

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MOHAMED BOUDIAF - M'SILA

FACULTE: Mathématique et Informatique

DEPARTEMENT : Informatique

N° :



DOMAINE : Mathématique et Informatique

FILIERE : Informatique

OPTION: Systèmes d'Informations Avancés

Mémoire présenté pour l'obtention
Du diplôme de Master Académique

Par : BELKORCHIA MOUAOUIA

Intitulé

**PROPOSITION D'UN OUTIL DATAMINING POUR
LES BASES DE DONNEES SPATIALES MULTIPLES**

**EXTRACTION DES CIRCONSTANCES COMMUNES
DES ACCIDENTS ROUTIERS**

Soutenu devant le jury composé de :

Pr BOUDREH BRAHIM

Université de M'sila

Président

Dr MEHENNI TAHAR

Université de M'sila

Rapporteur

Mr KADRI SAID

Université de M'sila

Examineur

Année Universitaire : 2016 /2017

Dédicace

Avant tout je remercie le Seigneur Tout-Puissant de m'avoir accordé « volonté » et « patience », pour pouvoir accomplir ce modeste travail.

Je dédie le présent projet à :

** Ma chère mère qui m'a indiqué le bon chemin à entreprendre.*

** Ma petite famille qui m'a encouragé et soutenu tout au long de mon parcours quotidien.*

** Mon encadreur « Mehenni Tahar ».*

** Tous mes amis, surtout « Benaïssa Lakhdar ,Saadi Khalil, Deffous Ahmed» qui m'ont aidé à reprendre mes études.*

** Tous les enseignants du département informatique, surtout Pr « Boudrah Brahim » qui m'a aidé à reprendre mes études.*

ainsi que tous les responsables du département Informatique.

Et enfin,

** A tous mes camarades de promotion 2016/2017.*

Remerciements

En premier lieu je remercie « ALLAH » pour m'avoir guidé et donné la force pour finaliser ce Mémoire.

Je tiens à présenter mes remerciements les plus distingués à ma famille, mes amis qui m'ont aidé à poursuivre mes études et atteindre l'objectif souhaité.

Je remercie aussi mes professeurs pour leurs soutiens et les efforts déployés à mon profit, ainsi que leur contribution pour le bon déroulement de ma formation notamment mon encadreur monsieur MEHENNI Tahar qui m'a réservé une partie de son temps pour réaliser mon projet de Fin de « Cycle Master SIA » conformément aux modalités prévues à cet effet.

Je remercie les membres du Jury de ce mémoire.

Je remercie aussi les responsables du Département Informatique.

INDEX DES FIGURES

Figure 1.1 : Différentes couches spatiales	1
Figure 1.2 : Mode Raster et Mode Vecteur	3
Figure 2.1 : Le cheminement théorique d'une jointure spatiale.....	13
Figure 2.2: Architecture d'un SIG exemple est ARC-INFO.....	15
Figure 2.3: Architecture hybride	17
Figure 2.4 : Architecture hybride	18
Figure 4.1 : Carte du réseau routier d'Algérie	33
Figure 4.2 : Image via Google Earth région d'Alger	34
Figure 4.3 : Image via Google Earth région d'Alger avec une route découpée	34
Figure 4.4 : Découpage des routes par Wilaya	35
Figure 4.5 : Carte de M'Sila et différentes routes découpées	35
Figure 4.6 : Image via Google Earth région M'Sila	36
Figure 4.7 : Image via Google Earth région M'Sila.....	37
Figure 4.8: Image via Google Earth région M'Sila importée à ArcGis	37
Figure 4.9 : MAP région de M'sila sur ArcGis avec différents reliefs	38
Figure 4.10 : Différentes bases de données sur Arc Catalogue	39
Figure 4.11: Routes « intersection sur ArcGis »	41
Figure 4.12 : Routes « intersection sur ArcGis »	41
Figure 4.13 : Nouvelle table de la requête intersection	42
Figure 4.14 : Routes « voisinage sur ArcGis »	42
Figure 4.15 : La jointure et relate des tables	43
Figure 4.16 : La jointure et relate des tables	43
Figure 4.17 : Relate permanente des tables	44
Figure 4.18 : L'affichage des différentes tables jointées	45
Figure 4.19 : L'affichage des différentes tables jointées avec la table Accident sur Access... ..	49
Figure 4.20 : List view de l'ensemble des accidents avec attributs.....	50
Figure 4.21 : Forme principale de l'application	50
Figure 4.22 : Forme d'identification	51
Figure 4.23 : Forme ajouter accident	52
Figure 4.24 : Forme de toutes les routes partielles avec leurs reliefs	52
Figure 4.25 : Forme charger tous les accidents de la Base pour le traitement	53
Figure 4.26 : Affichage de tous les accidents de la Base pour le traitement	54
Figure 4.27 : Affichage de toutes les règles d'Association après le traitement.....	54

INDEX DES TABLES

Table1.1 :Comparaison entre mode vecteur et raster	4
Table 4.1 :Base de données routes partielles.....	39
Table 4.2 : Base de données Rivière.....	39
Table4.3 :Base de données spatiales et attributaires.....	41
Table4.4 :Nouvelle base de données créée « intersection »	42
Table4.5 :Nouvelle base de données créée « voisinage »	42
Table4.6 :La table spatiale route partielle intersection forêt.....	45
Table4.7 : La table attributaire accident.....	46
Table 4.8 : La table attributaire et spatiale accident.....	48

TABLE DES MATIERES

Dédicace	I
Remerciements :	II
Introduction générale	VII
CHAPITRE 1 - LES BASES DE DONNEES SPATIALES ET LES SYSTEMES D'INFORMATIONS GEOGRAPHIQUES.....	1
1- Introduction	1
2- L'information géographique	2
2.1- Composants de l'information spatiale	2
2.1.1- composante spatiale	2
2.1.2- composante attributaire	2
2.1.3- composante temporelle	2
2.2- Collection des informations spatiales	2
3- L'analyse des données spatiales	2
4- Mode de représentation des données spatiales	3
4.1- Mode Vecteur.....	3
4.2- Mode Raster.....	3
5- Comparaison entre mode raster et vectoriel	4
6- Les systèmes d'informations spatiales SIG.....	4
6.1- La composition d'un SIG	4
6.1.1- Matériel	4
6.1.1.1 - Les données	4
6.1.1.2 - Les méthodes	4
6.1.1.3 - Les logiciels	4
6.1.1.4 - Les réseaux	5
6.1.1.5 - Le Hardware	5
6.1.2- Humaine	5
6.2- Apport du SIG	5
6.3- Les domaines d'utilisation du SIG.....	5
6.4- Les technologies SIG	5
6.4.1- Les outils côté client	6
6.4.1.1- MapInfo	6
6.4.1.2- AutoCAD MAP 3D	6

6.4.1.3- Quantum GIS	6
6.4.2- Les outils côté serveur	7
6.4.2.1- Map Server	7
6.4.2.2- Arc GIS-Server	7
6.4.2.3- Géo Server	8

CHAPITRE 2 : BASES DE DONNEES SPATIALES MULTIPLES ET LEURS

SYSTEMES DE GESTION.....	9
1- Introduction	9
2- Les Bases de données spatiales multiples	9
3- Gestion des données spatiales	10
4- L'extension du modèle relationnel au modèle spatial.....	11
5- La jointure spatiale.....	13
6- Architecture des SIG	14
7- Architecture hybride	16
8- Outils de développement.....	18
9- Relations spatiales.....	19
10- Méta-informations géographiques.....	20

CHAPITRE 3 : LE DATAMINING DANS LES BASES DE DONNEES

SPATIALES MULTIPLES.....	21
1- Introduction	21
2- Fouilles de données « Datamining »	21
2.1- Définition de la fouille de données ou datamining	22
2.2- Les méthodes de datamining	22
2.2.1- Les méthodes supervisées/prédictives	22
2.2.1.1- Classification	23
2.2.2- Les méthodes non supervisées/prédictives.....	24
2.2.2.1- Règles d'association	24
2.2.2.2- Clustering	25
3- Le Datamining spatial.....	25
3.1- Le Clustering spatial	26
3.2- La classification spatiale.....	26
3.3- Les règles d'association spatiales.....	27
4- Le datamining dans les bases de données spatiales multiples	28
4.1- Problématique	28
4.2- Solutions	29

5- Conclusion.....	31
CHAPITRE 4 : UN ALGORITHME DE DATAMINING POUR LES BASES DE DONNEES SPATIALES MULTIPLES.....	32
1- Introduction	33
2- Problématique.....	34
3- Algorithme de datamining « règles d'association» pour les accidents routiers.....	36
4- L'exportation des images vers Arc Gis	37
5- La création de la carte Map avec Arc Catalogue	37
6- Approches d'utilisation des Bases de données spatiales multiples	40
6.1- Regrouper toutes les tables	40
6.2- Travailler sur les liens.....	40
7- La construction des tables spatiales	40
8- La construction des tables attributaires	46
9- La jointure spatiale des tables	46
10- L'exportation des bases spatiales vers C Sharp pour extraire les règles	48
10.bis-Présentation de l'application	50
11- Conclusion	55
CONCLUSION GENERALE	56
BIBLIOGRAPHIE	57

INTRODUCTION GENERALE

L'espace d'un monde réel peut se présenter par des cartes spatiales (géographiques), on peut associer à ces dernières un ensemble des bases de données spatiales collaborées entre elles pour donner des réponses bien précises sur nos différents besoins comme par exemple les voisinages, les limites des villes, les mesures des superficies relatives aux différentes villes etc...

L'étude de ces phénomènes avec l'indépendance de ces bases nous crée quelques difficultés telles que l'incohérence entre les bases qui nous donne des résultats différents, et comme les bases représentent les différentes couches de l'espace du monde réel alors on ne peut pas faire une analyse constituée de plusieurs couches.

Pendant les 20 dernières années les chercheurs ont donné une grande importance à ce sujet ,plusieurs approches ont été proposées pour résoudre les problèmes d'incohérence, la complexité, la capacité énorme des données spatiales reçus, la nature des liens des tables qui se diffèrent intégralement au liens des tables relationnelles simples.

Pour atteindre cet objectif on a fait la fusion des bases de données, le résultat est une base de données commune large et riche rassemblant les différentes bases et pouvant répondre au différents besoins.C'est dans ce cadre qu'on a pensé à réaliser une base de données à référence spatiale contenant les différentes routes d'une région avec leurs reliefs en lui ajoutant une base de données contient les statistiques des accidents.

Dans notre mémoire on va procéder à extraire les circonstances communes des accidents, cette connaissance nous permet à séparer ou isoler ces circonstances pour diminuer le nombre des accidents.Pour cela on a choisi les règles d'association comme méthode de datamining pour l'appliquer sur notre base de données et faire sortir les règles d'association qui représentent en quelque sorte les circonstances communes de l'accident. **Le premier chapitre** porte sur la définition de l'information spatiale, les systèmes d'informations spatiales. **Le deuxième chapitre** porte sur la définition des bases de données spatiales multiples, et les systèmes de gestion des bases de données spatiales multiples. **Le troisième chapitre** porte sur le datamining sur les bases de données spatiales multiples avec deux approches1-assembler toutes les données dans une seule table2-laisser les bases et travailler sur les liens selon leurs utilités. **Le quatrième chapitre** porte sur un algorithme datamining Règles d'association.

Résumé

Le projet que nous avons étudié a porté sur la réalisation d'un outil datamining pour les Bases de données spatiales multiples qui va nous permettre de prendre des décisions aidant à diminuer le nombre d'accident sur le territoire national, en outre la réservation des puissances pour les lieux les plus touchés par les accidents. Pour atteindre cet objectif, nous avons entrepris la réalisation d'une application regroupant les différentes bases de données des secteurs concernés à savoir la protection civile, les services de sécurité fournissant les différentes statistiques sur les accidents qui ont eu lieu dans chaque région, ajoutant les bases de données spatiales sur les lieux concernés par ces accidents. Cette collaboration nous permet de regrouper les données et les traiter en utilisant le logiciel Arc Gis permettant d'analyser les données spatiales et les transférer au logiciel C Sharp pour extraire les règles d'association.

Mots clés :

Logiciel ARCGIS, Base de données spatiales multiples, localisation spatiale, Google Maps.

ملخص

المشروع الذي قمنا بدراسته هو انشاء تطبيق للتنقيب عن المعلومات في قواعد البيانات الجغرافية المتعددة الذي من شأنه السماح لنا باتخاذ القرارات المساعدة في تقليل حوادث المرور في البلاد، إلى جانب حجز القبول للمناطق الأكثر تضرراً من هذه الأخيرة. ولتحقيق هذا الهدف، بدأنا تنفيذ تطبيق يضم مختلف قواعد بيانات الحماية المدنية، الأمن التي توفر لنا مختلف الاحصائيات المتعلقة بالحوادث التي وقعت في منطقة ما بالإضافة الى قواعد البيانات الجغرافية مترابطة بينها موفرة لنا قاعدة بيانات كبيرة تسمح لنا وتحويلها الى برنامج بتحليلها في برنامج ArcGIS سي شارب للقيام بعملية التنقيب واستخراج القواعد.

كلمات البحث:

قاعدة بيانات جغرافية متعددة، ArcGIS برنامج نظام المواقع المكانية، خرائط جوجل.

Abstract

The project we have studied has focused on the realization of a data mining tool for multiple spatial databases which will help us to make decisions concerning the knowledge of the most current path between the different centers of civil protection and places Was held of an accident and also the reservation of the powers for the places more affected by the accidents. To achieve this objective, we have undertaken the implementation of an application regrouping the different databases of the sectors concerned, namely civil protection, transport management, etc. ... This collaboration allows us to consolidate the data in a single table and proceed to use the GIS to record information on the detected damage of an area, to update the global database and data mining to extract knowledge On the damage and to plan the necessary devices for the future. We use ArcGIS software to process and manipulate the spatial data in question.

Keywords :

ARCGIS software, multiple spatial database, spatial localization, Google Maps.

CHAPITRE 1

BASES DE DONNEES SPATIALES(GEOGRAPHIQUES) ET SYSTEMES D'INFORMATIONS GEOGRAPHIQUES

1- Introduction :

Le monde géographique est relativement complexe pour étudier profondément ses entités telles que les mers, les forêts, les montagnes, les routes, les bâtiments etc. ...

Il faut simplifier toutes ces entités dans une carte géographique prise soit par satellite, par avion ou par d'autres moyens... Et l'adapter de telle sorte à contenir que les objets qui nous intéressent et qu'on peut les représenter par des bases de données géographiques, et omettre les objets qui ne nous intéressent plus.

L'accès à l'information spatiale ou géographique nous oblige de consulter la carte géographique qui a une échelle, cette dernière comporte des objets concrets comme les bâtiments, les forêts et des objets abstraits comme les limites administratives mais avec l'absence des détails (les forêts sont présentées sans la présence des arbres), pour cela on est obligé d'accompagner la carte géographique par une base de données spatiale et l'utilisation de système d'information géographique pour pouvoir collecter ,sauvegarder, faire des mises à jour indépendamment de l'échelle des symboles.

Ces bases se composent d'un ensemble de tables relationnelles qui présentent l'ensemble des couches superposables du monde réel comme le montre la figure 1.1

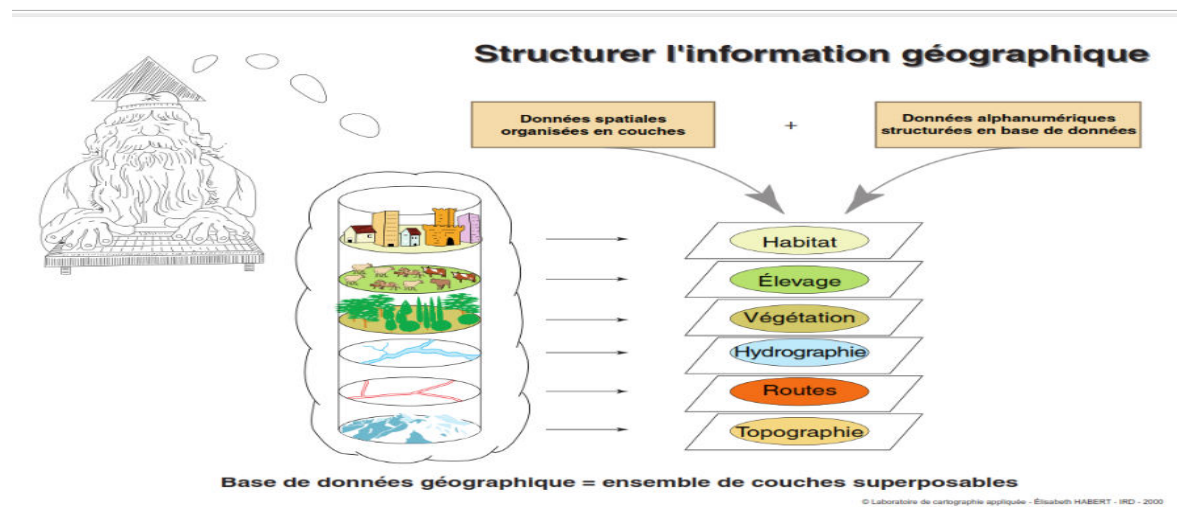


Figure1.1 Différentes couches spatiales

2- L'information géographique :

Désigne toute information sur des objets localisés à la surface de la Terre.

