

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

Université Abderrahmane Mira de Bejaia
Faculté des Sciences Exactes
Département d'Informatique

ECOLE DOCTORALE RESEAUX ET SYSTEMES DISTRIBUES

Mémoire de Magister

En Informatique

Option : Réseaux et Systèmes Distribués

Thème

Vers un langage de haut niveau pour la recherche d'information multimédia

Présenté par

AMRANE Abdesalam

Devant le jury composé de :

Président	TARI Abdelkamel	M.C.A	Université de Bejaïa
Examineur	NOUALI Omar	Directeur de recherche	CERIST, Alger
Examineur	CHALAL Rachid	M.C.A	ESI, Alger
Rapporteur	AMGHAR Youssef	Professeur	INSA de Lyon, France
Invitée	MELLAH Hakima	Chargée de recherche	CERIST, Alger

Mars 2013

Remerciements

Tout d'abord je remercie Dieu tout puissant de m'avoir donné le courage et la patience pour achever ce modeste travail.

J'adresse mes remerciements à mes directeurs de recherche, pour avoir accepté de m'encadrer et m'avoir aidé à conduire ce travail jusqu'au bout : Monsieur AMGHAR Youssef, Professeur des universités et Directeur du département informatique de l'INSA de Lyon, ainsi que Mme MELLAH Hakima, Chargée de recherche au sein du CERIST.

Je remercie: Mr. TARI Abdelkamel, Mr. NOUALI Omar et Mr. CHALAL Rachid pour avoir accepté de juger ce modeste travail.

Je remercie tous les membres de ma famille pour m'avoir soutenu et supporté durant toute cette période de travail, en commençant par ma mère Ouardia, mon épouse Anissa, et surtout mes petits enfants, Amira, Ikram, Ferial, Amine, Anes et Mohamed qui exprimes un grand respect envers moi.

Je remercie tous les membres de ma famille ainsi que ma belle famille pour m'avoir encouragé.

Je n'oublierai pas de remercier mes collègues du CERIST, pour m'avoir encouragé, motivé et surtout soutenu pour finaliser mon mémoire.

Table des matières

Introduction générale.....	1
1. Contexte	1
2. Problématique	1
3. Contribution	2
4. Organisation du mémoire	3
Partie I : Etat de l’art	4
Chapitre 1 : Recherche d’information classique	5
1.1. Introduction	5
1.2. Processus de recherche d’information	5
1.2.1. Processus d’indexation.....	6
1.2.2. Pondération des termes	7
1.2.3. Appariement requête-document.....	9
1.2.4. Reformulation de requêtes	9
1.3. Modèles de recherche d’information	10
1.3.1. Modèle booléen.....	11
1.3.2. Modèle vectoriel	11
1.3.3. Modèle probabiliste	12
1.3.4. Modèle de langue	13
1.4. Evaluation des systèmes de recherche d’information	14
1.5. Conclusion.....	17
Chapitre 2 : Recherche d’information multimédia.....	18
2.1. Introduction	18
2.2. Approches d’indexation et de recherche multimédia.....	19
2.3. Recherche multimédia basée sur le texte	19
2.3.1. Indexation manuelle.....	20
2.3.2. Indexation automatique.....	22
2.4. Recherche multimédia basée sur le contenu	24
2.4.1. Indexation par le contenu visuel	25
2.4.1.1. Descripteurs de couleur	25
2.4.1.2. Descripteurs de texture	27

