation et l'installation des modules).

3.1.5.2.

Compilation

Avertissement

Cette opération est nécessaire seulement si vous installez depuis les sources obtenues depuis l'entrepôt CVS de Cyberdocs. Si vous installez depuis la distribution, cette opération a déjà été effectuée pour vous. Elle peut cependant s'avérer nécessaire si vous utilisez une version de Java antérieure à 1.6 (dans ce cas supprimez le répertoire « `build` » à la racine des sources).

Plusieurs classes Java ont été écrites pour le module de conversion, et elles sont nécessaires dès l'installation des outils. Ainsi leur compilation est-elle la première chose à effectuer lors de l'installation de Cyberdocs.

Pour cela, il suffit simplement d'exécuter le script `build-ant.bat` ou `build-ant.sh` en fonction du système d'exploitation que vous utilisez.

Si la compilation vous renvoie une erreur que vous ne comprenez pas, vérifiez votre fichier `pcd.properties` et consultez la FAQ.

Si la compilation se termine correctement, vous pouvez maintenant passer à l'installation des modules.

3.1.5.3. Installation des modules

Important : avant d'installer les modules, nous vous rappelons qu'il est nécessaire d'avoir lu et respecté les « pré-requis » matériels et logiciels, ainsi que les trois étapes précédentes de ce manuel d'installation : téléchargement, configuration du fichier `pcd.properties` et compilation.

L'installation complète de Cyberdocs se déroule en trois étapes :

- Installation du module de conversion
- Installation du module de gestion
- Installation du module de publication

L'installation du **module de gestion** requière au moins l'installation du **module de conversion**. Mais vous n'êtes pas obligé d'utiliser le module de gestion pour faire fonctionner le module de conversion en ligne de commande.

En ce qui concerne les modules de publication et de conversion, ils sont **indépendants**.

3.1.5.3.1. Installation du module de conversion

Cette étape va créer le dossier de production du module de conversion, contenant toutes les sources nécessaires à la conversion des documents ainsi que des exemples. Vous pouvez effectuer cette installation en exécutant simplement le script `installation-oo2xml.bat` ou `installation-oo2xml.sh`. Par la même occasion, les outils de conversion, de même que les fichiers du module de gestion, seront copiés.

Ce script aura pour effet de créer un dossier qui sera le dossier de base des modules de conversion et de gestion. La localisation de ce dossier est précisée dans le fichier `pcd.properties`. Dans ce dossier, on trouvera trois sous-dossiers :

- `outils` : contient les outils du module de conversion, soit les processeurs XSLT ou XSL-FO, transformations XSLT, l'environnement ANT et son fichier de directives, etc.
- `php` : contient les scripts PHP du module de gestion ainsi que certains binaires pour l'appel au module de conversion.
- `production` : constitue le répertoire de base pour la production des documents à l'aide du module de conversion.

3.1.5.3.2. Installation du module de gestion

Avertissement :

Cette partie ne concerne que les utilisateurs désirant utiliser le module de gestion en php. Si vous souhaitez utiliser la plate-forme Cyberdocs en ligne de commande, passez cette étape.

Les fichiers du module de gestion sont copiés automatiquement lors de l'installation du module de conversion. Si vous ne désirez pas utiliser le module de gestion, vous pouvez supprimer l'intégralité du répertoire « gestion ».

Avant d'utiliser le module de gestion, procéder aux vérifications et modifications suivantes :

1. Droits d'accès aux fichiers

Le module de gestion est un site web en php pilotant le module de conversion. Pour réaliser cela, les scripts php doivent pouvoir exécuter certaines commandes à l'intérieur des répertoires d'installation de la plate-forme Cyberdocs.

Il vous faut donc autoriser votre serveur web à accéder aux commandes de conversion de la plate-forme. Le plus simple est d'effectuer les actions suivantes :

Sous Linux et Unix-like : placez-vous dans le répertoire contenant le dossier `pcd-install` et attribuez tous les fichiers de la plateforme au compte et groupe utilisateur utilisé par le serveur web.

Par exemple : `chown -Rh apache:apache pcd-install`.

Pour connaître le nom du compte utilisateur et son groupe, vous pouvez éditer votre fichier `httpd.conf` et regarder les paramètres *User* et *Group*.

Sous Windows : vérifiez que les fichiers sont accessibles en lecture, écriture et exécution au compte utilisateur servant à votre serveur web. C'est normalement le cas par défaut.

2. Paramètres de PHP/Apache

Editez le fichier `php.ini` présent dans `/etc` sous Linux, `c:/windows` ou le répertoire d'installation de Apache sous Windows et vérifiez les paramètres suivants (ces valeurs sont données à titre indicatif, adaptez-les selon vos besoins) :

- Maximum execution time of each script, in seconds : (possibilité de modifier uniquement pour php-cli- cf doc de php-cli et son fichier `php.ini` propre.)
`max_execution_time = 300000`
- Maximum amount of time each script may spend parsing request data :
`max_input_time = 6000`
- Maximum amount of memory a script may consume
`memory_limit = 1600M`
- Maximum size of POST data that PHP will accept.
`post_max_size = 1500M`
- Show all errors, except coding standards warnings (vous pouvez aussi désactiver les erreurs !)
`error_reporting=E_ALL & ~E_NOTICE`
- Whether to allow HTTP file uploads.
`file_uploads = On`
- Maximum allowed size for uploaded files.
`upload_max_filesize = 1400M`

Note : pour les utilisateurs de Redhat Linux 8.0/9.0 avec Apache 2.x.x, il peut être nécessaire de modifier le fichier `/etc/httpd/conf.d/php.conf` et de changer la valeur de `LimitRequestBody`.

Ajouter la variable d'environnement suivante dans `httpd.conf` :

SetEnv JAVA_HOME repertoire_installation_java.

Par exemple : `SetEnv JAVA_HOME /usr/java/j2sdk1.4.2_01`

Relancer Apache après modification des fichiers.

3. Accès HTTP au module de gestion

Créer un accès web à la plateforme dans la configuration de votre serveur web (il s'agit du fichier `httpd.conf` pour Apache), pointant dans le dossier `pcd-install/gestion/`. Pour des raisons de sécurité, ne PAS donner d'accès HTTP au répertoire « production ».

Si certains documents sont confidentiels, nous vous recommandons d'utiliser des connexions sécurisées (HTTPS) pour accéder au module de gestion.

Si nécessaire, modifier également le fichier `.htaccess` à la racine du dossier `pcd-install/gestion/` afin de configurer les redirections automatiques permettant de télécharger les fichiers du répertoire « production » : par exemple, décommentez et modifiez « `RewriteBase` » avec le chemin de base de l'URL vers le module de gestion - notamment si vous utilisez un « alias » dans votre configuration d'Apache.

Aller sur la page d'accueil (`http://nom_host/php/`) du module de gestion, et s'identifier avec le nom d'utilisateur `root`, sans mot de passe. La première action à réaliser sera donc de mettre un mot de passe en allant dans la rubrique `utilisateurs`. Voir la [documentation du module de gestion](#) pour en savoir plus.

4. Configuration avancée du module de gestion

Le module de gestion peut utiliser un serveur LDAP pour l'authentification, être traduit aisément dans la langue de votre choix et être paramétré différemment. Pour cela, consultez la partie intitulée : [#3.1.6.8. Personnaliser le module de gestion](#)

5. Avertissement

Il est fortement recommandé de limiter l'accès au module de gestion à votre intranet. Vous pouvez également opter pour un accès en https.

3.1.5.3.3. Installation du module de publication

Pour installer le module de publication, quelques étapes peuvent être nécessaires. Tout d'abord, on doit exécuter le script `installation-web.bat` ou `installation-web.sh`. Cela aura pour effet de copier les fichiers sources dans le dossier approprié sous SDX, tout en filtrant certains paramètres pour contextualiser votre installation.

Ensuite, il est également préférable de copier les fichiers de configuration des institutions pour lesquelles vous souhaitez publier des documents. Le script `copie-institutions.{bat | sh}` effectue cette opération. En particulier, il copie le fichier exemple fourni avec la plate-forme Cyberdocs, pour l'université Lyon 2. Si vous souhaitez publier des documents provenant d'autres institutions, il faut créer des fichiers semblables à ceux qu'on retrouve dans le dossier `exemples/configurations/web/lyon2` dans les sources Cyberdocs.

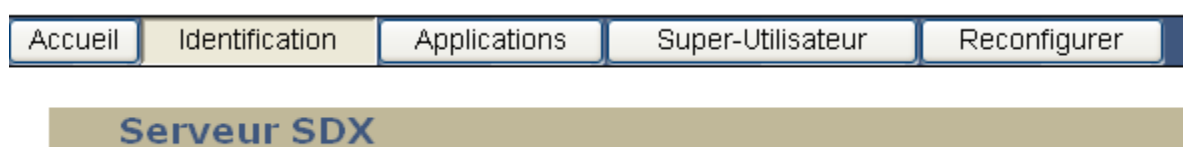
Une fois toutes les sources de Linux 8.0/9.0 avec Apache 2.x.x : modifier le `ouvrir` l'application dans SDX. Cela signifie qu'il faut indiquer au serveur SDX qu'il existe une nouvelle application installée, et qu'on veut rendre cette application visible et utilisable.

Pour cela, il suffit de se rendre à l'adresse de la page d'administration du serveur SDX (par exemple <http://localhost:8080/sdx/sdx/admin/loginsu.xsp>), au formulaire d'identification tel que présenté ci-dessous, et s'identifier en tant que super-utilisateur :

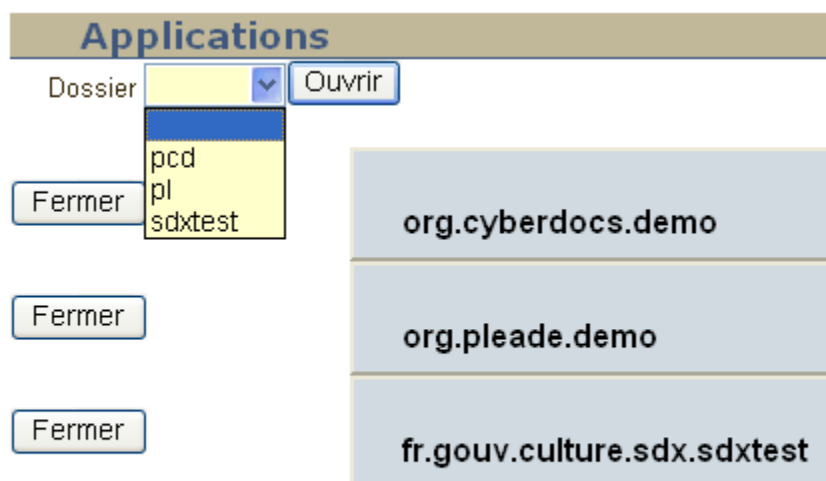


The image shows a login form titled "Super-Utilisateur". It has a blue header and a white body. There are two input fields: "Code" and "Mot de passe". Below the fields is a button labeled "Entrer".

Cette démarche effectuée, il suffit ensuite de cliquer sur le bouton **applications** en haut de l'écran :



Sur cette page, une liste déroulante propose l'ensemble des applications présentes sur le serveur :



En sélectionnant l'identifiant de votre application (valeur `sdx.application.path` dans le fichier `pcd.properties`), un lien vers l'application apparaît alors. Elle est désormais ouverte, il ne reste plus qu'à cliquer sur ce lien.

Cette opération d'ouverture de l'application n'est à effectuer qu'une seule fois.

3.1.6. Personnaliser cyberdocs

3.1.6.1. Créer une nouvelle institution

L'interface Web du module de publication suppose que chaque document est publié par une institution. Des informations sur cette institution doivent être fournies afin de rendre l'interface

claire et efficace. Si un document d'une nouvelle institution doit être ajouté, il est donc nécessaire de configurer le module pour inclure cette institution. Le plus simple est de partir de l'exemple fourni avec l'institution *lyon2*.

Les informations sur les institutions sont placées dans le dossier *institutions* dans l'application Web. Dans ce dossier, chaque institution possède son propre dossier, et le nom de ce dossier doit être le code donné à l'institution. Dans ce dossier, on retrouve trois éléments d'information importants :

1. Un fichier de configuration pour l'institution, absolument essentiel.
2. Une feuille de style CSS pour les documents publiés par l'institution, essentielle mais peut être vide.
3. Un logo, affiché dans le cadre supérieur de la fenêtre de consultation, lorsqu'on consulte un document publié par cette institution.

C'est le document de configuration, qui doit être nommé *config.xml*, qui contient les informations importantes. L'exemple ci-dessous permet de documenter ce fichier.

Exemple 1. Exemple de fichier de configuration d'une institution

```
<!-- Le code d'institution doit être dans l'attribut "id" -->
<ins:institution id="lyon2"
  xmlns:ins="http://cyberdocs.org/institutions"
>
  <ins:informations>

    <!--le nom de l'institution-->
    <ins:nom>Université Lumière Lyon 2</ins:nom>
    <ins:logos>
      <!--le logo-->
      <ins:logo role="consultation" height="20" width="10" src="logos/logo.png"/
    >
    </ins:logos>
    <ins:description/>

    <!--les adresses ip incluses dans l'intranet-->
      <ins:intranet>
        <ins:masque ip="192.168.123.*"/>
        <ins:masque ip="127.0.0.*"/>
      </ins:intranet>
    </ins:informations>

    <!--les intitulés des champs de métadonnées-->
    <ins:metadonnees>
      <ins:champ nom="dc:creator">
        <ins:libelle xml:lang="fr">Auteur</ins:libelle>
      </ins:champ>
      <ins:champ nom="dc:title">
        <ins:libelle xml:lang="fr">Titre</ins:libelle>
      </ins:champ>
      . . .
    </ins:metadonnees>
    <ins:messages>
      <!--l'intitule de la dédicace-->
      <ins:intitule code="dedicace" xml:lang="fr">Dedicace</ins:intitule>
    </ins:messages>
  </ins:institution>
```

3.1.6.2. styles

Ajout d'un nouveau code de

Soit le style *auteur-nouveau* qui permet de styler les auteurs dans le code de styles '*nouveau-code*'.

