



UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENE

FACULTE DE MATHEMATIQUES

THESE

Présentée pour l'obtention du grade de :

MAGISTER

En : **MATHEMATIQUES**

Spécialité : RECHERCHE OPERATIONNELLE

Par : **KOUICI SALIMA**

Thème

Réseaux de neurones artificiels et analyse en composantes principales pour l'amélioration de la recherche d'information

Soutenue le 17/12/2002, devant le jury :

Mr BERRACHEDI Abdelhafid	Maître de Conférence, USTHB	Président
Mr ABBAS Moncef	Professeur, USTHB	Directeur de thèse
M. BOUCHEMAKH Isma	Maître de Conférence, USTHB	Examinatrice
Mr. MAAMRA M. Saïd	Chargé de cours, USTHB	Examinateur
Mr MOULAI Mustapha	Chargé de cours, USTHB	Examinateur
Mr MOULAI Mustapha	Chargé de cours, USTHB	Examineu

Sommaire

Partie I : Introduction et état de l'art

I- Introduction.....	6
1- Problématique.....	6
2- Objectifs.....	7
II- Notions de base.....	8
1- L'information.....	8
1.1-Typologie de l'information.....	9
1.2- L'information scientifique et technique.....	9
2- Le document.....	9
2.2-Typologie des documents.....	10
III- Représentation de l'information.....	11
1- L'indexation.....	11
2- Base de données bibliographiques.....	12
3- Base de données en texte intégral.....	12
4- L'indexation automatique.....	13
4.1- Méthodes d'extraction des termes d'indexation	13
4.1.1- Méthode statistique.....	13
4.1.2- Méthode lexicale.....	14
5- Pondération des termes d'indexation.....	15
IV- Système de recherche d'information	16
V- Modèles de recherche d'information.....	17
1- Modèles classiques de recherche d'information	18
1.1- Modèle booléen.....	18
1.2- Modèle utilisant la logique floue.....	18
1.3- Modèle vectoriel.....	20
1.4- Modèle probabiliste.....	21
2- Nouveaux modèles.....	22
2.1- Modèle de recherche basé sur le profil des utilisateurs.....	22
2.2- Modèle de recherche basé sur la reformulation de la requête....	25

Partie II : Modèles neuronaux, analyse des données et distribution Ziphienne pour la classification, la cartographie et le découpage des données documentaires

I- Les réseaux de neurones artificiels.....	27
1- Introduction.....	27
2- Neurone biologique.....	27
3- Neurone formel.....	28
4- Réseau de neurones artificiels.....	29
5- Typologie des Réseaux de neurones artificiels	30
5.1- Réseaux non bouclés.....	30
5.2- Réseaux bouclés.....	31
6- L'apprentissage dans les réseaux de neurones.....	31
7- Problèmes résolus grâce aux réseaux de neurones artificiels.....	32
II- La méthode de classification K Means Axiales.....	33
1- Introduction.....	33
2- La méthode K Means.....	33
3- La loi d'apprentissage d'Oja	34
4- La loi d'apprentissage d'Oja modifiée.....	35
5- La méthode K Means Axiales.....	36
6- Version non adptative (iterative) de la méthode K means axiales.....	37
7- Application de la méthode K means axiales pour le traitement des données documentaires.....	38
7.1- Spécifités des données documentaires.....	38
7.2- Choix de la méthode K means axiales.....	38
7.3- Principe de la méthode pour la classification des données documentaires.....	38
7.4- Etapes de la méthode.....	39
7.5- Algorithme détaillé de la méthode.....	40
7.6- Organigramme de la méthode.....	42
7.7- Conclusion.....	44
III- L'analyse de données pour la représentation des classes thématiques.....	45
1- Analyse de données.....	45
2- Analyse en composantes principales.....	45

2.1- Tableau de données.....	46
2.2- Pondération des individus.....	46
2.3- Centre de gravité du nuage des individus.....	46
2.4- Tableau centré des données.....	47
2.5- Matrice de variance covariance des variables.....	47
2.6- Matrice de corrélation.....	47
2.7- Tableau centré réduit des données.....	47
2.8- Espace vectoriel des individus	48
2.9- L'inertie totale du nuage des individus.....	48
2.10- Espace des variables.....	48
3- Principe de la méthode ACP.....	50
3.1-Axes principaux.....	51
3.2-Facteurs principaux.....	52
3.3- Composantes principales.....	52
4- L'Analyse en composantes principales pour la représentation des classes thématiques.....	53
IV- Distribution Ziphienne pour le découpage de l'information.....	54
1- Distribution Ziphienne.....	54
2- Découpage en zones.....	56
3- Conclusion.....	57

Partie III : Réalisation d'un prototype de système de recherche d'information en se basant sur la classification thématique et la cartographie des données

I- Présentation du système.....	59
II- Facteurs à prendre en compte pour la réalisation du système.....	60
1- Diversité des systèmes de gestion des bases de données.....	60
2- Diversité de la structure des données.....	60
3- Volume important des données.....	61
4- Diversité des traitements à effectuer par le système.....	61

II- Traitements effectués par le système	61
1- Chargement des données.....	61
2- Génération de données descriptives.....	63
3- Découpage des données.....	63
4- Classification thématique.....	63
5- Représentation des classes thématiques sur une carte.....	63
6- Recherche d'information.....	63
IV- Schéma général du système.....	65
V- Réalisation du système.....	67
1- Environnement de développement.....	67
2- Le système WINISIS.....	67
3- L'interface ISIS-DLL.....	68
4- Interface hypertextuelle.....	68
4.1- L'hypertexte	69
VI- Description générale des Menus.....	70
1- Fenêtre principale.....	70
2- Fonctionnalités du système	73
2.1- Balayage des données.....	73
2.2- Impression des données.....	73
2.3- Génération des données descriptives.....	74
2.4- Découpage des données.....	75
2.5- Classification thématique des données.....	76
2.6- Cartographie des classes thématiques.....	78
2.7- Recherche bibliographique.....	78
3- Perspectives.....	82
VIII- Contexte d'application.....	83
Bibliographie.....	84