2.1 Composants de l'information spatiale :

L'information géographique a trois composantes :

2.1.1 Une composante spatiale :

- a) description de la forme de l'objet spatiale.
- b) sa localisation dans un référentiel cartographique.

2.1.2 Une composante attributaire :

Caractéristiques décrivant l'objet (description géométriques, Caractéristiques thématiques).

2.1.3 Une composante temporelle :

Parfois on est dans le besoin de connaître le temps d'apparition des phénomènes et le suivie de leurs changements.

2-2 Collection des informations spatiales : La phase de collection des données est la phase la plus essentielle, durant et couteuse dans le processus des systèmes d'information spatiales, et elle est composée de six phases :

- la planification.
- la préparation.
- la capture des informations.
- la transformation des informations.
- l'amélioration des informations.
- l'évaluation des informations.

Citons quelques outils de saisis ou collection d'informations spatiales par le moyen :

- des satellites
- de la photographie aérienne.
- du GPS.
- du Web.

3-L'analyse des données spatiales

Définition: L'analyse des données spatiales c'est la transformation de ces mêmes données en informations utile de signification pour l'exploiter dans les études scientifiques et la prise de décisions.

On distingue deux types d'analyse :

- a-L'analyse spatiale qui traite les informations spatiales.
- b- L'analyse statistique qui traite les informations attributaires.

Comme il y a énormément d'informations qu'il faut les organiser, ce besoin oblige les informaticiens à créer des systèmes de gestion qui traitent les informations spatiales qu'on va aborder ultérieurement.

4-Mode de représentation des données spatiales

On peut distinguer deux modes de représentation de l'espace le mode vecteur et le mode raster.

4.1-Mode Vecteur :

Dans ce mode on représente l'espace par des entités des points (X, Y), des arcs liant les points (X, Y) et des polygones constitués des arcs qui représentent les frontières d'un espace. Les données spatiales en mode vecteur peuvent prendre plusieurs formats.

Exemples :- Shape file dans ARCGIS.

-DWG dans AUTOCAD.

Si on prend ARCGIS qu'on va le voir en détail les données sont stockées dans le fichier :- SHP qui contient les vecteurs (polygones, lignes, points)

-DBF qui contient la base de données correspondant aux attributs.

-(SHX) un fichier index de la géométrie des objets.

4.2- Mode Raster :

Dans le mode RASTER, l'image est représentée par une grille contenant des pixels dont leurs tailles égales mesurées par l'unité de distance, définissent la résolution de cette image. La figure 1.2 illustre la représentation via les deux modes raster et vecteur.

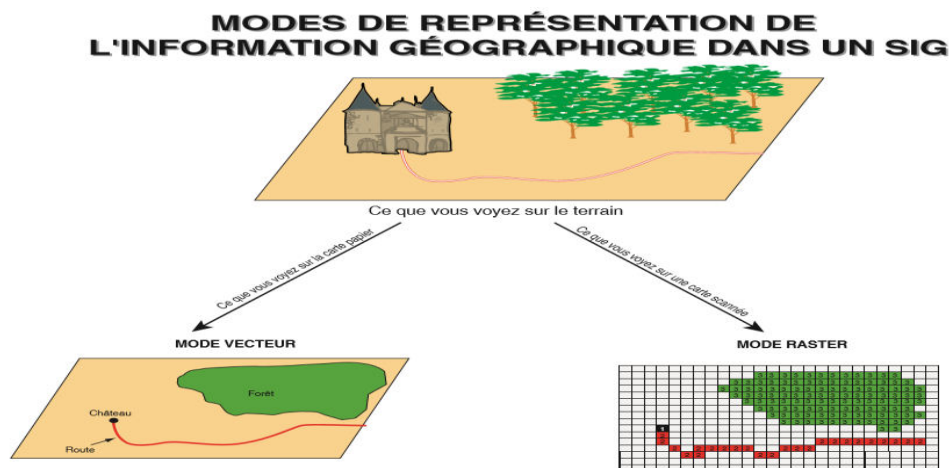


Figure1.2: Mode Raster et Mode Vecteur

5-Comparaison entre mode raster et vectoriel [11]

VECTORIEL	RASTER
<ul style="list-style-type: none"> -Petit espace de stockage -La structure de donnees est complexe -N'utilise pas le volume de pixel dans la precision -la collection demande un grand temps -un grand effort dans l'analyse spatiale -la substitution du monde reel par des symboles -une grande precision spatiale 	<ul style="list-style-type: none"> -Grand espace de stockage - La structure de donnees est simple -utilise le volume de pixel dans la precision - la collection ne demande pas un grand temps. -un petit effort dans l'analyse spatiale -dans la majorite du temps les images representent le monde reel -une petite precision spatiale.

Table 1.1 Comparaison entre mode vecteur et raster

6-Les systèmes d'informations spatiales SIG

« Un système informatique permettant, à partir de diverses sources, de rassembler et d'organiser, de gérer, d'analyser et de combiner, d'élaborer et de présenter des informations localisées géographiquement, contribuant notamment à la gestion de l'espace ». [15]

6.1- La composition d'un SIG :

Un système d'informations spatiales est composé de deux types de ressources ;

6.1.1-Matériel : à savoir

6.1.1.1- Les données : les données collectées.

6.1.1.2- Les méthodes : le processus suivie pour l'organisation de l'étude.

6.1.1.3- Les logiciels : on compte plusieurs logiciels de SIG à savoir les logiciels d'utilisation personnelle, ainsi autre plus compliquées destinées au grandes sociétés et qui sont chargées sur des serveurs.

6.1.1.4- Les réseaux : la majorité des SIG maintenant exploitent le réseau, le web pour l'échange des données entre les utilisateurs.

6.1.1.5- Le Hardware : représente tout matériel exploité par le SIG soit dans la saisie, l'analyse ou l'exploration des données.

6.1.2- Humaine : la composante la plus essentielle du SIG qui regroupe les différents utilisateurs humains du système.

6.2-Apport d'un SIG

On cite quelques apports d'un SIG :

- le stockage des informations spatiales.
- la gestion du partage de l'information.
- la gestion de multiplicité d'informations attributaires sur des objets.
- comprendre les phénomènes.
- localiser dans l'espace et dans le temps.
- réagir rapidement après les évènements ayant une influence sur le territoire.
- prévoir les risques.

6.3-Les domaines d'utilisation du SIG

On compte plusieurs domaines d'application :

- la biologie
- les forêts
- l'urbanisme
- l'hydraulique
- le transport
- la géologie
- la télécommunication
- la protection civile

6.4- Les technologies SIG [30]

Les logiciels SIG sont classés en deux catégories : les logiciels SIG client et les logiciels SIG Serveur.

Les logiciels SIG client s'exécutent au niveau d'une machine cliente. Ces logiciels permettent ainsi de manipuler les données, de les charger sur les serveurs ou encore d'interroger les serveurs.

Les Logiciels SIG serveur sont surtout les serveurs de données, ils permettent de rendre une carte après l'interrogation de la base du serveur par le client.

Il existe plusieurs outils SIG, nous ne présenterons que quelques-uns d'entre eux, afin d'avoir une première vue du marché des logiciels SIG.

6.4.1-Les outils côté client

6.4.1.1-MapInfo

MapInfo Professional est un Système d'information géographique (SIG) à l'origine Bureautique créé dans les années 1980 aux Etats-Unis. C'est un logiciel qui permet de réaliser des cartes en format numérique. MapInfo est conçu autour d'un moteur d'édition de cartes qui permet la superposition de couches numériques. Il permet de représenter à l'aide d'un système de couches des informations géo-localisées : points, polygones, image raster ... Il incorpore un grand nombre de formats de données, de fonctions cartographiques et de gestion de données. MapInfo est ouvert vers le Web et les globes virtuels ; il permet de publier sur le web des cartes réalisées sur un PC, de faire de la cartographie interactive, d'incorporer des informations des globes virtuels.

6.4.1.2-AutoCAD MAP 3D

AutoCAD Map 3D est une excellente plateforme SIG pour la création et l'édition de données géo-spatiales. Il relie la conception assistée par ordinateur et les systèmes d'informations géographiques (SIG). AutoCAD Map 3D offre un accès direct aux données, quel que soit leur mode de stockage, et permet d'utiliser les outils AutoCAD pour gérer une grande variété d'informations géo spatiales. La technologie d'accès aux données FDO en code libre permet à Map3D d'accéder de manière native aux données spatiales stockées dans des bases de données relationnelles, des fichiers et des services Web. Ainsi, il peut gérer facilement un grand nombre de jeux de données géo spatiales tout en rationalisant les procédures dans leur intégralité. La fluidité de l'intégration de la technologie Autodesk MapGuide fait d'AutoCAD Map 3D le moyen le plus rapide pour publier des données sur le Web ou sur un réseau intranet [10].

6.4.1.3- Quantum GIS

Quantum GIS, également appelé plus simplement QGis, est un système d'information géographique (SIG) libre multi plateforme publié sous licence GPL. Il gère les formats d'images matricielles (raster) et vectorielles, ainsi que les bases de données. En outre il gère l'extensionspatiale de PostgreSQL, PostGISet prend en charge les archives vectorielles Shapefile, les couvertures d'autres logiciels tel que GRASS GIS. Il utilise un nombre important de types d'archives matricielles. L'un de ses avantages majeurs est la possibilité de l'utiliser comme interface graphique du SIG GRASS, utilisant toute la

puissance d'analyse de ce logiciel en un environnement de travail plus convivial. Les fonctionnalités GRASS sont optionnelles, et passent par un module d'extension [11].

6.4.2-Les outils côté serveur

6.4.2.1-MapServer

MapServer est un serveur cartographique qui se place dans la partie inactive d'un serveur WEB. Quand une requête est envoyée à MapServer, il utilise les informations passées dans l'URL de la requête et le fichier de configuration fourni (MapFile) pour créer une image de la carte demandée.

Cette image peut être affichée dans un navigateur. La requête peut aussi fournir une légende, une barre d'échelle, une carte de référence et des valeurs passées comme variables CGI. Il est aussi disponible sous forme d'une API dans plusieurs langages de programmation. Le programmeur a alors accès à des fonctions lui permettant entre autres de manipuler la carte, d'effectuer des requêtes spatiales, puis d'ajouter des objets géo référencés [30].

6.4.2.2-ArcGIS-Server

ArcGis-Server est un serveur cartographique commercial développé par ESRI (Environmental System Research Institute). Il ouvre l'accès aux informations géographiques. De nombreuses Organisations l'utilisent pour diffuser sur le Web des cartes et des fonctionnalités SIG, de manière à améliorer la productivité de l'organisation, à communiquer des éléments vitaux et à engager les décideurs.

ArcGIS-server permet :

- le contrôle des données spatiales et des sources d'images, grâce à un système centralisé de gestion des données.
- l'amélioration de la prise de décisions et la productivité à l'aide de services Cartographiques Web et des applications qui peuvent être mises à la disposition des utilisateurs.
- La prise du meilleur parti d'une architecture informatique actuelle en intégrant un serveur SIG et des données spatiales avec d'autres systèmes professionnels [30].

6.4.2.3-GeoServer

GeoServer est un serveur informatique open source écrit en Java qui permet aux utilisateurs de partager et de modifier des données géographiques. Conçu pour l'interopérabilité, il publie les données de toutes les sources principales de données spatiales qui utilisent des normes ouvertes.

GeoServer a évolué pour devenir une méthode simple de connexion d'informations existantes à des globes virtuels tels que Google Earth et NASA, World Wind, ainsi que pour les cartes à base de services Web tels qu'OpenLayers, Google Maps et Bing Maps. Geoserver fonctionne en tant qu'implémentation de référence pour la mise en œuvre du standard du Web Feature Service de l'Open Geospatial Consortium ; il implémente aussi le Web Map Service.

CHAPITRE 2

BASES DE DONNEES SPATIALES MULTIPLES ET LEURS SYSTEMES DE GESTION

1-Introduction :

L'information géographique est relative à un objet ou un phénomène du monde terrestre est décrite plus ou moins complètement par :

- la nature et les attributs de l'objet (sémantique)
- la localisation et la forme de l'objet (géométrie)
- Les relations de voisinage entre objets (topologie)

Ces objets qui représentent les différentes couches du monde réelle telles que les routes les villes les forêts les fleuves etc. vont former une base de données spatiales d'une région définie, prenant un objet de cette région terrestre qui possède plusieurs tables, il y'a des tables qui sont au début complète comme par exemple la base des forêts existante dans cette région les fleuves etc., mais d'autres tables sont démunis de cette spécification, autrement dit pour avoir ces tables intégralement il faut regrouper plusieurs tables par exemple le nombre d'accidents qui ont eu lieu dans cette région, il faut regrouper les statistiques des différents services de sécurité dont chaque service est indépendant des autres services comme le service de la gendarmerie, la police, la protection civile, la santé etc., donc avec la base d'un seul service l'analyse reste incomplète, pour dépasser cette incomplétude et palier à ce manque, il nous faut une collaboration entre les différents service pour avoir une vision générale sur le phénomène étudié.

Cela nous conduit à étendre notre base de données à plusieurs bases reliées entre elle et par conséquent nous sommes pratiquement dans une nouvelle situation nommée les bases de données spatiales multiples.

2- Les bases de données spatiales multiples

Comme les bases de données spatiales est un ensemble organisé d'objets géographiques on définit les bases de données spatiales multiples comme un ensemble des bases de données spatiales distribuées sur plusieurs sites, communiquant entre elle pour donner une large base pouvant répondre suffisamment aux besoins.

« Si la masse d'information statistique territoriale aujourd'hui disponible à toutes les échelles géographiques, elle permet d'envisager le développement de scénarios prospectifs

sur le long terme pour l'aménagement du territoire, il n'en demeure pas moins que la manipulation de cette information spatio-temporelle pose de nombreux problèmes. Il apparaît que les supports, les définitions, les modalités de classification, et le niveau de fiabilité de ces données ne sont pas homogènes, ni dans l'espace, ni dans le temps. De ce fait, les données sont difficilement comparables. Cette hétérogénéité est au cœur de la problématique de notre thèse». [24]

« Les données des bases de données et d'images de l'environnement générique, sont extraites des différentes applications, ainsi ils sont non seulement de types divers, mais également de sources et de formats différents. Ce sont des données complexes et hétérogènes. Dans notre plate-forme nous avons spécifié deux types de données ; des données factuelles et des données images satellitaires.

Nous aurons ainsi deux sortes de bases de données :

- Des bases de données factuelles constituées de données provenant de nombreux organismes, et sont de natures diverses : fichiers, documents, bases de données, etc. Les documents sont de formats différents ; RTF, DOC, PDF, HTML, PPT, XLS, XML, etc. et peuvent contenir des données telles que : les données vérité du terrain, les divisions administratives, le Plan d'Occupation des Sols (POS), les données sociodémographiques, etc.
- Des bases d'images (HRS et BRS) qui sont constituées d'un ensemble d'images provenant de plusieurs types de satellites : images à BRS (données NOAA/AVHRR, VEGETATION), images à HRS (données SPOT, LANDSAT, IRS), images de classification valides sur le terrain, images HRS ou BRS calibrées, images MNT à 3D (Modèle numérique du terrain), données cartographiques, etc. Ces bases de données doivent être structurées de façon à supporter une variété de données hétérogènes. » [33]

3- Gestion des données spatiales

La difficulté de la gestion des données géographiques provient de leurs doubles natures, géométriques et thématiques. Si l'on imagine aisément l'enregistrement d'attributs alphanumériques quelconques, caractérisant la thématique, dans une simple table (relation d'une base de données relationnelle, ou à la rigueur une simple feuille de calcul d'un tableur), il n'en va pas de même des coordonnées spatiales et des relations géométriques entre les entités. Les solutions envisagées pour résoudre cette difficulté ont évolué au

cours du temps, suivant en cela les générations successives des modèles logiques des bases de données et les capacités croissantes des ressources informatiques.

Les changements les plus importants étant apparus durant la dernière décennie, des solutions relevant de générations distinctes continuent à coexister aujourd'hui. Les logiciels informatisant ces solutions sont intitulés, un peu abusivement (cf. Infra), « systèmes d'information géographique »[7][22].

On peut définir le SGBDS comme étant :

1- un SGBD

2- il offre un type de donnée spatiale dans son modèle de données et son langage de requêtes

3- il implémente ce type et ses opérateurs, fournissant au moins l'indexation spatiale et des algorithmes efficace pour la jointure spatiale.

4- L'extension du modèle relationnel au modèle spatial

Les principes de l'extension du modèle :

Le modèle relationnel ne concerne que les types d'attributs simples. Ces attributs prennent leurs valeurs dans un ensemble fini ou ordonné de dimension une. Les opérations classiques de l'algèbre relationnelle sont limitées aux attributs de type simple, et sont liées à la structure d'ordre canonique des ensembles de définition.

La localisation spatiale ne rentre pas dans le cadre des attributs simples, puisque cet attribut prend ses valeurs dans un espace de dimension 2 ou 3. Il faut donc étendre le modèle relationnel classique pour qu'il puisse gérer des objets localisés en en conservant la structure multidimensionnelle [28] [20].