1. Il faut tout d'abord faire passer 'nouveau-code' en paramètre dans le script de conversion. (cf : [informations de conversion](#))
2. Dans le fichier dossier_de_production/outils/xslt/utiles/styles.xml, il est nécessaire de rajouter un élément : `<institutions> <institution code="nouveau-code"/> </institutions>` (comme celui existant pour lyon2)
3. Pour **chacun** des styles , il est nécessaire de déclarer un élément nom, indiquant le nom du style dans le code de styles 'nouveau-code', par exemple :

```
<style code="auteur"> <nom code="lyon2" xml:lang="fr">1|Auteur</nom> <nom code="nouveau-code" xml:lang="fr">auteur-nouveau</nom> </style>
```

ou encore :

```
<style code="text"> <nom code="lyon2" xml:lang="fr">Text</nom> <nom code="nouveau-code" xml:lang="fr">Text</nom> </style>
```

3.1.6.3. Conversion dans une autre langue

1. Pour utiliser l'anglais, il faut tout d'abord faire passer le parametre 'en' dans le [script de conversion](#)
2. Dans chaque fichier xml de dossier_de_production/outils/xslt/utiles les éléments du type : `<element xml :lang= 'fr'>----</element>` doivent avoir leur équivalent dans la langue souhaitée : `<element xml :lang= 'en'>----</element>`
3. Il en va de même pour le fichier dossier_de_production/outils/xslt/metadonnees/metadonnees.xml

3.1.6.4. Les feuilles de styles CSS

La consultation des documents, dans le module de publication ou en statique, passe par le format HTML ou XHTML, et celui-ci permet de préciser l'affichage à l'aide de feuilles de styles CSS. Dans Cyberdocs, nous utilisons beaucoup ces feuilles de styles, permettant notamment d'en utiliser plusieurs :

1. Une feuille de styles générale, appliquée à tous les documents.
2. Une feuille de styles par institution, appliquée à tous les documents publiés par cette institution.
3. Une feuille de styles spécifique à un document.

Ces trois CSS sont présentées dans l'ordre inverse de priorité ; en effet, si deux sélecteurs CSS s'appliquent, c'est celui que l'on trouve dans la feuille de styles spécifique, puis la feuille de styles pour l'institution, et enfin la générale.

Pour ajouter une feuille de styles CSS spécifique à un document, il suffit de nommer le fichier [code du document].css et de le placer dans le dossier sources du document si on utilise le module de conversion, ou de le placer directement à la racine du dossier du document.

Pour ajouter une CSS propre à une institution, on doit créer un fichier [code de l'institution].css, et le placer dans le dossier de l'institution du module de conversion ou du module de publication.

3.1.6.5. Publier d'autres documents

Il est possible d'utiliser le module de publication sans que le document soit au format XML TEILite. Par exemple, il pourrait s'agir d'un document PDF.

Ces documents doivent aussi avoir un code unique et être organisés par institution et par année de publication. Pour les traiter, il suffit de placer le fichier source dans le dossier [code du document]/sources et de nommer le fichier avec le code de document et une extension représentative de son format.

On doit ensuite absolument avoir un fichier de métadonnées, toujours nommé [code du document]-md.xml. L'élément dc:identifier doit à ce moment contenir l'URL complète du document.

3.1.6.6. Modifier les pages statiques

L'application Web dynamique offre à l'utilisateur une interface ergonomique qu'il est souhaitable de présenter dans sa langue. Les contenus générés par cette application sont associés à un *habillage* spécifique, ce qui permet d'offrir plusieurs langues pour une même apparence et un même contenu statique. Ces contenus générés sont stockés dans les documents XML suivants :

- src/web/habillages/pcd/messages/global.xml: messages communs à la plupart des pages de l'application
- src/web/habillages/pcd/messages/menu-general.xml: définition du menu de navigation générale
- src/web/habillages/pcd/messages/messages-avancee.xml: messages du formulaire de recherche avancée
- src/web/habillages/pcd/messages/messages-cadre.xml: messages du cadre supérieur de consultation dans un document
- src/web/habillages/pcd/messages/messages-chargeement.xml: messages reliés au formulaire d'indexation d'un document
- src/web/habillages/pcd/messages/messages-contenu.xml: messages reliés au cadre principal de consultation d'un document
- src/web/habillages/pcd/messages/messages-fiche.xml: messages reliés à la fenêtre de métadonnées d'un document
- src/web/habillages/pcd/messages/messages-id.xml: messages reliés au formulaire d'identification
- src/web/habillages/pcd/messages/messages-imprimer.xml: messages reliés à la fenêtre de choix du format pour l'impression
- src/web/habillages/pcd/messages/messages-rchargeement.xml: messages concernant la page de résultat de l'indexation
- src/web/habillages/pcd/messages/messages-resultats.xml: messages reliés aux résultats de recherche
- src/web/habillages/pcd/messages/messages-termes.xml: messages reliés aux listes de termes d'index

Afin de modifier le contenu statique (notamment la page index), éditez et créez des fichiers dans le répertoire src/web/habillages/pcd/statique.

Vous noterez que l'habillage par défaut est « pcd » : il vous est bien sûr possible de créer un nouvel habillage totalement différent en vous basant le modèle de base. Pour cela, dans les sources, créez un nouveau répertoire dans le répertoire habillage, placez-y vos fichiers personnalisés et modifiez le paramètre « habillage » du fichier pcd.properties afin d'y spécifier le nom de votre nouvel habillage. Puis relancer le script « installation-web » pour que vos nouveaux fichiers soient copiés dans le répertoire de travail de SDX.

3.1.6.7. Modifier l'application Web

Afin de modifier l'application Web, il est possible de distinguer quatre niveaux de modifications :

- **Modifier la feuille de styles `src/web/css/structure.css`** : il s'agit de la solution la plus simple qui vous permet de changer la mise en page des documents. La feuille de styles contient des sections pour les différentes parties du site, pensez à lire les commentaires du fichier pour plus d'information.
- **Modifier les pages statiques du site** : consulter la partie de cette documentation traitant spécialement de ces modifications.
- **Modifier les fichiers XSL du répertoire `src/web/xsl`** : ne modifiez ces fichiers que si vous ne pouvez pas obtenir ce que vous souhaitez avec les feuilles de styles et les pages statiques. Il s'agit de modifications en profondeur, notamment pour changer le HTML généré dynamiquement.
- **Modifier les fichiers XSP du répertoire `src/web`** : écrits en JAVA, ces pages sont l'articulation de l'application Web. Editez ces pages si vous souhaitez modifier le comportement du site (recherches, filtrage, nouvelles fonctionnalités... etc.).

A noter : il vous faudra relancer le script d'installation « installation-web » après chaque modification des sources. Sinon les fichiers modifiés ne seront pas copiés dans le répertoire de de SDX.

3.1.6.8. Personnaliser le module de gestion

Le module de gestion est une application Web dynamique écrite en PHP (<http://www.php.net>). Vous pouvez modifier aisément son comportement en éditant les fichiers PHP. Cependant, certains fichiers de configuration ont été prévus afin d'éviter des modifications en profondeur.

Voici un rapide descriptif de l'architecture du module de gestion :

- **Le fichier de configuration général** : il s'agit de fichier « `pcd.properties` » global à cyberdocs. Consultez la documentation d'installation pour plus d'information.
- **Le fichier de « configuration » `src/gestion/gestion/config.inc.php`** : les langues disponibles, configuration facultative d'un serveur LDAP.
- **Le fichier de « constantes » `src/gestion/gestion/inc/const.php`** : retrouvez dans ce fichier tous les constantes du module de gestion (répertoire de document, langues, étapes de conversion... etc.)
- **Les fichiers « d'affichage » dans le répertoire « `src/gestion/gestion/inc` »** : il s'agit de l'habillage pour le module de gestion. C'est ici que vous pouvez personnaliser les pages ou les traduire dans une autre langue.
- **Les fichiers « applicatifs » à la racine du répertoire `src/gestion/gestion`** : il s'agit des fichiers de traitement pour la gestion des conversions et le stockage des données. A ne modifier que si vous souhaitez changer le comportement du module de gestion.

A noter : il vous faudra relancer le script d'installation « installation-oo2xml » après chaque modification des sources. Sinon les fichiers modifiés ne seront pas copiés dans le répertoire d'installation.

Traduction : il est possible de traduire le module gestion sans modifier le code PHP. Pour cela :

- Dans l'onglet « configuration » du module de gestion, cliquez sur « Extraire le fichier de traduction vierge ». Le fichier extrait est un fichier « `.po` », éditable avec un logiciel de gestion de traduction, tel que « [poEdit](#) » « [Lokaliz](#) », « [KBabel](#) » ...
- Une fois traduit, renommez le fichier avec le suffixe de la langue (par exemple « `en.po` » pour l'anglais)
- Placez le fichier dans le répertoire « `pcd-install/gestion/po/` »
- Ajouter la langue au tableau php « `$GLOBALS["TRANSLATIONS"]` » du fichier de configuration : « `pcd-install/gestion/config.inc.php` »

3.1.6.9. Lier Apache et Tomcat : installer le module JK

Le module JK permet de faire communiquer Apache et Tomcat : le principal avantage se trouve dans la possibilité d'utiliser les deux serveurs en même temps comme s'ils étaient tous les

deux sur le port 80. En effet, le port 8080 (par défaut pour Tomcat) n'est pas recommandé pour un serveur Web et certains administrateurs réseaux en bloquent l'accès pour des raisons de sécurité. De plus, les URL de vos documents seront moins compliquées à retenir et surtout plus « conformes » à ce que les utilisateurs d'internet connaissent. Enfin, vos statistiques de consultations d'Apache contiendront également les appels aux documents et les recherches via SDX.

Avertissement : cette page d'aide ne se veut pas exhaustive sur la démarche à suivre afin d'installer le module JK. Nous vous conseillons de lire la documentation officielle disponible sur [le site du module JK](#).

Cette page est un exemple d'installation : nous nous baserons sur l'utilisation d'une distribution Linux Fedora 7.

Vous devez avoir installé et configuré vos serveurs Apache et Tomcat. Nous vous conseillons d'installer préalablement SDX afin de pouvoir tester plus facilement votre nouvelle configuration.

1) Dépendances

Vérifiez que vous avez installé les packages suivants :

- http-devel (yum install http-devel)

2) Télécharger les sources du module JK

Vous trouverez les sources du dernier module JK sur le site officiel : <http://tomcat.apache.org/connectors-doc/>

Ou, par exemple, directement en ligne de commandes :

```
wget http://www.apache.org/dist/tomcat/tomcat-connectors/jk/source/jk-1.2.25/tomcat-connectors-1.2.25-src.tar.gz
```

Décompressez l'archive dans /usr/local/src/

3) Compilation et installation

Trouvez le répertoire d'installation de APX :

```
# which apxs
/usr/sbin/apxs
```

Compilation :

```
cd /usr/local/src/tomcat-connectors-1.2.25-src/native/
./configure --with-apxs=/usr/sbin/apxs
make
su -c 'make install'
```

4) Configuration d'Apache

Créez le fichier de configuration /etc/httpd/conf/workers.properties avec le contenu suivant, en adaptant selon vos chemins d'installation de tomcat et de java.

```
# tomcat home
workers.tomcat_home=/data/tomcat

#localisation de la jvm
workers.java_home=/data/jdk1.6.0

# Define 1 real worker using ajp13
worker.list=worker1

# Set properties for worker1 (ajp13)
worker.worker1.type=ajp13
```

```
worker.worker1.host=127.0.0.1
worker.worker1.port=8009
```

Ajoutez la configuration du module JK à votre fichier httpd.conf :

```
# ##### MOD JK #####
# Load mod_jk module
# Update this path to match your modules location
LoadModule    jk_module    modules/mod_jk.so

# Where to find workers.properties
# Update this path to match your conf directory location (put workers.properties
next to httpd.conf)
JkWorkersFile "/etc/httpd/conf/workers.properties"

# Where to put jk shared memory
# Update this path to match your local state directory or logs directory
JkShmFile     "/etc/httpd/logs/mod_jk.shm"

# Where to put jk logs
# Update this path to match your logs directory location (put mod_jk.log next to
access_log)
JkLogFile     "/etc/httpd/logs/mod_jk.log"

# Set the jk log level [debug/error/info]
JkLogLevel    info

# Select the timestamp log format
JkLogStampFormat "[%a %b %d %H:%M:%S %Y] "

# Send everything for context /examples to worker named worker1 (ajp13)
JkMount    /sdx/* worker1
```

Ici, nous voyons sur la dernière ligne que toutes les requêtes commençant par « sdx » seront redirigées vers tomcat.

Rechargez votre serveur web :

```
/etc/init.d/httpd reload
```

Testez : http://votre_serveur/sdx/

Vous devez voir s'afficher la page d'administration de SDX.

3.1.6.10.

Utiliser une base MySQL

Par défaut, SDX travaille avec une base de documents stockée dans une arborescence de répertoires située dans son espace de travail. Ce système est acceptable pour une petite base de documents, mais il devient vite très peu performant pour une base volumineuse et une montée en charge importante.

Il est donc préférable, une fois l'installation de Cyberdocs terminée et testée, de changer de système de stockage pour un moteur de bases de données plus performant. Si vous souhaitez faire ce changement sur une base en production, notez qu'il vous faudra ré-indexer tous vos documents.

Nous verrons ici l'exemple d'utilisation de MySQL, disponible en licence GPL et proposé avec la plupart des distributions Linux. Bien évidemment, l'utilisation d'un autre système de base de données est tout à fait possible.

Voici les modifications de configuration à apporter pour utiliser MySQL :

1. Arrêtez votre Tomcat (cd tomcat/bin, ./shutdown.sh)

2. Téléchargez le connecteur java pour MySQL à l'adresse suivante : <http://dev.mysql.com/downloads/connector/j/3.1.html>. Copiez le fichier « .jar » dans le répertoire webapps/sdx/WEB-INF/lib. Le fichier « .jar » se trouve normalement à la racine de l'archive téléchargée. Par exemple : mysql-connector-java-3.1.7-bin.jar.
3. Créez une nouvelle base de données MySQL. Pour l'exemple, nous l'appellerons « pcd ». Nous vous conseillons également de créer un utilisateur MySQL « pcd » ayant uniquement des droits complets sur cette nouvelle base.
4. Modifiez votre configuration de MySQL pour autoriser des requêtes de taille importante. Pour cela, ajoutez les lignes suivantes au fichier /etc/my.cnf :

```
# Taille maximum des paquets
max_allowed_packet=1024M
```

Relancez votre serveur MySQL (/etc/init.d/mysqld restart).

