

**Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique**

**Institut National de formation en Informatique (I.N.I)
Oued-Smar Alger**

Mémoire de fin d'études

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

Option : Systèmes d'Informations

Thème :

**Conception et implémentation d'un outil
de DataMining Spatial**

« Application à l'analyse du risque lié au territoire »

Réalisé par :

Ilhem Zarga GHALAMALLAH

Lamine KHENTACHE

Proposé par :

Mme.F.ADMANE

Mr.K.ELARABA

Melle.C.ZEGHACHE

Promotion : 2003/2004

RESUME

Dans le domaine de l'informatique décisionnelle, l'évolution importante de ces dernières années a été le développement du DataMining. Initialement, appliquées à des données tabulaires simples, ces recherches s'étendent aujourd'hui à des données plus complexes et notamment aux données spatiales, d'où la naissance du DataMining spatial.

En effet, l'extraction de connaissances des données géographiques répond à un besoin réel des applications en géomatique pour tirer profit de la disponibilité croissante de données localisées et de la richesse potentielle en informations de ces données. C'est le cas de l'analyse du risque lié à un territoire comme l'étude de la diffusion d'épidémies ou de la distribution d'accidents de la circulation sur un réseau.

Mots clés : Systèmes d'Informations Géographiques (SIG), Extraction de Connaissances, Extraction de Connaissances Spatiales, Analyse spatiale, Relations Spatiales.

Table des matières

Introduction générale

1	Contexte général	1
2	Problématique	2
3	Objectifs initiaux de l'étude	3
4	Organisation du mémoire	4

Première partie : Etat de l'art

Chapitre I : Systèmes d'Information Géographiques et Analyse Spatiale

I.1.	Introduction	5
I.2.	Définition des données à référence Spatiale	5
I.3.	Définition d'un SIG	6
I.4.	Utilité d'un SIG	6
I.5.	Les composants d'un SIG	7
I.6.	Architecture des SIG	8
I.7.	Fonctionnalités d'un SIG	10
I.8.	Concepts liés à l'information géographique	12
I.8.1.	Définition	12
I.8.2.	Représentation de l'information géographique	13
I.8.3.	Mode d'acquisition des données géographiques	14
I.8.4.	Mode de représentation des données géographiques	14
I.9.	Base de données géographiques	16
I.10.	Organisation de l'information géographique	17
11.	Modélisation et interrogation de l'information géographique	18
I.11.1.	Modélisation de l'information géographique	18
I.11.2.	Indexation spatiale	21
I.11.3.	Interrogation de l'information géographique	22
I.12.	Analyse spatiale	22
I.12.1.	Définition	22
I.12.2.	Capacités analytiques des SIG	23

I.12.3. Limites des capacités analytiques des SIG	23
I.13. Domaines d'application et utilisateurs	24
I.13.1. Domaines d'applications	24
I.13.2. Utilisateurs	24
I.14. Conclusion	25

Chapitre II : Extraction de Connaissances à partir des Bases de Données

II.1. Introduction	26
II.2. Définition du DataMining	26
II.3. Le DataMining et la décision	27
II.4. Le DataMining et le Data Warehouse	29
II.5. Les principaux domaines d'applications	30
II.6. Le DataMining et le processus KDD	30
II.6.1. Etapes du processus KDD	31
II.7 Architecture typique d'un système de DataMining	36
II.8. Les tâches génériques du DataMining	37
II.9. Les techniques du DataMining	39
II.10. Présentation des techniques	40
II.10.1. Les associations	40
II.10.2. Les arbres de décisions	42
II.10.3. Les réseaux de neurones	44
II.11. Les techniques et méthodes associées à chaque tâche	48
II.12. Conclusion	48

***Chapitre III : Extraction de Connaissances à partir des Bases de Données
Géographiques***

III.1. Introduction	49
III.2. La fouille de données spatiales	50
III.2.1. Définition & intérêts	50
III.2.2. Principales tâches du DMS	50
III.3. Spécificités de la fouille de données spatiales	51

III.3.1. Matrice et graphe de voisinage	51
III.3.2. Matrice de Contiguïté	52
III.3.3. Les relations de voisinages	52
III.4. Les jointures spatiales	54
III.5. Index de Jointures Spatiales	55
III.5.1. Définition.....	55
III.5.2. Structure de l'index	55
III.5.3. Construction de l'index	56
III.5.4. Utilisation de l'Index	57
III.6. Panorama des méthodes de fouille de donnée spatiale	58
III.6.1. Phase exploratoire	59
III.6.2. Phase décisionnelles : Analyse Explicative.....	61
III.7. Conclusion	62

Deuxième partie : Approche proposée

Chapitre IV :

Approche proposée :

vers une architecture combinée des SIG et du DataMining

IV.1. Introduction	64
IV.2. Cadre conceptuel de l'approche proposée	64
VI.3. Conception du système projeté	70
IV.3.1. Méthodologie de conception	71
IV.3.2. Architecture du Système	71
IV.4. Analyse du système	75
IV.4.1. Identification des classes	76
IV.4.2. Diagramme des classes	78
IV.4.3. Description des classes	80
IV.4.4. Description des relations entre les classes	83
VI.5. Conclusion	85

Chapitre V : Etude de cas « Analyse des risques routiers »

V.1. Introduction	86
V.2. Positionnement & Objectifs de l'étude	87
V.3. Processus d'extraction	88
V.3.1. Intervenants internes et externes	88
V.3.2. Analyse du problème	88
V.3.2. Préparation des données	89
V.3.3. Consultation des données	91
V.3.4. Création d'un index de jointure	92
V.3.5. Choix de l'algorithme de calcul	94
V.3.6. Exécution de l'algorithme de calcul	96
V.3.7. Visualisation des résultats	99
V.3.8. Evaluation des résultats	100
V.3.9. Validation des résultats	100
V.3.10. Mémorisation des résultats.....	100
V.4. Conclusion.....	101

Chapitre VI: Illustration du prototype développé

VI.1. Introduction	102
VI.2. Environnement de développement	102
VI.3. Présentation	103
VI.3.1. Problématique	103
VI.3.2. Contraintes	103
VI.3.3. Données disponibles	104
VI.3.4. Résolution	105
VI.4. Conclusion	121

Conclusion générale et perspectives

1 Conclusion générale	122
2 Apports de l'approche proposée	123
3 Perspectives	123

Références bibliographiques

ANNEXES :

ANNEXE A : Le Système d'Information Géographique « MapInfo »

ANNEXE B : Conception Orienté Objet « UML »

ANNEXE C : La Terminologie OLE