Il n'existe pas de structure d'ordre canonique dans un espace de dimension supérieure à 1 qui puisse être utilisée dans la résolution de problèmes spatiaux liés à la proximité ou la distance : la structure d'ordre canonique de la dimension 1 doit être remplacée par une structure topologique ou une structure métrique. Les opérations de l'algèbre relationnelle liées à l'attribut de localisation ne sont plus définies dans un espace à une dimension, mais à deux ou trois.

Les relations entre objets ne s'expriment donc plus à partir d'une relation d'ordre, mais à partir d'une topologie ou d'une distance. Ainsi, un domaine de restriction ne sera plus une valeur ou un intervalle, mais un domaine de l'espace métrique. Un critère de jointure géographique ne sera plus une relation entre les valeurs des objets eux-mêmes, mais une

relation ensembliste ou une relation entre les valeurs des distances entre les objets [28], [20].

D'une manière classique, la localisation des objets géographiques s'exprime soit par des éléments de l'espace (pour les objets dont la localisation peut être ramenée à un point), soit par des sous-ensembles de l'espace (lignes ou zones). L'extension du modèle relationnel ne concerne donc pas seulement le domaine de définition des attributs (l'espace de dimension 2 ou 3) : il introduit également des types dans la définition des relations, se rapprochant ainsi du modèle objet. Les contraintes d'intégrité ne seront pas les mêmes suivant le type des objets.

La schématisation géométrique de la localisation des objets géographiques en zone, ligne, point, pixel établit une classification naturelle sur laquelle repose à la fois l'extension du modèle relationnel et la définition de classes d'objets géographiques pour le modèle objet. Cette classification est très discriminante, elle sera donc présente très rapidement comme membre des objets géographiques. Pour l'extension du relationnel, elle intervient au niveau de la définition des relations, puisqu'elle dépend de la mise en œuvre des opérations de l'algèbre relationnelle.

Pour le modèle objet, peu de classes d'objets non localisés peuvent prétendre encapsuler des classes d'objets géographiques localisés.

Toutes les considérations ultérieures seront donc basées sur la classification suivante, que nous avons déjà vue et qui permet de schématiser la géométrie des objets géographiques et de définir des collections d'objet de même type :

- les objets point, dont la localisation est un point de l'espace,
- les objets ligne, dont la localisation est un sous-ensemble de dimension 1 de l'espace, en général schématisé par un ou par plusieurs arcs (ligne brisée entre des points),
- les objets zone, dont la localisation est un sous-ensemble de dimension 2 de l'espace, en général défini par un ensemble d'arc (schématisant le contour de la zone),
- les objets pixel, dont la localisation est un sous-ensemble de dimension 2 de l'espace définie par un point (tous les pixels ont une géométrie identique, maille carrée, hexagonale, etc.).

Un arc n'est pas un objet géographique, mais seulement un objet géométrique. Les contraintes d'intégrité géométriques seront exprimées sur les points et les arcs, c'est-à-dire sur les objets géométriques qui permettent de définir la localisation des objets géographiques. [28].

5- La jointure spatiale

De toutes les opérations algébriques, la jointure est certes la plus importante en pratique, car, comme nous l'avons souligné, c'est elle qui permet la mise en relation de deux tuples par le biais d'un attribut commun, en créant ainsi un nouvel objet ayant les caractéristiques des deux objets répondant à la qualification de jointure, et correspondant dans la nouvelle relation à un tuple formé de valeurs d'attributs des deux tuples mis en relation.

La jointure spatiale est donc l'opération de jointure portant sur l'attribut de localisation. La qualification de jointure s'exprime soit par rapport à la distance entre deux objets :

- $a1 \leq d(O1, O2) \leq a2$

Soit par rapport à une relation ensembliste entre les objets :

- $O1 \cap O2 = \emptyset$
- $O1 \cap O2 \neq \emptyset$
- $O1 \subset O2$
- $O1 \not\subset O2$

Par exemple, $a2 = 0$ ($d(O1, O2) = 0$) signifie que deux points de l'espace sont mis en relation s'ils ont la même localisation. On parle alors d'équijointure géométrique.

Pour bien définir cette opération, il est nécessaire de revenir à l'espace géométrique de référence et de repasser du type de représentation des objets en zones, lignes, points aux objets élémentaires que sont les points de l'espace euclidien avec leur localisation intrinsèque.

L'opération de jointure spatiale s'effectue, en théorie, en considérant que les seuls objets logiques sont les points de l'espace.

Le résultat de l'opération de jointure pourra à nouveau faire l'objet d'un changement de représentation, du type passage entité principale-entité dérivée.

Le cheminement théorique d'une jointure spatiale est donc :

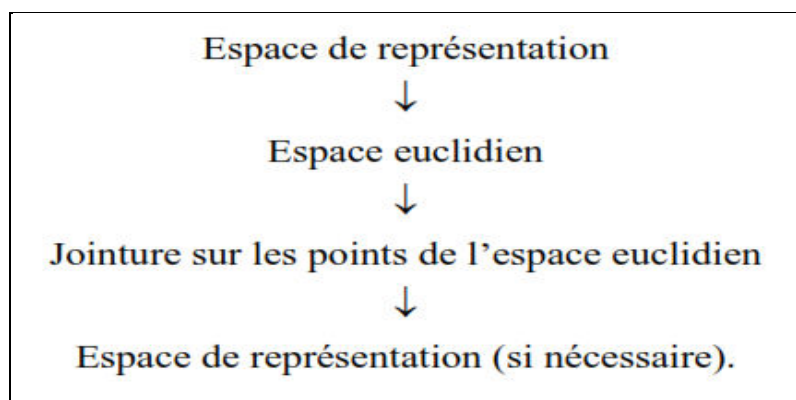


Figure2.1 Le cheminement théorique d'une jointure spatiale

Espace de représentation (si nécessaire).

La jointure spatiale crée de nouveaux objets.

La description de la localisation de ces objets pose les mêmes problèmes que la description de la localisation des objets initiaux : les nouveaux objets sont des points ou des ensembles de points de l'espace euclidien, qu'il est nécessaire de regrouper et de décrire d'une manière ensembliste avec les choix possibles que nous avons vus. Bien sûr, les méthodes de réalisation de la jointure et de la description de la localisation des résultats seront fonction du type de description des objets initiaux.

La jointure spatiale est une opération très intéressante dans la mesure où elle permet de comparer des objets sur leur localisation, sans avoir à spécifier et décrire cette localisation. Le résultat peut d'ailleurs être considéré d'avantage au niveau de cette mise en relation, plutôt qu'au niveau du résultat graphique formé par les tuples de la nouvelle relation. Ainsi par exemple, on pourrait mettre en relation la relation parcelle (numéro, nom de propriétaire, nombre de construction,) et la relation bornes d'incendie (numéro, débit) sur la qualification $d(O1, O2) < 100$ mètres pour avoir la liste des numéros de parcelles et le nombre de constructions associées à une borne plutôt qu'une carte des parcelles ou portions de parcelles et bornes résultant de la jointure.

Les jointures du type $d(O1, O2) = 0$ sont appelées des équijointures géométriques. Si la qualification de jointure est $d(O1, O2) < a$, le résultat est plus complexe puisqu'il correspond à un ensemble d'objets pour lesquels la localisation ne constitue plus une clé : un point peut apparaître plusieurs fois avec des descripteurs différents s'il répond à la qualification de jointure pour plusieurs objets de chaque relation. La notion de clé de localisation n'est pas conservée par la jointure spatiale, comme c'est d'ailleurs le cas pour le résultat d'une jointure classique.

Dans ce cas, le résultat sera une relation non localisée, et le résultat sera impossible à représenter cartographiquement car nous n'avons plus de dépendance fonctionnelle des attributs descriptifs par rapport à la localisation : on peut avoir des tuples formés du même ensemble de points mais ayant des valeurs descriptives différentes. [20]

6- Architecture des SIG

Pour évaluer les différents SIG du marché, il est important de considérer leur architecture qui impacte leurs performances. Il existe principalement (figure 2.2) trois grandes familles d'architecture pour la gestion des données sémantiques et des données spatiales [33],[4],

[16]. Ces architectures influent sur les performances, la fiabilité et la cohérence dans la gestion des deux types de données. Ainsi, suivant la figure 2.2 :

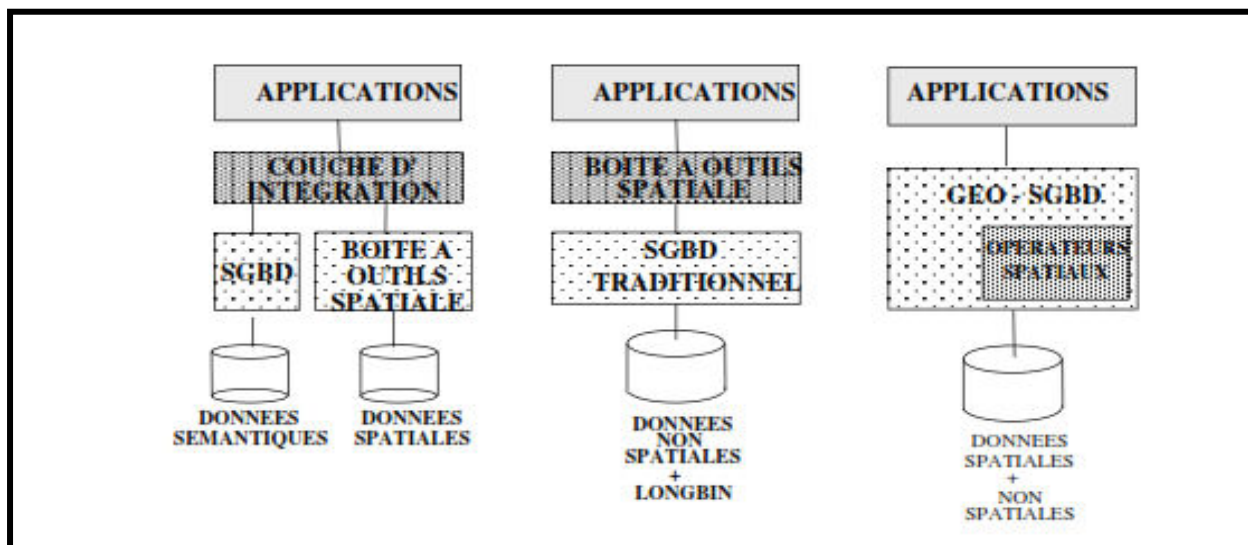


Figure 2.2 architecture d'un SIG exemple est ARC-INFO [33]

– La gauche traduit la première génération de SIG qui place côte à côte un SGBD pour la gestion des données sémantiques et un système propre (appelé également boîte à outils) pour gérer les données spatiales. Ce type de couplage nécessite une couche d'intégration qui doit ventiler les demandes (requêtes) des applications vers l'une des deux parties.

-L'avantage est de pouvoir s'intégrer à n'importe quel SGBD du commerce dont il est impossible d'adapter le noyau en raison de l'inaccessibilité des codes sources. Le SGBD est alors un serveur du système qui est sollicité pour gérer ce qu'il sait faire au mieux, à savoir les données alphanumériques.

-Le désavantage est un couplage faible. Bien qu'un objet spatial soit constitué de données sémantiques et de données spatiales, le stockage dans les deux parties pose le problème de cohérence entre les deux systèmes.

Par exemple, la restauration après une panne d'un côté ne garantit pas la cohérence avec l'autre côté. L'autre désavantage est de nécessiter que la couche au-dessus joue le rôle de médiateur pour synchroniser la gestion d'objets géographiques de part et d'autre. Ce rôle de médiateur se traduit par de nombreuses requêtes qui dégradent les performances globales du système.

Un exemple de système très connu et basé sur cette architecture est Arc-Info.

C'est un précurseur qui offre néanmoins de nombreuses fonctionnalités et outils intéressants à l'heure actuelle.

Le milieu de la figure montre une architecture qui résout en partie les problèmes cités précédemment. Les données spatiales et alphanumériques sont stockées uniformément dans le SGBD. Ceci a été rendu possible grâce aux nouvelles possibilités des SGBD de stocker des données binaires longues et non interprétées (appelé également BLOB). La construction d'un SIG consiste alors à placer une surcouche chargée, comme dans le cas précédent, du traitement des opérations spatiales.

L'avantage est un stockage plus intégré des données spatiales. Cependant, l'exécution d'une requête spatiale reste peu performante car toute requête doit être interprétée par la couche supérieure qui doit ainsi solliciter de manière importante la couche du bas pour évaluer toute requête. Signalons que cette approche est récemment mise en œuvre dans Oracle 8 avec les notions de « Cartridges ».

– La droite de la figure montre un système totalement intégré. Un tel système résout les problèmes évoqués. Cependant, l'offre est pauvre dans cette approche, car elle nécessite de concevoir un SGBD entier, ce qui requiert des ressources de développement très importantes. Il existe des prototypes de recherches dans cette catégorie d'architecture comme le SIG Géo Sabrina[33] [16].

En résumé, l'architecture peut influencer sur la fiabilité des données gérées et sur les performances. Les évolutions de ces architectures ont permis d'aller vers un couplage de plus en plus fort entre la partie sémantique et la partie spatiale. Ces évolutions garantissent des performances et une fiabilité accrue.

7- Architecture hybride

Pendant longtemps, le stockage et la gestion des données géographiques ont exploité des solutions dites « hybrides », c'est-à-dire traitant de façon distincte la géométrie et la sémantique. Le modèle le plus simple consiste à utiliser un système de gestion de base de données relationnel (SGBD-R) pour les attributs alphanumériques décrivant la thématique, et, en parallèle, un système de gestion de fichiers (SGF) pour les géométries des entités géographiques.

Les attributs sont donc stockés dans des tables selon un schéma relationnel classique, et sont susceptibles d'être traités (mises à jour, requêtes) par le langage de transaction normalisé SQL (StructuredQueryLanguage), pourvu que ces transactions ne portent que sur la thématique. Les coordonnées des points fixant les géométries des entités

géographiques sont enregistrées dans des fichiers de formats distincts selon les géométries, de manière à faciliter leur accès et leur gestion.

A cet égard, cependant, les différents producteurs de logiciels présentent des solutions informatiques différentes et les formats de fichiers sont qualifiés de "propriétaires" dans la mesure où l'accès aux données est strictement dépendant du logiciel utilisé.

C'est également ce logiciel qui prend en charge l'association entre une géométrie particulière et les valeurs d'attributs correspondantes, reconstituant ainsi l'information complète relative à une entité géographique. Cette association utilise un "identifiant géographique commun pour la géométrie en question et l'enregistrement correspondant dans la table d'attributs" (Figure2.3). Le logiciel étant capable de reconstituer l'information géographique complète, il peut la soumettre à différents types de traitements relevant de l'analyse spatiale et/ou de la cartographie, mais on constate que les traitements sont dépendants des formats des données.

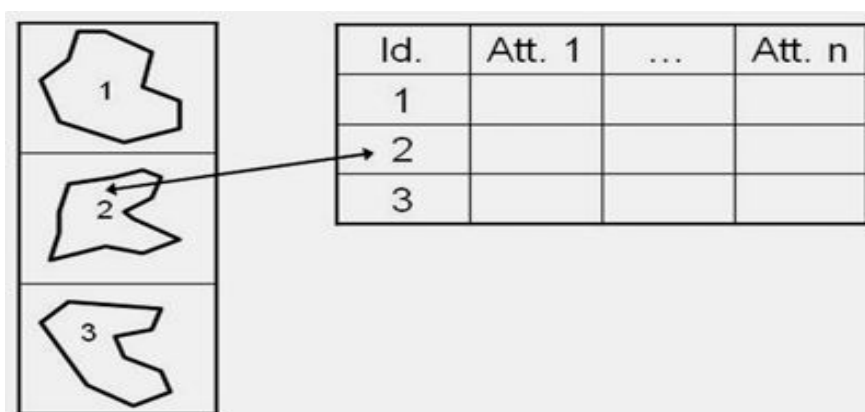


Figure2.3 Architecture hybride [7]

Fig2.3: A gauche, le fichier des géométries ; à droite, la table des attributs. La liaison passe par l'identifiant commun (ici, un numéro) des géométries et des enregistrements.

Les systèmes hybrides les plus simples ne conservent pas, et par conséquent ne peuvent être exploités, les relations entre les entités géographiques, en particulier les relations spatiales. Par contre, les systèmes dits "géo-relationnels" constituent une avancée dans ce domaine, en ce qu'ils identifient, conservent et permettent à l'utilisateur d'exploiter les principales relations topologiques entre entités géographiques.

Les géométries et les attributs sont toujours gérés par des sous-systèmes distincts, respectivement un SGF et un SGBD-R, mais en plus, sur base des coordonnées, le logiciel identifie les relations spatiales entre les géométries.

Ces relations sont stockées dans son système de gestion de fichiers, toujours dans un format propriétaire, mais le logiciel en fournit une vue à l'utilisateur, sous la forme de tables reprenant les principaux indicateurs de connexité et de connectivité pour chaque entité spatiale (Figure2.4).

Ces tables sont engendrées automatiquement dans le SGBD-R et peuvent bien sûr être associées, par exemple par jointure, aux tables d'attributs définissant la thématique des entités géographiques.