5. Editez le fichier webapps\sdx\WEB-INF\web.xml, pour ajouter ou modifier les lignes suivantes :

```
<init-param>
  <param-name>load-class</param-name>
  <param-value>com.mysql.jdbc.Driver</param-value>
</init-param>
```

6. Editez le fichier webapps\sdx\WEB-INF\cocoon.xconf pour ajouter ou modifier les lignes suivantes - en y modifiant l'url, le login et le password mysql comme il convient :

```
<datasources>
  <jdbc name="id_mysql" logger="sdx.rdbms.id_mysql">
    <pool-controller min="5" max="10"/>
    <dburl>jdbc:mysql://localhost:3306/pcd?autoReconnect=true</dburl>
    <user>myuser</user>
    <password>mypassword</password>
  </jdbc>
</datasources>
```

7. Modifiez le fichier src/web/conf/application.xconf :

- Au début du noeud « <sdx:userDocumentBase> », ajoutez : « <sdx:database type="MYSQL" dsi="id_mysql"/> »

- Remplacez (ou commentez) dans « <sdx:repositories> » du noeud « <sdx:userDocumentBase> » :

```
<sdx:repository id="users" type="FS" baseDirectory="users/xml" depth="0" extent="1000"/>
Par :
<sdx:repository id="users" type="MYSQL" dsi="id_mysql"/>
```

- Au début du noeud « <sdx:documentBase id="documents" type="lucene" default="true" maxFieldLength="100000"> » situé dans « <sdx:documentBases> », ajoutez : « <sdx:database type="MYSQL" dsi="id_mysql"/> ».

- Remplacez (ou commentez) dans « <sdx:repositories> » du noeud « <sdx:documentBases> » :

```
<sdx:repository type="FS" id="rdocuments" baseDirectory="documents" default="true"/>
Par :
<sdx:repository type="MYSQL" id="rdocuments" dsi="id_mysql" default="true"/>
```

Pour vous aider, vous pouvez télécharger un exemple de fichier application.xconf après modifications.

8. Relancez le script « installation-web » pour copier le fichier application.xconf dans le répertoire de SDX.

9. Relancez Tomcat (cd tomcat/bin, ./startup.sh)

Vous devez voir apparaître 4 tables dans votre base MySQL : documents, rdocuments, sdxuserdb, users. De plus, les noms des tables seront préfixés par l'identifiant de l'application spécifié dans votre fichier « pcd.properties ».

A noter :

- La base de données peut être d'une taille conséquente (plusieurs Go) : prévoyez donc un espace disque suffisant.
- Vous pouvez supprimer le répertoire webapps\sdx\pcd\conf\document.

3.1.6.11. Intégrer le module de publication à votre site web

Il n'existe pour le moment aucune solution préétablie pour intégrer le module de publication à un site déjà existant. Cependant, des projets basés sur XML-RPC/SOAP sont à l'étude.

Pour le moment, dans le cadre du [programme Cybertheses](#), l'université Lyon 2 a modifié les fichiers XSL du module de publication pour **l'utiliser avec le CMS Drupal**. Cependant, il est tout à fait possible d'utiliser ces sources « modifiées » avec d'autres sites dynamiques, en apportant quelques modifications aux scripts PHP.

Nous tenons à préciser qu'il s'agit d'une solution temporaire en attendant des développements plus poussés.

Le principe est le suivant :

- Ne plus utiliser directement les interfaces de SDX pour les pages statiques et pour rechercher dans les documents.
- Intégrer le moteur de recherche à un site déjà existant.

Voici les conditions pour utiliser ces sources « remodifiées » :

- Vous possédez déjà une installation de Cyberdocs fonctionnelles.
- Votre configuration serveur accepte PHP.
- Vous posséder des connaissances en PHP
- Vous disposer de temps pour tester et modifier les scripts fournis.
- Votre institution dispose déjà d'un site dynamique auquel vous souhaitez intégrer le module de publication de Cyberdocs (notamment pour plus de visibilité et pour conserver un accès unique).

Installation et utilisation :

Devant la complexité de l'installation et les risques de trou de sécurité en cas de mauvaise utilisation des sources, nous avons choisi de ne pas diffuser directement ces modifications : contactez-nous via la rubrique « contact » pour plus d'information.

3.1.7. Outils

Vous trouverez sur le site [Cyberdocs.org](#) où télécharger les outils documentés ci-dessous.

3.1.7.1. Réindexer tous vos documents

En cas de problème ou suite à une modification de configuration, il peut être nécessaire de réindexer tous les documents de la base.

La méthode proposée n'est sans doute pas la meilleure, mais elle possède l'avantage d'être relativement simple, pour un coup de développement presque nul.

Pré-requis : le script de ré-indexation utilise « wget » et « awk ».

Réindexer vos documents, étape par étape :

- Téléchargez [cette archive](#) et décompressez-la : vous y trouverez un répertoire « reindex » contenant les outils de ré-indexation.
- Dans le répertoire « reindex/web », vous trouverez trois fichiers permettant d'ajouter à votre application SDX une page XSP de ré-indexation simplifiée. Copiez les trois fichiers dans l'arborescence des sources Cyberdocs comme indiqué ci-dessous :
 - reindex/web/rsimplechargement.xsp : placez ce fichier dans votre répertoire src/web/
 - reindex/web/xsl/rsimplechargement.xsl : placez ce fichier dans votre répertoire src/web/xsl/
 - reindex/web/habillages/pcd/xsl/rsimplechargement.xsl : placez ce fichier dans le répertoire xsl de votre habillage.
- Exécutez à nouveau le script « installation-web » afin d'installer les nouveaux fichiers dans votre application SDX. Relancez Tomcat.
- Si vous utilisez le module de gestion, copiez le fichier « index.dat » de votre institution (par exemple production/lyon2/index.dat) dans le répertoire « reindex »
- Si vous n'utilisez pas le module de gestion, il vous faut créer un fichier texte « index.dat » contenant la liste de tous les espaces de document à ré-indexer, ayant pour chaque ligne le format suivant : « nom_repertoire_doc:annee_sur_4_chiffre: ». Par exemple : « grosjean_mf:2003: ».
- Placez-vous dans le répertoire « réindex » et exécutez la commande « reindex.sh » comme indiqué ci-dessous, afin de fabriquer le fichier de commandes de ré-indexation globale :

```
./reindex.sh URL_APPLICATION_PCD CHEMIN_REPERTOIRE_PRODUCTION LANGUE  
ID_INSTUTION REPERTOIRE_TOMCAT REPERTOIRE_GESTION|all FICHER_INDEX_DAT >  
reindex_all.sh
```

Par exemple, pour réindexer toute l'institution lyon2 :

```
./reindex.sh http://mon_server/sdx/pcd /data/pcd-install/production fr  
lyon2 /data/tomcat all index.dat > reindex_all.sh
```

Par exemple, pour réindexer uniquement les documents de l'institution lyon2 classés dans le répertoire « archive » du module de gestion :

```
./reindex.sh http://mon_server/sdx/pcd /data/pcd-install/production fr  
lyon2 /data/tomcat arch index.dat > reindex_all.sh
```

- Vérifiez le fichier « reindex_all.sh » : il doit contenir des appels à « wget ».
- Attribuez les droits en exécution à tous les fichiers « sh » du répertoire « reindex » et lancez « reindexall.sh » : `./reindexall.sh > resultat.txt &`
- Vous pouvez surveiller la ré-indexation en consultant le fichier resultat.txt (tail resultat.txt). Si des erreurs apparaissent, notez les espaces de document en question et traitez ces cas manuellement.

Important :

- **Tous les 50 chargements de documents, le server Tomcat est arrêté et relancé automatiquement.**

- Une fois les documents réindexés, pensez à supprimer les fichiers « rsimplechargement.xml/xsp » : ils ne sont plus utiles.

3.1.7.2. Modifier vos métadonnées en masse

Il peut être utile de modifier certaines métadonnées de façon globale. Cependant, il n'existe pas d'outil intégré à Cyberdocs pour effectuer ce genre d'opération.

Le plus simple est sans doute d'utiliser les expressions régulières et les outils de remplacements automatiques. Nous prendrons ici l'exemple de la suppression du port 8080 dans les URL de référence des documents, suite à une installation du module JK.

Tout ceci est bien sûr à titre d'exemple.

1) Création d'un script de remplacement

Créer un fichier « remplace.sh » :

```
cp $1 $1.orig
sed 's/mon_serveur.fr:8080/mon_serveur.fr/g' $1.orig > $1
```

Une copie de sauvegarde du fichier sera créée avant le remplacement.

2) Trouver les fichiers à modifier

Un exemple pour trouver les fichiers à modifier.

```
find pcd-install/production/ -name "*.xml" -exec grep "mon_serveur.fr:8080" -l
'{}' \; > list_fichier.txt
```

```
find pcd-install/production/ -name "*.dat" -exec grep "mon_serveur.fr:8080" -l
'{}' \; >> list_fichier.txt
```

3) Effectuer le remplacement

```
awk '{ print "./remplace.sh "$0 }' list_fichier.txt > changeall.sh
chmod +x changeall.sh
./changeall.sh
```

Une fois les vérifications effectuées, vous pouvez supprimer les fichiers de sauvegarde.

```
find pcd-install/production/ -name "*.xml.orig" -exec rm -f '{}' \;
find pcd-install/production/ -name "*.dat.orig" -exec rm -f '{}' \;
```

3.2. Aide aux utilisateurs

3.2.1. Approche technologique

Le module de conversion de la plate-forme Cyberdocs vise à automatiser un processus de conversion depuis un format traitement de texte vers un document structuré en format XML, selon la DTD TEILite. Une telle opération consiste donc à identifier le plus de *structure* possible dans le document original pour rendre le document XML le plus riche possible. Les éléments de structure présents dans un document traitement de texte peuvent être :

- De l'enrichissement typographique : caractères gras ou italiques, couleurs, polices particulières, etc.
- Des structures déjà gérées par les traitements de texte : tableaux, illustrations, liens hypertexte, notes de bas de page, etc.

- Des styles (de paragraphes ou de caractères) ayant une sémantique commune à tous les documents : Titre 1, Titre 2, etc.
- Des styles ayant une sémantique particulière pour la plate-forme Cyberdocs : auteur, titre du document, année de publication, etc.

Le module de conversion de la plate-forme Cyberdocs ne se contente pas de produire ce document XML de référence. En effet, une fois celui-ci obtenu, le module peut produire des versions du document prêt à une publication statique, en format HTML, XHTML ou PDF. De plus, le module va préparer un ensemble de documents qui alimenteront le module de publication pour rendre l'interface de consultation encore plus riche.

Une série de traitements seront effectués, en séquence, dans le but de tirer le meilleur parti de ces informations. Ces traitements font intervenir essentiellement cinq technologies :

- [XSLT](#) (*eXtensible Stylesheet Language Transformations*) : ce langage de programmation a été conçu spécialement pour traiter de l'information en format XML. Il s'avère parfaitement approprié pour les traitements effectués par le module de conversion, et à part l'étape initiale qui fait appel à OpenOffice.org, tous les autres développements ont été effectués à l'aide de ce langage normalisé. Le module de conversion inclut un processeur XSLT : [Saxon](#) version 6.5.2, réalisé par Michael Kay et disponible librement.
- [XSL-FO](#) (*eXtensible Stylesheet Language Formatting Objects*) : il s'agit de l'autre volet de la norme XSL, qui consiste en un format (XML) de mise en page, permettant ainsi de préparer des documents adaptés à l'impression. La production des documents PDF dans le module de conversion passe préalablement par la production d'une version en format XSL-FO, qui ensuite sera traitée à l'aide d'un processeur XSL-FO, inclut dans la distribution Cyberdocs. Ce processeur est [FOP](#) du projet Apache XML.
- [SVG](#) (*Scalable Vector Graphics*) : ce format de graphique vectoriel est utilisé lors de la conversion vers le format XSL-FO, afin de permettre une représentation plus fine et plus précise des images incluses dans les documents traités. Le processeur XSL-FO FOP, lorsqu'il traite des documents, donne un meilleur rendu lorsque les images sont incluses dans des documents SVG.
- [ANT](#) : ce projet Apache fournit un environnement souple et puissant pour exécuter différentes opérations. Il inclut déjà une série de tâches, dont le pilotage des transformations XSLT ou XSL-FO, la copie de fichiers, etc. C'est pourquoi ANT est utilisé pour piloter les différentes opérations de conversion proposées par le module de conversion Cyberdocs.
- [Java](#) : l'environnement ANT étant une application Java, une machine virtuelle Java est utilisée pour supporter toutes les opérations effectuées par le module de conversion, à l'exception du traitement initial effectué par OpenOffice.org.

Un opérateur de conversion de documents à l'aide de Cyberdocs n'a pas nécessairement besoin de connaître tous ces détails techniques. Toutefois, en cas de problèmes, ou lorsqu'une installation doit être correctement effectuée, il peut être important de comprendre le rôle joué par les différentes technologies et leur implémentation.

L'utilisation du module de gestion dispense de connaître précisément les noms des étapes de conversion et notamment d'utiliser le module de conversion en ligne de commande. Nous avons donc séparé la documentation en deux parties distinctes : une [documentation plus visuelle pour les utilisateurs du module de gestion](#) et une [documentation très détaillée pour les utilisateurs en ligne de commande](#).

3.2.2. Convertir un document via le module de gestion

L'outil **Cyberdocs** permet de convertir des documents texte en un format XML garantissant un archivage pérenne des documents (qui ainsi ne dépendent plus d'un logiciel particulier), et à partir duquel sont générés des formats de consultation (HTML, XHTML) et d'impression (PDF).

3.2.2.1.

S'identifier sur Cyberdocs

Pour utiliser Cyberdocs, il faut que l'administrateur vous ait attribué un nom d'utilisateur et un mot de passe qui vous donnent accès à votre institution. Pour exemple ici, nous prendrons l'Université Lumière Lyon 2.



Cyberdocs - Module de gestion

Utilisateur :

Mot de passe :

Institution :

Vous pouvez alors sélectionner l'institution dont vous allez traiter les documents :



Cyberdocs - Module de gestion

Accueil Utilisateurs Conversions

Bienvenue dans l'outil d'administration de Cyberdocs.

◆ Listes des institutions disponibles:

- ◇ Université Lumière Lyon 2

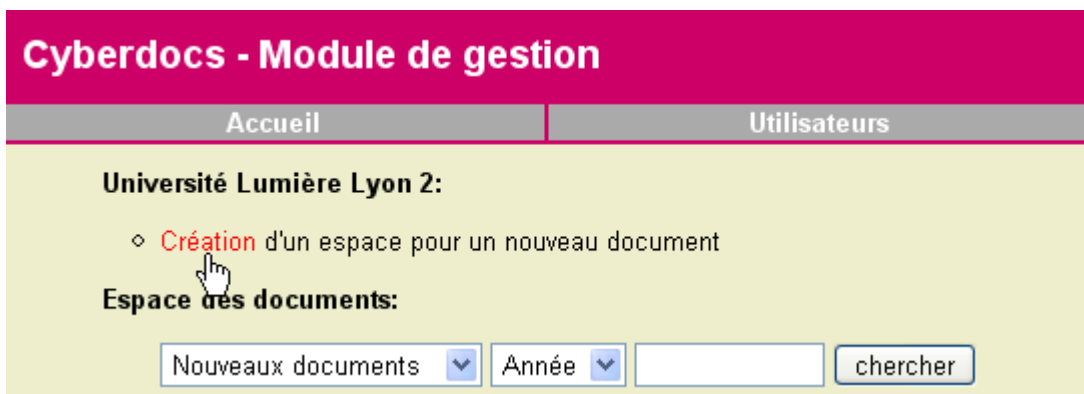
L'écran suivant vous permet d'accéder aux documents qui ont déjà été traités sur le serveur, ou de créer de nouveaux documents.

