

UNIVERSITE CHEIKH ANTA DIOP DE DAKAR

BIBLIOTHEQUE CENTRALE

PROGRAMME DE NUMERISATION, D'ARCHIVAGE ET DE DIFFUSION ELECTRONIQUE DES TRAVAUX ACADEMIQUES

ARCHIVAGE ET DIFFUSION ELECTRONIQUE DES THESES ET MEMOIRES :

L'expérience du programme CYBERDOC-UCAD de numérisation, d'archivage et de diffusion électronique des travaux académiques de l'Université Cheikh Anta Diop de Dakar.

Communication présentée par Jean Baptiste SARR, à l'occasion de la conférence sur la publication et la diffusion électronique, organisée par le CODESRIA, Dakar, 1^{er} au 2 septembre 2004

SOMMAIRE

INTRODUCTION

CHAPITRE I- PRESENTATION DU PROGRAMME CYBERDOC-UCAD

CHAPITRE II- MISSION ET OBJECTIFS

CHAPITRE III- TACHES ET REALISATIONS DE CYBERDOC-UCAD

III.1-INSTALLATION DE LA PLATE FORME ET DE SES COMPOSANTES

III.2- REDEFINITION DE LA PROCEDURE

DE DEPOT DES THESES ET MEMOIRES

III.3. TRAITEMENT ELECTRONIQUE DES DOCUMENTS

III.3.1- LE STYLAGE

III.3.2- LA CONVERSION

III.3.3-INDEXATION ET MISE EN LIGNE

CONCLUSION

INTRODUCTION

Depuis deux ans déjà, l'université Cheikh Anta Diop de Dakar, à travers la plate forme CYBERTHESES, devenue CYBERDOCS, dispose d'un programme d'archivage et de diffusion électronique des travaux académiques et de recherche (thèses, mémoires, rapports...).

Cyberthèses ou Cyberdoc est une plate forme qui permet de (re) structurer, de convertir en XML (format pivot), puis en HTML et en PDF (formats d'archivage et de diffusion) des documents traitement de texte et de les publier sur Internet.

La plate forme Cyberthèses, initiée conjointement en 1998 par les presses de l'Université de Montréal et l'université Lyon 2, avec le soutien du fonds francophone des inforoutes présente le double avantage :

➤ **d'être entièrement gratuite .**

En effet , elle est libre et est constituée de logiciels libres protégés par la licence GPL, téléchargeables à partir de son site de développement .

Son installation ne nécessite donc aucun investissement en dehors du matériel informatique indispensable à son fonctionnement .

➤ **d'être modifiable ;**

les sources qui la composent peuvent être modifiées, améliorées et adaptées de manière à permettre , non seulement de convertir et de mettre en ligne, en dehors des thèses et mémoires, d'autres types de documents tels que : les articles de périodiques, les revues, les rapports, les livres, etc. , mais aussi à plusieurs institutions de partager le même serveur pour traiter et diffuser leurs documents .

L'intérêt d'un tel programme pour des universités africaines dont les moyens financiers sont très limités comme celle de Dakar est sans aucun doute très grand .

C'est ainsi que l'Université Cheikh Anta Diop de Dakar, par le biais de la bibliothèque centrale a saisi l'occasion et a adhéré dès le 20 mai 2002 au programme Cyberthèses . Depuis Cette date, CYBERDOC-UCAD, programme de numérisation, d'archivage et de diffusion des travaux académiques de l'Université Cheikh Anta Diop de Dakar est né.

CHAPITRE I- PRESENTATION DU PROGRAMME CYBERDOC-UCAD

Le Programme CYBERDOC-UCAD est un produit de la bibliothèque centrale de Dakar . Il est piloté par une équipe de quatre personnes dont :

☛ un bibliothécaire, responsable du programme.

Il chargé de la formation , de la conception et de la supervision de toutes les opérations relatives aux fonctionnement du Programme :

- formation des secrétaires chargées du stylage,
- initiation des doctorants aux techniques de rédaction de leurs travaux académiques (thèse et ou mémoire),
- vérification et correction des documents déjà stylés,
- conversion, indexation et mise en ligne des documents déjà stylés.