L'utilisateur a donc la possibilité d'interroger la base de données au moyen du langage de Requête SQL, tant sur la thématique que sur les relations spatiales. [7]

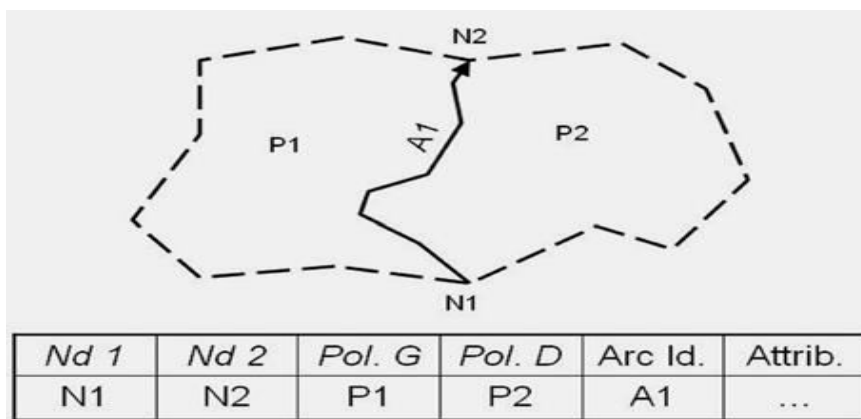


Figure2.4 Architecture hybride[7]

Figure2.4 : Vue engendrée par un système géo-relationnel, des relations topologiques (champs en italique) d'un arc délimitant une frontière commune entre deux polygones Nd 1 : nœud de départ – Nd 2 : nœud d'arrivée – Pol. G : polygone à gauche Pol. D : polygone à droite – Arc Id. : identifiant de l'arc – Attrib. : attribut(s) de l'utilisateur.

8- Outils de développement

Le SIG GRASS (<http://www.baylor.edu/grass/> et <http://grass.itc.it/index2.html>)

Le SIG Grass Est un système d'information géographique ouvert (libre) utilisé dans un grand nombre d'applications. En effet, il permet la gestion des données, l'analyse spatiale, le traitement d'images, la production graphique, la génération de cartes, la modélisation spatiale, et la visualisation de données à 2D, 2.5D et à 3D.

GRASS peut aussi être utilisé pour sauvegarder des mises à jour de cartes.

Il permet la manipulation et l'utilisation de plusieurs types de données géographiques; ces données peuvent être de sources, de satellites et de formats divers.

Les utilisateurs de ce SIG peuvent ainsi produire des graphiques, des statistiques, des rapports ainsi que des palmarès de données modélisées et calibrées.

GRASS se présente ainsi comme un système géographique très puissant pour la gestion des données spatiales de nature très diverses, ce qui semble très intéressant vu l'hétérogénéité des données que nous voulons intégrer au sein de la plate-forme ciblée. Ainsi, ce système sera utilisé, d'une part, pour stocker les données des bases d'images de la plate-forme.

De plus, bien que le point faible de GRASS est qu'il ne soit pas couplé en interne à un SGBD, il peut facilement s'interfacer à n'importe quelle base de données en utilisant par exemple un ODBC, entre autres aux bases de données spatiales et spatiotemporelles que nous nous proposons d'intégrer dans notre plate-forme.

GRASS dispose d'un ensemble de modules intégrant un nombre important d'outils, de programmes et de méthodes permettant de traiter les données géographiques.

Ce qui pourra constituer le noyau de départ pour la réalisation de la boîte à outils de la plateforme.

De plus, étant donné l'architecture ouverte et multi niveaux de ce système, il est aisément possible d'intégrer de nouveaux programmes et de nouvelles méthodes.

Notons également que les résultats-sorties de ces méthodes peuvent être aussi bien des données factuelles que des données images.

Ces résultats peuvent être visualisées sous plusieurs formes. [27]

9- Relations spatiales

Les relations spatiales sont des informations qui traduisent des propriétés essentielles dans le monde géographique.

Tout géographe s'accorde à dire que tout phénomène en un endroit est lié à l'influence du voisinage et cette influence décroît avec l'éloignement (1^{ère} loi en géographie [33] [29]).

Ces informations sur les relations spatiales peuvent exister explicitement ou implicitement dans une base de données géographique.

Dans le cas explicite, on utilise des modèles de type topologique qui permettent de décrire ces relations à l'aide de graphes.

Dans le cas implicite, les relations spatiales les plus simples sont déduites par des calculs géométriques.

Ce calcul revient à effectuer le produit cartésien entre tous les couples d'objets spatiaux et à déterminer les couples qui répondent aux critères.

Dans le cas d'une décomposition en couches thématiques, ces relations spatiales peuvent être intra ou inter thèmes.

10- Méta-informations géographiques

Pour exploiter les données d'une couche thématique, il est nécessaire d'avoir des méta-informations (information sur les données contenues).

Ce sont typiquement les informations d'échelle, d'emprise, de référentiel géographique (système de projection), de qualité, de datation, ...

Ces méta-informations s'appliquent le plus souvent sur un lot d'objets géographiques dit « lot homogène ».

Ces informations servent à différents niveaux d'un SIG pour l'exploitation des données, soit par le système, soit pour renseigner l'utilisateur. En général, elles sont gérées dans un dictionnaire de données contenu dans le SIG. [33]

CHAPITRE 3

DATAMINING DANS LES BASES DE DONNEES SPATIALES MULTIPLES

1- Introduction :

Avec le développement de la cartographie numérique, le volume de données dans les bases de données spatiales ne cesse d'augmenter. Ces données sont de plus en plus utilisées dans des applications décisionnelles, surtout depuis le développement d'outils de géocodage permettant la localisation par l'adresse. C'est le cas en géomarketing, dans l'analyse de la criminalité ou dans l'analyse de risques d'accidents ou d'épidémies. Seulement, la nature et le volume de données de base dépassent les capacités humaines d'analyse. D'où l'intérêt d'appliquer des techniques d'extraction automatique de connaissances telles que le Data Mining aux bases de données géographiques.

Le Data Mining spatial (DMS) est né du besoin d'exploitation dans un but décisionnel de données à caractère spatial produites, importées ou accumulées, susceptibles de délivrer des informations ou des connaissances par le moyen d'outils exploratoires (de fouille de données). Il constitue un domaine à part car il considère les interactions des objets dans l'espace. Ce domaine intègre des techniques provenant à la fois des bases de données spatiales et des SIG, du Data Mining et des statistiques spatiales [33].

La diversité des sources des informations spatiales nous oblige d'étendre notre Base de données spatiales qui va contenir plusieurs bases spatiales interagissant entre elle formant une large base utilisée par le Data Mining pour extraire les connaissances dans les différents domaines.

L'objectif de ce mémoire est de proposer deux approches le premier d'assembler toutes les données dans une seule base spatiale, et le deuxième de laisser les bases dans les sites où elles existaient et travailler sur leurs liens selon leurs utilités.

2-Fouille de données Datamining

2.1- Définition de la fouille de données ou datamining :

La fouille de données constitue l'étape clé du processus de découverte de connaissances, étape sur laquelle bien des acteurs se focalisent d'emblée au détriment des autres étapes qui rappellent le sont toutes aussi importantes.

Depuis son apparition, la fouille de données a connu un essor fulgurant. A la faveur du développement technologique et des besoins du marché, elle s'est trouvée utilisée dans bien des domaines (Marketing, Télécommunication, détection de fraude, domaine manufacturier, etc.).

La fouille de données en elle-même consiste en l'utilisation des techniques et algorithmes afin d'extraire des connaissances implicites et potentiellement utiles, pouvant servir de support au processus de décision de l'entreprise.

Son objectif principal est d'être soit prédictive, soit descriptive. Prédictive dans le sens de prédire la valeur future des variables étudiées et descriptive dans le sens de la production de modèle expliquant les données sous étude.

Elle est composée d'un grand nombre de techniques et algorithmes que l'on peut caractériser selon plusieurs points de vue : supervisé vs non-supervisé, prédictif vs descriptif, transparent vs opaques.

Dans les lignes qui suivent, nous décrirons les différentes caractéristiques de la fouille de données. Notons toutefois que les différentes caractéristiques peuvent avoir des points communs dans la mesure où certaines techniques ou algorithmes possèdent plus d'une caractéristique. [23]

2.2- Les méthodes de datamining :

On distingue deux méthodes de datamining :

2.2.1- Les méthodes supervisées/prédictives

Il s'agit de méthodes qui requièrent généralement de l'utilisateur la définition d'une variable cible dont on veut par exemple prédire la valeur. Il pourra s'agir par exemple, de prédire la valeur d'attributs dénommés variables dépendantes en fonction d'autres variables appelées variables explicatives ou indépendantes. Les algorithmes désignés comme supervisés fonctionnent généralement sur la base de trois (3) jeux de données :

- jeu de données d'essai ou training set : contient l'ensemble des attributs y compris les valeurs de la variable à prédire. Ces valeurs aident à la supervision du processus en mettant à nus les erreurs quand l'algorithme utilise le modèle pour prédire les résultats.
- jeu de données test ou test set : contient les différentes variables exceptées les valeurs de l'attribut à prédire. Ce jeu de données contient après le lancement de l'algorithme, les valeurs prédites de la variable cible.
- jeu de données de validation ou validation set : est semblable au test set avec toutefois les vraies valeurs de l'attribut cible. Ce jeu de données sert à faire une

confrontation avec les valeurs prédites de la variable cible afin d'estimer le pourcentage d'efficacité de l'algorithme utilisé.

On peut également subdiviser les méthodes supervisées/prédictives en deux classes d'algorithmes : les algorithmes de classification et ceux de régression.

La classification consiste en la prédiction de catégories de valeurs (valeurs discrètes) (exemple : quelle sera la réponse d'un client à une offre ? la réponse pouvant être « j'accepte sans réserve », « j'accepte avec réserve », « Je refuse », etc.) Contrairement à la régression qui prédit plutôt des valeurs continues (numériques) (exemple : quelle sera la valeur d'une maison ou le revenu d'une personne (25milles dollars par an, 10075.99 \$, etc.). La classification et la régression diffèrent également quant à leur mode d'évaluation du résultat de la prédiction. [23]

2.2.1.1-Classification

Le terme classification pose souvent une ambiguïté en raison de la confusion possible avec regroupement ou clustering. La classification est de loin l'une des tâches de fouille de données la plus utilisée car intervenant dans plusieurs domaines d'activité Banque, Médecine ... (Larose, 2005). La finalité d'une tâche de classification est d'assurer la prédiction d'un attribut cible nominal (type chaîne de caractère) sur la base d'une connaissance préalable des données qui leur sont fournies en entrée (training set). Bien que largement utilisé, le problème majeur avec la classification demeure la nécessité de faire un compromis entre un modèle qui minimise le taux d'erreur et la variance.

Comme algorithmes accomplissant cette tâche, on note : le KNN, les arbres de décision, les réseaux de neurones, les réseaux de Bayés.

L'estimation ressemble beaucoup à la classification à la différence que la variable cible est numérique en lieu et place d'être nominale (type chaîne de caractère) comme dans le cas de la classification. Comme algorithmes, on note la régression linéaire simple ou multiple. Les réseaux de neurones peuvent être utilisés à cette fin. Un exemple concret d'une tâche de classification est la prévision météo. En effet, pour prévoir le temps qu'il fera dans une zone géographique donnée, on prend en considération un certain nombre de paramètres nommés variables explicatives pouvant être l'humidité, la vitesse du vent, la température de la veille... et éventuellement la position de la zone géographique étudiée par rapport aux pôles. Ces paramètres ainsi que la variable à prédire (la température par exemple) pourraient être modélisés dans une fonction mathématique. Chaque fois qu'on aura alors besoin de prédire la température, il suffira de passer à cette fonction ou modèle les paramètres appropriés pour obtenir les résultats désirés. [23]

2.2.2- Les méthodes non supervisées/descriptives

Contrairement aux méthodes supervisées, celles non-supervisées n'utilisent pas de cible. Elles fonctionnent plutôt sur la base de recherche de structures intrinsèques, des relations, ou affinités dans le jeu de données fourni en entrée. En d'autres termes, il s'agit de trouver des tendances et corrélations qui résument les relations entre données. Ces méthodes peuvent servir par exemple dans des domaines tels la détection de fraude, la réduction de dimensionnalité. La plus connue des tâches d'apprentissage non supervisé est le clustering. Le caractère descriptif de ces méthodes réside dans le fait qu'elles décrivent de manière concise et résumée un jeu de données en présentant les propriétés intéressantes de ces données. [23]

2.2.2.1-Règles d'association

Encore connue sous le nom d'analyse d'affinité ou analyse du panier de consommation, la tâche d'association en fouille de données vise à voir quelles sont les variables qui vont ensemble.

Il s'agit de trouver des règles du type *si X alors Y* avec un certain niveau de probabilité. Deux métriques sont utilisées pour caractériser généralement la qualité d'une règle d'association : le support et la confiance. Le support décrit la probabilité d'existence de X et Y au sein du jeu de données. La confiance décrit quant à elle la probabilité d'existence de Y dans l'ensemble de données contenant X. Le problème majeur avec la tâche d'association est le nombre d'association possible entre attribut.

En effet, avec k attributs prenant une valeur binaire, le nombre maximal d'associations est $2^k - 1$.

A titre d'exemple pour les algorithmes effectuant une tâche d'association, on note : GRI (General Rule Induction) et l'algorithme Apriori .

A titre d'exemple, on peut mettre à profit les techniques d'association pour déceler les liens éventuels entre les différents produits vendus dans un supermarché. On peut ainsi noter que chaque fois que de la viande hachée est achetée, à 80% les pâtes sont également achetées. On note donc une certaine association entre les produits Viande et Pâtes avec 80% de taux de confiance.

On pourra entreprendre comme action de disposer le rayon « Pâtes » à proximité de celui concernant la « Viande » afin d'amener le client à ne pas fournir d'effort pour aller dans le rayon « Pâte » ou que celui-ci se rappelle qu'il doit acheter des pâtes au cas où il aurait oublié. [23]

2.2.2.2-Clustering

Selon (Mirkin, 2005), l'idée de clustering renvoie tout simplement à l'utilisation de mesures de similarité (ou dissimilarité) entre les entités de sorte à regrouper ensemble celles similaires et celles dissimilaires dans un autre groupe. En d'autres termes, il s'agit d'une organisation des données en un ensemble de groupes homogènes: les clusters regroupés de telle sorte à minimiser la variance intra-classe et à maximiser celle inter-classe. Plusieurs algorithmes de clustering existent selon qu'ils utilisent différentes techniques d'évaluation de la similarité ou selon les objectifs visés :

□ *Approche basée sur les partitions* : partitionne l'ensemble de données en un nombre K de clusters. Ces k partitions sont corrigées jusqu'à obtenir une similarité satisfaisante. On note comme algorithmes : k -Means, k -Médoids, ...

□ *Approche basée sur la densité* : utilise un modèle probabiliste pour déterminer les entités denses au sein de l'ensemble de données. Les objets sont groupés selon que la valeur de la densité avoisine une certaine limite (voir les algorithmes DBSCAN, DENCLUE, ...)

□ *Approche hiérarchique* : qui récursivement construit les clusters sous la forme d'une structure hiérarchique selon une approche top-down ou bottom-up (agglomérative ou divisive). (Voir les algorithmes CURE, STING, ...)

□ *Approche par grille* : l'espace est divisé en cellules pour former une grille multi-niveaux. Par la suite, les cellules voisines sont groupées en fonction de la distance (CLIQUE,)

□ *Approche par modèle* : modélise les groupes et utilise le modèle pour classer les points. Exemple d'algorithmes : COBWEB, réseaux de neurones.

A titre d'exemple, on peut grâce au clustering, se pencher sur l'étude de la typologie des étudiants – toutes années confondues – d'une université ; avec comme question de fond : est-ce que la qualité de l'enseignement a baissé ou augmentée ? Où est-ce plutôt le niveau des étudiants qui a baissé ou non ? [23]

3- Le Datamining spatial

La fouille de données spatiales est définie comme l'extraction de connaissances implicites de relations spatiales ou d'autres propriétés non explicitement stockées dans les bases de données spatiales. [27] [1]

Avec le développement de la cartographie numérique, le volume de données dans les bases de données spatiales ne cesse d'augmenter. Le développement d'outils de géocodage permettant la localisation par l'adresse d'un côté et la banalisation des moyens de

localisation de l'autre genre de nouvelles sources de données spatiales. Ces données sont de plus en plus utilisées dans des applications décisionnelles. Seulement, la nature et le volume de données de base dépassent les capacités humaines d'analyse.

La fouille de données est reconnue comme un moyen très efficace d'analyse avancée de données, permettant d'extraire des connaissances cachées depuis des grandes masses de données. Etant donné le volume croissant des données spatiales, la fouille de données spatiales permet d'extraire des propriétés spatiales cachées dans ces données et présente donc un intérêt certain pour les applications spatiales décisionnelles. La fouille de données spatiale (FDS) est aujourd'hui identifiée comme un domaine de la fouille de données à part entière. Elle résulte de la combinaison de la fouille de données et des bases de données spatiales [34].