3.2.2.2.

thèse

Créer un espace pour une

Si vous souhaitez convertir un document que vous venez de préparer, il faut d'abord créer un espace pour ce nouveau document :



Cyberdocs - Module de gestion

Accueil Utilisateurs

Université Lumière Lyon 2:

- ◇ Création d'un espace pour un nouveau document

Espace des documents:

Cyberdocs - Module de gestion

Accueil Utilisateurs **Conversions** Se déconnecter

Université Lumière Lyon 2:

Création d'un nouvel espace

- Identifiant du document:
- Année de soutenance:
- Langue du document:

Créer un nouvel espace

Retour

L'**identifiant** correspond au nom de l'auteur suivi de l'initiale du prénom dont il est séparé par un tiret bas, le tout en minuscules ; par exemple, pour la thèse d'Edith Jaillardon, on aura l'identifiant suivant : **jaillardon_e**. Cet identifiant peut être modifié, mais en règle générale ce n'est pas nécessaire.

Le nouvel espace que vous venez de créer dans Cyberdocs a l'apparence suivante :

Cyberdocs - Module de gestion

Accueil Institution **Gestion de fichiers** Métadonnées Traitements

Université Lumière Lyon 2: jaillardon_e (2006)

- ◊ Répertoire du document: **Nouveaux documents**
 - Modifier le répertoire du document:
- ◊ Commentaire du document:
 - Modifier le commentaire du document:

Vous devez transférer un document pour commencer le traitement. Pour cela, cliquez sur [Gestion de fichiers](#)

- ◊ [Supprimer](#) ce document

3.2.2.3. Procédure de conversion

La navigation à l'intérieur de l'espace se fait à l'aide des onglets **Gestion de fichiers**, **Métadonnées** et **Traitements**. Procédez dans cet ordre :

3.2.2.3.1. Transfert des fichiers stylés et des fichiers associés (onglet Gestion de fichiers)

Chargez les fichiers qui doivent être traités par la chaîne dans les champs appropriés. Le fichier principal doit être un document texte ayant une extension **.odt** ou **.sxw**, créé à l'aide d'OpenOffice. Les fichiers associés au document principal peuvent être des fichiers image (possédant une extension **.jpg**, **.gif** ou **.png**, qui peuvent être regroupés dans une archive **.zip**) ou des fichiers **.pdf** (qu'il n'est pas possible de regrouper dans une archive **.zip**).

Université Lumière Lyon 2: jaillardon_e (2006)

◊ Liste des fichiers transférés:

	Date	Taille	Type	Nom du fichier
<input type="checkbox"/>	2006 Sep 18 16:18	290332	Document	these_jaillardon_e.sxw
<input type="checkbox"/>	2006 Sep 18 16:18	209170	PdfAmont	jaillardon_e_annexes.pdf
<input type="checkbox"/>	2006 Sep 18 16:18	45233	PdfAmont	jaillardon_e_planches.pdf

Supprimer

◊ Transférer le fichier **SXW** contenant le document:

Parcourir...

◊ Transférer la feuille de style du document:

Parcourir...

◊ Transférer une image ou un fichier ZIP contenant des images:

Parcourir...

◊ Transférer des fichiers PDF générés en amont:

Parcourir...
 Parcourir...
 Parcourir...
 Parcourir...

Transférer

◊ Importer le fichier contenant le document:

Importer

*Entrez le nom d'un fichier déjà présent sur le serveur,
en chemin absolu ou relatif à l'espace du document*

3.2.2.3.2. Saisie des métadonnées (onglet Métadonnées)

- Il est indispensable de saisir et valider les métadonnées avant la conversion pour que celles-ci soient associées au document.

Cyberdocs - Module de gestion

Accueil	Institution	Gestion de fichiers	Métadonnées
---------	-------------	---------------------	--------------------

Université Lumière Lyon 2: jaillardon_e (2006)

◊ Métadonnées du document:

- [Editer](#)
- [Visualiser](#)

Cyberdocs - Module de gestion

Accueil	Institution	Gestion de fichiers	Métadonnées	
---------	-------------	---------------------	--------------------	--

Métadonnées du document - Université Lumière Lyon 2: jaillardon_e (2006)

Les champs avec une * doivent être renseignés obligatoirement

Auteur* (NOM Prénom) JAILLARDON Edith

Titre* Français
La suppléance parlementaire sous la Cinquième République

Français
Français
Français

- Lorsqu'une thèse possède 2 auteurs, les noms et prénoms des deux auteurs doivent être indiqués dans le même champ « Auteur » :

Cyberdocs - Module de gestion

Accueil	Institution	Gestion de fichiers	Métadonnées	Traitements
---------	-------------	---------------------	-------------	-------------

Métadonnées du document - Université Lumière Lyon 2: gerfaud-tourrel (2000)

Les champs avec une * doivent être renseignés obligatoirement

Auteur* (NOM Prénom) TOURREL Jean-Paul & GERFAUD Jean-Pierre

- Lorsqu'une thèse a été rédigée dans une langue autre que le français (cas de certaines thèses en co-tutelle écrites en anglais, italien, portugais, ...), on indique, dans le même champ « Titre français », le titre en français et dans la langue étrangère :

Auteur* (NOM Prénom) MARIÉTHOZ Johnny

Titre* Français
Discriminant Models for Text-independent Speaker Verification / Algorithmes d'apprentissage discriminants en vérification du locuteur

- Pour les thèses au format PDF (mais pas celles contenant uniquement certaines parties en PDF), on indique, dans la rubrique « Commentaires de diffusion », la mention suivante : « Thèse diffusée au format PDF ».



Afin de valider ces métadonnées, il faut **impérativement** cliquer une première fois, en bas du premier écran, sur **Modifier**, puis au bas du second écran sur **Continuer**.

Métadonnées du document - Université Lumière Lyon 2: jaillardon_e (2006)	
Auteur (NOM Prénom)	JAILLARDON Edith
Titre	Français : La suppléance parlementaire sous la Cinquième République
Sous-titre	
Ecole doctorale	Sciences des sociétés et du droit
Faculté de soutenance	
Département	
Discipline	Science politique
NOM Prénom du directeur de thèse	DEMICHEL Francine
Résumé	
Mots-clés (séparés par des ;)	
Mots-clés temporels (séparés par des ;)	
Mots-clés géographiques (séparés par des ;)	
NOM prénom d'un membre du jury	
Date de soutenance (aaaa-mm-jj)	1976-01-01
Nature du contenu	
Langue principale de la thèse	Français
Copyright	
Type de diffusion	Intranet
Commentaires de diffusion	
Références à d'autres ressources	
URL de référence	http://demeter.univ-lyon2.fr:8080/sdx/theses/lyon2/2006/jaillardon_e

Continuer

Les métadonnées sont à présent enregistrées et l'on peut lancer la conversion des documents.

3.2.2.3.3.

Conversion (onglet Traitements)

Lorsqu'on travaille sur un nouveau document, on lance un **Traitement complet** qui va générer un document XML et des produits dérivés (HTML, PDF).

The screenshot shows the 'Cyberdocs - Module de gestion' interface. At the top, there is a navigation bar with tabs: 'Accueil', 'Institution', 'Gestion de fichiers', 'Métadonnées', and 'Traitements'. The 'Traitements' tab is active. Below the navigation bar, the document title is 'Université Lumière Lyon 2: jaillardon_e (2006)'. The main content area is divided into several sections:

- ◇ Répertoire du document: **Nouveaux documents**
 - Modifier le répertoire du document: Nouveaux documents (dropdown menu)
- ◇ Commentaire du document:
 - Modifier le commentaire du document: [text input field] [Modifier button]
- ◇ Commandes disponibles:
 - **Traitement complet** (highlighted with a mouse cursor)
 - [Production du document XML de référence](#)
 - [Validation du document XML de référence](#)
 - [Génération des produits dérivés \(html, pdf...\)](#)
 - [Autres traitements - document non XML](#)
- ◇ Résultat de la conversion: *Aucun fichier*
- ◇ Liste des fichiers produits par l'outil de conversion:
 - [PDF en Amont](#)
- ◇ [Effacer](#) tous les fichiers produits par les conversions.
- ◇ [Supprimer](#) ce document

Le temps nécessaire à la conversion d'un document dépend de sa taille et de la complexité de son contenu (présence d'illustrations par exemple) ; pour une thèse, il faut compter en général quelques minutes (moins d'un quart d'heure) mais certains documents peuvent prendre jusqu'à une heure de traitement et plus. La rubrique **Résultat de la conversion** : Voir permet de suivre les étapes de la conversion, de s'informer des problèmes rencontrés et de vérifier son résultat :

Cyberdocs - Module de gestion

Accueil Institution Gestion de t

Université Lumière Lyon 2: jaillardon_e (2006)

- ◊ Répertoire du document: **Nouveaux documents**
 - Modifier le répertoire du document: Nouveaux documents ▾
- ◊ Commentaire du document:
 - Modifier le commentaire du document:
- ◊ Conversion : **en cours**
 - Utilisateur : **jerome**
 - Date de lancement : **19/09/2006 08:32:33**
 - Date de fin :
 - Statut : **En cours**
 - Action : [Annuler](#)
- ◊ Résultat de la conversion: [Voir](#)
- ◊ Liste des fichiers produits par l'outil de conversion:
 - [OpenOffice](#)
 - [PDF en Amont](#)
- ◊ [Effacer](#) tous les fichiers produits par les conversions.
- ◊ [Supprimer](#) ce document

Cyberdocs - Résultat de la conversion

Note : si vous laissez cette page ouverte, elle se rechargera automatiquement tout

lyon2 - 2006 - jaillardon_e

Buildfile: oo-vers-tei.xml

initialisation

Conversion du fichier de traitement de texte vers OpenOffice.org

Décompresse le fichier produit par OpenOffice

```
[unzip]Expanding: /data/oo2xml/production/lyon2/2006/jaillardon_e/oo/jaillardon_e.sxw into /data/oo
[touch]Creating /data/oo2xml/production/lyon2/2006/jaillardon_e/oo/xml/office.dtd
[touch]Creating /data/oo2xml/production/lyon2/2006/jaillardon_e/oo/xml/META-INF/Manifest.dtd
```

- Si la conversion aboutit, le message « BUILD SUCCESSFUL » est affiché en bas de la page ; avant de passer à l'étape de vérification, il faut s'assurer que, dans le haut de la page, sous la rubrique **Première étape XSLT de conversion, depuis OpenOffice.org vers un format XML intermédiaire** n'apparaisse aucun message annonçant que le document traité contient un ou plusieurs *styles non déclarés* :

```

Première étape XSLT de conversion, depuis OpenOffice.org vers un format XML intermédiaire

[style]Processing /data/oo2xml/production/lyon2/2004/teil_g/oo/xml/content.xml to /data/oo2xml/production/lyon2.
[style>Loading stylesheet /data/oo2xml/outils/xslt/01-validation/pilote.xsl
[style]INFO:100:Le traitement de l'étape débute à 2006-09-25T08:45:19
[style]INFO:200:Attention, on ne traite pas office:script
[style]AVERTISSEMENT:400:Il manque le style l|EcoleDoct, l|Auteur,
[style]AVERTISSEMENT:300:Attention, le style Contents 5 n'est pas déclaré dans le fichier utiles/styles.xml
[style]AVERTISSEMENT:300:Attention, le style Contents 2 n'est pas déclaré dans le fichier utiles/styles.xml
[style]AVERTISSEMENT:300:Attention, le style Contents 2 n'est pas déclaré dans le fichier utiles/styles.xml
[style]AVERTISSEMENT:300:Attention, le style Contents 1 n'est pas déclaré dans le fichier utiles/styles.xml
[style]AVERTISSEMENT:300:Attention, le style Contents 1 n'est pas déclaré dans le fichier utiles/styles.xml
[style]AVERTISSEMENT:300:Attention, le style Contents 1 n'est pas déclaré dans le fichier utiles/styles.xml
[style]AVERTISSEMENT:300:Attention, le style Contents 1 n'est pas déclaré dans le fichier utiles/styles.xml
[style]AVERTISSEMENT:300:Attention, le style Contents 2 n'est pas déclaré dans le fichier utiles/styles.xml
[style]AVERTISSEMENT:300:Attention, le style Contents 3 n'est pas déclaré dans le fichier utiles/styles.xml

```

La présence de styles non déclarés dans le document ne bloque pas la conversion, mais tout paragraphe auquel est associé un style non déclaré est **supprimé** lors de la conversion. Si vous voyez apparaître des messages vous signalant la présence de styles non déclarés, il faut considérer que la conversion n'est pas correcte. Il vous faut retourner au document OpenOffice, remplacer ces styles incorrects par des styles appartenant au modèle **Thèse Lyon2**, et recommencer la conversion en suivant la procédure décrite plus bas pour une conversion qui a échoué.

Il s'avère quelquefois impossible de remplacer un style dans Word. Dans ce cas, plusieurs solutions s'offrent à vous :

- **appliquer le style « Police par défaut »** au paragraphe concerné ;
- le **couper** et de le **recoller en tant que texte sans mise en forme** (menu Édition/collage spécial) ;
- **supprimer totalement le style.**



Sans oublier dans tous les cas de reprendre manuellement les enrichissements de texte qui auront été supprimés.

- Si aucun style non déclaré n'apparaît, vous pouvez passer à l'étape suivante : la vérification.
- Si la conversion n'aboutit pas, le message « BUILD FAILED » est affiché en bas de la page ; les informations figurant sur la page vont alors vous permettre d'identifier le problème qui a empêché la conversion.
- Si l'erreur s'est produite plus tôt dans la conversion, c'est qu'il existe un problème dans les documents soumis à la conversion, qui empêche le bon déroulement de celle-ci (liste des problèmes de conversion courants à établir).

Dans ce second cas, la conversion est à refaire. Procédez comme suit :

- Retournez dans vos documents source et corrigez les problèmes qui ont bloqué la conversion.

