

République Algérienne Démocratique et Populaire

Université des Sciences et de la Technologie Houari Boumediène

USTHB

Institut d'informatique

**MEMOIRE DE FIN D'ETUDES POUR L'OBTENTION DU DIPLOME
D'INGENIEUR D'ETAT EN INFORMATIQUE**

Thème

**CONCEPTION D'UNE APPLICATION POUR
LE DEVELOPEMENT D'UNE COLLECTION
NUMERIQUE**

Proposé par :

M^{ell} RAHILE HAYET

M^{eu} BAKILLI YAHIA

Jury :

M^d ALIMASIGHI

M^d MAHDAOUI

M^{eu} AKLOUF

Réalisé par :

BOUARISSA KARIMA.

NECHEME AMINA.

Organisme d'accueil : CERIST

Promotion : 2003 / 2004

RESUME :

La problématique du projet est fondée sur l'idée d'introduire le concept de collection numérique et la technique de numérisation dans les unités documentaires et les bibliothèques algériennes, un tel investissement étant engagé afin d'optimiser et d'élargir l'accès aux contenus documentaire locaux.

Un objectif central du projet consiste à développer un environnement intégré devant assister et permettre à ces différentes unités documentaires de développer des collections numériques et de les mettre à la disposition des utilisateurs.

La démarche de cet mémoire commence par une étude comparative des solutions internationales dans le développement des collections numériques (Greenstone, Sdx) ensuite le développement d'une application qui réalise :

- + La capture et la numérisation d'un corpus électronique.
- ♦ Génération automatique des métadonnées selon les standards internationaux (Dublin Core).
- + Indexation automatique du corpus.
- ♦ Archivage sous le format XML.
- ♦ Diffusion de la collection aussi bien en ligne qu'au local.
- + Administration et organisation des collections..

SOMMAIRE

CHAPITRE 1 : LES CONCEPTS DE COLLECTION ELECTRONIQUE

| | |
|--|-----------|
| 1Introduction | 1 |
| 2-Definition de collection numériques..... | 1 |
| 3-Objectifs des collections numériques..... | 2 |
| 3.1)- La preservation | 2 |
| 3.2)- Le partage..... | 2 |
| 3.3)- Explorer la technologie..... | 2 |
| 4-Cadre de developpement des collections..... | 2 |
| 4.1)- Bibliothèques, archives, institutions savantes..... | 3 |
| 4.2)- projets de programme de numérisation régionale et international..... | 4 |
| 4.3)- réservoirs de texte et de corpus électronique..... | 4 |
| 5-Operations traitees par une collection..... | 5 |
| 5.1) Capture..... | 7 |
| a-1)- scannairisation..... | 7 |
| a-2)- reconnaissance optique des caractères OCR..... | 7 |
| a-3)- comment numérisé...7 | |
| 5.2) Indexation..... | 8 |
| 5.2.1) Definition..... | 8 |
| 5.2.2) les composants de l'indexation | 9 |
| 5.2.3)Les méthodes d'indexation..... | 10 |
| 5.3) Archivage | 11 |
| 5.4) Exploitation..... | 11 |
| 6-Les différentes solutions de collection électronique..... | 12 |
| 6.1) CONTENTdm | 12 |
| 6.1.1) Definition | 12 |
| 6.1.2) Traitement des collections | 12 |
| 6.1.2.1j Capture | 12 |
| 6.1.2.2)Indexation | 12 |
| 6.1.2.3j Archivage | 13 |
| 6.1.2.3) Exploitation..... | 14 |
| 6.1.3j Quelques exemples de collection..... | 14 |

| | |
|---|-----------|
| 6.2) VTLS | 14 |
| 6.2.1j Définition | 14 |
| 6.2.2) Traitement des collections | 14 |
| 6.2.2.1) Capture | 14 |
| 6.2.2.2) Indexation | 14 |
| 6.2.2.3) Archivage | 14 |
| 6.2.2.3)Exploitation | 14 |
| 6.2.3j Quelques exemples de collection | 14 |
| 6.3) SDX | 15 |
| 6.3.1) Historique | 15 |
| 6.3.2) Définition | 15 |
| 6.3.3) Traitement des collections | 15 |
| 6.3.3.1) Capture | 15 |
| 6.3.3.2) Indexation | 16 |
| 6.3.3.3) Archivage | 16 |
| 6.3.3.4)Exploitation | 16 |
| 6.3.3) Quelques eseniples de collection | 17 |
| 6.4) GREENSTONE | 17 |
| 6.4.1) Définition | 17 |
| 6.4.3) Traitement des collections | 17 |
| 6.4.3.1) Capture | 17 |
| 6.4.3.2) Indexation | 18 |
| 6.4.3.3) Archivage | 18 |
| 6.4.3.4)Exploitation | 20 |
| 6.2.3) Quelques esemples de collection | 21 |
| 6.5) Tableau comparatif entre les trois projet | 22 |
| 7. Conclusion | 26 |
| Chapitre 2: langage extensible de balisage | |
| 1.Introduction | 27 |
| 2. Historique | 27 |
| 3.Définition | 27 |
| 4. les avantages de XMI. | 28 |
| 4.1)- La perennite de l'information | |
| 4.2)- Echange de donnee | |
| 4.3)- harmonisation des sources dc données disparates | |
| 5-la Syntaxe des documents XMI. | 29 |

