

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

جامعة هواري بومدين للعلوم والتكنولوجيا  
UNIVERSITE DES SCIENCES ET TECHNOLOGIES HOUARI BOUMEDIENNE

FACULTE DE GENIE ELECTRIQUE ET D'INFORMATIQUE  
DEPARTEMENT INFORMATIQUE

*Mémoire de fin d'études pour l'obtention du diplôme  
D'ingénieur d'Etat en Informatique*

Thème :

**La recherche d'information personnalisée  
dans un environnement distribué**

Thème proposé et encadré par : Mr S. KECHID (USTHB)

Réalisé par : Mr AMRANE Abdeslam & Mr EL MARAI Oussama

Le jury:

Président du jury : Mr KADDOURI

Membre du jury : Mr SELMOUNE

Membre du jury : Mme ALEB

Promotion 2004/2005 – N° 13

# SOMMAIRE

<b>Introduction générale</b> .....	5
------------------------------------	---

## **Chapitre I: Introduction aux systèmes de recherche d'informations**

I.1 Introduction .....	8
I.2 Problématique de la recherche d'information .....	9
I.3 Représentation de l'information textuelle "Indexation" .....	9
I.3.1 Extraction des mots simples .....	10
I.3.2 Normalisation des mots extraits .....	10
I.3.3 Pondération des mots normalisés .....	11
I.4 Les modèles de recherche d'information .....	13
I.4.1 Le modèle "Matching score" .....	14
I.4.2 Le modèle booléen .....	15
I.4.2.1 Le modèle booléen standard .....	15
I.4.2.1.1 Avantage .....	16
I.4.2.1.2 Inconvénients .....	16
I.4.2.2 Le modèle booléen basé sur les ensembles flous .....	16
◆ Reformulation de requête .....	17
◆ Associer une importance aux termes de la requête .....	18
I.4.3 Le modèle vectoriel (VSM) .....	18
I.4.3.1 Principe .....	18
I.4.3.2 Avantages .....	21
I.4.3.3 Inconvénients .....	21
I.4.4 Le modèle probabiliste .....	21
I.4.4.1 Le modèle probabiliste de base .....	22
I.4.4.2 Le modèle de recherche indépendant .....	23
I.4.4.3 Le modèle de Robertson et Spark-Jones .....	24
I.4.4.4 Avantages .....	25
I.4.4.5 Inconvénients .....	25
I.5 Conclusion .....	26

## **Chapitre II: Recherche d'information personnalisée**

II.1 Introduction .....	28
II.2 La recherche classique .....	29
II.3 La recherche adaptative .....	29
II.3.1 Reformulation de requête .....	29
II.3.1.1 Reformulation directe .....	30
II.3.1.2 Reformulation par injection de pertinence .....	30
II.3.1.3 Reformulation dans le modèle vectoriel .....	30
II.3.2 Filtrage d'information .....	31
II.3.2.1 Les modes de filtrage .....	32
II.3.2.2 Limites de la recherche d'information adaptative .....	32

II.4 La recherche personnalisée .....	33
II.4.1 Modélisation de l'utilisateur .....	34
II.4.1.1 Le contenu du profil .....	34
II.4.1.2 Représentation du profil .....	35
II.4.1.2.1 Représentation vectoriel .....	35
• Représentation vectoriel personnalisée .....	35
• Représentation par des stéréotypes .....	35
II.4.1.2.2 Représentation sémantique .....	35
II.4.1.2.3 Représentation multidimensionnelle .....	36
II.4.1.3 Construction et mise à jour du profil .....	36
II.4.1.3.1 Analyse statistique des termes .....	37
II.4.1.3.2 Algorithmes de filtrage collaboratif .....	37
II.4.1.3.3 Techniques d'apprentissages .....	38
a) Les techniques de classification .....	38
• Le raisonnement par cas .....	38
• Les classificateurs bayésiens .....	38
b) Les méthodes de clustering .....	38
II.4.2 Agents personnels de recherche d'information .....	39
II.5 Conclusion .....	40