- Dans Cyberdocs, effacez tous les fichiers produits par la conversion à partir de la page **Traitements** :

Université Lumière Lyon 2: jaillardon_e (2006)

- ◊ Répertoire du document: **Nouveaux documents**
 - Modifier le répertoire du document: Nouveaux documents ▼
- ◊ Commentaire du document:
 - Modifier le commentaire du document:
- ◊ Commandes disponibles:
 - [Traitement complet](#)
 - [Production du document XML de référence](#)
 - [Validation du document XML de référence](#)
 - [Génération des produits dérivés \(html, pdf...\)](#)
 - [Autres traitements - document non XML](#)
- ◊ Conversion : **terminée**
 - Utilisateur : **jerome**
 - Date de lancement : **19/09/2006 08:32:33**
 - Date de fin : **19/09/2006 08:36:38**
 - Statut : **Terminé**
 - Action : [Fermer](#)
- ◊ Résultat de la conversion: [Voir](#)
- ◊ Liste des fichiers produits par l'outil de conversion:
 - [OpenOffice](#)
 - [XML TEI Lite](#)
 - [HTML](#)
 - [XHTML](#)
 - [PDF](#)
 - [PDF en Amont](#)
- ◊ **Effacer** tous les fichiers produits par les conversions.
- ◊ [Supprimer](#) ce document

Supprimez les documents source que vous avez chargés, à partir de l'onglet **Gestion de fichiers** » :

Université Lumière Lyon 2: censier_c (2007)

◊ Liste des fichiers transférés:

	Date	Taille	Type	Nom du fichier
<input checked="" type="checkbox"/>	2007 Jul 30 11:29	218081	Document	censier_c these.sxw
<input checked="" type="checkbox"/>	2007 Jul 30 11:29	2237057	Images	illustrations.zip

Chargez les fichiers corrigés à partir de l'onglet **Gestion des fichiers**.

Lancez un nouveau traitement complet à partir de l'onglet **Traitements**.

En cas de nouvel échec de la conversion, cette procédure est à répéter autant de fois que nécessaire, jusqu'à l'obtention d'un message « BUILD SUCCESSFUL ».

N.B. La chaîne Cyberdocs peut traiter plusieurs documents en parallèle.

3.2.2.3.4. Vérification du résultat de la conversion

Lorsque la conversion a été menée à terme, il faut encore vérifier que les documents produits ne comportent pas de problèmes résiduels, avant de les diffuser sur le site des thèses en ligne. Pour cela, on effectue une vérification à partir des fichiers HTML produits par la conversion, qui donnent un aperçu des documents lorsqu'ils seront mis en ligne et accessibles par les internautes.

Pour cela, dans l'onglet **Traitements**, rubrique **Liste des fichiers produits par l'outil de conversion** : activez le lien **HTML** et choisissez le document **index-frames.html** :

Cyberdocs - Module de gestion

Accueil

Institution

Gestion de fichiers

Université Lumière Lyon 2: jaillardon_e (2006)

◊ Répertoire du document: **Nouveaux documents**

- Modifier le répertoire du document: Nouveaux documents ▼

◊ Commentaire du document:

- Modifier le commentaire du document:

◊ Commandes disponibles:

- [Traitement complet](#)
- [Production du document XML de référence](#)
- [Validation du document XML de référence](#)
- [Génération des produits dérivés \(html, pdf...\)](#)
- [Autres traitements - document non XML](#)

◊ Conversion : **terminée**

- Utilisateur : **jerome**
- Date de lancement : **19/09/2006 08:32:33**
- Date de fin : **19/09/2006 08:36:38**
- Statut : **Terminé**
- Action : [Fermer](#)

◊ Résultat de la conversion: [Voir](#)

◊ Liste des fichiers produits par l'outil de conversion:

- [OpenOffice](#)
- [XML TEI Lite](#)
- [HTML](#)
- [XHTML](#)
- [PDF](#)
- [PDF en Amont](#)

Cyberdocs - Module de gestion

Accueil

Institution

Université Lumière Lyon 2: jaillardon_e (2006)

◊ Fichiers HTML (html) :

- [TH.1.html](#) - 22.8 Ko
- [index-frames.html](#) - 1.4 Ko
- [TH.2.html](#) - 92.5 Ko
- [TH.3.html](#) - 502.6 Ko
- [TH.4.html](#) - 50.3 Ko
- [TH.5.html](#) - 23.3 Ko
- [TH.6.html](#) - 10.1 Ko

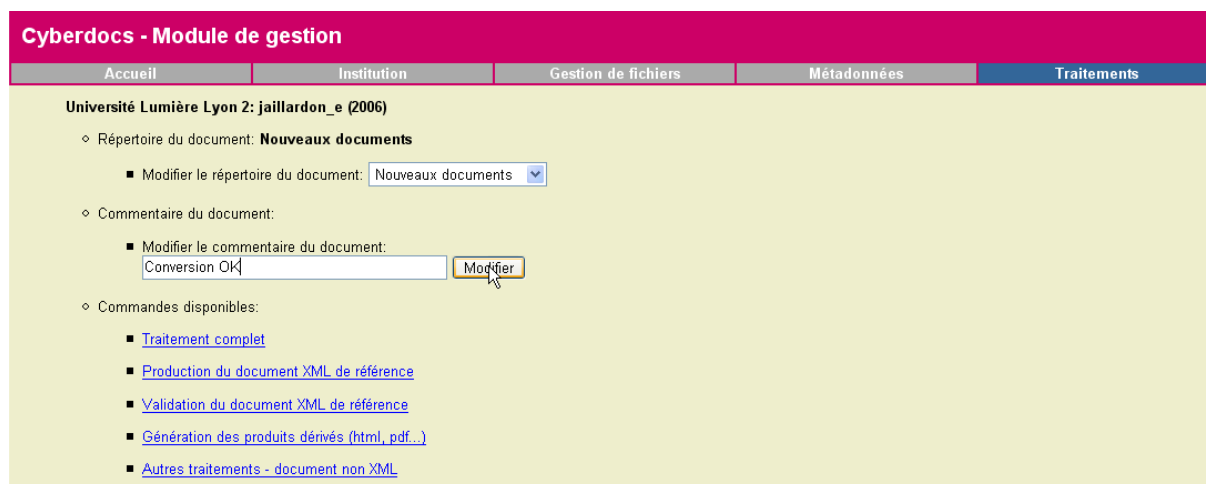
Vous accédez alors au document tel qu'il apparaîtra sur le site des thèses en ligne (hormis certaines différences de mise en forme) ; la table des matières active, dans le cadre de gauche, permet de naviguer à l'intérieur du document ; des onglets permettent d'accéder directement à la liste des tableaux et des figures contenus dans le document :

The image shows a digital interface for a thesis document. On the left, there is a table of contents (TOC) with three tabs: 'Contenu', 'Illustrations', and 'Tableaux'. The 'Contenu' tab is active, showing a list of document sections with expandable icons. The main area on the right displays the title page of a thesis. The title is 'LOGIQUES DE DIFFÉRENCIATIONS : ANCIENNES LYCÉENNES D'ORIGINE ALGÉRIENNE'. The page includes the university name 'UNIVERSITÉ LUMIÈRE LYON 2', the faculty 'faculté de sociologie et d'anthropologie', and the department 'sociologie et sciences sociales'. It also lists the presenter 'Monique DESCHAMPS', the degree 'THÈSE pour obtenir le grade de : DOCTEUR DE L'Université LUMIÈRE LYON 2', the director 'Yves GRAFMEYER', and the date 'Présentée et soutenue publiquement le 7 juillet 2006'. A list of jury members is provided, including Isabelle BERTALUX-WIAME, Yves GRAFMEYER, Yves LEQUIN, André PETITJEAN, and Olivier SCHWARTZ. The copyright notice at the bottom reads 'Copyright DESCHAMPS Monique et Université Lumière - Lyon 2 - 2006.'.

Les points à vérifier en particulier sont les suivants :

- o La page de titre comporte toutes les informations requises ; ces informations sont conformes à celles du bordereau de métadonnées ;
- o Les parties et sous-parties s'affichent correctement et ne comportent pas d'erreur de hiérarchisation ; les titres sont placés en minuscules accentuées, avec une capitale à l'initiale ;
- o Tableaux et illustrations :
- o Tester tous les liens accessibles depuis les onglets et s'assurer que les tableaux et illustrations s'affichent correctement ;
- o Lorsque les listes sont incrémentées (ex. Figure 1, Tableau 1), s'assurer qu'aucun numéro ne manque ;
- o Si différentes figures ou différents tableaux sont repérés par le même intitulé, il faut les distinguer en précisant la légende (par ex. « Tableau 1 » dans le corps de texte et « Annexe - Tableau 1 »).
- o Les caractères spéciaux s'affichent correctement ;
- o Les illustrations sont visibles et ne dépassent pas du cadre (sauf dans le cas d'illustrations de grande taille comportant beaucoup de détails, qu'on ne peut pas réduire sans nuire à la lisibilité) ;
- o Les fichiers liés (PDF, sons, vidéos) s'ouvrent sans problème.

Si le document est correct, on signale que la vérification a été faite en ajoutant un commentaire dans l'onglet **Traitements**, rubrique **Commentaire du document** :



On peut alors procéder à la diffusion du document. Cette diffusion s'effectue à l'aide de la plate-forme SDX (**pour installer le module de publication, reportez-vous au site www.cyberdocs.org/**)

3.2.3. Convertir un document sans le module de gestion : en ligne de commandes

Le module de conversion Cyberdocs effectue globalement une tâche de conversion finalement assez complexe, car la variabilité des documents en entrée suppose une grande souplesse dans les traitements à effectuer, d'où une certaine complexité. C'est pourquoi cette tâche de globale de conversion a été divisée en plusieurs opérations ou étapes qui sont expliquées ici.

Bien comprendre ces opérations est essentiel pour l'opérateur de conversion, car même si on cherche à l'éviter, il peut lui être nécessaire d'intervenir manuellement sur un document produit à l'une ou l'autre des étapes.

3.2.3.1. Opérations et étapes de conversion

Chapeau : Toutes les opérations effectuées par le module de conversion, expliquées étape par étape.

Une manière un peu plus détaillée de présenter la conversion est de préciser les étapes effectuées. Dans la liste ci-dessous, le nom de code qui figure correspond à la tâche ANT qui permet d'exécuter cette étape seulement. Les dossiers mentionnés sont toujours créés à l'intérieur du dossier de travail d'un document.

etape_oo

Cette étape va permettre de convertir le fichier source en un fichier XML grâce à l'outil OpenOffice.org. Le dossier oo est créé et contient le résultat de ce traitement.

etape_01

Cette étape va récupérer la sortie de l'étape précédente et faire un premier traitement qui permet de bien identifier ce qui sera traité ou exclu. Ainsi, on va donner les bons noms de style aux éléments, vérifier que les styles obligatoires sont bien présents, exclure les éléments OpenOffice.org non traités. Le résultat est envoyé dans le dossier prod/01 .

etape_02

Cette étape va récupérer la sortie de l'étape précédente et transformer chaque élément en élément TEILite, en fonction de son nom ou du style qui lui est associé. Le produit de cette opération, prod/02, est ensuite traité par l'étape **etape_02-bis** qui va simplement numéroter chaque élément afin de pouvoir les regrouper ultérieurement. Le répertoire résultat est placé dans le dossier prod/02-bis.

etape_03

Cette étape va récupérer la sortie de l'étape précédente et hiérarchiser les éléments en fonction des niveaux de titre. Le dossier prod/03 contient le résultat de cette étape.

etape_04

Enfin, cette étape va transformer la sortie de l'étape précédente en un document TEI, et générer le document XML final dans le répertoire xml.

Une fois l'étape **etape_04** effectuée, nous avons un document XML TEILite qui servira de référence ; il s'agit en quelque sorte du véritable produit de ce module de conversion. Toutefois, les traitements peuvent se poursuivre, afin de valider ce document ou d'obtenir des produits dérivés. Les étapes décrites ci-dessous effectuent ces traitements.

metadonnees

Cette tâche permet de créer le fichier de métadonnées en format texte, qui pourra être récupéré par le module de gestion. Le fichier produit est dans le dossier xml.

html

Cette tâche permet de générer deux versions HTML statiques du document. Une première, située dans le dossier html, est une version qui peut être utilisée dans n'importe quel contexte, pour une publication statique. La seconde, semblable mais sans la table des matières dynamique, est une version destinée à l'impression, et elle sera utilisée par le module de publication. Elle se trouve dans le dossier html/sdx.

Toute cette opération est effectuée une seconde fois, pour produire deux versions en format XHTML. Ces versions sont parfaitement en phase avec les standards du W3C, avec notamment des équations mathématiques incluses en format MathML, et le navigateur Mozilla permet de les consulter aisément. Ces versions sont dans le dossier xhtml.

pdf

Cette tâche permet de générer une version PDF du document, qui sera placée dans le dossier pdf.

web

Cette tâche permet de créer l'ensemble des documents qui seront utilisés par le module de publication (version XML TEILite, version HTML pour impression, version PDF).

Les principales opérations à connaître

La section précédente décrit toutes les étapes de conversion, mais les outils prévoient également des *opérations*, qui sont des séquences de tâches permettant d'effectuer un traitement, depuis un point de départ (par exemple le document source) vers un point d'arrivée (le document de référence par exemple). Le fichier bin/taches.xml dans le dossier d'installation du module de conversion (ou le fichier src/oo-vers-tei/bin/taches.xml dans les sources Cyberdocs) constitue la référence ultime des opérations de conversion (et des étapes également), mais nous reprenons les principales informations ici. En début d'item, en caractère gras, se trouve le nom de la tâche ANT qui effectue l'opération.

tout : Traitement complet, depuis la source vers le document de référence, ainsi que les produits dérivés.

Production du document de référence :

- **document-reference** : depuis le document de traitement de texte

- **depuis-oo** : depuis le document OpenOffice.org
- **depuis-xslt-01** : depuis la fin de la première étape de transformation XSLT
- **depuis-xslt-02** : depuis la fin de la seconde étape de transformation XSLT
- **depuis-xslt-03** : depuis la fin de la troisième étape de transformation XSLT
- **depuis-xslt-04** : depuis la fin de la quatrième étape de transformation XSLT
- **validation** : validation du document de référence

A partir du document de référence :

- **produits-derives** : génération de tous les produits dérivés
- **metadonnees** : génération d'un fichier de métadonnées minimal en format texte
- **html** : génération des versions HTML et XHTML statiques
- **pdf** : génération des versions PDF
- **web** : production de tous les fichiers nécessaires pour le module de publication

3.2.3.2. Préparer un espace de travail

Chapeau : Les dossiers et documents à préparer avant d'utiliser d'appliquer les opérations du module de conversion sur un document.