3.1-Le clustering spatial :

Le Clustering est une méthode en fouille de données permettant le regroupement d'objets par classes homogènes. Donc il s'agit de trouver une métrique qui peut détecter les points similaires vis-à-vis un phénomène et visualisé les concentrations. Le point faible de cette technique est le fait qu'elle repose sur un seul critère qui est la localisation alors que la réalité est multicritère. La méthode STING comme les méthodes à base de Clustering, vise à identifier des classes d'objets fortement similaires. Elle ne fournit aucune description sur la relation des classes et leurs relations avec les caractéristiques des objets. La méthode d'hierarchie et généralisation de concept repose sur l'intervention préalable d'un expert pour définir la hiérarchie de concept. Ainsi nous pouvons remarquer que les méthodes précédentes n'exploitent qu'un seul thème à la fois, alors que les données géographiques sont souvent représentées par plusieurs couches thématiques qui cachent des relations topologiques entre les objets qu'elles contiennent.

Notre contribution dans ce domaine, en prenant en considération le caractère spatial de l'interaction avec l'environnement géographique, est la conception d'un système d'extraction et d'optimisation des règles de classification appliqué à la sécurité routière. [1]

3.2-La classification spatiale :

La tâche de la classification consiste à assigner un objet à une classe donnée basées sur les valeurs d'attributs de cet objet. Dans la *classification spatiale* les valeurs d'attribut des objets voisins sont également considérées. [1]

Contrairement au clustering qui cherche à identifier des classes, la classification supervisée permet d'affecter les objets à une classe parmi celles prédéfinies. Elle peut être utilisée pour prédire les classes de nouveaux objets ou simplement pour décrire ou expliquer les

liens entre les propriétés de l'objet et sa classe. Parmi les méthodes les plus utilisées, car fournissant des règles interprétables par l'analyste, on trouve les arbres de décision [34] [17], [35]. Un arbre de décision est construit par l'application successive de critères de subdivision sur une population d'apprentissage afin d'obtenir des sous-populations plus homogènes.

Le critère de subdivision est déterminé au niveau de l'attribut dans l'arbre ID3 [34] [25] et au niveau d'une valeur d'attribut dans CART [34] [5]. Ester et al. [34] [8] ont proposé une méthode de classification spatiale basée sur ID3 et utilisant le concept de graphe de voisinage. Pour intégrer les propriétés du voisinage, les auteurs définissent une notion d'attribut généralisé par l'ajout d'un degré de voisinage. Ils intègrent cet attribut dans l'algorithme ID3 comme critère possible de subdivision. Cette méthode ne distingue pas explicitement des couches thématiques et considère que tous les objets possèdent les mêmes attributs. Néanmoins, les auteurs supposent dans l'exemple donné en illustration l'existence d'un attribut « type d'objet », qui implicitement reflète la notion de thème.

Comparativement la méthode de classification de Koperski [34] [25] considère des thèmes de référence et des relations de voisinage précises.

Elle représente toutes les propriétés par une liste de prédicats. Ainsi le prédicat noté attribut (x, valeur) est l'équivalent de "x.attribut = valeur" et close to (x, catégorie) représente le voisinage à une catégorie d'objets. De plus, elle intègre des mesures agrégées par extension au voisinage (ex : la population totale aux alentours) et dont le périmètre d'extension est déterminé algorithmiquement.

Il suffit ensuite de calculer le gain d'information pour chacun des prédicats. Afin d'optimiser les traitements, ce calcul est fait sur une généralisation des données selon des hiérarchies données par l'utilisateur.

D'autres optimisations sont proposées pour le calcul approximatif des prédicats spatiaux et pour le filtrage préalable des prédicats pertinents de l'objet. [21]

3.3-Les règles d'association spatiales :

L'extension de la découverte de règles d'association de [34] [2] aux données spatiales a été proposée par Koperski et al. [34] [13], [14] et permet de générer des règles de type : $X \rightarrow Y (s, c)$ avec un support s et une confiance c telles que X et Y sont des ensembles de prédicats, dont au moins un est un prédicat spatial. En d'autres termes, ces règles correspondent à des associations entre des propriétés des objets et celles de leurs « voisins ». Cette méthode génère pour chaque objet d'un thème cible de l'analyse, une liste de prédicats comprenant sa description, la description d'objets liés d'autres thèmes et leurs

relations spatiales avec l'objet. Cette liste constitue des items d'une transaction sur laquelle un algorithme de type APRIORI peut opérer pour générer des règles d'association spatiales de type :

$is_a(x, school) \wedge close_to(x, sport\ center) \rightarrow close_to(x, park) (s, c)$

Ici, le premier prédicat est non spatial tandis que les trois autres prédicats sont spatiaux. Cette méthode permet, en outre, de générer des règles multi-niveaux en exploitant à la fois la généralisation des attributs sur la base d'une hiérarchie de concepts et la généralisation de la relation spatiale de voisinage. Un avantage est de permettre la découverte de règles qui n'apparaissent qu'à un niveau général. De plus, cette approche permet d'optimiser la recherche car elle utilise la propriété d'anti-monotonie qu'une combinaison de prédicats détaillés n'est fréquente que si le niveau général correspondant l'est. La transformation en prédicats mène naturellement à l'application de la programmation logique. C'est ainsi qu'une méthode similaire a été développée dans le cadre de la fouille de données multi-relationnelles [34] [19], [6] par une adaptation de l'algorithme de [34] [13] aux données spatiales exprimées en logique du premier ordre. [34]

4-Le datamining dans les bases de données spatiales multiples

4.1-Problématique :

La découverte de la connaissance devrait être comprise comme découverte de l'information combinée avec la création de la connaissance. La création de la connaissance de l'information peut être favorisée par les représentations appropriées d'information qui rendent la structure logique inhérente de l'information transparente. Puisque les concepts sont les unités de base de la pensée humaine et par conséquent des structures de base de la logique, la structure logique d'information est basée sur des concepts et des systèmes de concept. Par conséquent, les treillis de concept en tant qu'abstraction mathématique des systèmes de concept peuvent soutenir des humains pour découvrir l'information et puis pour déduire la connaissance. Dans ce contexte, nous utilisons les treillis de concept pour l'extraction de la connaissance spatiale, dans l'objectif de générer les règles d'association spatiale.

Le but de l'utilisation des règles d'association, comme méthode de fouille de données est l'identification des relations significatives entre les données de taille très importante. Dans le cas des bases de données transactionnelles, étant donné un ensemble d'articles, le but est de découvrir si l'occurrence de cet ensemble dans une transaction est associée à l'occurrence d'un ensemble d'articles. Par exemple, « 90% des clients qui achètent du thé

achètent le sucre » est une règle d'association, sous la forme, associant l'article thé à l'article sucre. Sucrethé.

Dans le cas des bases de données géographiques ou spatiales, la nature et le volume de données dépassent les capacités humaines en terme d'analyse. D'où le besoin de l'utilisation des méthodes simples et faciles à comprendre, parmi ces méthodes on peut citer les règles d'association spatiale. Ces dernières, appliquées dans plusieurs domaines d'activités, présentent des résultats claires, faciles à interpréter. Cependant, deux problèmes majeurs liés à l'utilisation des règles d'association spatiales sont posés, à savoir le problème des temps d'extraction des règles d'association spatiales à partir du jeu de données et le problème de la pertinence et de l'utilité des règles d'association spatiales extraites. [21]

4.2-Solutions :

Les algorithmes d'extraction des règles d'association spatiales sont une extension des algorithmes des règles d'association, incluant la composante spatiale. A ce titre avant de présenter les travaux de recherches traitants les règles d'association spatiales, nous allons présenter un bref aperçu sur les algorithmes d'extraction des règles d'association. Plusieurs travaux de recherches ont proposé des solutions pour les deux problèmes majeurs de l'extraction des règles d'association, à savoir :

1. le temps de réponse de l'extraction des règles d'association qui dépend principalement des temps d'extraction des items ets fréquents, car plusieurs balayages du contexte doivent être réalisés. En comptant pour chaque items et fréquent potentiel le nombre d'objets du contexte dans lesquels il est contenu.
2. l'utilité des règles d'association extraite, afin d'obtenir les règles d'association pertinentes à partir de celles qui sont extraites.

Les algorithmes d'extraction des règles d'association peuvent être classés en trois grandes catégories : l'extraction des fréquents, l'extraction des maximum et l'extraction des fermés. Les algorithmes de la première catégorie, sont basés sur la propriété d'antimonotonie :

1. tout sous-ensemble d'un ensemble d'article fréquent est fréquent ;
2. tout sur ensemble d'un ensemble non fréquent est non fréquent.

Parmi les algorithmes de cette catégorie, on peut citer l'algorithme Apriori dans lequel, deux paramètres, support minimum (minsup) et confiance minimale(minconf) sont introduits et deux phases principales résumant l'algorithme :

1. la première phase a pour objet la recherche de tous les sous-ensembles d'articles fréquents (fréquent Itemset) tel que : $\text{support}(X) > \text{minsup}$, autrement dit c'est la recherche des règles d'association utiles ;

2. la deuxième a pour objet la recherche de toutes les règles d'association intéressantes et pertinentes à partir des sous-ensembles X fréquents déterminés lors de la première phase tel que la confiance $(X) > \text{minconf}$. Les algorithmes de la deuxième catégorie, extraction des maximums, sont basés, en plus de la propriété d'antimonotonie, sur la définition des itemsets fréquents maximum, c'est à dire les itemsets dont tous les sur-ensembles sont infréquents, forment une bordure au-dessous de laquelle tous les itemsets sont fréquents.

Parmi les algorithmes qui utilisent cette approche, nous citons Pincer-Search, Max Clique, Max-Eclat et MaxMiner. L'avantage de ces algorithmes par rapport à ceux présentés auparavant est qu'ils réalisent deux types de parcours : de bas en haut et de haut en bas pour la découverte des itemsets fréquents maximums lors de chaque itération. Ce nouveau mode de parcours anticipatif ou en avance de haut en bas, « lookahead » permet de réduire le nombre d'itérations et par conséquent de diminuer le nombre de balayage du contexte réalisé.

En ce qui concerne la dernière catégorie, relative à l'extraction des fermés. Elle est basée sur la fermeture de la connexion de Galois, les itemsets sont des itemsets fréquents qui sont fermés selon l'opérateur de fermeture de la connexion de Galois. Les itemsets fermés fréquents, selon cet opérateur de fermeture, forment un ensemble générateur non redondant minimal pour tous les itemsets fréquents et leurs supports. Tous les itemsets fréquents et leurs supports, peuvent donc être déduits efficacement, sans accéder au jeu de données. Cette propriété découle du fait que le support d'un itemset fréquent est égal au support de sa fermeture. Les itemsets fermés fréquents forment un treillis dont la taille est bornée par la taille du treillis des itemsets fréquents. Toutefois, en pratique, la taille de ce treillis est en moyenne bien inférieure à la taille du treillis des itemsets (treillis des parties).

Koperski et Han ont prolongé la méthode de l'extraction des règles d'association aux bases de données spatiales.

Une règle d'association spatiale est de la forme :

$P_1 \text{ et } P_2 \text{ et } \dots \text{ et } P_n \rightarrow Q_1 \text{ et } Q_2 \text{ et } \dots \text{ et } Q_m (s\%, c\% \dots)$

Où au moins un des prédicats P_1, P_2, \dots, P_n et Q_1, Q_2

Q_m est un prédicat spatial, $s\%$ est le support de la règle et $c\%$ est la confiance de la règle.

Plusieurs types de prédicats spatiaux peuvent être utilisés au niveau des règles d'association spatiales.

Ils peuvent être utilisés pour représenter des relations topologiques entre objets spatiaux, comme : disjoint, intersectés, à l'intérieur de, contient, adjacent à, couvre, égal, ou pour représenter un ordre ou une orientation spatial, comme : à gauche, à droite, Nord, Sud, ou apporter une information de distance : près de, loin de, etc. [21]

5-Conclusion :

Le regroupement des bases spatiales avec des liens nous a permis de faire une extension à l'extraction des règles d'association appliquée précédemment sur une seule base vers plusieurs bases spatiales et attributaires.

Cette extension devenue une obligation tant que notre étude ne concerne pas une seule base, comme sachant que chaque couche spatiale représente une ou plusieurs bases.

Avec l'utilisation des requêtes de liaison spatiales sur Arc Gis et relationnelles sur Access n'importe quelle sorte de nouvelle table composée de toutes les tables attributaires et spatiales est devenue prête à être une matrice des items pour l'extraction des règles d'association, ce qu'on va le voir détaillé dans le quatrième chapitre

CHAPITRE 4

ALGORITHME DATAMINING

POUR LES BASES SPATIALES MULTIPLES

1-Introduction :

La fouille de données est un condensé de plusieurs disciplines dont la finalité est l'extraction de connaissances depuis des entrepôts de données volumineux. A cette fin, plusieurs outils ont été mis en œuvre au sein desquels sont implémentés plusieurs algorithmes de fouilles. De plus en plus, ces outils renferment d'autres fonctionnalités notamment celles d'analyse des données, de prétraitement et d'interprétation des résultats, qui vont au-delà de la fouille proprement dite.

Il est bon, dans une démarche de « spatialisation » d'un outil de fouille « traditionnelle », de choisir un outil offrant diverses fonctionnalités et ce pour plusieurs raisons. Premièrement parce que l'extraction de connaissances ne se limite pas à la tâche de fouille de données, contrairement à ce que l'on a pu penser de par le passé. L'extraction couvre d'autres étapes toutes aussi importantes que la fouille elle-même.

La seconde raison qui pourrait être un corollaire de la première, serait que l'existence de telles fonctionnalités, aiderait à gagner en temps de traitement d'autant plus dans un contexte de fouille géo-spatiales quand on sait que ces données sont complexes. En effet, passer par les étapes de compréhension des données, d'analyse exploratoire, etc. permettrait de s'intéresser aux sous-ensembles des données géo-spatiales « digne d'intérêt ». Ceci pour dire que le choix d'un outil offrant diverses fonctionnalités est une étape importante dans la démarche de « spatialisation ». Et plus, ce choix ne devrait pas être fondé que sur les fonctionnalités de l'outil mais doit aller au-delà pour couvrir d'autres critères tels l'extensibilité, la modularité, la licence d'utilisation, la robustesse, les fonctionnalités offertes. [23]

Dans ce chapitre on a proposé une méthode hybride qui fait la fouille dans une base mixte contenant deux types de bases spatiales et classiques.

La première contenant les différentes couches spatiales telles que les routes, les montagnes, les forêts et les autres reliefs, la deuxième contenant les statistiques des accidents ont eu lieu une région.

2- Problématique :

Pour l'étude de la fouille de données sur la base de données de la carte d'Algérie comme exemple en particulier les accidents routiers sur les différentes routes, on voit qu'on ne peut pas faire des études bien précises sur une carte globale cause de deux problèmes.

* le premier : Concerne la grande capacité de la base de données spatiale contenant l'énorme quantité des objets spatiaux tels que lessivages des routes et les autres couches spatiales (montagne, lac, mer,) dont la capacité de leurs bases de données dépasse les dizaines de TO qui pose deux problèmes fondamentaux :

- l'insuffisance de la capacité mémoire pour traiter ces données.
- la grande complexité qui dépassées capacités actuelles des processeurs.

* le deuxième : Comme la petite tranche d'une route peut aller jusqu'à 50 km sur la carte globale, comme il est montré sur la figure 4.1.

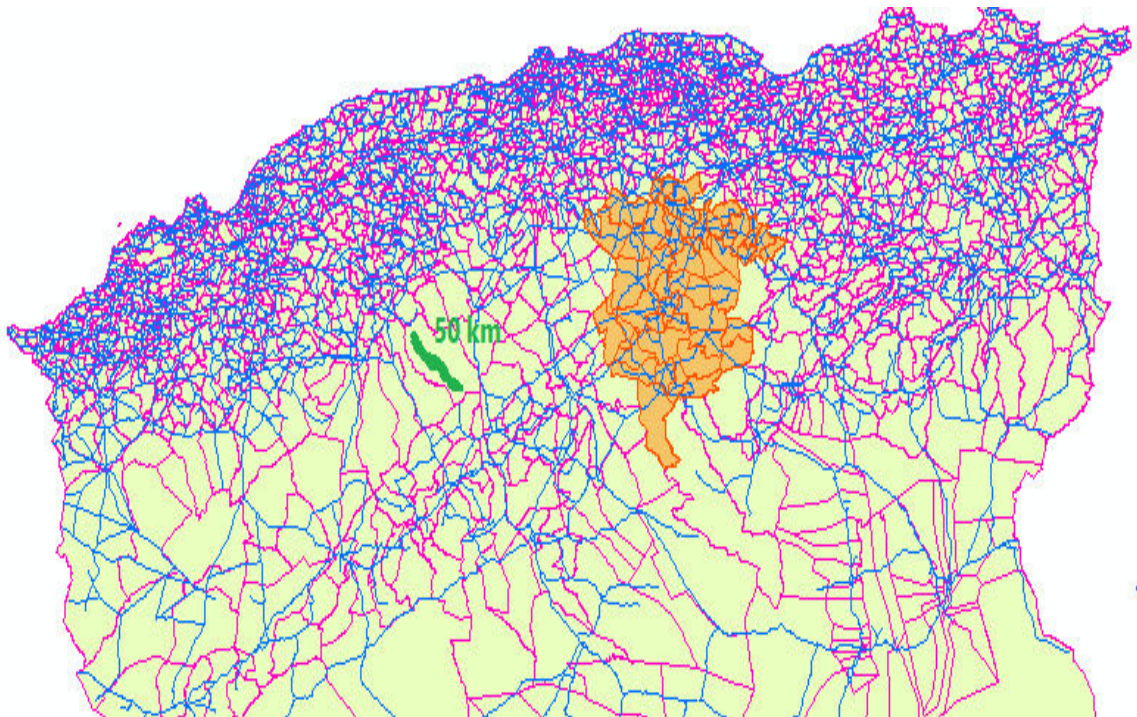


Figure4.1 Carte du réseau routier d'Algérie[32]

Et comme sur 50 Km on peut rencontrer plusieurs reliefs (montagne, plat, barrage, descente...) comme on voit sur la figure4.2.

