

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE
HOUARI BOUMEDIENE
USTHB/ALGER



Mémoire pour l'obtention
du diplôme d'Ingénieur d'Etat en Informatique

Sujet :

*Développement d'un environnement de syndication et
d'agrégation de contenu Web
basé sur RSS*

Présenté par :
KHALAF Zineb
ZIZI Nissette

Promoteur **Mme El-Maouhab**
Co-Promoteur **Mme Hesas**
Mme Bouali

Organisme d'accueil : C.E.R.I.S.T
(Centre de recherche sur l'Information Scientifique et Technique)

Soutenu le : /10/2005

Président du jury: **Mme N.KAMEL**
Membre du jury : **Mr KECHID**
Membre de la sous commission : **Mme BOUAGGAR**

Promotion 82/2005

Résumé

La masse d'information disponible sur les sites web croit sans cesse, ses formes sont diverses et l'information véhiculée devient difficilement perceptible, ce qui rend fastidieux la tâche de retrouver l'information recherchée, d'en assurer sa véracité et sa cohérence avec le reste des sources disponibles.

Dans le but de simplifier cette tâche aux éditeurs et consommateurs d'information, nous avons développé un environnement (client/serveur) basé sur les spécifications RDF Site Summary (RSS) comprenant un serveur fournisseur de nouvelles, des client consommateur de nouvelles et des outils d'édition de nouvelles, tout en incluant la gestion des droits d'utilisateurs afin de garantir la confidentialité des informations produites par tout un chacun.

Abstract

The masse of information available on Web sites grows without cease, its forms are various and the information transported becomes perceptible with difficulty, which makes boring the task to find the information searched, to assure its truthfulness and its coherence with the rest of available sources.

With the aim to simplify this task to information publisher and consumer, we have developed a (Client/Server) environment, based on RDF Site Summary (RSS) specifications which is composed of a news supplier server, news consumer clients and news edition tool, including user rights management in order to guarantee the confidentiality of information produced by every one of us.

ملخص

إن كمية المعلومات الموجودة على مواقع الانترنت في ازدياد مستمر، تحت أشكال مختلفة ومتعددة. مما صعب مهمة إيجاد المعلومة أو البحث عنها.

وبهدف تسهيل هذه المهمة لكتاب وناشري ومستهلكي المعلومة، طورنا محيط (Client/Server) مبني على مبدأ اختيار المعلومة حسب مضمونها وخاصيتها، لتوزيعها بين المواقع أو تجميعها أو إدماجها في المواقع وذلك بالاعتماد على RSS (إدماج بسيط وسهل) مكون من Server (مقدم خدمات) ممول للأخبار ومستهلكي للأخبار و أدوات لكتابة ونشر الأخبار، وذلك مع حفظ حقوق المستعمل (كاتب، ناشر، مستهلك) بهدف ضمان سرية المعلومات المنتجة من طرف كل واحد.

SOMMAIRE

Introduction générale.....	1
<u>Chapitre I : Introduction aux system de gestion de contenu</u>	
1. Introduction.....	3
2. Qu'est ce que le contenu ?.....	4
2.1 Les données non structurées.....	5
2.2 Les données structurées.....	5
3. Qu'est ce q'un système de gestion de contenu Web ?.....	6
4. Une évolution de plus en plus dynamique !.....	6
4.1 Les sites statiques.....	6
4.2 Les sites dynamiques.....	7
5. Domaines d'application d'un system de gestion de contenu.....	7
5.1 Les sites institutionnels.....	7
5.2 Les sites éditoriaux.....	8
5.3 Les communautés en ligne.....	8
5.4 Les sites catalogues et marchands.....	8
5.5 Les Intranet / Extranet.....	8
6. Fonctionnalités d'un système de gestion de contenu.....	9
6.1 Le stockage de contenu.....	9
6.2 La gestion de l'interface utilisateur.....	9
6.3 Planification et validation du contenu.....	9
6.4 Gestion des utilisateurs et contrôle d'accès.....	10
6.5 Indexation / Recherche.....	11
6.6 Publication.....	11
6.7 Syndication.....	11
7. Objectif d'un système de gestion de contenu.....	11
7.1 Eliminer les goulots d'étranglement.....	11
7.2 Faciliter la production du contenu.....	11
7.3 Organiser la production de contenu.....	12
7.4 Multidiffusion, mutualisation.....	12
8. Conclusion.....	13