2.4.1.3. Descripteurs de forme	28
2.4.1.4. Segmentation d'image.....	29
2.4.1.5. Points d'intérêt	30
2.4.2. Indexation par concepts sémantiques.....	30
2.5. Recherche multimédia multimodale	35
2.6. Tableau comparatif des différentes approches	37
2.8. Conclusion.....	38
Chapitre 3 : Les langages de requêtes	39
3.1. Introduction	39
3.2. Langages textuels	40
3.2.1. SQL	40
3.2.2. XQuery.....	41
3.3. Langages graphiques	42
3.3.1. XQBE.....	43
3.3.2. GMQL.....	44
3.4. Langages visuels	46
3.5. Synthèse	48
3.6. Conclusion.....	52
Partie II : Contribution	53
Chapitre 4 : Modèle de recherche sémantique d'information multimédia.....	54
4.1. Introduction	54
4.2. Motivations	56
4.3. Modèle MMSemSearch proposé.....	60
4.3.1. Indexation des images.....	61
4.3.1.1. Indexation par le contexte	61
4.3.1.2. Indexation par le contenu	65
4.3.1.3. Pondération des concepts et non-concepts	69
4.3.2. Processus de recherche	70
Chapitre 5 : Evaluations expérimentales	71
5.1. Introduction	71
5.2. Indexation par le contexte	71
5.3. Indexation par les concepts	73
5.4. Langage de requêtes.....	74
5.5. Evaluation du SRI multimédia	76

5.6. Conclusion.....	79
Conclusion générale et perspectives.....	80
1. Conclusion générale	81
2. Perspectives.....	82
Bibliographie et Annexes	83
Bibliographie	84
Annexes	91
1. Description de <i>WordNet</i>	91

Liste des figures

Figure 1. Processus de recherche d'information [CHA 05]	6
Figure 2. Répartition des documents d'une collection suite à une interrogation [HO 04].....	16
Figure 3. Approches de recherche multimédia [BAN 09]	19
Figure 4. Interface de l'outil Video-Annex [LIN 03].....	21
Figure 5. Interface de l'outil M-OntoMat-Annotizer [PET 06].....	22
Figure 6. Exemple de segmentation en bloc d'une page web [FAE 11]	24
Figure 7. Architecture du processus de classification supervisée [AYA 07].....	32
Figure 8. Exemples d'annotation automatique par le système ALIPR [JIA 08].....	34
Figure 9. Architecture globale du modèle de RI multimodale [MOU 10].....	36
Figure 10. Composants de XQBE	43
Figure 11. Requête XQBE	44
Figure 12. Icônes graphiques utilisées par GMQL	45
Figure 13. Exemple de requête graphique GMQL	46
Figure 14. Une requête par sketch [IMG 09]	47
Figure 15. Scénario d'utilisation du langage MPQF [TOU 08]	49
Figure 16. Exemple de requête MPQF [TOU 08].....	50
Figure 17. Exemple de résultat de requête MPQF [TOU 08]	51
Figure 18. Résultat de la requête « horses » retourné par Google.....	56
Figure 19. Exemple de bruit dans la recherche d'images.....	57
Figure 20. Résultat de la requête « helicopter site : www.globalhopenetwork.org »	58
Figure 21. Exemple de silence dans la recherche d'images	59
Figure 22. Architecture proposée pour le Système d'Indexation Automatique et de Recherche Sémantique d'Information Multimédia (MMSemSearch)	60
Figure 23. Exemple d'image avec les informations de contexte.....	62
Figure 24. Schéma de l'algorithme Lesk [VAS 03].....	64
Figure 25. Algorithme Lesk simplifié [VAS 03]	65
Figure 26. Hyperplan SVM.....	66
Figure 27. Exemple de classifieur SVM	66
Figure 28. Architecture du système d'indexation d'images par le contenu	67
Figure 29. Exemple de contenu de la base de données exploitée.....	75
Figure 30. Un exemple de résultat de recherche retourné par le système proposé	76
Figure 31. Courbes Rappel/Précision pour les trois modalités de recherche	78

Liste des tableaux

Tableau 1. Comparaison des approches de recherche d'information multimédia	37
Tableau 2. Tableau comparatif des langages de requêtes	48
Tableau 3. Résultats d'évaluation de l'indexation par le contexte.....	72
Tableau 4. Données d'apprentissage et de tests pour la classification des images	73
Tableau 5. Résultats d'évaluation de l'indexation par les concepts.....	74
Tableau 6. Résultats d'évaluation obtenus selon le score de confiance α	77
Tableau 7. Récapitulatif des résultats obtenus pour les différentes modalités.....	78