Il a par conséquent suivi une formation à l'utilisation de la plate forme cyberthèses ; formation sous la forme de séminaires au nombre de deux : le premier séminaire-atelier s'est déroulé à Dakar au RESAFAD (*Réseau Africain de Formation à Distance*) du 20 au 25 mai 2002, le second, sous forme de journées de mise à niveau, s'est tenu à Lyon, précisément à l'Université de Lyon 2 du 23 au 27 mars 2003 . Ce dernier a été l'occasion de la présentation de la nouvelle plate forme Cyberdoc qui est celle utilisée par le Programme **CYBERDOC-UCAD** d'archivage et de diffusion des travaux de recherche de l'Université Cheikh Anta Diop de Dakar . Ces deux séminaires ont été financés par le fonds francophone des inforoutes pour le premier (RESAFAD-Dakar) et par l'Agence Intergouvernementale de la Francophonie pour le deuxième séminaire de Lyon.

☛ trois secrétaires chargées de la restructuration ou stylage des documents (thèses et mémoires) déposés à la bibliothèque et destinés à la mise en ligne. Ces trois secrétaires ont été formées aux techniques de stylage des documents (thèses et mémoires .) et sont chargées de la préparation des documents traitement de texte en vue de leur conversion , indexation et mise en ligne .

Cette équipe est appuyée par le service informatique de la bibliothèque pour les problèmes techniques.

CHAPITRE II- MISSION ET OBJECTIFS DU PROGRAMME

Dès sa mise sur « pieds » le Programme de numérisation , d'archivage et de diffusion électronique des travaux académiques de l'Université Cheikh Anta Diop de Dakar s'est défini une mission et s'est fixé des objectifs qui sont :

- l'initiation des doctorants et maîtrisards aux techniques de rédaction d'une thèse ou d'un mémoire et à l'utilisation des logiciels qui leur permettront à l'avenir de mieux structurer et présenter leurs documents (thèses, mémoires, rapports,...).
- le traitement (stylage, conversion, indexation et mise en ligne) mais aussi la recherche, l'exploitation et la gestion des documents produits par la communauté scientifique de l'université et des instituts qui lui sont rattachés ;
- la sensibilisation des auteurs (doctorants, maîtrisards, chercheurs) aux avantages de la diffusion électronique de leurs travaux de recherche ;
- la production des copies d'archive des thèses et mémoires
- le renseignement des doctorants et maîtrisards sur les modalités de dépôt des résultats de recherche .

A travers son programme d'archivage des thèses, mémoires, rapports, revues, ...

CYBERDOC-UCAD a donc pour ambition de permettre à l'ensemble de la communauté scientifique de l'Université Cheikh Anta Diop de Dakar :

- de combler son retard informationnel
- de résoudre les problèmes d'archivage, de gestion et de conservation de sa littérature grise et d'en assurer ainsi la pérennité
- de rendre visibles, exploitables et accessibles, ses travaux académiques, tant à l'intérieur qu'à l'extérieur du pays ;
- de valoriser et promouvoir par une très large diffusion, l'ensemble de ses travaux de recherche ;
- de réduire les dépenses engendrées par la recherche et l'accès à l'information scientifique et technique, car CYBERDOC-UCAD sera bientôt accessible, grâce au protocole OAI-PMH , depuis le portail francophone des thèses et mémoires, ce qui aura pour avantage de permettre à tous les usagers du monde entier d'accéder gratuitement, mais de manière contrôlée, non seulement aux documents (thèses, rapports, articles, revues, mémoires, etc.) produits par l'Université de Dakar mais aussi à l'ensemble de la production documentaire des universités des autres pays

membres de CYBERTHESES (France, Chili, Venezuela, Brésil, Pérou, Espagne, Canada, Suisse, Etats Unis d'Amérique, Mali , plus tard, Côte d'Ivoire, Maroc, Madagascar, Burkina Faso, ...). Ceci grâce à des serveurs moissonneurs tels que ceux de l'Université de Lyon 2 et de l'université du Chili, dont le rôle est de retrouver, dans les entrepôts OAI (Open Archives Initiative) , les différents documents (thèses, mémoires, ..) archivés sur les serveurs des universités participant au programme CYBERTHESES .

