

République Algérienne Démocratique & Populaire

Ministère de l'Enseignement Supérieure et de la Recherche
Scientifique

INSTITUT NATIONAL D'INFORMATIQUE
I.N.I
Oued Smar - ALGER

MEMOIRE

Pour l'obtention du Diplôme
d'Ingénieur d'Etat en Informatique
Option : Systèmes Informatiques

THEME

**Utilisation des cartes auto-organisatrices de
Kohonen dans la recherche documentaire**

Encadré par :
M^{me}. Bessai F.Z.

Réalisé par :

**Guergab Mohamed
Krelifaoui Mohamed**

Organisme d'accueil :
**CENTRE de Recherche sur l'Information Scientifique et Technique
(CERIST)**

Promotion 1999

Résumé

La quantité toujours croissante de données textuelles nécessite un effort constant pour la mise au point de méthode de stockage et de consultation afin que la totalité des informations conservées reste accessible de manière satisfaisante.

Dans ce travail nous nous intéressons à la conception et la réalisation d'un système de consultation et de classification automatique de données bibliographiques.

Ce système utilise la théorie des réseaux de neurones et plus précisément les cartes auto-organisatrices de Kohonen. Ces cartes permettent le classement d'objets tout en produisant des cartes de densités qui forment les fondements du système de consultation et de navigation à travers la base de données.

Mots-clés

Système de recherche documentaire, SRD, réseaux de neurones, apprentissage, carte auto-organisatrice de Kohonen, classement de documents, navigation, carte de densité, recherche documentaire, carte réactive.

SOMMAIRE

RÉSUMÉ	I
REMERCIEMENTS	II
SOMMAIRE.....	III
LISTE DES FIGURES	VII
LISTE DES TABLEAUX.....	IX
INTRODUCTION GÉNÉRALE	1
I. LES SYSTÈMES DE RECHERCHE DOCUMENTAIRE CLASSIQUES	3
I.1. Introduction.....	3
I.2. Les Bases Documentaires	4
I.3. Les systèmes de recherche documentaire (SRD).....	4
1.3.1. L'indexation.....	7
1.3.2. les modèles de recherche	10
1.3.3. Evaluation des systèmes de recherche documentaire	12
I.4. Limites des SRD classiques	14
I.5. Nouvelle approche des SRD	14
I.5.1. Recherche approchée	15
I.5.2. Les graphes conceptuels	15
I.5.3. Hypertexte et hypermédia	15
I.5.4. SRD utilisant un réseau de neurones	16
I.6. Conclusion	17
II. RÉSEAUX DE NEURONES	18

II.1. introduction.....	18
II.2. Fondement biologique	19
II.2.1. Le neurone	19
II.2.2. structure des neurones	19
II.2.3. Fonctionnement des neurones	21
II.2.4. le cortex visuel	21
II.3. le neurone formel	22
II.4. La structure des connexions	25
II.4.1. Les réseaux à couches	25
II.4.2. Les réseaux entièrement connectés	25
II.5. La règle de plasticité synaptique	26
II.6. Quelques types de réseaux de neurones	27
II.6.1. Réseaux à apprentissage supervisé.....	27
II.6.2. Réseaux à apprentissage non supervisé.....	29
II.7. Conclusion.....	32
III. CARTE AUTO-ORGANISATRICES :L'ALGORITHMES SOM.....	33
III.1. Introduction.....	33
III.2. La quantification vectorielle (QV).....	34
La tessellation de Veronoi.....	35
III.3. L'algorithme SOM	35
III.4. L'effet de la fonction de voisinage.....	40
III.5. Le taux d'apprentissage optimal.....	42
III.5.1. Le taux d'apprentissage optimisé pour la QV réursive.....	42
III.5.2. Le taux d'apprentissage optimal pour la SOM.....	43
III.5.3. Le taux d'apprentissage semi-empirique	44
III.6. initialisation de l'algorithme SOM	44

III.6.1. Initialisation au hasard	44
III.6.2. Utilisation des échantillons initiaux	45
III.6.3. Initialisation linéaire.....	45
III.7. Conseils pratiques pour la construction des SOM.....	46
III.7.1. La forme du réseau.....	46
III.7.2. Apprentissage avec un petit nombre d'échantillons.....	46
III.7.3. Mise en valeur des cas rares	46
III.7.4. Forcé la représentation à un endroit souhaité sur la carte	47
III.7.5. La qualité d'apprentissage	47
III.8. Quelques exemples d'utilisation des cartes de Kohonen dans la recherche documentaire	48
III.8.1. L'approche de Xia Lin	48
III.8.2. L'approche proposée par Poincot.....	49
III.9. Conclusion	50
IV. CONCEPTION DU PROTOTYPE SOMSEARCH (UN SYSTÈME DE RECHERCHE DOCUMENTAIRE RÉALISÉ À PARTIR DES SOM).....	51
IV.1. Introduction.....	51
IV.2. Architecture général du système « SOMSEARCH ».....	52
IV.3. Le sous système d'indexation	53
IV.3.1. La sélection des mots-clés.....	53
IV.3.2. la construction des descripteurs de documents.....	56
IV.4. le sous système de classification	57
IV.4.1. L'implémentation et le choix des paramètres.....	58
IV.4.2. Les cartes à deux niveaux.....	61
IV.4.3. Les cartes de densité	64
IV.4.4. Le format de sauvegarde des images	64
IV.5. Le sous système de recherche.....	65
IV.6. Conclusion	67

V. MISE EN ŒUVRE.....	68
V.1. Introduction	68
V.2. L'environnement de programmation	69
V.3. Description des modules réalisés	69
V.3.1. Le gestionnaire des fichiers	70
V.3.2. Le gestionnaire de la mémoire.....	71
V.3.3. Le gestionnaire des cartes.....	71
V.4. L'interface d'utilisation	74
V.5. L'interface de recherche	76
V.5.1. Le HTML.....	76
V.5.2. Une session de recherche.....	79
V.6. Conclusion	83
CONCLUSION	84
ANNEXES	86
Annexe A Le format des images GIF	86
Annexe B Le format bibliographique UNIMARC	90
Annexe C Simulateur des cartes de Kohonen.....	92
BIBLIOGRAPHIE	96