Chapitre II : Etude du concept d'agrégation et de syndication de contenu Web

1.	Introduction	14
2.	Qu'est ce que la syndication de contenu ?.....	15
2.1	Définition de la syndication de contenu.....	16
2.1.1	Mise à disposition de l'information.....	16
2.1.2	Récupération de fils d'information sur d'autres sites.....	16
2.2	Intérêt de la syndication.....	17
2.3	Pour quel type de contenu utilise t-on la syndication ?.....	17
3.	Agrégation de contenu	17
4.	Structure d'un "FIL DE NOUVELLES" ?.....	18
5.	Les Formats de syndication existants	19
6.	Comparaison des deux formats de syndication.....	23
7.	Conclusion	24

Chapitre III : Introduction aux Langage XML et XSLT

1.	Introduction	25
2.	Présentation de XML.....	26
3.	Mise en page de XML.....	26
4.	Décodage d'un document XML.....	27
5.	Le XML et le HTML.....	27
6.	HTML, XML et XHTML.....	28
7.	Les avantages de XML.....	29
8.	Qu'est ce que le XML ?.....	30
9.	Structure d'un document XML.....	30
10.	La syntaxe des éléments en XML.....	31
11.	Le rôle de la DTD	33
12.	Espaces de noms.....	33
13.	Feuilles de style.....	33
14.	XSLT.....	33
15.	Conclusion	39

Chapitre IV : Etude des spécifications RSS

1.	Introduction.....	40
2.	RSS, un format XML pour l'échange de données.....	41
	2.1 Définition.....	41
	2.2 Historique.....	41
	2.3 Pourquoi utiliser un format XML pour l'échange de données?.....	42
	2.4 Les différents standards RSS.....	44
3.	RDF Site Summary (RSS 1.0).....	45
	3.1 Introduction.....	45
	3.2 Structure de base.....	45
	3.3 Les modules RSS 1.0.....	52
4.	Conclusion.....	58

Chapitre V : Conception et mise en œuvre de l'application

I.	Introduction au le Processus de développement RUP.....	59
II.	Démarche de conception.....	60
II.1	Expression des besoins et cas d'utilisation.....	60
	1. Contexte du problème.....	61
	2. Identification et description sommaire des acteurs.....	62
	3. Identification des cas d'utilisations.....	62
	4. Diagrammes et description des cas d'utilisation.....	63
	5. Diagrammes de séquences.....	69
	6. Identification des classes candidates.....	74
II.2	Analyse.....	75
	1. Développement du modèle statique.....	75
	1.1 Découpage en catégories.....	75
	1.2 Dépendances entre catégories.....	76
	2. Développement du modèle dynamique.....	78
	2.1 Diagrammes de collaboration.....	78

II.3 Conception	83
1. Conception architecturale du système.....	83
1.1 Architecture des sous-systèmes.....	84
1.1.1 Sous-système 1 : Serveur fournisseur de flux RSS.....	84
1.1.2 Sous-système 2: Editeur de flux RSS.....	85
1.1.3 Sous-système 3: Lecteur de flux RSS.....	85
2. Conception détaillée	86
2.1 Coté Serveur	86
2.2 Coté client	87
3. Les classes du domaine.....	88
4. Réalisation des cas d'utilisation.....	89
5. Le modèle relationnel.....	92
II.4 Implémentation	94
4.1 Les technologies utilisées.....	94
4.1.1 Plate forme.....	94
4.1.2 La distribution avec les Web Services	94
4.1.3 Environnement de développement	95
4.1.4 Environnement d'implémentation	95
4.1.5 Serveur de base de données	95
4.2 Implémentation des modules.....	96
4.3 Modèle physique de données.....	103
4.4 Sécurité des données.....	105
..	
Conclusion générale.....	107