CHAPITRE III- TACHES ET REALISATIONS DE CYBERDOC-UCAD

Depuis son adhésion au programme Cyberthèses, la bibliothèque centrale universitaire de Dakar s'est évertuée à tout mettre en œuvre pour l'aboutissement de son programme d'archivage et de diffusion électronique des travaux académiques de l'université Cheikh Anta Diop de Dakar. Ce fut un long « chemin de croix » parsemé d'embûches et composée de plusieurs « stations » ou étapes dont les principales sont :

III.1-INSTALLATION ET DE LA PLATE FORME CYBERDOC ET DE SES COMPOSANTES

Signalons que le Programme CYBERDOC-UCAD d'archivage et de diffusion électronique des travaux de recherche de l'Université Cheikh Anta Diop de Dakar utilise actuellement la nouvelle plate forme cyberdoc qui est distribuée sous la forme de sources, c'est-à-dire qu'on y trouve toutes les sources de l'application qui doivent être préparées ou compilées avant leur installation. Pour fonctionner, la plate forme et ses composantes doivent être installés sur un serveur.

C'est ainsi que la bibliothèque centrale Universitaire de Dakar a dépêché le responsable de son service informatique à Lyon du 15 au 18 septembre pour y suivre un stage de formation à l'utilisation de la nouvelle plate forme Cyberdoc, mais surtout pour configurer un serveur offert par l'Agence intergouvernementale de la francophonie, qui est destinée à remplacer l'ancien qui malheureusement n'a jamais fonctionné.

La plate forme cyberdoc-ucad et ses composantes (java, openOffice, SDX, Tomcat) a donc été installée sur un serveur Linux Redhat 9 et comprend trois modules :

1°) Un module de gestion qui, grâce à un interface Web, permet de convertir des documents vers un format XML TEILite et de les publier par l'intermédiaire du

module de publication. Le module de gestion comprend plusieurs rubriques dont les principales sont :

- ◆ la rubrique « Utilisateurs » qui permet la gestion des utilisateurs ;
- ◆ la rubrique « Accueil » qui permet d'afficher la liste des institutions et d'en créer de nouvelles ;
- ◆ la rubrique « gestion des fichiers » qui permet de transférer des documents vers le serveur ou d'importer des documents déjà présents sur le serveur en vue de leur conversion ,
- ◆ la rubrique « métadonnées » qui permet de visualiser et d'éditer les métadonnées d'un document déjà transféré sur le serveur ;
- ◆ la rubrique « traitement » qui permet d'exécuter les différentes étapes de la conversion du document ;

2°) un module de conversion qui automatise la conversion de tout document traitement de texte préalablement structuré vers un format XML respectant la DTD TEILite et permet d'obtenir des produits dérivés XHTML, HTML ET PDF .

3°) un module de publication qui est une application SDX (Système de recherche documentaire) qui permet d'indexer, de mettre en ligne et de rechercher des documents. Les documents une fois indexés sont donc consultables et exploitables via l'interface SDX.

L'ensemble de ces trois modules permet de convertir tout document traitement de texte vers un format XML TEILite et d'obtenir des produits dérivés (HTML et PDF) diffusables et exploitables sur Internet .

III.2- REDEFINITION DE LA PROCEDURE DE DEPOT DES THESES ET MEMOIRES

Rappelons avant tout que Cyberthèses est un programme qui permet de structurer des documents traitement de texte avant de les convertir en formats HTML et PDF diffusables et consultables sur Internet .

Pour se faire une version électronique des documents (thèses ou mémoires) est indispensable. C'est la raison pour laquelle la bibliothèque centrale universitaire, par l'intermédiaire de son Directeur, a saisi les autorités universitaires, pour d'une part, les informer sur ce nouveau type de traitement des thèses et mémoires qu'offre la plateforme CYBERDOC, et d'autre part, leur faire part de la nécessité d'instaurer un nouveau type de dépôt de ces documents à la bibliothèque centrale universitaire .

L'arrêté rectoral n° 797 du 16 août 1990 relatif aux modalités d'autorisation de soutenance des thèses et diplômes à l'université Cheikh Anta Diop de Dakar a donc subi des modifications .C'est ainsi qu'à coté du dépôt traditionnel des thèses et mémoires qui faisait obligation aux doctorants et maîtrisards de soumettre trois (3) exemplaires « papier » de leur travail , la bibliothèque centrale universitaire de Dakar à décidé d'instaurer un dépôt mixte qui consiste maintenant à déposer deux exemplaires « papier » et un exemplaire électronique (CD-ROM. ou disquettes).C'est cette version électronique qui sera traité et diffusé sur le Net par le Programme d'archivage et de diffusion électronique des travaux de recherche de l'université Cheikh Anta Diop de Dakar.

