

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université des Sciences et de la Technologie Houari Boumediène
Faculté d'électronique et d'informatique

Département d'Informatique



MEMOIRE

Présenté pour l'obtention du diplôme de MAGISTER
en INFORMATIQUE

Spécialité : Systèmes Intelligents et Ingénierie des Logiciels

Par : BOUKHEDOUMA Hocine

Sujet

Approche de Conception d'un Entrepôt de Données Multimédia à base Ontologique

Soutenu publiquement le 16 septembre 2012 devant le jury composé de :

M. BELKHIR A.	Professeur à l'USTHB	Président
Mme. ALIMAZIGHI Z.	Professeur à l'USTHB	Directrice de mémoire
M. BOUKRA A.	Maitre de conférences / A à l'USTHB	Examineur
M. BOUKHALFA K.	Maitre de conférences / A à l'USTHB	Examineur

Remerciements

Je tiens à adresser mes sincères remerciements à madame ALIMAZIGHI ZAIA, professeur à la FEI- USTHB pour m'avoir proposé le sujet et orienté dans la réalisation de ce travail. Je la remercie aussi pour sa gentillesse, sa patience et ses conseils.

Je tiens à remercier monsieur BELKHIR Abdelkader, professeur au département informatique de l'USTHB de m'avoir fait l'honneur de présider le jury de ce mémoire.

Je tiens aussi à remercier messieurs BOUKRA Abdelmadjid, BOUKHALFA Kamel et BOUYAKOUB fayçal, maîtres de conférences au département informatique de l'USTHB d'avoir accepté d'examiner et juger mon travail.

Je ne saurais terminer sans remercier vivement ma sœur SOUAD pour son aide précieuse, sa gentillesse, sa patience, ses conseils et sa disponibilité.

Je remercie aussi tous les membres de ma famille pour leur patience et leur soutien moral.

Je remercie Leila pour son soutien moral.

Enfin, Je remercie tous ceux qui ont participé, de près ou de loin, à l'aboutissement de ce travail.

Résumé

Les *entrepôts de données (ED)* ou encore *Data warehouse (DW)* ont été conçus dans le but d'aider les décideurs dans leur prise de décisions, à partir d'analyses effectuées sur un grand volume de données historisées.

La mise en œuvre d'un entrepôt de données passe par trois phases principales à savoir : la planification, la conception et la maintenance. La phase de *conception* est une phase très importante dans les projets d'entreposage de données. En effet, différentes études ont montré que le manque ou l'absence d'une phase de modélisation conceptuelle est une cause possible d'échec des projets d'entreposage.

Plusieurs approches de conception ont été proposées. Ces dernières peuvent être classifiées en trois catégories : les approches *directes*, les approches *traditionnelles* et les approches *ontologiques* assurant une conception à base d'ontologies.

Par ailleurs, dans certains domaines (telles que la médecine, la biologie, la géographie), les données se présentent sous diverses formes (texte, image, son, vidéo) et sont qualifiés de données multimédias. Ces données ont la particularité d'être volumineuses et complexes nécessitant ainsi des traitements et des outils particuliers pour leur stockage et leur exploitation. Les entrepôts qui traitent ce type de données sont appelées *entrepôt de données multimédia (EDM)*.

Notre problématique consiste à proposer une démarche de conception d'entrepôts de données multimédia sur la base d'approches existantes. Pour cela, nous avons adopté une approche de conception d'un EDM prenant en compte les sources de données et les besoins des utilisateurs en même temps. Nous optons pour une approche à base *ontologique* permettant ainsi de régler les conflits sémantiques dans la phase d'intégration des données. De plus, l'apport de notre approche par rapport aux approches existantes réside dans la prise en compte des données *multimédia*.

Mots clés

Données multimédia- Entrepôt de données multimédia – Ontologie – Approche de conception – Intégration.

