

Ministère de l'Enseignement Supérieur et de
La Recherche Scientifique
Université El-Hadj Lakhdar - BÂTNA
Faculté des Sciences de l'Ingénieur
Département d'Informatique



وزارة التعليم العالي والبحث العلمي
جامعة الحاج لخضر باتنة
كلية الهندسة
قسم الإعلام الآلي

PROJET

pour l'obtention du diplôme d'Ingénieur d'Etat en Informatique
Option : Systèmes d'Information Avancés

par

IGUELMAMENE Yacine
BENSEDIRA Farid

TRAITEMENT DES DOCUMENTS XML SOUS ORACLE

Encadreur:
DEKHENET Abdelhamid

Promotion 2007

Résumé

La publication électronique des thèses est un domaine en forte croissance, la plupart des universités développent ou réfléchissent à un projet de publication sur internet. Différent format de fichier a été adopté : HTML, SGML, XML, PDF, offrant différent confort de recherche, de navigation et de visualisation. L'apport de tel projet est très important pour la communauté scientifique cela permet un accès très facile à toute cette littérature, et lui permet d'être connu du grand public. Le projet de traitement des documents XML sous oracle en est à sa première phase de développement. Ses principales caractéristiques sont une diffusion des thèses en texte intégral au format Adobe PDF. Notre projet démarre une nouvelle réflexion sur les possibilités offertes par le langage XML.

Mot clefs : thèse, Document électronique, serveur, Texte intégral, Internet, TEI Lite, XML, Oracle, XSQL, XSL, Base de données.

Abstract

Electronic thesis is an increasing research field. Most of the universities are currently developing a project to put their thesis on internet. Different file format exist: HTML, SGML, XML, PDF, providing different features to search, browse or view on electronic document. Such project allows an easier access to scientific document for the research community and the general public. The project of treatment of XML documents under oracle is with its first phase of development. Its principal characteristics are a diffusion of the thesis in complete text to the format Adobe PDF. Our project starts a new reflexion on the possibilities offered by XML language.

Keywords: Thesis, Electronic Document, Server, Full text, Internet, TEI Lite, XML, Oracle, XSQL, XSL, Data base.

Sommaire

Introduction générale

CHAPITRE 1 : ORACLE

1. Définition d'oracle.....	1
1.1. les fonctions d'Oracle.....	1
1.2. les outils d'oracle.....	1
1.2.1. les outils d'administration d'Oracle.....	2
1.2.2. les outils de développements d'Oracle.....	2
1.2.3. les outils de programmations.....	3
2. le choix d'Oracle.....	3
2.1. L'entreprise Oracle.....	3
2.2. Les préférences d'Oracle.....	4
3. Composants de l'architecture Oracle.....	6
3.1. Le serveur Oracle	6
3.1.1. Tache de l'administrateur Oracle.....	6
3.1.2. Connexion à un serveur Oracle.....	7
3.1.2.1. Connexion locale	7
3.1.2.2. Connexion deux tiers.....	7
3.1.2.3. Connexion multi tiers.....	7
3.2. Instance Oracle et base de données	9
3.2.1. Système Global Area (SGA).....	9
3.2.2. Processus d'arrière plan.....	10
3.3. Traitement de requêtes.....	11
4. la base de données d'Oracle	13
4.1. La structure logique d'une BD.....	13
4.2. La structure physique d'une BD.....	16
5. Dictionnaire de données Oracle.....	18
6. Démarrage et arrêt d'une instance.....	19
7. En conclusion.....	22

CHAPITRE 2 : Oracle et XML.

PARTIE 1 : XML (eXtensible Markup Language).

1. Définition de XML.....	23
1.1 Les documents XML.....	23
1.2 Les document type definition (DTD).....	25
1.3 Les domaines de noms (namespace).....	26
1.4 Les parsers XML.....	27
2. Définition de XSLT	28
2.1 Pourquoi transformer un document XML.....	29
3. Langage XSL.....	30
3.1 XSL : le langage de transformation des données.....	31
3.2 Quelques notion de base sur la grammaire XSL.....	32
4. XPATH.....	33
5. XML schéma.....	34
6. En conclusion.....	36

PARTIE 2 : Les produits Oracle XML

1. Parser XML et processeur XSLT.....	37
2. Les utilitaires XML SQL.....	39
3. Servelet XSQL.....	41
4. En conclusion.....	45

PARTIE 3 : Stockage des documents XML.

1. Introduction.....	46
2. Définition de XMLType.....	47
3. L'utilisation de XMLType.....	48
4. Stockage des données dans une colonne ou table.....	48
5. Stockage des données XML de type XMLType dans une BD Oracle.....	49
6. Accéder aux données stockées dans des colonnes ou des tables.....	50

7.	XPath.....	50
7.1	Naviguer dans XML avec XPth.....	50
7.2	Expressions de XPath.....	52
8.	Introduction au schéma XML.....	53
8.1	XML schéma et la BD Oracle.....	54
8.2	Nos besoins du schéma XML.....	54
8.3	Comment le stockage des données XML basée sur les schémas est structuré.....	55
9.	En conclusion.....	56

CHAPITRE 3 : Le balisage

PARTIE 1 : La TEI (Text Encoding Initiative).

1.	Introduction.....	57
2.	Origine de la Text Encoding Initiative.....	57
2.1.	La valeur d'un texte sur le support Electronique.....	57
2.2.	Balisage.....	58
2.3.	SGML.....	59
2.4.	Définition de la Text Encoding Initiative (TEI).....	59
3.	But de la TEI.....	60
3.1.	Normalisation des données.....	60
3.2.	Proposition de la TEI.....	61
4.	La DTD de la TEI et lignes directrices.....	62
4.1.	Guidelines for Electronic Text Encoding and Interchange.....	62
4.2.	Modèle de la DTD de la TEI.....	62
4.3.	TEI Lite.....	63
5.	Structuration générale d'un fichier TEI.....	63
6.	Le futur de la TEI.....	67
7.	En conclusion.....	67

PARTIE 2 : Le balisage proposé

1.	Principe de balisage.....	68
2.	Description formelle du balisage proposé.....	68

3. En conclusion.....	74
-----------------------	----

Chapitre 4 : Analyse et mise en œuvre

1. Introduction.....	75
2. Scénario de fonctionnement.....	75
3. Présentation de l'interface d'application.....	76
4. Architecture logicielle.....	77
5. Présentation de l'exemple de démonstration.....	78
6. Conclusion.....	83
Conclusion générale.....	84

Les annexes, La bibliographie, et les Abréviations

1. Annexe A : La configuration des XSQL servlets.....	85
2. Annexe B : La configuration du package http.....	89
3. Annexe C : Code source du schéma des thèses.....	94
4. Glossaire.....	99
5. Bibliographie.....	101