Il faut toutefois rappeler qu'aucune obligation n'est faite aux étudiants maîtrisards ou doctorants de respecter ce dépôt mixte . Seuls ceux qui souhaitent voir leur travaux mis en ligne sont soumis à ce type de dépôt, les autres, sont tenus de se conformer uniquement au dépôt traditionnel (trois exemplaires papier).

Des prospectus sont envoyés au niveau des facultés et départements pour informer les doctorants et thésards sur les modalités de dépôt et des recommandations relatives à la structuration de leur travaux leur sont faites . Au moment du dépôt des thèses et mémoires, les explications nécessaires sont aussi données aux thésards et maîtrisards quant au traitement qui sera fait de leur travail.

C'est seulement à l'issue de cet entretien, que ceux qui émettent un avis favorable, remplissent un formulaire de dépôt et signent une déclaration autorisant le responsable du Programme CYBERDOC-UCAD de mettre en ligne leur travaux. Après quoi le dépôt est accepté, enregistré et reçu par le service des acquisitions de la bibliothèque centrale universitaire de Dakar.

Il faut signaler que le programme cyberdoc-Ucad est très bien accueilli par les chercheurs et étudiants de l'université Cheikh Anta Diop de Dakar . Depuis 2002, seuls deux individus ont refusé la mise en ligne de leurs travaux : l'un parce qu'il compte faire éditer sa thèse, l'autre soit disant qu'il a été financé par un organisme de recherche qui ne souhaite pas que certaines informations contenues dans la thèse soient publiées sur le Net.

Depuis donc l'année 2002 (date d'adhésion de la bibliothèque au programme Cyberthèses ou Cyberdoc) , la bibliothèque centrale universitaire de Dakar a reçu sous forme de dépôt électronique :

❖ **451 Thèses dont :**

- 341 en médecine, pharmacie et odonto-stomatologie
- 11 en Lettres et sciences humaines
- 1 en sciences juridiques
- 35 en Sciences et techniques
- 63 en médecine vétérinaire

❖ **115 Mémoires de maîtrise dont :**

- 107 en Lettres et sciences humaines
- 6 en Sciences et techniques
- 2 en Médecine

❖ **150 Mémoires de DEA dont :**

- 12 en Lettres et sciences humaines
- 130 en Sciences et techniques
- 18 en Médecine vétérinaire

Ces documents, déposés au service des acquisitions et du développement des collections de la bibliothèque sont transmis au responsable du programme CYBERDOC-UCAD pour être traités.

III.2. TRAITEMENT ELECTRONIQUE DES DOCUMENTS

Dans cette phase de traitement des documents électroniques, il est à souligner que nous avons défini un programme de travail consistant à traiter dans une première phase uniquement les thèses produites par les différentes facultés de l'université Cheikh

Anta Diop de Dakar ainsi que celles de l'Ecole Inter-Etats des sciences et médecine vétérinaires (EISMV) .

Les autres documents tels que les mémoires de maîtrise et de DEA seront traités dans une deuxième phase dont la date reste encore à déterminer .

Le processus de traitement des thèses comprend trois étapes qui sont :

- * le stylage,
- * la conversion ;
- * l'indexation et mise en ligne .

Ces étapes sont interdépendantes, c'est-à-dire que :

- * la conversion ne peut réussir que si le document traitement de texte présenté par l'auteur est bien structuré ou stylé ;
- * l'indexation ou la mise en ligne dépend de la réussite de la conversion .