Table des matières

Page

Introduction générale	1
1. Contexte	1
2. Problématique et contribution	2
3. Organisation du document	2
Chapitre 1 : Les entrepôts de données (ED)	4
1. Introduction.....	4
2. Définition d'un ED.....	5
3. Architecture et composants.....	5
3.1 Données opérationnelles.....	6
3.2 Magasin de données opérationnelles.....	6
3.3 Administrateur de chargement	7
3.4 Administrateur de l'entrepôt	7
3.5 Administrateur de requêtes	7
3.6 Données détaillées	7
3.7 Données légèrement ou fortement résumées	8
3.8 Données archivées et sauvegardées	8
3.9 Métadonnées	8
3.10 Outils d'accès des utilisateurs finaux	8
4. Mini-entrepôt de données	9
4.1 Définition d'un mini-entrepôt de données	9
4.2 Raisons de créer un mini-entrepôt de données	10
4.3 Architecture d'un mini-entrepôt de données	10
5. La modélisation multidimensionnelle	12
5.1 Définitions	13
5.2 Les implémentations d'un cube de données	13
5.2.1 L'approche MOLAP	13
5.2.2 L'approche ROLAP	14
a. Le modèle en étoile	15
b. Le modèle en flocon de neige	16
c. La constellation de faits	16
5.2.3 L'approche HOLAP.....	17
5.3 Opérations sur les cubes de données	18
6. Maintenance d'un ED	18
7. Outils et technologies de l'entreposage de données	20
7.1 Outils d'extraction de purification et de transformation.....	20
7.2 SGBD de l'entrepôt	20
7.3 Outils d'administration et de gestion	21

8. Evolution d'un ED	21
9. Avantages apportés par les ED	22
10. Problèmes liés aux ED	23
11. Conclusion	23
Chapitre 2. Approches de conception d'ED.....	24
1. Introduction.....	24
2. Cycle de vie d'un ED.....	25
a. Planification	25
b. Conception et implémentation	25
c. Maintenance et évolution	25
3. Conception d'un ED	25
3.1 Analyse des besoins	26
3.2 Modélisation conceptuelle	28
3.3 Modélisation logique	28
3.4 Processus ETL	29
3.5 Modélisation physique	29
4. Approche de conception des projets d'entreposage	29
4.1 Les approches directes	30
4.1.1 Principe et travaux connexes.....	30
4.1.2 Analyse	32
4.2 Les approches traditionnelles	33
4.2.1 Les méthodes orientées sources	33
4.2.2 Les méthodes orientées besoins utilisateurs	35
4.3 Les approches ontologiques	40
5. Conclusion	43
Chapitre 3 : Modélisation des données multimédias	44
1. Introduction.....	44
2. Les données multimédias.....	45
2.1 Le multimédia	45
2.2 Utilité du multimédia.....	45
2.3 Types de données multimédias	46
2.3.1 Les textes.....	46
2.3.2 Les images	46
2.3.3 Les sons	47
2.3.4 La vidéo	48
2.4 Spécificités des supports	49
2.5 Spécificités des outils et méthodes de développement	50
2.6 Spécificités des outils de traitement	50

3. Problématique de la modélisation des données multimédias	51
3.1 Acquisition	51
3.2 Compression et stockage	51
3.3 Accès	51
3.4 Composition et présentation	52
4. XML et les données multimédias	52
4.1 Qu'est ce que XML et pourquoi l'utiliser ?	52
4.2 Avantages de XML	52
4.3 Comment utiliser XML	53
4.4 Applications de XML	53
4.5 Description des données multimédias avec XML	54
4.5.1 XML et le format SVG	54
4.5.1.1 Objectifs de SVG	54
4.5.1.2 Définition du format	55
4.5.2 Utilisation de XML avec FLASH pour représenter des mouvements.....	55
4.5.3 Le standard SMIL	56
5. La norme MPEG-7	57
5.1 Principe de description de MPEG-7.....	58
6. Modélisation des données dans les BD multimédia	59
6.1 Caractéristique d'un SGBD multimédia	59
6.2 Architecture fonctionnelle d'un SGBD multimédia	59
6.3 Gestion de données multimédias avec Oracle 8	60
6.3.1 Stockages des données	60
6.3.1.1 Stockage interne des données	60
6.3.1.2 Stockage externe des données	60
6.3.2 Chargement de données	61
6.3.3 Gestion des images avec Oracle	62
6.3.3.1 Comparaison d'images	62
6.3.3.2 Quelques opérateurs utilisés pour la comparaison d'images	65
6.4 Gestion de données multimédia avec MySQL	66
6.4.1 Stockage d'un fichier dans une table MySQL	66
7. Conclusion.....	67
Chapitre 4 : Les entrepôts de données multimédias (EDM).....	68
1. Introduction	68
2. Qu'est ce qu'un ED multimédia ?.....	69
3. Architecture d'un ED multimédia	69
4. Domaines d'utilisation des ED multimédias	70
4.1 La médecine	71
4.2 Le Web	71