Afin de simplifier le nombre de paramètres lors d'une opération de conversion, le module de conversion a été programmé en fonction d'une structure de répertoires précise. De plus, les documents qu'il va produire seront toujours situés dans les mêmes répertoires par rapport au répertoire de travail d'un document. Nous allons explorer ici l'organisation générale du dossier de production puis l'organisation spécifique du dossier d'un document.

1. Organisation du dossier de production

Le dossier de production est un dossier nommé `production` dans le dossier d'installation du module de conversion. Ce dossier contiendra tous les documents à convertir, de même que tous les documents issus du processus de conversion.

L'organisation de ce dossier de conversion suit une logique immuable : les documents sont organisés sur trois niveaux, le premier étant une institution, le second une année d'édition, et enfin le troisième un document à traiter. Cette organisation relativement rigide pourrait évoluer dans le futur.

Exemple 1. Contenu simplifié d'un dossier de production

```
production up.bat up.sh lyon2 lyon2.css 2000 dieng_sa.bat
           dieng_sa.sh dieng_sa sources dieng_sa-md.xml dieng_sa.css dieng_sa.doc images
           images-dieng_sa.zip
```

L'exemple ci-dessus nous montre une partie du dossier de production tel qu'il est lorsqu'on installe le module de production. Nous montrons dans cet exemple les fichiers et dossiers relatifs à un seul document. Directement dans le dossier de production se trouve un script `up.{bat|sh}` qui va appeler le script de démarrage du module de conversion.

On voit tout d'abord un premier niveau, le dossier `lyon2`, qui correspond au **code d'institution** de l'éditeur du document. Ce dossier contient une feuille de style CSS qui lui est spécifique ; cette feuille de style doit être présente, quitte à copier celle qui est fournie avec la plate-forme. Le nom du fichier doit être le code de l'institution avec l'extension `.css`.

Au second niveau, on retrouve un dossier `2000` qui correspond à cette **année de production**. S'y retrouve également un script `dieng_sa.{bat|sh}` qui permet de démarrer la conversion du document dont le code est le même que le nom du script.

Au troisième niveau se trouve un dossier `dieng_sa` qui correspond au **code du document** à traiter. Dans ce dossier se trouve un dossier `sources` où doit nécessairement se trouver la source de traitement de texte à convertir. Si ce document possède des illustrations externes, elles doivent être dans le sous-dossier `images`, soit regroupées dans un ou des fichiers ZIP ou directement dans ce dossier.

Lorsque l'un ou l'autre des étapes de conversion aura été exécutée, d'autres dossiers et fichiers seront ajoutés, mais ils seront tous situés à l'intérieur du dossier du document, soit le dossier `dieng_sa` dans l'exemple ci-dessus.

2. A faire avant de traiter un nouveau document

Lorsque l'on souhaite traiter un nouveau document, les opérations suivantes doivent être effectuées :

- S'il s'agit d'un document provenant d'une nouvelle institution, créer un dossier pour cette institution dans le dossier `production`, et à l'intérieur de celui-ci ajouter un fichier de style CSS qui porte le même nom que le dossier mais avec une extension `.css`.
- S'il s'agit d'une nouvelle année de publication, créer un dossier pour cette année à l'intérieur du dossier de l'institution qui édite le document.

Dans le dossier de l'année de publication, créer un dossier dont le nom sera l'identifiant du document tout au long de son parcours dans la plate-forme Cyberdocs. Cet identifiant est très important, et il ne devrait pas contenir de caractères spéciaux (accents, ponctuation, etc.).

Dans le dossier du document, créer un dossier `sources` et déposer à cet endroit le document de traitement de texte qui sera converti.

- Si vous avez des métadonnées pour la publication, placer un fichier `[identifiant du document]-md.xml` dans le dossier `sources`.
- S'il y a des illustrations externes au document, créer un dossier `images` à l'intérieur du dossier `sources` et placer les illustrations à cet endroit.

Dans le dossier de l'année d'édition, copier des scripts d'exécution fournis avec les exemples (par exemple `essai.{bat|sh}`) et modifier ces scripts pour y passer les bons paramètres, en particulier l'identifiant du document qui doit être le second paramètre. Pour cet aspect, une [section ultérieure](#) donne plus d'indications.

Nous rappelons que si vous avez installé le module de gestion, ces manipulations sont grandement simplifiées, puisque vous avez seulement à demander la création d'un espace pour un document et ensuite d'y télécharger les fichiers sources.

3.2.3.3. Préparer un document

Chapeau : Description et rôle des différents styles utilisés dans les documents.

La plate-forme Cyberdocs est en mesure de traiter tout document de traitement de texte. Toutefois, afin de rendre les outils encore plus intéressants, une phase de préparation des documents peut être envisagée. Cette phase consiste essentiellement à appliquer des styles aux différents paragraphes du document, et à nettoyer quelque peu leur contenu.

Pour Cyberdocs, un style est identifié par un *code* et non par un nom. C'est pourquoi il est possible d'utiliser des feuilles de styles différentes dans le traitement de texte, tout en obtenant un traitement adéquat avec le module de conversion. Pour utiliser votre propre feuille de styles, vous devez tout simplement modifier le fichier `outils/xslt/utiles/styles.xml` pour y inclure les noms de vos styles pour chaque style supporté par Cyberdocs (voir liste ci-dessous).

Liste des styles supportés par Cyberdocs

Cette liste donne le code (défini dans le fichier outils/xslt/utiles/styles.xml), la structure XML produite par la présence de ce style, ainsi que le nom du style correspondant dans la feuille de styles *lyon2*.

Tableau 1. Liste des styles

Code de feuille de style	Élément xml correspondant	Nom correspondant pour la feuille de style lyon2
auteur	<docAuthor>	1 Auteur
copyright	<titlePart type="copyright">	1 Copyright
dedicace	<div type="dedicace">	1 Dedicace
depot	<docDate>	1 Depot
dept	<titlePart type="dept">	1 Dept
directeur	<titlePart type="directeur">	1 Directeur
discipline	<titlePart type="discipline">	1 Discipline
ecole-doct	<titlePart type="ED">	1 EcoleDoct
epigraphe	<epigraph>	1 Epigraphe
faculte	<titlePart type="faculte">	1 Faculte
grade	<titlePart type="grade">	1 Grade
jury	<titlePart type="jury">	1 Jury
no-officiel	<titlePart type="Reference">	1 NoOfficiel
sous-titre	<docTitle>/<titlePart type="sub" lang="fr">	1 Sous-titre
titre-front	<div type="***">	1 TitreFront
titre-these	<docTitle>/<titlePart type="main" lang="fr">	1 TitreThese
universite	<titlePart type="univ">	1 Universite
ann-titre	----- -----	3 Ann_titre
ann-titre1**9	<div type="appendix">	3 Ann_titre1**9
bibli-item	<bibl>	3 Bibli_item
bibli-tit	----- -----	3 Bibli_tit
bibli-tit1**4	<div type="bibl">	3 Bibli_tit1**4
citation	<q rend="block">	Citation
citation-bloc1**2	<q rend="block">	CitatioBloc1**2
closer	<closer>	closer
conclu	<div type="conclusion">	Conclu
entree	<list type="gloss">/<item>	Entree
figure	<figure>	Figure
intro	<div type="***">	Intro
legende-fig	<* id="fig">/<head>	LegendeFig
legende-tab	<* id="tab">/<head>	LegendeTab
liste-num	----- -----	ListeNum
liste-num1	<list type="ordered">/<item>	ListeNum1
liste-num2	<list type="ordered">/<item>	ListeNum2
liste-puce	----- -----	ListePuce
liste-puce1**8	<list type="bulleted">/<item>	ListePuce1**8

Code de feuille de style	Élément xml correspondant	Nom correspondant pour la feuille de style lyon2
liste-simple	<list type="simple">/<item>	ListeSimple
liste-titre	<list>/<head>	ListeTitre
partie	<div type="part">	Partie
source	<note place="interlinear">	Source
caption	<head>	Caption
heading1**9	<div>/<head>	heading 1**9
image-ligne	<figure>	ImageLigne
image-tab	<figure>	ImageTab

3.2.3.4. Effectuer une conversion

Chapeau : Les scripts à exécuter pour convertir un document, ainsi que leurs paramètres potentiels.

Le module de conversion peut être utilisé à l'aide d'une cascade de scripts qui s'appellent les uns les autres. Le plus simple est de copier un exemple existant, mais nous allons présenter ici les détails en référence.

La cascade peut être schématisée ainsi :

- Un **script spécifique à un document** appelle un script général de conversion en lui passant des paramètres qui déterminent où se trouvent les documents à convertir.
- Le **script général de conversion** appelle le script de démarrage de ANT en spécifiant la tâche à exécuter ainsi que des variables d'environnement qui correspondent aux paramètres du script précédent contextualisés.
- Le **script de démarrage de ANT** va préparer l'environnement Java en définissant un CLASSPATH approprié, puis va démarrer une machine virtuelle Java en lui demandant d'exécuter ANT.

Comme on peut le constater, seul le premier script doit être modifié lorsqu'on traite un nouveau document ; les deux autres sont génériques et n'ont pas à être modifiés d'un document à l'autre. C'est pourquoi, lors de la préparation d'un document, il est seulement nécessaire de copier un script spécifique à un document et de le modifier pour passer les bons paramètres.

1. Le script de démarrage de ANT

Ce script est correspond au fichier `outils/bin/oo-vers-tei.{bat|sh}` dans le dossier d'installation du module de conversion. Ce script va lui-même appeler le script `ant.{bat|sh}` qui est fourni avec ANT.

Si on regarde la ligne de commande principale de ce script, on remarque qu'elle est très simple :

```
call ant -logger org.cyberdocs.ant.HTMLLogger -buildfile
oo-vers-tei.xml %1 %2 %3 %4 %5 %6 %7 %8 %9
```

Essentiellement, le script définit deux paramètres importants et passe les autres paramètres qu'il reçoit. Ces deux paramètres sont :

-logger

Il s'agit du type de journalisation (*logging*) que l'on souhaite obtenir. En effet, afin de rendre le module de gestion très intuitif et auto-documenté, un type de journalisation spécifique à la plate-forme Cyberdocs, en format HTML, a été défini. C'est d'ailleurs celui qui est utilisé par défaut, tel qu'on le voit dans l'exemple ci-dessus. Si on souhaite utiliser le type de journalisation

par défaut (qui consiste à sortir des messages standards en format texte), on peut supprimer la partie **-logger org.cyberdocs.ant.HTMLLogger** dans cette commande.

-buildfile

ANT fonctionne à partir d'un document XML où se trouvent toutes les tâches possibles et leurs dépendances. Ce paramètre permet de le spécifier, et dans le cas de Cyberdocs il n'y aucune raison de le modifier.

Sauf pour modifier le type de journalisation, il n'y a donc aucune raison de modifier ce script.

2. Le script générique de conversion

Ce script est le fichier `production/up.{bat|sh}` situé dans le dossier d'installation du module de conversion. Ce script va appeler le script de démarrage de ANT avec une commande telle que celle-ci :

```
call oo-vers-tei.bat
    "-Ddossier.production=..\production" "-Dnom.fichier=%2"
    "-Dcode.institution=%3" "-Dcode.doc=%4" "-Dcode.styles=%5" "-Dlangue.doc=%6"
    "-Dannee.soutenance=%7" %1
```

Dans cet exemple, la ligne de commande a été scindée en plusieurs lignes pour faciliter la lecture.

Comme on peut le constater, ce script va envoyer essentiellement des paramètres de la forme `-D[nom du paramètre]=[valeur du paramètre]`. Cette syntaxe est importante, car elle est exploitée lors du démarrage de Java et permet de spécifier des variables d'environnement. Ces variables peuvent ensuite être exploitées dans un script ANT, ce qui est effectivement le cas dans `outils/bin/oo-vers-tei.xml`. Ainsi, si vous voulez envoyer d'autres paramètres au script ANT, ou même si vous voulez redéfinir des propriétés utilisées dans ce script, vous pouvez le faire à la ligne de commande comme dans l'exemple ci-dessus, sans modifier le fichier `outils/bin/oo-vers-tei.xml`. Pour en savoir plus, consulter la [documentation de ANT](#).

La signification de ces variables d'environnement sera indirectement expliquée dans la section suivante.

3. Le script spécifique à un document

Ce script est par exemple le fichier `production/lyon2/2002/essai.{bat|sh}`. La principale ligne de commande est semblable à celle-ci :

```
call up.bat tout essai.doc lyon2 essai lyon2 fr 2002
```

On voit donc que ce script ne fait qu'appeler le script générique, en lui passant sept paramètres qui sont :

- *tout* est l'une des tâches du module de conversion, tel qu'expliqué dans une [autre section](#) ou détaillé dans le fichier `outils/bin/taches.xml` dans les sources Cyberdocs.
- *essai.doc* est le nom de fichier du document à convertir, fichier qui doit se trouver dans le dossier sources.
- *lyon2* est le code de l'institution qui édite le document, et doit donc correspondre au nom d'un dossier parent.
- *essai* est l'identifiant du document, ce qui doit correspondre au nom du dossier propre à ce document.
- *lyon2* (le second) représente le code de feuille de styles utilisée, et ce code doit être défini dans le fichier `outils/xslt/utiles/styles.xml`.
- *fr* est le code ISO de la langue principale du document.

- 2002 correspond à l'année de publication du document, ce qui doit correspondre au nom du dossier où se trouve le dossier du document.

Afin que le script fonctionne correctement, il faut veiller à ce que l'ordre de ces arguments ne soit pas modifié.

3.2.3.5. Résultats de la conversion

Chapeau : Description des différents documents et fichiers produits lors de la conversion d'un document.

A la suite d'une opération de conversion, un certain nombre de fichiers et de dossiers seront ajoutés ; ces fichiers et dossiers se retrouvent tous à l'intérieur du dossier du document, et toujours à l'extérieur du dossier contenant les sources (dossier sources. Voici quelques informations à propos de ces dossiers et de leur contenu.

- Le répertoire oo contient le résultat de la conversion du document traitement de texte par OpenOffice.org.
- Le répertoire prod lui-même découpé en 4 sous-répertoires : 01, 02, 02-bis, 03, contenant respectivement le résultat de la conversion après les première (01), deuxième (02 et 02-bis) et troisième (03) étapes de conversion.
- Un répertoire xml contenant :
 - o le document final, résultat de la conversion, au format xml (dans notre exemple dieng_sa.xml)
 - o la DTD TEILite associée
 - o un fichier texte contenant quelques métadonnées (dans notre exemple dieng_sa-meta-auto.txt)
 - o un fichier notes.xml contenant les notes de bas de page du document, ce document étant utilisé par le module de publication pour optimiser l'affichage.
 - o un fichier config.xml indiquant des informations de configuration pour ce document, essentiellement pour l'instant le nombre de niveaux de la table des matières
 - o le fichier de métadonnées XML situé au départ dans le répertoire sources (dans notre exemple dieng_sa-md.xml)
 - o un fichier Javascript tocTab.js contenant la table des matières du document
 - o un répertoire ressources dans lequel sont stockées les images du document
 - o un répertoire formules contenant les formules présentes dans le document au format MathML.
- Le répertoire html contient une version HTML statique du document, la page de départ étant index-frames.html, et des versions HTML imprimables - du document et de ses parties - dans le sous-répertoire sdx.
- Le répertoire xhtml contient une version XHTML statique du document, avec les mêmes remarques que pour la version HTML.
- Le répertoire pdf contient une version PDF du document et de ses parties, en plus des fichiers XSL-FO qui ont servi à les produire.