### Chapitre III: La recherche distribuée d'information

III.1 Introduction .....	42
III.2 Les systèmes de recherche distribuée d'information .....	43
III.2.1 Les outils de recherche d'information existant.....	43
III.2.1.1 Les annuaires.....	44
III.2.1.2 Les moteur de recherche plein texte.....	44
III.2.1.3 Les métas moteurs.....	45
III.2.1.4 Les agents intelligents.....	45
III.2.1.5 Les outils graphiques.....	45
III.2.1.6 Les assistants personnels.....	45
III.2.1.7 Les outils collectifs.....	45
III.2.2 classement de serveurs des recherche .....	46
III.2.2.1 L'environnement coopératif .....	46
III.2.2.2 L'environnement non coopératif .....	46
III.3 Sélection de serveurs .....	46
III.3.1 L'environnement coopératif .....	46
III.3.1.1 GLOSS .....	47
III.3.1.2 CORI .....	49
III.3.2 L'environnement non coopératif .....	51
III.3.2.1 Acquisition de modèle de langage par échantillonnage .....	51
III.4 Reformulation de requête .....	51
III.4.1 Langage de requête en méta recherche .....	52
III.4.2 Subsumption de requête .....	54
III.4.3 Fonction de filtrage .....	55

III.4.4 Les prédicats des termes .....	55
III.4.5 Les prédicats des champs .....	56
III.4.6 Algorithme de reformulation de requête .....	57
III.4.7 Exemple de subsomption de prédicats .....	58
III.5 Fusion des rangs .....	60
III.5.1 Cas des serveurs coopératifs .....	60
III.5.2 Cas des serveurs non coopératifs .....	63
III.6 Conclusion .....	65

## Chapitre IV: Conception

IV.1 Introduction .....	67
IV.2 Notre approche .....	68
IV.2.1 Gestion du profil utilisateur .....	69
IV.2.1.1 Dimension du profil persistant .....	69
a) Les données personnelles .....	69
b) L'ontologie du domaine .....	69
IV.2.1.2 Dimension du profil évolutif .....	69
a) Le centre d'intérêt .....	69
b) Les préférences personnelles .....	69
IV.2.2 Utilisation du profil dans notre travail .....	70
IV.2.3 Reformulation de la requête .....	70
IV.2.4 L'environnement choisi pour la recherche distribuée .....	71
IV.3 Diagramme des cas d'utilisation .....	72
IV.3.1 Les acteurs .....	72
IV.3.2 Les cas d'utilisation .....	73
IV.3.2.1 Inscription .....	73
IV.3.2.2 Identification .....	73
IV.3.2.2 Recherche personnalisée .....	73
IV.3.2.3 Recherche adaptative .....	74
IV.4 Diagrammes de séquences .....	74
IV.4.1 Inscription .....	75
IV.4.2 Identification .....	75
IV.4.3 Recherche personnalisée .....	76
IV.4.4 Recherche adaptative .....	77
IV.5 Diagrammes de collaboration .....	78
IV.5.1 Inscription .....	78
IV.5.2 Identification .....	79
IV.5.3 Recherche personnalisée .....	79
IV.5.4 Recherche adaptative .....	80
IV.6 Diagrammes d'activité .....	80
IV.6.1 Inscription .....	81
IV.6.2 Identification .....	82
IV.6.3 Recherche personnalisée .....	83
IV.6.4 Recherche adaptative .....	84

IV.7 Diagramme de classe .....	85
IV.7.1 Classe MétaMoteur .....	85
IV.7.2 Classe Serveur .....	86
IV.7.3 Classe Utilisateur .....	87
IV.7.4 Classe TypeUtilisateur.....	88
IV.7.5 Classe Domaine .....	88
IV.7.6 Classe MotClé.....	89
IV.7.7 Classe Langue .....	89
IV.7.8 Classe TypeDocument .....	90
IV.7.9 Classe Période .....	90
IV.7.10 Diagramme de classe .....	92
IV.8 Conclusion .....	93

## Chapitre V: Implémentation

V.1 Introduction .....	95
V.2 L'environnement de travail .....	95
V.2.1 Description du langage de développement utilisé .....	95
V.2.2 Le serveur web Apache .....	96
V.2.3 Le SGBD utilisé .....	96
V.2.4 Les services web .....	96
V.2.4.1 Architecture des services web .....	97
V.2.4.2 Les normes des services web .....	98
▪ Simple Object Acces protocole .....	98
V.3 Configuration requise .....	99
V.4 Interfaces utilisateur .....	100
V.5 Conclusion .....	103

<b>Conclusion générale</b> .....	105
----------------------------------	-----

## Annexe

## Bibliographie