5. Problèmes liés aux ED multimédias	72
6. Modélisation des données dans les ED multimédias	72
6.1 Travaux liés à la modélisation multidimensionnelle multimédia	73
6.2 Cube de données multimédias	73
7. Quelques travaux liés aux ED multimédias	74
7.1 XML et les ED multimédias	74
7.1.1 Le générateurs ' <i>Meta view generator</i> '	74
7.1.2 Le gestionnaire ' <i>Meta knowledge manager</i> '	75
7.1.3 Les ' <i>wrappers</i> '	76
7.2 Aspect sémantique dans les ED multimédias	77
7.2.1 Relation entre les objets	77
7.2.2 Versionnement de schémas basé sur la sémantique	79
7.3 Traitement des requêtes dans les ED multimédias	79
7.3.1 Utilisation du vecteur de domaine	80
7.3.2 Avantage apportés par la structure du vecteur de domaine	80
8. Conclusion.....	80
Chapitre 5 : Les ontologies : concepts, principes et application	81
1. Introduction	81
2. Motivation	82
3. Définitions	83
4. Eléments de base d'une ontologie	83
5. Construction d'une ontologie	84
5.1 Approche générale	85
5.2 Processus de développement d'une ontologie d'application.....	89
5.2.1 Description du processus	90
5.2.2 Limites de l'approche proposée	93
5.2.3 Discussion	93
6. Ontologies et multimédia.....	94
6.1 Analyse de l'image et conceptualisation.....	94
6.2 Exploitation d'un consensus	95
6.3 Applications des ontologies multimédias	96
6.3.1 Le domaine de la biologie : « les réseaux de neurone)	96
6.3.2 Les ontologies et l'imagerie satellitaire	102
- Les projets du domaine satellitaire	102
- Le projet DAFOE	103
6.4 Une méthode interactive et incrémentale de construction de concepts	104
7. Conclusion	105

Chapitre 6 : Démarche de conception d'EDM proposée.....	106
1. Introduction	106
2. Analyse des approches par rapport aux problèmes des ED et aux données multimédias	107
2.1 Analyse par rapport aux problèmes des EDs	107
2.2 Analyse par rapport aux données multimédias	108
2.3 Comment choisir l'approche appropriée ?	109
3. Approche d'intégration des données	110
3.1 La représentation des données intégrées.....	110
3.1.1 Approche virtuelle	110
3.1.2 Approche matérialisée	110
3.2 Le sens de la mise en correspondance entre le schéma global et les Schémas locaux	111
3.2.1 Approche GaV	111
3.2.2 Approche LaV	111
3.3 La nature du processus d'intégration	111
3.3.1 Les approches manuelles	112
3.3.2 Les approches semi-automatiques	112
3.3.3 Les approches automatiques	112
4. Analyse des approche par rapport à notre cas.....	113
5. Approche proposée	113
5.1 Architecture de l'approche proposée	113
5.2 Etapes de la démarche proposée	115
6. Exemple d'illustration	118
7. Conclusion	124
Conclusion générale	125
Bibliographie	127
Annexe A : Compléments sur les standards de description de données multimédia	
Annexe B : Langages et outils d'édition d'ontologies	