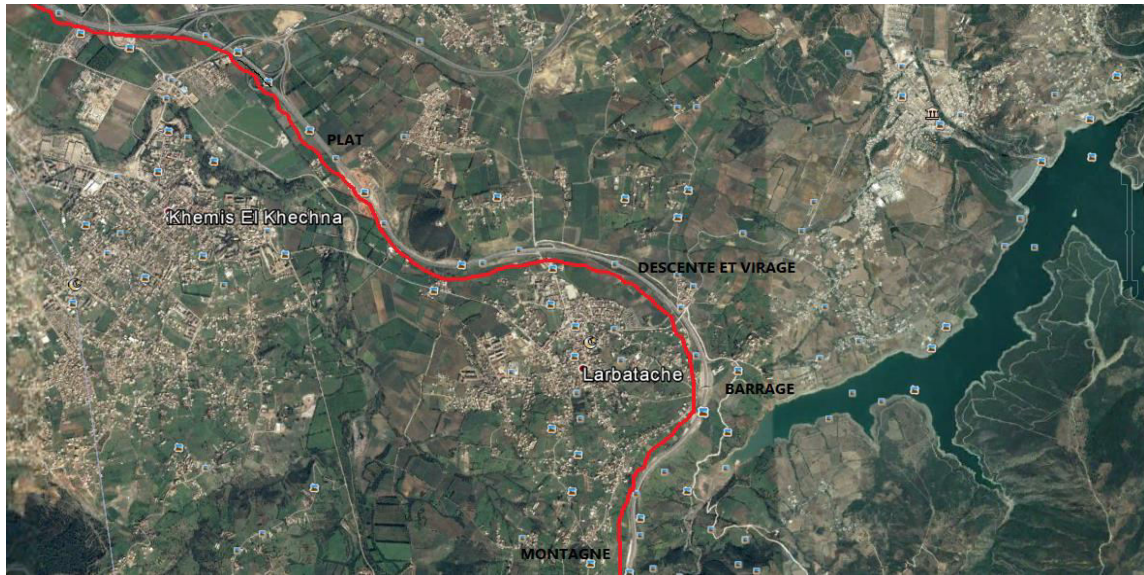


Figure4.2Image via Google Earth région d’Alger [9]

Ce qui contredit la première loi spatiale (Les objets voisins s’agissent entre eux) donc on a procédé à un découpage de la carte en plusieurs régions (wilaya) et on découpe les routes de chaque région en plusieurs morceaux de 5 à 10 km selon les changements de la région, et les nommer R_1, R_2, \dots, R_n , pour assurer une seule caractéristique pour chaque tranche, cela va régler le problème de voisinage des objet spatiaux ,autrement dit sur une tranche quelconque R_m on voit pas le plat ,la montagne en même temps,et ce découpage nous conduit à diviser la base de données globales a des sous bases permettant de faire le traitement en reliant ces sous bases entre elles avec une jointure,ce qu’on va le détailler ultérieurement.Ceci est montré sur la figure4.3

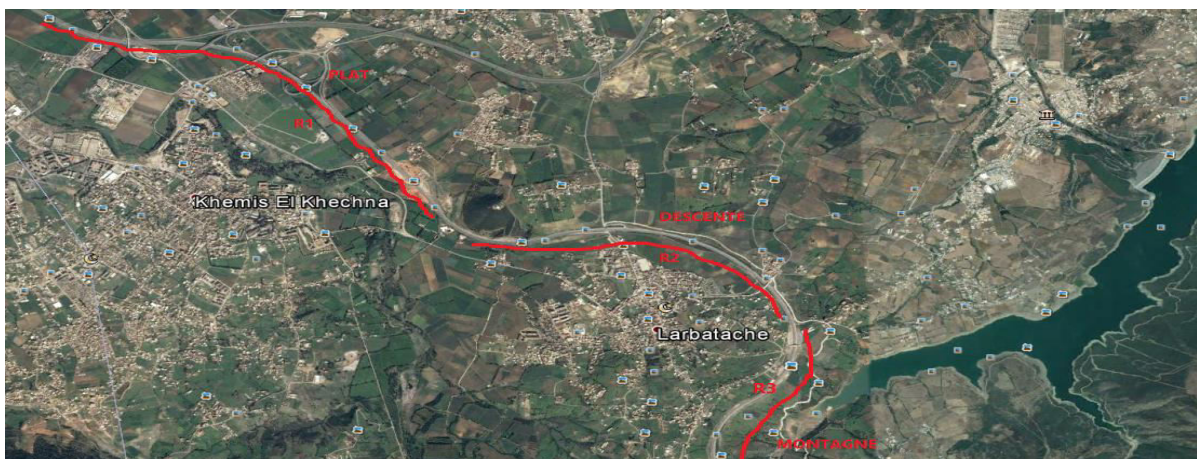


Figure4.3Image via Google Earth région d’Alger avec une route découpée[9]

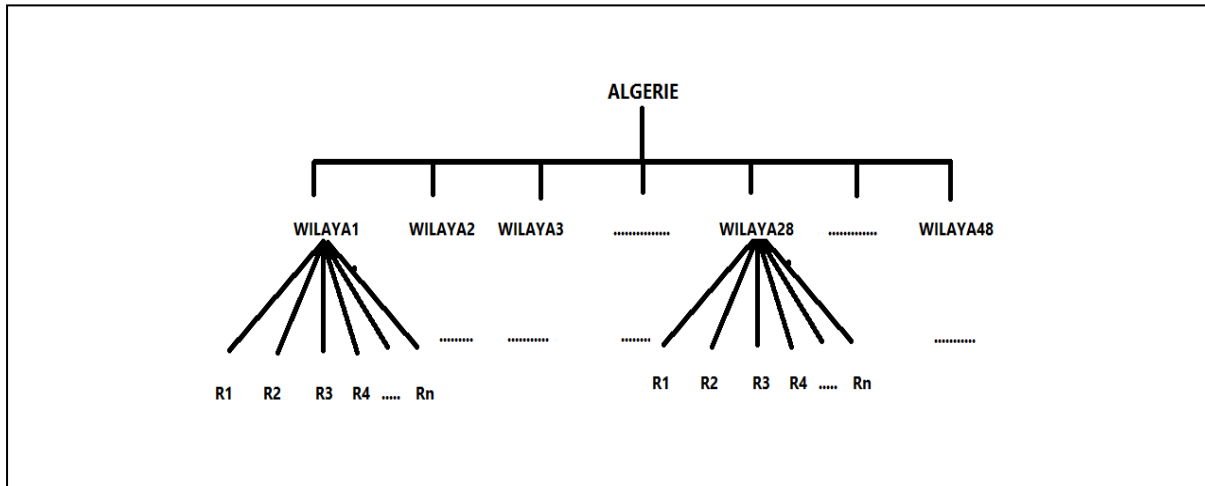


Figure4.4 Découpage des routes par Wilaya

Selon une hiérarchie comme il est montré sur la figure 4.4 le pays est divisé en 48 wilaya et chaque wilaya se divise en N routes et chaque morceau R_m et son voisinage (montagne, barrage, mer, forêt,) construisent une base de données BD_n , le regroupement de ces bases avec les bases des autres wilayas construisent la base de données mère de la carte d'Algérie Comme il est montré sur la figure 4.5. Exemple la wilaya de M'sila la route nationale entre M'sila et Boussaâda est découpée en 7 morceau d'environ 8Km R_1, R_2, R_8 .

Et la même chose pour les autres routes (figure 4.5)

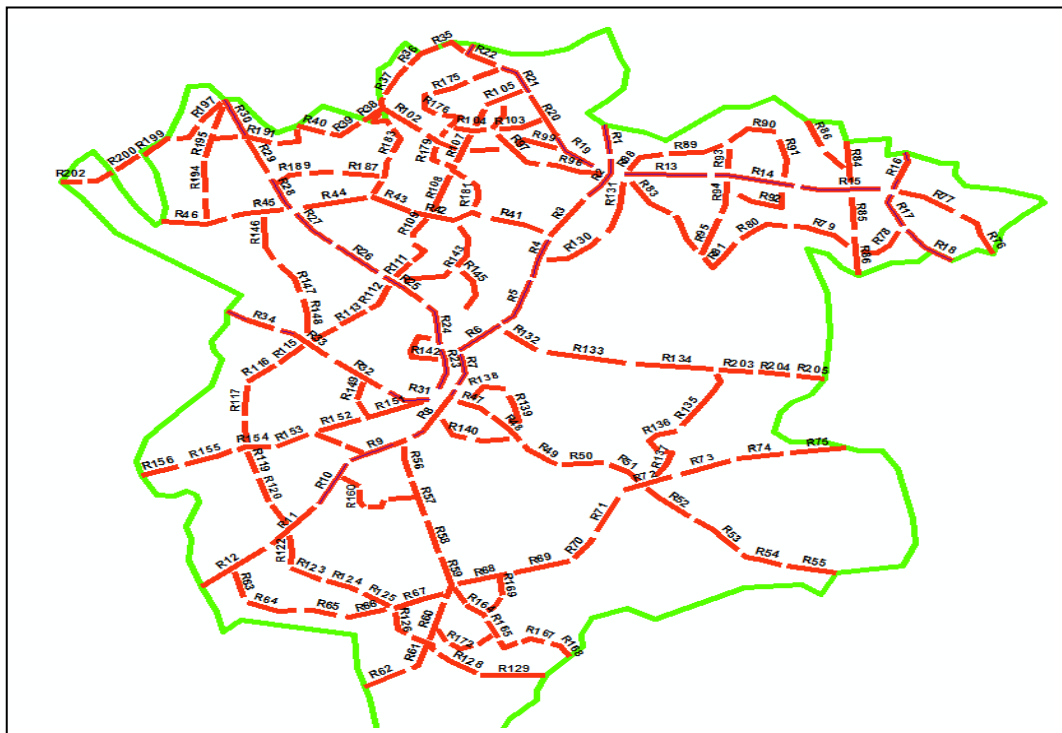


Figure4.5: Carte Map de M'sila et les différentes routes découpées

2-Algorithmme de datamining « règles d'association » pour les accidents routiers

Les accidents routiers en Algérie abattent les centaines des personnes chaque année, pour déterminer les causes de ses derniers on utilise un des outils de data Mining « La méthode règle d'association » qui nous convient à découvrir exactement les différents facteurs lors de leur réunion causent des accidents.

Pour cela on va former une base de données composée de deux bases une spatiale contenant les routes et les autre couches spatiales, l'autre est la base des statistiques contenant l'ensemble des accidents ont eu lieu les différentes routes, précisément les Rn routes de la wilaya de M'Sila comme échantillon.

Chaque accident a eu lieu une tranche Rm cette tranche a un voisinage exemple une montagne, une forêt, une mer, un lac etc. ... En plus de ses attributs spatiaux a comme attributs attributaires le type de véhicule léger, lourd, le chauffeur a un permis nouveau ou ancien, la vitesse est élevée ou normale, etc.

Comme on a plusieurs accidentons peut former une table de n ligne(accidents) et m colonnes (contexte de l'accident) pour ce faire on prépare la base spatiale.

La capture des images spatiale :

En premier lieu on va capturer une image satellitaire pour la wilaya de M'sila via GoogleArth. [Figure4.6]

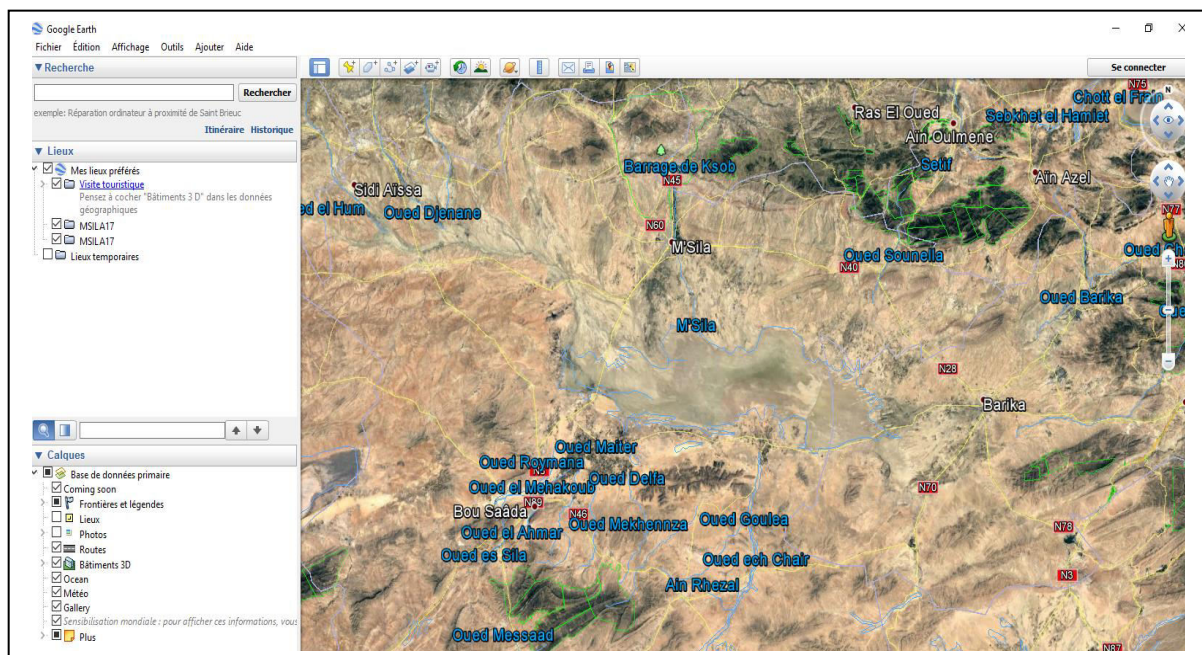


Figure4.6 Image via Google Earth région M'Sila [9]

4- L'exportation des images vers Arc Gis

On va exporter l'image capturée par Google Arth vers le logiciel Arc GIS pour construire la carte Mapet les shapfiles des différents objets de la base spatiale tels que les Rn routes qu'on va les appeler Routes partielles, les montagnes, les barrages, les mers, les lacs, les rivières, les villes et les forêts. [Figure4.7]

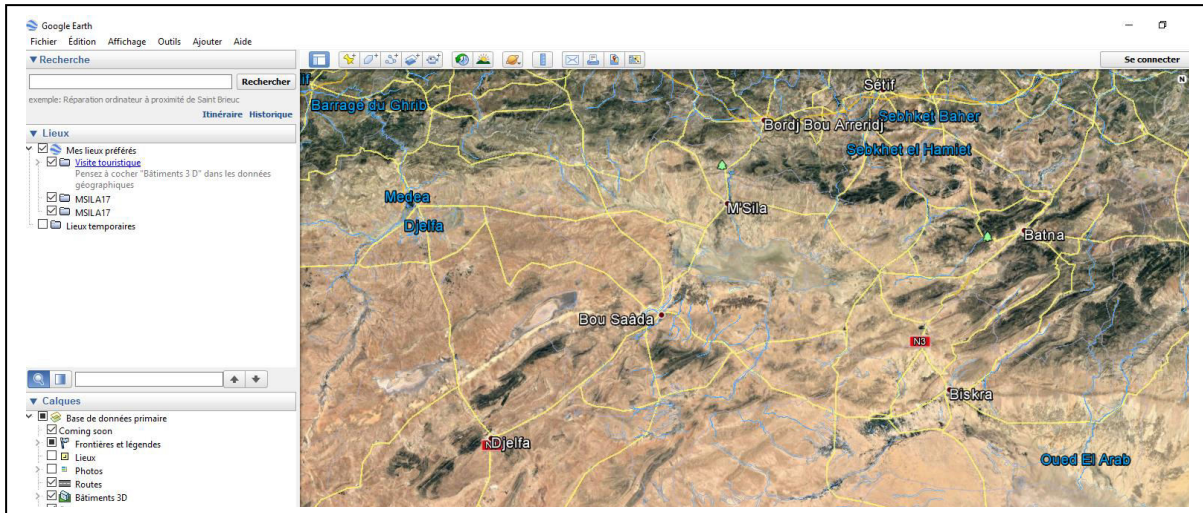


Figure4.7 Image via Google Earth région M'Sila [9]

5-La création de la carte Map avec Arc Catalogue

Avec l'outil Arc Catalogue on va ajouter à notre carte les différents shapfiles avec leur base de données (figure 8.4)

