

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari BOUMEDIENE

INSTITUT D'INFORMATIQUE

Mémoire du Projet de fin d'études

Pour l'obtention du diplôme
D'ingénieur d'état en informatique

SUJET :

**Construction Automatique de Profils par une
Méthode de Classification Non Supervisée**

Encadreur :

Mr. Omar Nouali

Réalisé par :

Mlle. Ourari Meriem

Mlle. Harrouni Bétra

Soutenu devant le jury composé de :

Mme. F.Khellaf

Présidente du jury

Mme. S.Benchouk

Membre du jury

Melle. Hank

Membre du jury

Organisme d'accueil : Laboratoire des Logiciels de Base, CERIST.

PROMOTION : 2003/2004

N° 33/04

Sommaire

Introduction Générale.....	1
-----------------------------------	----------

Chapitre I

Recherche d'Information et Filtrage d'Information

1. Introduction.....	4
2. La Recherche et le Filtrage d'Information	5
2.1. La Recherche d'Information IR.....	7
2.1.1. Définition.....	7
2.1.2. Modèles conceptuels des systèmes IR.....	8
2.1.3. Opérations sur les mots.....	9
2.1.4. La norme opérationnelle pour les systèmes IR.....	10
2.1.5. Evaluation de systèmes IR.....	11
2.2. Le Filtrage d'Information IF	11
2.2.1. Définition.....	11
2.2.2. Caractéristiques du Filtrage d'Information	12
2.2.3. Méthodes de Filtrage.....	13
2.2.4. Modélisation du profil de l'utilisateur.....	14
2.2.5. Evaluation de système de IF	15
3. Conclusion	16

Chapitre II

Classification Documentaire

1. Introduction.....	18
2. Le clustering de documents	19
3. Domaines d'application	19
4. Un système de classification.....	20

5. Méthodes de Classification	21
5.1. Méthodes d'explication et de prédiction (supervisées).....	21
5.1.1. Arbre de décision	22
5.1.2. Réseaux de neurones.....	25
a) Exemple de réseaux de neurones	26
b) Etapes de construction d'un réseau de neurone.....	27
5.1.3. Naïve Bayes (ou Simple Bayes)	28
5.1.4. K plus proches voisins	28
5.2. Méthodes de structuration (non supervisées).....	31
5.2.1. Méthodes monothétiques	32
• La segmentation de Williams et Lambert	32
5.2.2. Méthodes polythétiques	33
• Méthodes de classification hiérarchique (dendrogramme)	33
• Les Nuées Dynamiques.....	35
5.2.3. Carte auto organisatrice de Kohonen.....	36
6. Conclusion	38

Chapitre III

Conception du Système

1. Introduction.....	40
2. Architecture globale du système.....	41
3. Prétraitement des documents	43
4. Traitement sur les documents (textes)	43
4.1. Filtrage de mots vides (stoplist).....	43
4.2. Lemmatisation.....	44
4.2.1. Lemmatiseur pour l'anglais	44
Algorithme de Porter.....	44
4.2.2. Lemmatiseur pour le français.....	46
Algorithme pour le français	46
4.3. Représentation des documents	47
5. La Classification	49
5.1. Calcul des distances	49
5.2. Représentation d'une classe.....	50

5.3. Distance document - classe	50
5.4. Classification.....	50
5.4.1 Classification hiérarchique.....	51
5.4.2. Nuées dynamiques	51
6. Interprétation des classes	52
7. Conclusion	53

Chapitre IV

Mise en Œuvre et Evaluation

1. Introduction.....	56
2. Choix du langage	56
3. Implémentation des différents modules du système	57
3.1. Suppression des caractères spéciaux.....	57
3.2. Filtrage de mots vides (stoplist).....	57
3.3. Construction du vecteur V (vocabulaire).....	58
3.4. Calcul des fréquences	60
3.5. Calculer_ IDF.....	60
3.6. Algorithme de représentation des documents	61
3.7. Matrice des distances	63
3.8. Calcul de la Distance	64
3.9. Calcul des représentants d'une classe.....	65
3.10. Distance entre deux classe	65
3.11. Calcul de la Matrice des Représentants	66
3.12. Classification hiérarchique.....	67
3.13. Nuées Dynamiques	68
3.14. Interprétation.....	69
4. L'interface graphique.....	69
4.1. Fenêtre d'accueil.....	70
4.2. Fenêtre de saisie des données	70
4.2. Fenêtre résultat.....	72
5. Evaluation	73
❖ Expérience 1.....	73
❖ Expérience 2.....	79

❖ Expérience 3.....	83
6. Conclusion	88
Conclusion Générale.....	90
Références Bibliographiques	
Annexe A	
Annexe B	
Annexe C	
Annexe D	
Liste des Figures	
Liste des Tableaux	
Liste des Algorithmes	
Glossaire	