III.2.1- Le Stylage

Une thèse ou un mémoire doit respecter une structure dont les éléments sont définis à l'avance. La structure d'une thèse ou d'un mémoire doit comprendre les éléments suivants :

- ✍ **La page de titre qui comprend** : le nom de l'université de soutenance, la faculté ou institut, l'école doctorale, le département, le titre du document, le nom de l'auteur , la discipline, le grade , le nom du directeur de recherche, les membres du jury, la date de soutenance, le numéro national de soutenance.
- ✍ **Les préliminaires qui correspondent aux** : dédicaces, épigraphes, remerciements, table des illustrations, table des annexes, résumé.
- ✍ **Le corps du texte qui comprend** : l'introduction, les titres de parties, les chapitres, les sections correspondant à des niveaux de titres, les paragraphes et la conclusion.
- ✍ **Les citations qui comprennent** : les citations en ligne, les citations en bloc simple, les citations en bloc en vers.
- ✍ **Les listes** : listes titre, listes à puces, listes numérotées,
- ✍ **Les illustrations qui correspondent aux** : figures, tableaux , précédés de leur légende.
- ✍ **les postliminaires** : bibliographie, annexes.

Une thèse ou mémoire qui ne présente pas ces caractéristiques ou qui les présente mal pose de sérieux problèmes de stylage, c'est à dire le choix et l'application d'un modèle de saisie appelé «Feuille de styles » dont le but est de corriger les erreurs de styles, de structuration, ... commises lors de la rédaction de la thèse ou du mémoire. Le modèle de feuilles de styles utilisé par Cyberdoc-UCAD est celui de Lyon 2 .

Le stylage est donc une opération qui consiste à appliquer un style adéquat aux différentes parties d'un document traitement de texte. Il permet de restructurer ce document pour en rendre la conversion plus facile. C'est l'étape la plus importante du processus de traitement électronique des documents car le document issu du stylage constitue la matière première et le succès de la chaîne d'édition électronique en dépend.

Le stylage est à la fois un travail répétitif , mécanique, mais aussi et surtout intellectuel car il exige des choix et une prise de décision basés à la fois sur l'application des normes de structuration du document et le respect de la logique de l'auteur de ce document.

Au niveau du programme CYBERDOC-UCAD, le stylage est de loin l'opération qui pose le plus de problèmes et qui prend plus de temps . Ceci est dû au fait que l'initiation des doctorants et maîtresards à l'utilisation de la feuille de style, indispensable à une bonne structuration des documents n'a pas encore été entreprise . Ce qui fait que les thèses et mémoires reçus sont soit très mal structurés, soit tout simplement inexploitable parce que enregistrés sous des formats inadaptés ou sur des supports (disquettes ou CD-ROM) infectés par des virus ou de très mauvaise qualité.

Depuis donc le 23 mars 2004 , sur un total de quatre cents cinquante et un (451) thèses reçues, le département de stylage de la Chaîne d'archivage et de diffusion électronique des thèses et mémoires de la bibliothèque centrale universitaire de Dakar a réussi à en styler quatre-vingts seize (96) et enregistré quarante et un (41) qui sont inexploitable ; soit une moyenne de vingt (20) thèses stylées par mois pour l'ensemble des trois secrétaires chargée du stylage et d'environ sept (07) thèses pour chacune. C'est une moyenne très appréciable compte tenu, d'une part du fait qu'elles ne se consacrent au stylage qu'à temps partiel (elles sont le plus clair du temps confinées à

d'autres tâches), et d'autre part que les thèses sont le plus souvent très mal structurées.

En attendant donc de mettre en place un programme de formation des doctorants et thésards, prévu pour la rentrée 2004-2005, le programme CYBERDOC-UCAD a entrepris depuis 2004 , au niveau des facultés et départements de l'université Cheikh Anta Diop de Dakar, une vaste campagne d'information et de recommandations relatives à une meilleure présentation des travaux de recherche déposés à la bibliothèque dans le but d'alléger le stylage . Le stylage est une étape très importante car les erreurs commises ont des conséquences néfastes dans les phases ultérieures du traitement .

III.2.2- La Conversion

La conversion est la deuxième étape dans le processus de la chaîne d'édition numérique CYBERDOC-UCAD. Après le stylage, le document est converti à travers le module de gestion de la plate forme Cyberdoc ; ceci grâce à une interface Web qui comporte une barre de menus. Ce module de gestion facilite grandement les opérations de conversion.

En effet, toutes les étapes de la conversion des documents se trouvent simplifiées : il suffit tout simplement, après s'être identifié à travers la rubrique « utilisateurs » de créer un espace pour le document à convertir et ensuite d'y transférer le fichier contenant ce document. Une fois ce fichier transféré sur le serveur, le formulaire permettant d'éditer les métadonnées renseigné, la phase de Conversion proprement dite peut commencer . Si le processus de conversion réussit, on obtient un document XML et des produits dérivés HTML et PDF qui peuvent dès lors être indexés et mis en ligne.