| | |
|--|----|
| 5.1)- prologue..... | 29 |
| 5.1.1) identification XML..... | 30 |
| 5.1.2)- l'encodage | 30 |
| 5.1.3)- la DTD..... | 30 |
| 5.1.4)- le schema..... | 31 |
| 5.2)- les elements..... | 31 |
| 5.3) les attributs..... | 32 |
| 5.4) LES ENTITES..... | 32 |
| 6-grammaire et schema..... | 33 |
| 6.1) les DTDs..... | 33 |
| 6.1.1) Definition | 33 |
| 6.1.2) Composantes d'une DTD..... | 33 |
| 6.1.2.1) Les déclarations d'éléments..... | 34 |
| 6.1.2.2) Les declarations d'attributs | 34 |
| 6.2)Les schema. Xml | 34 |
| 6.3) les espaces de nommage..... | 35 |
| Les outils XML..... | 35 |
| 7.1) langage de presentation..... | 35 |
| 7.1.1) les transformation XSL..... | 35 |
| 7.1.1.1) definition..... | 35 |
| 7.1.1.2) Aspects fondamentaux de XSL..... | 36 |
| 7.1.2) feuille de style en cascade..... | 38 |
| 7.2) les langage de lien et d'adressage..... | 38 |
| 7.2.1) Xlink..... | 38 |
| 7.2.1.1) définition | 38 |
| 7.2.1.2) objectif... .. | 38 |
| 7.2.1.3) principe... .. | 39 |
| 7.2.1) XPath..... | 40 |
| 7.2.1.1) définition | 40 |
| 7.2.1.2) objectif... .. | 40 |
| 7.2.1.3) principe... .. | 41 |
| 7.2.1) Xpointer | 42 |
| 7.2.1.1) definition..... | 42 |
| 7.2.1.2) objectif... .. | 42 |
| 7.2.1.3) principe... .. | 42 |
| 7.3) les langages de traitement | 43 |
| 7.3.1) le modele objet de document..... | 43 |
| 7.3.2) simple API pour XML SAX..... | 45 |

| | |
|-------------------|-----|
| 8.conclusion..... | .42 |
|-------------------|-----|

CHAPITRE III : Les Metadonnees

| | |
|---|------|
| 1.Introduction..... | .47 |
| 2.Historique..... | .48 |
| 3.DEFINITION..... | .48 |
| 4.pourquoi les metadonnées sont-elle si importante?..... | .49 |
| 5. ou sont les metadonnées?..... | .50 |
| 5.1) Metadonnees externes aux ressources - Bases de données | .50 |
| 5.2)Métadonnées internes aux ressources | .50 |
| 6.metadonnées "metiers"et la necessite de standard..... | .50 |
| 7.La norme Dublin core | 52 |
| 7.1) definition..... | 52 |
| 7.2) la force du dublin core | ..52 |
| 7.3) Eléments du Dublin Core | 54 |
| 9.Conclusion..... | .57 |

CHAPITRE VI : Conception et développement de l'application

| | |
|---------------------------------------|-----|
| 1.Presentation..... | 58 |
| 2-schema conceptuel de DigiCol..... | 59 |
| 3-Les principales orientation..... | 62 |
| 4.architecture web..... | .63 |
| 5-pourquoi windows 9x?..... | 63 |
| 6.pourquoi php..... | .64 |
| 7.pourquoi mysql..... | .64 |
| 8.pourquoi apache..... | .64 |
| 9-langage de programmation..... | .64 |
| 10.les modules de digicol..... | .65 |
| 10.1) module creation..... | .65 |
| 10.1.1)Importation des documents..... | .65 |
| 10.1.2)Format intermediaire..... | .65 |
| 10.1.3)Indexation | .67 |
| 10.1.4)Format de donnee | .68 |
| 10.1.5)Archivage..... | .69 |
| 10.2)Module modification..... | .70 |

| | |
|---|------------|
| 10.3)Module administration..... | 71 |
| 10.4)Module Diffusion..... | 73 |
| 11. Conception de la base de donnée..... | 73 |
| CHAPITRE 5 : test et mise en oeuvre..... | 76 |
| Annexe A (php)..... | 84 |
| Annexe B (sdx)..... | .86 |
| Annexe C (greenstone)..... | 94 |