Parmi tous ces fichiers, un certain nombre peuvent être supprimé après conversion, soit :

- Le dossier prod car il contient des fichiers de travail pour les étapes intermédiaires de conversion.
- Les fichiers .fo dans le dossier pdf car ils ont seulement servi à produire le PDF.

3.2.4. Module de publication

Le module de publication de la chaîne Cyberdocs est une application Web dynamique bâtie avec la plate-forme SDX. Les principales fonctionnalités sont la recherche et la consultation des documents publiés. Dans cette documentation, nous allons décrire brièvement les fonctionnalités présentes et leurs principales caractéristiques.

3.2.4.1. La gestion et l'identification des utilisateurs

Chapeau : La gestion et l'identification des utilisateurs dans SDX et dans le module de publication Cyberdocs en particulier.

Le module de publication nécessite une authentification des utilisateurs pour permettre d'ajouter (ou *indexer*) un document. Dans cette partie, nous présentons comment créer des utilisateurs et comment s'identifier.

Dans la configuration initiale de l'application, il existe un utilisateur qui fait partie du groupe `admins` : son code est `admin` et il n'a pas de mot de passe. Comme tous les utilisateurs de ce groupe, il a le droit d'ajouter ou supprimer des documents dans le module de publication.

1) Gestion des utilisateurs

Le module de publication ne présente pas de fonctionnalités propres à la gestion des utilisateurs. Pour ce faire, il faut plutôt utiliser l'interface d'administration de SDX, qui possède de telles fonctionnalités. Pour y accéder, allez à l'URL **`sdx/admin/index.xsp`** de votre installation SDX (attention, cela peut donner une URL avec deux fois *sdx*, c'est normal!), puis cliquez sur le bouton *Administrer* à côté de l'application Cyberdocs. Vous pouvez à ce moment saisir le code et le mot de passe d'un administrateur (groupe `admins`) , puis cliquer sur le bouton *Identification* pour gérer les comptes utilisateurs.

2) Identification

Pour s'identifier dans l'application Cyberdocs et avoir des privilèges administrateurs, il faut aller à l'URL **`identification.xsp`** de l'application.

3.2.4.2. La gestion des documents

Chapeau : La démarche à suivre pour indexer et supprimer des documents dans le module de publication.

Le module de publication de Cyberdocs est une application SDX. Pour cet outil de recherche, les documents publiés n'ont une signification que s'ils sont *indexés*, c'est pourquoi nous allons parler de l'indexation et de la suppression des documents.

Ces fonctions sont bien entendu réservées aux administrateurs de l'application. Pour mettre en place cette restriction, le module de publication utilise les fonctionnalités de [gestion des utilisateurs](#) disponibles avec la plate-forme SDX. Pour indexer ou supprimer des documents, vous devez préalablement vous identifier comme un administrateur de l'application.

1) Indexer des documents

L'administrateur de l'application, une fois identifié, a la possibilité de charger (d'indexer) les documents, afin qu'ils soient recherchables et consultables via l'application SDX.

Pour ce, deux formulaires sont à sa disposition, ces deux formulaires étant situés dans la même page accessible depuis l'URL **`chargement.xsp`** :

Pour indexer des documents qui sont stockés sur un disque directement accessible par le serveur

Formulaire d'indexation par institution, année, code de document :

Répertoire de base contenant les documents :

Code de l'institution :

Année de publication :

Identifiant du document :

Document non XML :

Dans ce cas, une zone de texte pré-remplie indique l'emplacement du dossier de production du module de conversion. Vous pouvez conserver cette valeur ou la modifier si vos documents sont ailleurs. Il n'y a plus qu'à renseigner le code de l'institution, l'année de publication et l'identifiant du document. Ces informations doivent correspondre à des sous-dossiers dans le dossier de production.

Si le document n'est pas disponible en format XML (par exemple s'il est en format PDF), il faut cocher le bouton **document non-xml**.

Pour indexer des documents situés sur un serveur distant:

Formulaire d'indexation par URL complète :

URL du document à indexer :

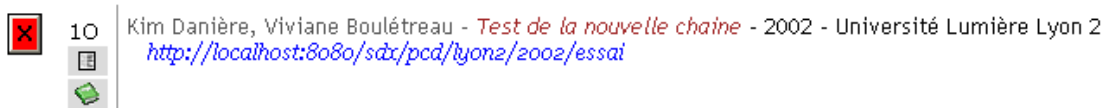
Dans ce cas, il faut saisir complètement l'adresse URL complète du document XML, par exemple **http://monserveur.org/mes-documents/essai/xml/essai.xml**. S'il s'agit d'un document non disponible en format XML, il faut fournir l'URL de son fichier de métadonnées, par exemple : **http://monserveur.org/mes-documents/essai/xml/essai-md.xml**.

Il est à noter qu'il est nécessaire que le document soit situé dans un dossier qui respecte la structure du dossier xml produit par le module de conversion.

2) Supprimer des documents

Un administrateur peut supprimer un document indexé, cela ayant pour conséquence qu'il ne peut plus être trouvé par une recherche ni afficher. Pour supprimer, il suffit de cliquer sur la petite croix qui apparaît à côté d'un document dans un résultat de recherche ; vous devez donc

préalablement effectuer une recherche, par la méthode qui semble la plus opportune, pour trouver le document à supprimer.



A noter que ce bouton de suppression n'est disponible que pour les documents complets, par pour les parties de documents (les chapitres par exemple). Par contre, si on supprime un document, toutes ses parties sont également supprimées, pour conserver ainsi toute la cohérence souhaitée.

3.2.4.3. L'utilisation des métadonnées du document

Chapeau : Le format de métadonnées attendu par le module de publication et les traitements effectués avec ces informations.

Les métadonnées associées au document sont exploitées par le module de publication, et elles lui sont même essentielles. Peu importe comment elles sont produites (le module de gestion inclut une interface Web pour les saisir), elles doivent être sous la forme d'un document XML et ce document doit être dans un fichier situé dans le même dossier que le document XML principal, et ce fichier doit avoir le même nom que le document principal, mais avec un `-md` avant le `.xml`.

Les éléments de métadonnées font partie du [Dublin Core Metadata Element Set](#) ou du [Metadata Standard for Electronic Theses and Dissertations](#) (ETDMS). D'autres formats pourraient éventuellement être supportés.

1) Utilisation des métadonnées

Les métadonnées sont exploitées de différentes manières par le module de publication :

- La fiche sommaire du document, que l'on peut afficher dans sa propre fenêtre, est créée à partir des métadonnées.
- Les résultats de recherche sont constitués à partir des métadonnées.
- La plupart des champs de recherche du formulaire de recherche avancée exploitent l'une ou l'autre des métadonnées.
- La fourniture d'information par un entrepôt OAI se fait à partir des métadonnées.

2) Format des métadonnées

L'exemple ci-dessous constitue une référence pour les métadonnées traitées par Cyberdocs.

Exemple 1. Métadonnées supportées par Cyberdocs

```
<doc
  <!-- quelques namespaces -->
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:md-cyb="http://cyberdocs.org/md/1.0"
```

```

xmlns:etdms="http://www.ndltd.org/standards/metadata/etdms/1.0/"
>

<!-- l'identifiant unique (institution.annee.code du document) -->
<md-cyb:identifiant>lyon2.2002.essai</md-cyb:identifiant>

  <!-- Les trois parties de l'identifiant unique -->
  <md-cyb:identifiant-doc>essai</md-cyb:identifiant-doc>
  <md-cyb:annee>2002</md-cyb:annee>
  <md-cyb:code-institution>lyon2</md-cyb:code-institution>

    <!-- L'identification de l'outil utilisé pour produire / diffuser le
document. -->
      <md-cyb:plate-forme>Plate-forme                               CyberThèses
[http://sourcesup.cru.fr/cybertheses/]</md-cyb:plate-forme>

    <!-- Version de l'outil -->
    <md-cyb:version-plate-forme>1.0</md-cyb:version-plate-forme>

    <!-- L'auteur -->
    <dc:creator>Kim Danière, Viviane Boulétreau</dc:creator>

    <!-- Le titre dans la langue originale, avec son sous-titre -->
    <dc:title xml:lang="fr">Test de la nouvelle chaine</dc:title>

    <!-- Un titre dans une autre langue -->
    <dcterms:alternative xml:lang="en">Test</dcterms:alternative>

    <!-- Le directeur de thèse -->
      <dc:contributor etdms:role="directeur">Sous           notre           propre
direction</dc:contributor>

    <!-- Les membres du jury -->
    <dc:contributor etdms:role="jury">Martin           Sévigny</dc:contributor>
    <dc:contributor etdms:role="jury">Jean-Paul Ducasse</dc:contributor>

    <!-- Le diplôme, le lieu de soutenance, etc. -->
    <etdms:degree>

      <!-- Le diplôme -->
      <etdms:level>Thèse pour obtenir le grade de CyberThéséur</etdms:level>

      <!-- La discipline -->
      <etdms:discipline>Discipline : Edition Electronique</etdms:discipline>
      <!-- Le lieu de soutenance, en différents niveaux -->
      <etdms:grantor md-cyb:role="universite">Université Lumière Lyon
2</etdms:grantor>
      <etdms:grantor md-cyb:role="ecole-doctorale">Ecole Doctorale
tagada</etdms:grantor>
      <etdms:grantor md-cyb:role="faculte">Division ERAD</etdms:grantor>
      <etdms:grantor md-cyb:role="departement">Cellule Edition
Electronique</etdms:grantor>
    </etdms:degree>

    <!-- La date de soutenance -->
    <dcterms:dateAccepted>2002-11-27</dcterms:dateAccepted>

    <!-- Les résumés -->
    <dcterms:abstract xml:lang="fr">Un résumé en français</dcterms:abstract>
    <dcterms:abstract xml:lang="en">An English abstract</dcterms:abstract>

```

```

<!-- Les mots clés -->
<dc:subject xml:lang="fr">[Test ; conversion ; cyberthèses]</dc:subject>

<!-- L'éditeur -->
<dc:publisher>CyberThèses</dc:publisher>

<!-- Le format -->
<dc:format>text/xml</dc:format>

<!-- La langue -->
<dc:language>fr</dc:language>

<!-- Les différents droits -->
<dc:right md-cyb:role="copyright">[Copyright 2002 CyberThèses - Université
Lumière Lyon 2</dc:right>

<!--valeurs possibles pour la diffusion : rien ou [intranet]-->
<dc:right md-cyb:role="diffusion"></dc:right>
<dc:right md-cyb:role="commentaires-diffusion"> [Cette thèse est en accès
restreint pour
protéger la vie privée des personnes qui sont mentionnées]</dc:right>

<!-- L'URL de référence -->
<dc:identifiant>http://localhost:8080/sdx/pcd/lyon2/2002/essai</dc:identifiant>

<!-- Les mots clés temporels -->
<dcterms:temporal xml:lang="fr"/>

<!-- Les mots clés géographiques -->
<dcterms:spatial xml:lang="fr"/>

<!-- La taille du fichier (en octets) -->
<dcterms:extent>[13416]</dcterms:extent>
</doc>

```

3.2.4.4. La consultation des documents

Chapeau : Informations sur la fenêtre de consultation des documents et ses différentes options.

Le module de publication de Cyberdocs inclut une fonction de consultation relativement sophistiquée, particulièrement adaptée à la consultation de documents longs et structurés, par exemple des thèses de doctorat. Nous allons expliquer les principes derrière cette méthode de consultation, de même que les différentes fonctionnalités associées à la consultation.

1) Les modes de représentation des documents

Dans le module de publication Cyberdocs, ces différents modes de consultation des documents sont disponibles :

- Un format bref, utilisé dans les [résultats de recherche](#), qui contient les principales coordonnées bibliographiques du document (titre, auteur, date de publication, etc.).
- Les [métadonnées](#) du document, affichée dans une fenêtre spécifique sous la forme d'un tableau.
- Des [versions pour impression](#) des documents.
- La fenêtre de consultation structurée du document et de ses parties.

Dans cette partie de la documentation, c'est seulement ce dernier mode qui nous intéresse ; pour les autres aspects vous êtes invités à suivre les liens respectifs pour consulter d'autres parties de la documentation.

2) La fenêtre de consultation de structurée

La fenêtre de consultation structurée est divisée en trois cadres. Le cadre supérieur présente les informations bibliographiques principales du document, des informations sur la recherche effectuée pour arriver à ce document, de même qu'un formulaire de recherche simple.

A noter qu'il s'agit d'une fenêtre spécifique à cette fonctionnalité dans le module de consultation. Ainsi, lorsqu'une recherche est effectuée et que l'on souhaite consulter un document, celui-ci s'affiche dans une fenêtre différente de celle des résultats. Toutefois, tous les documents sont consultés dans la même fenêtre, pour éviter une multiplicité des fenêtres ouvertes.

2.1) Le cadre des informations bibliographiques

Les informations fournies dans ce cadre proviennent des métadonnées associées au document.

Ce cadre offre également des fonctionnalités de recherche et de navigation dans le document présenté. Premièrement, les mots recherchés sont mis en évidence dans le document présenté. Si on veut passer d'une occurrence d'un mot recherché à une autre, il suffit d'utiliser les deux boutons prévus à cet effet :



Ensuite, vous pouvez également effectuer une recherche dans le document ou l'une de ses parties en utilisant le formulaire de recherche simple que l'on retrouve dans ce cadre :



Cela aura pour effet de présenter une liste de résultats dans le cadre du document, résultats qui ne contiendront que des parties du document que vous consultez.

2.2) Le cadre de table des matières

Le cadre de table des matières se situe dans la partie gauche de la fenêtre de consultation. Il offre une table des matières interactives qui permet de sélectionner rapidement, en pliant ou dépliant des parties, la section du document que l'on souhaite consulter.

Normalement, trois tables des matières différentes sont présentées :

1. Une table des matières construite à partir des titres du document.
2. Une table des illustrations, construite à partir des titres d'illustrations.
3. Une liste des tableaux, construite à partir des titres de tableaux.