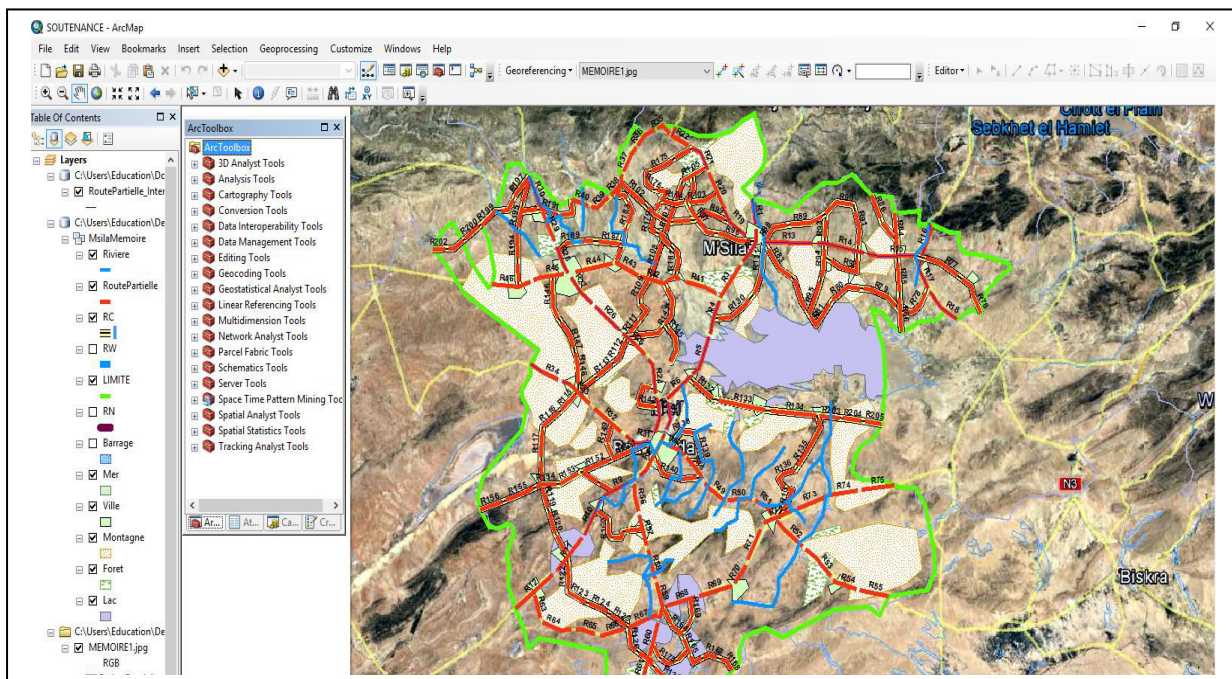


Figure4.8 Image via Google Earth région M'Sila importée à ArcGis [9]

Après cette opération on désactive la couche[layer] Memoire1 qui représente l'image spatiale de Google Earth on obtient la figure4.9

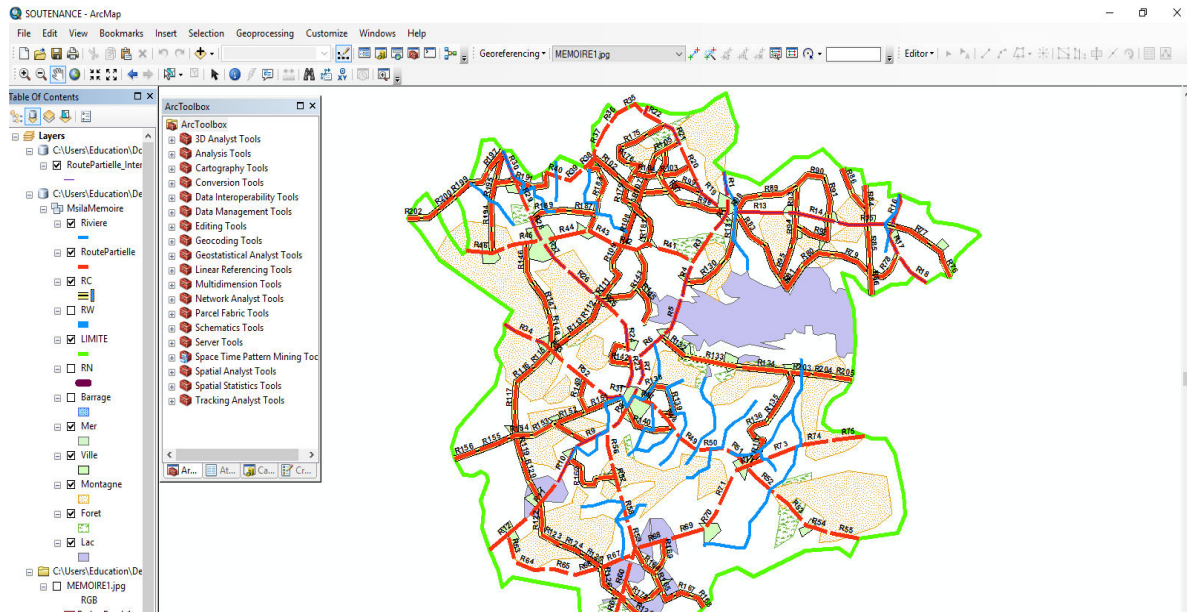
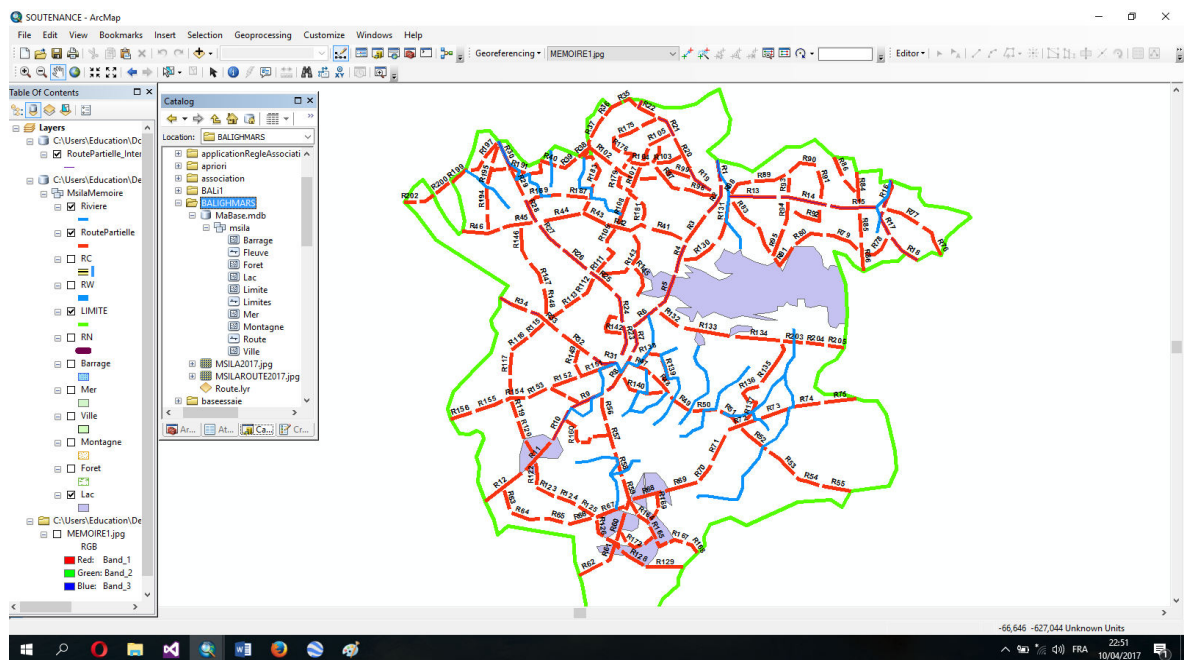


Figure4.9MAP région de M'sila sur ArcGis avec différents reliefs

On voit sur la figure4.9 les différentes couches sur la même carte qu'on peut les séparer, comme exemple on peut rassembler les routes partielles avec les rivières et les lacs seulement pour seulement pour ce faire en decoche sur (layer) les couches indésirables ; on voit aussi les bases de données de nos objets spatiaux sur Arc Catalogue. [Figure4.10]



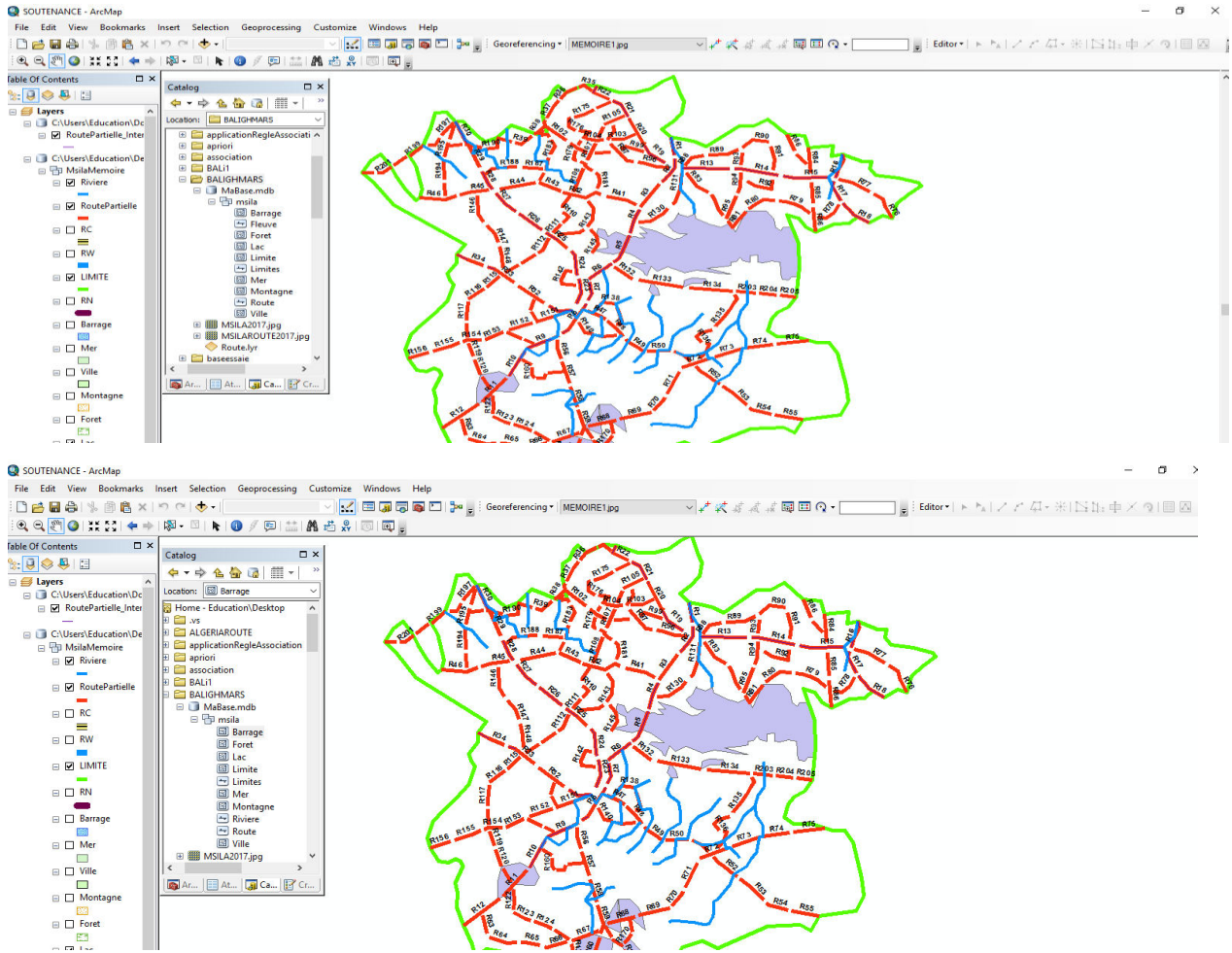


Figure4.10 Différentes bases de données sur Arc Catalogue

Maintenant voici deux échantillons de bases de données :

« RoutePartielles, Rivière »

OBJECTID *	SHAPE *	SHAPE_Length	IdRoutePartielle	NomRoutePartielle
12	Polyline	27,6500425	1	R1
13	Polyline	39,7992110	2	R2
14	Polyline	41,5878922	3	R3
15	Polyline	36,5419277	4	R4
17	Polyline	45,2630611	5	R5
18	Polyline	37,148883	6	R6
19	Polyline	29,218852	7	R7
20	Polyline	63,932813	8	R8
22	Polyline	45,0981184	9	R9
24	Polyline	42,272542	10	R10
25	Polyline	46,5641688	11	R11
30	Polyline	57,872896	12	R12
31	Polyline	58,872776	13	R13
42	Polyline	56,987224	14	R14
41	Polyline	57,229277	15	R15
43	Polyline	37,372744	16	R16
44	Polyline	31,784576	17	R17
45	Polyline	44,581344	18	R18
46	Polyline	37,929258	19	R19
47	Polyline	41,989314	20	R20
48	Polyline	31,899374	21	R21
49	Polyline	33,176988	22	R22
50	Polyline	38,585184	23	R23
52	Polyline	38,084931	24	R24
53	Polyline	38,467908	25	R25
54	Polyline	38,113519	26	R26
55	Polyline	37,188822	27	R27
56	Polyline	38,481431	28	R28
57	Polyline	34,539279	29	R29
58	Polyline	38,246168	30	R30
59	Polyline	32,01993	31	R31
60	Polyline	41,832188	32	R32
61	Polyline	41,682417	33	R33
62	Polyline	37,848968	34	R34
63	Polyline	20,120484	35	R35
64	Polyline	27,176927	36	R36
65	Polyline	29,563931	37	R37
66	Polyline	29,264078	38	R38
67	Polyline	26,308873	39	R39
68	Polyline	29,4438845	40	R40
69	Polyline	39,914556	41	R41
70	Polyline	52,352683	42	R42
71	Polyline	36,781283	43	R43
72	Polyline	45,872114	44	R44
73	Polyline	45,8188885	45	R45
74	Polyline	45,8188885	46	R46
75	Polyline	44,989891	47	R47
76	Polyline	38,632898	48	R48
77	Polyline	30,398298	49	R49
78	Polyline	39,721127	50	R50
79	Polyline	27,617471	51	R51
80	Polyline	49,851716	52	R52
81	Polyline	44,148989	53	R53
82	Polyline	44,148989	54	R54

Table4.1 Base de données routes partielles

OBJECTID *	SHAPE *	SHAPE_Length	IdRiviere	NomRiviere
1	Polyline	299,216857	17	RIV17
2	Polyline	138,409082	14	RIV14
3	Polyline	122,065061	13	RIV13
4	Polyline	80,381464	10	RIV10
5	Polyline	57,29504	11	RIV11
6	Polyline	136,763564	12	RIV12
7	Polyline	109,82142	9	RIV9
8	Polyline	135,827826	8	RIV8
9	Polyline	59,932985	2	RIV2
10	Polyline	61,711999	3	RIV3
11	Polyline	73,476079	1	RIV1
12	Polyline	100,401289	15	RIV15
13	Polyline	35,526393	16	RIV16
14	Polyline	75,049025	7	RIV7
15	Polyline	123,56564	6	RIV6
16	Polyline	46,60763	5	RIV5
17	Polyline	114,145685	4	RIV4
18	Polyline	123,341408	18	RIV18

Table4.2 Base de données Rivière

On va maintenant expliquer notre algorithme, comme tous savent l'accident a plusieurs causes une partie humaine comme l'excès de vitesse, le sommeil, le manque d'expérience de conduite etc. et l'autre est liée au reliefs et la nature de la région ,la qualité de la route etc., ...on essaie de regrouper et analyser ces causes dans une table virtuelle qui importe ses données via le lien Route Partielle(morceau découpé) entre les différentes tables à savoir les tables attributaires(causes humaines)et les tables spatiales(routes, montagne ,lac, ...)et extraire les règles d'association pour déterminer exactement les principales causes des accidents liées avec elles.

Dans cette situation on est bien dans un cas un peu compliqué pour extraire les connaissances on distingue deux approches.

6- Approches d'utilisation des bases de données spatiales multiples :

6.1-Regrouper toutes les tables dans une seule table qui est valable théoriquement ou d'une autre façon pour les tables de petite capacité.

Cette approche est très couteuse pour les base de grande capacité comme notre cas, la complexité de la fouille dans ce cas est très élevée, et aussi le problème de l'insuffisance de la mémoire pour traiter un cas pareil.

6.2-la deuxième approche et de laisser les tables dispersées et faire des liens entre elles et commencer le traitement.

On va détailler un peu notre algorithme en plusieurs étapes.