Il faut signaler que la conversion peut parfois échouer suite à une mauvaise installation des composantes de la plate forme ou à une mauvaise structuration des documents. Des messages d'erreurs (qui n'entravent pas toujours la conversion) sont parfois signalés au cours du traitement des documents. Ces erreurs sont le plus souvent relatives soit à la présence de styles existant dans le document et non déclarés au niveau de la feuille de styles (nous utilisons la feuille de styles de Lyon2), soit à

l'absence de styles dits obligatoires rencontrés généralement sur la page de titre (Université de soutenance, faculté, département, grade, directeur de recherche , etc...)

Nous n'avons jusqu'ici pas rencontré de problèmes majeurs au cours du processus de conversion ; tout se déroule pour le moment bien.

A ce jour, tous les documents (thèses) stylés ont été convertis presque sans problèmes à part trois (3) documents dont nous n'avons pas réussi à obtenir la version PDF . Nous obtenions alors un message du genre « fichier PDF endommagé » Nous les avons malgré tout indexés et mis en ligne.

Le nombre de thèses converties depuis la date du 23 mars 2004 est donc égal à celui des thèses stylées : quatre-vingts-seize (96) au total dont :

☞ soixante-cinq (65) thèses de 2002

☞ trente et un (31) de 2003

III.2.3-Indexation et mise en ligne

L'indexation et mise en ligne est la dernière étape du processus d'archivage et de diffusion électronique des travaux de recherche.

Le document (thèse ou mémoire) , une fois stylé et converti, est mis en ligne et publié en Intranet local ou sur Internet selon l'option choisie .

L'indexation et la mise en ligne des documents au niveau de la plate forme Cyberdoc se fait par le biais du module de publication qui est une application SDX de la manière suivante : l'administrateur de l'application SDX, une fois identifié, peut accéder à la page de chargement. xsp qui comprend un formulaire dont les zones (code de l'institution, année de publication, identifiant du document) sont renseignées . Il ne lui reste alors qu'à cliquer sur la zone « entrée » pour envoyer le document. Un message indique que le processus d'indexation est réussi. Le document est alors mis en ligne et est dès lors consultable.

Depuis le 23 mars 2004, date de démarrage effectif du dit programme , quatre-Vingt seize (96) thèses ont été traitées et mises en ligne .

Ces thèses sont d'ores et déjà accessibles et exploitables (pour le moment en Intranet local) à partir de tous les postes (ordinateurs) reliés au réseau de la Bibliothèque à l'adresse : <http://cyberdoc:8080/sdx/pcd/>

La base de données des thèses et mémoires de l'Université Cheikh Anta Diop de Dakar sera visible et consultable de l'extérieur sous peu .

CONCLUSION

Le Programme CYBERDOC-UCAD permet donc d'archiver , de diffuser , d'assurer une conservation pérenne et une consultation des travaux de recherche de la communauté universitaire

Il offre par la même occasion aux chercheurs la possibilité d'augmenter la visibilité de leurs travaux en leur permettant , en dehors de la version « papier » dont la diffusion est restreinte, de disposer d'une version électronique à durée de vie plus longue, plus facile à conserver et à exploiter à travers les possibilités qu'offrent les nouvelles technologies de la communication .

CYBERDOC–UCAD compte ainsi prendre part à **l'initiative de Budapest** pour l'accès libre aux publications scientifiques que leurs auteurs souhaitent mettre librement à la disposition du public (BOAI- Budapest Access Initiative)

Il assurera aussi et gratuitement une large publicité à la fois pour l'Université Cheikh Anta Diop de Dakar, pour les enseignants, étudiants et chercheurs, participant ainsi à l'amélioration du rayonnement de l'Université et de l'ensemble de la communauté universitaire .

Jean Baptiste SARR,
Responsable du programme CYBERDOC-UCAD de numérisation, d'archivage et de diffusion électroniques des travaux de recherche de l' Université Cheikh Anta Diop de Dakar.
Bibliothèque Centrale Universitaire de Dakar
BP 2006
Tel (221) 824 69 81
Fax (221) 824 23 79
jbsarr@hotmail.com