2.3) Le cadre du document

Le cadre principal, situé dans la partie droite de la fenêtre, permet de lire le document. A noter que si vous cliquez sur l'icône situé au début d'un titre, la table des matières sera synchronisée, c'est-à-dire que le titre en question sera affiché dans le cadre de table des matières.

3.2.4.5.

La recherche

Chapeau : Les différents modes de recherche, la manière de les utiliser, le type de résultats attendus.

La recherche constitue une fonctionnalité centrale dans le module de publication. Il existe essentiellement quatre façons d'effectuer des recherches, que nous présentons ci-dessous. Peu importe la manière de les effectuer, les [résultats de recherche](#) se présentent toujours sous la même forme.

1) Recherche *simple*

La recherche simple est celle qui consiste à utiliser la zone de texte présente en haut de chaque page pour y exprimer une requête de recherche. La plupart du temps, cette requête sera constituée d'un ou plusieurs mots, mais un [langage de recherche plus évolué](#) est en fait disponible. Par exemple, si vous effectuez la recherche *afrique agriculture*, les deux mots ne seront pas obligatoires (OU booléen implicite) mais si vous effectuez la recherche *afrique +agriculture*, le mot *agriculture* sera obligatoire.

Lorsqu'on effectue de telles recherches, la zone de recherche est le texte intégral des documents.

2) Recherche *avancée*

On accède à la recherche avancée en cliquant sur le lien *Recherche avancée* situé dans le bandeau supérieur de chaque page. On obtient alors un formulaire qui permet de saisir plusieurs critères de recherche et de les relier par les opérateurs booléens ET ou OU.

Les premières zones du formulaire permettent de saisir des mots et de les rechercher dans les principales métadonnées des documents. Ensuite, les zones suivantes permettent de rechercher dans des index précis.

3) Recherche *par listes*

La recherche par listes consiste à rechercher des documents en obtenant une liste en fonction de certains critères. Pour l'instant, le seul critère disponible est *Universités et facultés*, particulièrement bien adapté pour les thèses. Pour activer cette liste, il s'agit de choisir cette entrée dans le menu déroulant intitulé *Choisir un index*, présent dans le haut de toutes les pages.

Le résultat sera une liste d'universités et de facultés, et pour chacune les documents qui y sont publiés.

4) Recherche *dans un document*

Lorsqu'on consulte un document, on a la possibilité d'effectuer une recherche à l'intérieur de ce document seulement. Pour ce faire, il suffit d'utiliser la zone de texte dans le cadre supérieur de la fenêtre de consultation des documents. On peut y effectuer des recherches de la même manière que pour la recherche simple présentée ci-dessus.

Les résultats seront présentés dans le cadre principal de consultation du document.

3.2.4.6.

Les résultats de recherche

Chapeau : Informations au sujet de la page de présentation des résultats de recherche, de la navigation dans ces résultats, etc.

Les résultats de recherche sont toujours présentés de la même manière, que l'on arrive de la recherche simple, la recherche avancée, les listes ou la recherche dans un document.

Pour chaque résultat de recherche, une représentation sommaire du document incluant les informations suivantes :

- L'auteur du document
- Le titre du document, et ses éventuels sous-titres

- L'année de publication
- L'éditeur (par exemple l'université de soutenance pour les thèses)
- L'adresse du document

Cette dernière information, l'adresse, est cliquable pour accéder à la fenêtre de consultation.

Ensuite, à gauche de cette présentation sommaire, on retrouve le numéro de ce résultat dans la liste, de même qu'un lien pour consulter la fiche (métadonnées) et un autre pour consulter le document.

3.2.4.7. Les versions pour impression

Chapeau : Comment identifier des versions pour impression des documents, et aussi comment les obtenir à l'écran.

On peut obtenir, depuis le module de publication, des versions pour impression des documents que l'on consulte. Pour ce faire, on doit aller dans la fenêtre de consultation du document, puis cliquer sur l'icône en forme d'imprimante en haut à droite de la fenêtre. On obtient alors une fenêtre qui nous permet de choisir le document entier ou la partie du document à envoyer, de même que le format (HTML ou PDF).

Il ne reste plus qu'à utiliser les possibilités d'impression de votre navigateur et de votre poste de travail.

3.3. Aide aux développeurs

3.3.1. Signaler des bogues

Tout signalement d'erreurs est bien évidemment le bienvenu !

Pour cela, nous souhaitons que les discussions se fassent par l'intermédiaires des [listes de discussions](#). Vous n'avez pas besoin d'être un développeur confirmé pour signaler des erreurs : des problèmes dans la documentation, des soucis d'installation ou de conversion de document, des liens erronés sur notre site... les améliorations de Cyberdocs ne peuvent se faire sans ses utilisateurs.

3.3.2. Contribuer à Cyberdocs

Pour contribuer au développement de la plate-forme, vous devriez d'abord vous faire connaître auprès de la communauté des développeurs en vous abonnant à la liste de discussions [cybertheses-dev](#). Un [formulaire en ligne](#) vous permet de le faire aisément.

3.4. Téléchargements de Cyberdocs

4. Téléchargement

Cyberdocs est un logiciel libre qui peut être téléchargé sur Internet. Pour les impatientes, voici les deux liens essentiels pour récupérer la plate-forme, que ce soit une version distribuée ou les dernières sources :

- Télécharger une [distribution officielle de Cyberdocs](#)
- Récupérer les [dernières sources avec Subversion](#)

4.1. Distributions de Cyberdocs

Le site de développement collaboratif [Sourcesup](#) héberge le [projet Cyberdocs](#) et propose par le fait même une [page](#) où l'on peut télécharger les différentes distributions de la plate-forme. Cette page est disponible depuis l'onglet *Fichier* sur le site Sourcesup.

Cette page de téléchargement propose les différentes versions qui sont encore disponibles au téléchargement. Pour chaque version, différents formats pourront être accessibles, ou différents paquets.

Dans tous les cas, il est recommandé d'installer la dernière version disponible : le support et la documentation des anciennes versions ne sont plus assurés.

4.2. Récupérer les sources avec Subversion

Subversion est un logiciel de gestion de versionnement de sources. Les développeurs du projet Cyberdocs utilisent cet outil pour faciliter leur collaboration et la saine évolution des différentes versions de la plate-forme. Au-delà de cette utilisation en développement, tout utilisateur intéressé par les dernières sources de Cyberdocs peut le faire en utilisant un logiciel *client* Subversion.

Dans cette page, nous fournissons quelques informations sur Subversion, de même que les paramètres à utiliser pour récupérer les sources Cyberdocs, et enfin un exemple avec le logiciel TortoiseSVN sous Windows.

A propos de Subversion

Pour en savoir plus sur cet outil, nous vous proposons de lire [l'article correspondant dans l'encyclopédie Wikipedia](#). Vous y trouverez des explications sur l'origine de l'outil, sur les principales commandes disponibles, ainsi que des liens vers différents logiciels qui permettent de travailler avec Subversion.

Nous vous suggérons également de lire une [page d'aide sur Subversion](#) maintenue par le Comité Réseau des Universités (CRU, France), là où sont hébergées les sources Cyberdocs.

Si vous ne souhaitez que récupérer les sources de Cyberdocs, vous pouvez passer aux sections suivantes qui fournissent toutes les informations nécessaires.

Paramètres Subversion de Cyberdocs

Sur le site Sourcesup où sont hébergées les sources de Cyberdocs, [l'onglet Subversion](#) fournit les informations nécessaires pour accéder aux sources Cyberdocs via Subversion. Pour un accès en mode anonyme, qui permet uniquement de récupérer les sources, et non de les modifier, vous devez utiliser l'adresse suivante avec votre logiciel clien Subversion :

checkout <http://subversion.cru.fr/cyberdocs>

Selon l'outil que vous utilisez, la procédure exacte sera différente. Par exemple, si vous travaillez en mode ligne de commande, la commande sera :

```
svn checkout http://subversion.cru.fr/cyberdocs [dossier_local]
```

Où [dossier_local] est le chemin du dossier où vous souhaitez copier les sources sur votre ordinateur.

Utiliser TortoiseSVN avec Windows

Le logiciel [TortoiseSVN](#) est un outil qui s'intègre à l'explorateur Windows et qui vous permet d'utiliser très facilement toutes les fonctionnalités de Subversion en mode graphique. Voici la procédure à suivre pour installer TortoiseSVN puis pour récupérer les sources Cyberdocs.

Télécharger et installer TortoiseSVN

Une [page de téléchargement](#) vous permet de récupérer la dernière version. Vous devez télécharger l'installateur, et ensuite le *language pack* en français pour obtenir une interface en français.

L'installation est simple à partir de l'installateur. A noter que vous devrez redémarrer votre ordinateur pour bénéficier des icônes spécifiques installées par TortoiseSVN.

Récupérer Cyberdocs avec TortoiseSVN

Nous allons considérer que vous récupérez les sources pour la première fois. D'abord, vous devez créer un nouveau dossier sur votre ordinateur, qui contiendra les sources Cyberdocs. Une fois cela fait, avec votre explorateur Windows, cliquez sur le dossier avec le bouton droit de la souris et vous obtiendrez un menu contextuel semblable à celui-ci :

Dans le menu contextuel, cliquer sur l'entrée SVN Extraire... et vous obtiendrez une fenêtre telle que celle-ci :

(Image)

Dans la zone *URL du référentiel*, vous devez saisir la valeur suivante :

<http://subversion.cru.fr/cyberdocs/trunk>

Dans la zone *Extraction du répertoire*, le chemin du dossier où vous avez cliqué sera inscrit et vous pouvez le conserver. Pour les autres paramètres, les valeurs par défaut sont correctes, et sont celles illustrées dans l'image précédente.

Une fois ces informations fournies, vous n'avez qu'à cliquer sur le bouton *OK* et la récupération des sources va débuter. Une fenêtre vous indiquera en permanence où vous en êtes.

Mettre à jour les sources avec TortoiseSVN

Si jamais vous avez déjà récupéré les sources Cyberdocs et que vous voulez les mettre à jour (par exemple si les développeurs ont apporté des changements), vous pouvez utiliser l'explorateur Windows pour démarrer TortoiseSVN, un peu comme lors de la récupération initiale des sources.

Avec l'explorateur, ouvrez le dossier racine où se trouve les sources Cyberdocs. Puis vous cliquez avec le bouton droit de la souris, et vous obtiendrez un menu contextuel tel que celui-ci :

(Images)

Cette fois, dans le menu contextuel, choisissez l'entrée *SVN Mettre à jour* et TortoiseSVN se chargera de vérifier sur le serveur si des nouvelles sources sont disponibles (ajouts ou modifications) et il les récupèrera.

5. Listes

La collaboration entre les différents participants à Cyberdocs ou Cyberthèses se fait à travers plusieurs listes de discussion. Toutes ces listes sont hébergées par le [CRU](#), elles fonctionnent avec le logiciel multilingue [Sympa](#), leur archivage est réalisé par le logiciel [MhOnArc](#). **Les listes de discussions constituent le moyen privilégié d'entrer en contact avec les**

communautés Cyberdocs ou Cyberthèses, et toutes les questions ou remarques devraient d'abord passer par l'une ou l'autre de ces listes.

Toutes les listes reliées à Cyberdocs et Cyberthèses sont accessibles depuis une [même page](#) gérée par Sympa. En accédant à la page d'accueil d'une liste, vous pourrez effectuer différentes opérations, en particulier vous abonner, vous désabonner et consulter les archives de la liste. Veuillez noter qu'il est **fortement conseillé de consulter les archives d'une liste avant d'y poser des questions**.

Le reste de cette page vous expliquera le rôle et les modalités propres à chaque liste de discussions associée au projet. Nous vous invitons à bien lire les objectifs de chaque liste avant d'envoyer un message, pour être certain de choisir la bonne liste, la plus pertinente pour la question à poser ou la remarque à effectuer.

1) cybertheses-users

Cette liste est destinée aux utilisateurs de la plateforme. Y sont abordées toutes les questions relatives au traitement des documents par la plateforme Cyberdocs : stylage, traitement, compréhension des messages d'erreur, etc.

[Page d'accueil de la liste](#)

[S'abonner à la liste](#)

[Arrêter son abonnement à la liste](#)

[Consulter les archives de la liste](#)

Envoyer un message sur la liste : cybertheses-users@cru.fr

2) cybertheses-tech

Cette liste de discussion technique est destinée aux personnes qui doivent installer, gérer et maintenir une plateforme Cyberdocs, et qui donc sont susceptibles d'être intéressés par ses aspects techniques : comment ça marche, comment l'adapter ou la personnaliser, comment régler les problèmes d'installation ou de paramétrage, etc.

[Page d'accueil de la liste](#)

[S'abonner à la liste](#)

[Arrêter son abonnement à la liste](#)

[Consulter les archives de la liste](#)

Envoyer un message sur la liste : cybertheses-tech@cru.fr

3) cybertheses-dev

Cette liste est accessible en consultation à tout un chacun, son contenu est essentiellement technique puisque c'est là que les personnes participant activement aux développements de Cyberdocs s'entre-aident, échangent leurs informations, trucs et astuces, bouts de programmes, etc. L'abonnement est possible, mais il doit auparavant être autorisé par un des gestionnaires du projet.

[Page d'accueil de la liste](#)

[S'abonner à la liste](#)

[Arrêter son abonnement à la liste](#)

[Consulter les archives de la liste](#)

Envoyer un message sur la liste : cybertheses-dev@cru.fr

4) cybertheses-latex

Ce groupe a pour objet l'intégration de l'outil [LXir](#) qui permet une entrée LaTeX à la chaîne de conversion de la plateforme Cyberdocs, en particulier pour le projet Cyberthèses.

[Page d'accueil de la liste](#)
[S'abonner à la liste](#)
[Arrêter son abonnement à la liste](#)
[Consulter les archives de la liste](#)

Envoyer un message sur la liste : cybertheses-latex@cru.fr

5) cybertheses-multilingue

Ce groupe de travail animé par Gabriela Ortúzar de l'Universidad de Chile, va aborder le thème de l'amélioration de l'accès aux ressources CyberThèses à travers l'intégration de bases de données d'autorités multilingues.

[Page d'accueil de la liste](#)
[S'abonner à la liste](#)
[Arrêter son abonnement à la liste](#)
[Consulter les archives de la liste](#)

Envoyer un message sur la liste : cybertheses-multilingue@cru.fr

6. FAQ

Voir : <http://www.cyberdocs.org/fr/faq>

7. Contacts

Service Général de la Recherche et des Ecoles doctorales - ERAD
86 rue Pasteur
69007 Lyon
FRANCE

For any information, please write to:

cybertheses@univ-lyon2.fr