7-La construction des tables spatiales :

Débutant par la découverte des voisins et les autre relations comme les virages, les pentes de chaque route partielle, c'est-à-di refaire des requêtes pour chaque route partielle avec les autres couches pour avoir une table comme suit :

Id Route	R. Partielle	Voisin Montagne	Voisin Barrage	Voisine Rivière	Voisin LAC
1	R1	OUI	NON	OUI	OUI
2	R2	OUI	NON	NON	OUI
...	...	NON	NON	NON	OUI
N	Rn	NON	NON	NON	NON

VoisineForêt	VoisineVille	VoisineMer	RouteAvecVirage	RouteInclinée	RN	RW	RC	
OUI	NON	NON	OUI	NON	NON	OUI	NON	OUI
OUI	NON	NON	NON	OUI	NON	OUI	NON	OUI
NON	OUI	OUI	OUI	OUI	OUI	NON	NON	OUI
NON	OUI	OUI	NON	NON	OUI	NON	OUI	NON

Table4.3: Base de données spatiales et attributaires

Cela se fait via le logiciel ARC GIS on va construire les bases de données de chaque tranche Rm avec les autres couches voisines (forêt, montagne, barrage, agglomération, etc.). On peut sélectionner par exemple les noms des routes Rm qui traversent des montagnes, ce qui nous permet de remplir l'attribut **voisin Montagne** par oui(O) ou non (N).

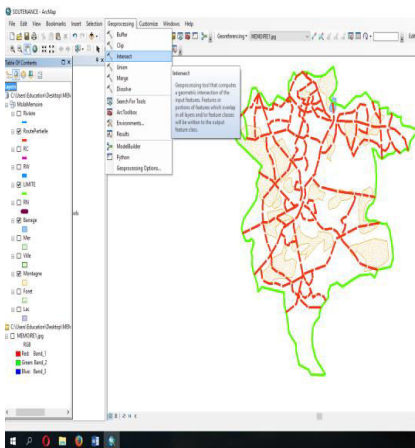


Figure4.11Routes intersection sur ArcGis

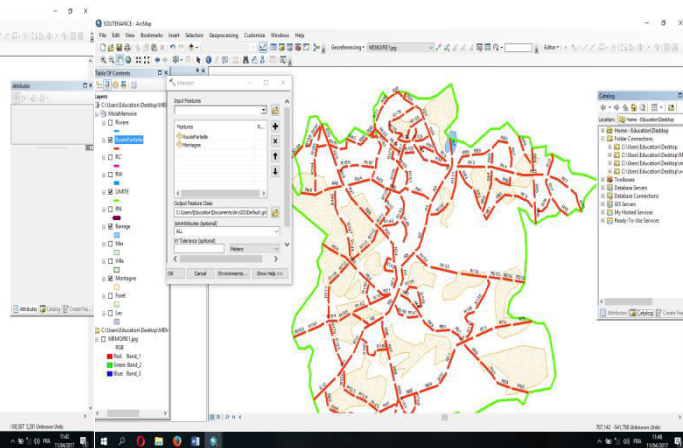


Figure4.12Routes intersection sur ArcGis

La requête se fait comme suit :

- 1-Geoprocessing
- 2-intersect
- 3-choisir les couches concernées (Route Partielle et Montagne)
- 4-Nommer la requête Route Partielle_Intersect_Montagne

Le résultat est :

- a) La création d'un nouveau Layer

RoutePartielle_Intersect_Montagne. Figure4.13

- b) L'ouverture de la table "RoutePartielle_Intersect_Montagne" nous donne les noms des routes partielles et les noms des montagnes qui ont une intersection. Table4.4

Exemple : R3 INTERSECT MON3

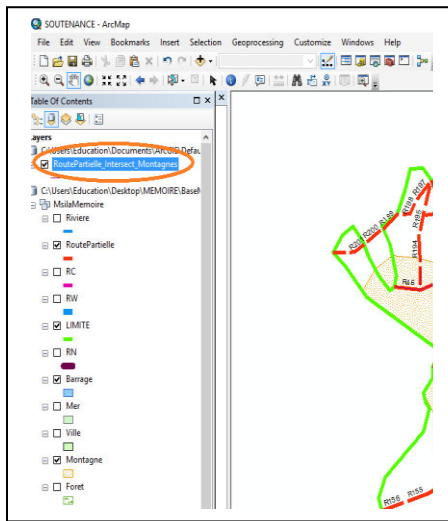


Figure4.13 Routes « intersection sur ArcGIS »

OBJETID	SHAPE	FID RoutePartielle	IdRoutePartielle	NomRoutePartielle	FID Montagne	SomMontagne	IdMontagne	SHAPE_Length
1	Polyline	40	13	R13	14	MON3	3	20,368363
2	Polyline	130	92	R92	14	MON3	3	20,872552
3	Polyline	132	94	R94	14	MON3	3	0,250454
4	Polyline	132	94	R94	14	MON3	3	4,44923
5	Polyline	171	120	R120	4	MON3	3	4,987353
6	Polyline	180	138	R138	3	MON7	7	16,559563
7	Polyline	217	175	R175	5	MON2	2	14,64234
8	Polyline	218	175	R175	5	MON2	2	25,210416
9	Polyline	219	177	R177	5	MON2	2	13,589729
10	Polyline	14	15	R15	4	MON3	3	8,437719
11	Polyline	42	15	R15	14	MON3	3	21,262448
12	Polyline	53	25	R25	8	MON4	4	29,967087
13	Polyline	54	26	R26	8	MON4	4	31,135109
14	Polyline	55	27	R27	8	MON4	4	17,452962
15	Polyline	61	33	R33	7	MON5	5	41,882417
16	Polyline	62	34	R34	7	MON5	5	33,362963
17	Polyline	74	43	R43	7	MON5	5	8,272795
18	Polyline	76	46	R46	7	MON5	5	46,969891
19	Polyline	83	53	R53	9	MON13	13	5,775946
20	Polyline	84	53	R53	9	MON13	13	18,374584
21	Polyline	86	54	R54	9	MON13	13	7,818135
22	Polyline	86	55	R55	9	MON13	13	23,259275
23	Polyline	87	56	R56	1	MON11	11	2,986214
24	Polyline	96	67	R67	1	MON11	11	28,54489
25	Polyline	96	63	R63	13	MON12	12	13,423817
26	Polyline	109	74	R74	10	MON9	9	27,918599
27	Polyline	110	75	R75	10	MON9	9	40,969691
28	Polyline	115	79	R79	14	MON3	3	7,504443
29	Polyline	116	80	R80	14	MON3	3	26,186753
30	Polyline	120	84	R84	14	MON3	3	25,869691
31	Polyline	124	87	R87	14	MON3	3	29,18823
32	Polyline	132	94	R94	14	MON3	3	13,231697
33	Polyline	133	95	R95	14	MON3	3	8,153516
34	Polyline	152	111	R111	8	MON4	4	1,571296
35	Polyline	153	112	R112	8	MON4	4	26,49679
36	Polyline	154	113	R113	8	MON4	4	21,935028
37	Polyline	165	114	R114	7	MON5	5	7,772821
38	Polyline	165	114	R114	8	MON4	4	3,015351
39	Polyline	166	115	R115	7	MON5	5	33,398325
40	Polyline	167	116	R116	7	MON5	5	25,839
41	Polyline	168	117	R117	7	MON5	5	3,472254

Table4.4 Nouvelle base de données créée intersection

Pour le voisinage exemple Les routes partielles qui ont des lacs comme voisins, on lance la requête comme suit :

- 1- Selection
- 2- Select by Location
- 3- Cocher Route Partielle
- 4- Cocher Lac
- 5- Déterminer la distance de voisinage

Après la validation on voit les Routes qui ont un voisinage LAC de distance inférieure à 4 Km change sa couleur du rouge au vert sur la carte figure4.14, et seront sélectionnées au niveau de la base de données Table4.5.

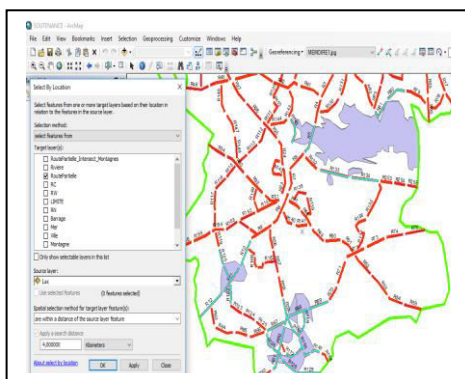


Figure4.14 Routes « voisinage »

OBJETID	SHAPE	SHAPE_Length	IdRoutePartielle	NomRoutePartielle
17	Polyline	46,263041	6	R6
18	Polyline	37,148883	6	R6
19	Polyline	26,218864	6	R6
20	Polyline	75,832813	6	R6
21	Polyline	44,696104	6	R6
22	Polyline	44,696104	6	R6
23	Polyline	46,564198	11	R11
24	Polyline	62,973696	16	R16
25	Polyline	59,677795	13	R13
26	Polyline	56,895244	14	R14
27	Polyline	37,176276	16	R16
28	Polyline	37,176276	17	R17
29	Polyline	44,581344	16	R16
30	Polyline	37,626268	16	R16
31	Polyline	37,176276	16	R16
32	Polyline	41,880314	20	R20
33	Polyline	41,880314	21	R21
34	Polyline	38,885194	23	R23
35	Polyline	38,885194	23	R23
36	Polyline	38,885194	23	R23
37	Polyline	38,885194	23	R23
38	Polyline	38,885194	23	R23
39	Polyline	38,885194	23	R23
40	Polyline	38,885194	23	R23
41	Polyline	41,632199	32	R32
42	Polyline	41,632199	32	R32
43	Polyline	37,176276	36	R36
44	Polyline	37,176276	36	R36
45	Polyline	37,176276	36	R36
46	Polyline	37,176276	36	R36
47	Polyline	37,176276	36	R36
48	Polyline	37,176276	36	R36
49	Polyline	37,176276	36	R36
50	Polyline	37,176276	36	R36
51	Polyline	37,176276	36	R36
52	Polyline	37,176276	36	R36
53	Polyline	37,176276	36	R36
54	Polyline	37,176276	36	R36
55	Polyline	37,176276	36	R36
56	Polyline	37,176276	36	R36
57	Polyline	37,176276	36	R36
58	Polyline	37,176276	36	R36
59	Polyline	37,176276	36	R36
60	Polyline	37,176276	36	R36
61	Polyline	37,176276	36	R36
62	Polyline	37,176276	36	R36
63	Polyline	37,176276	36	R36
64	Polyline	37,176276	36	R36
65	Polyline	37,176276	36	R36
66	Polyline	37,176276	36	R36
67	Polyline	37,176276	36	R36
68	Polyline	37,176276	36	R36
69	Polyline	37,176276	36	R36
70	Polyline	37,176276	36	R36
71	Polyline	37,176276	36	R36
72	Polyline	37,176276	36	R36
73	Polyline	37,176276	36	R36
74	Polyline	37,176276	36	R36
75	Polyline	37,176276	36	R36
76	Polyline	37,176276	36	R36
77	Polyline	37,176276	36	R36
78	Polyline	37,176276	36	R36
79	Polyline	37,176276	36	R36
80	Polyline	37,176276	36	R36

Table4.5 base de données créée « voisinage »

Dernièrement pour les routes possédant des virages, on fait une sélection des routes qui ont une distance supérieure à la distance directe (vol d’oiseau), et pour les route possédant une pente on fait une section des routes qui ont une altitude de départ supérieure à l’altitude à la fin.

La deuxième phase la concaténation des champs Rm avec les autres couches possédant une relation de voisinage ou intersection etc.,à savoir Voisin Montagne, Voisin Lac, RN,RW,RC

Pour se faire on procède à effectuer la requête de jointure dans ArcGis« RELATE » comme suit :

1-sélectionner Route Partielle, bouton droit

2-joins and relates

3-Relates

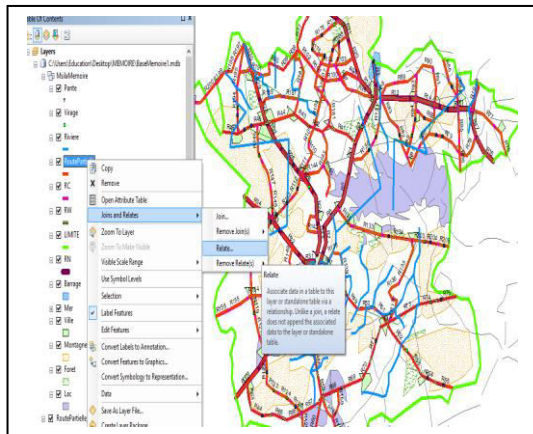


Figure 4.15: La jointure et relate des tables

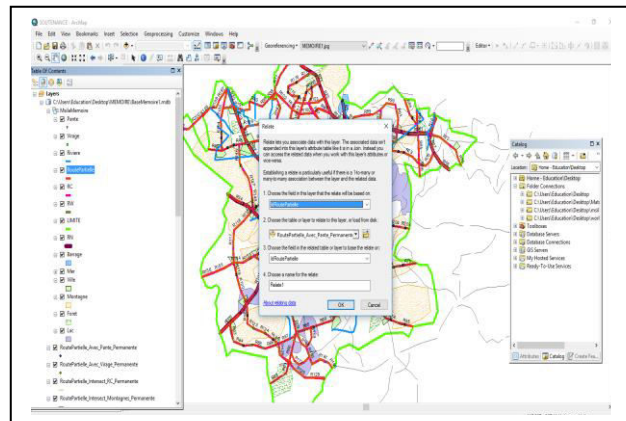


Figure4.16: La jointure et relate des tables

On fait cette opération avec toutes les couches on obtient des nouvelles couches qu’on va les exporter pour les mettre dans une base permanente, pourcela clique droit sur la couche créée, Data, puis Exporte Data comme il est montré sur la figure4.17

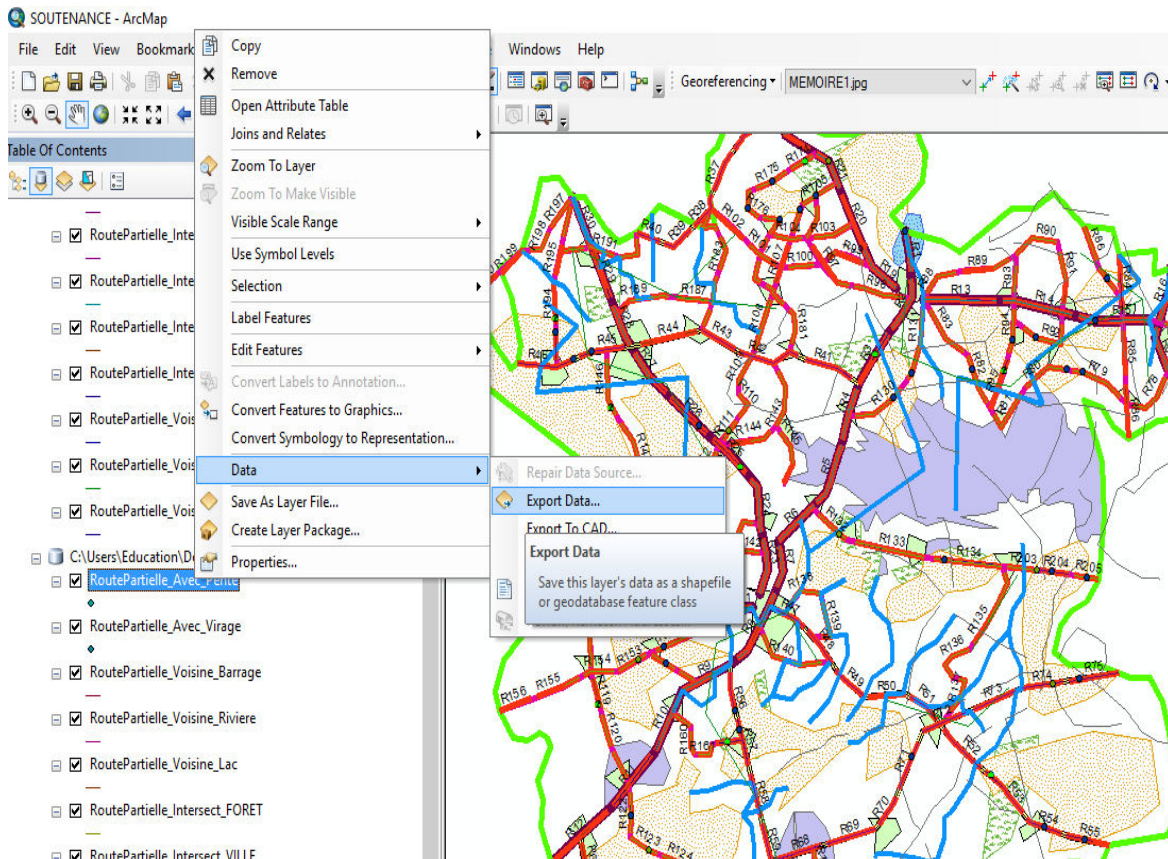


Figure4.17: La jointure et relate des table

La nouvelle base obtenue est une base contenant des nouvelles tables générées par des requêtes topologiques et métrique ce qu'on ne le trouve pas dans les requêtes classiques SQL, c'est-à-dire chaque table contient uniquement les routes partielles qui a une relation topologique ou métrique avec une des
 Autres couches, par exemple la table (Route_Partielle_Intersect_Forêt) c'est une table qui contient les m Routes partielles qui traversent des forêts, cette spécification conduit à une connaissance connue par rapport à un facteur majeur des accidents qui est le découpage de la route par les animaux, le stationnement des passagers au bord de la route ,etc.

Les nouvelles tables générées comme le montre la figure4.18

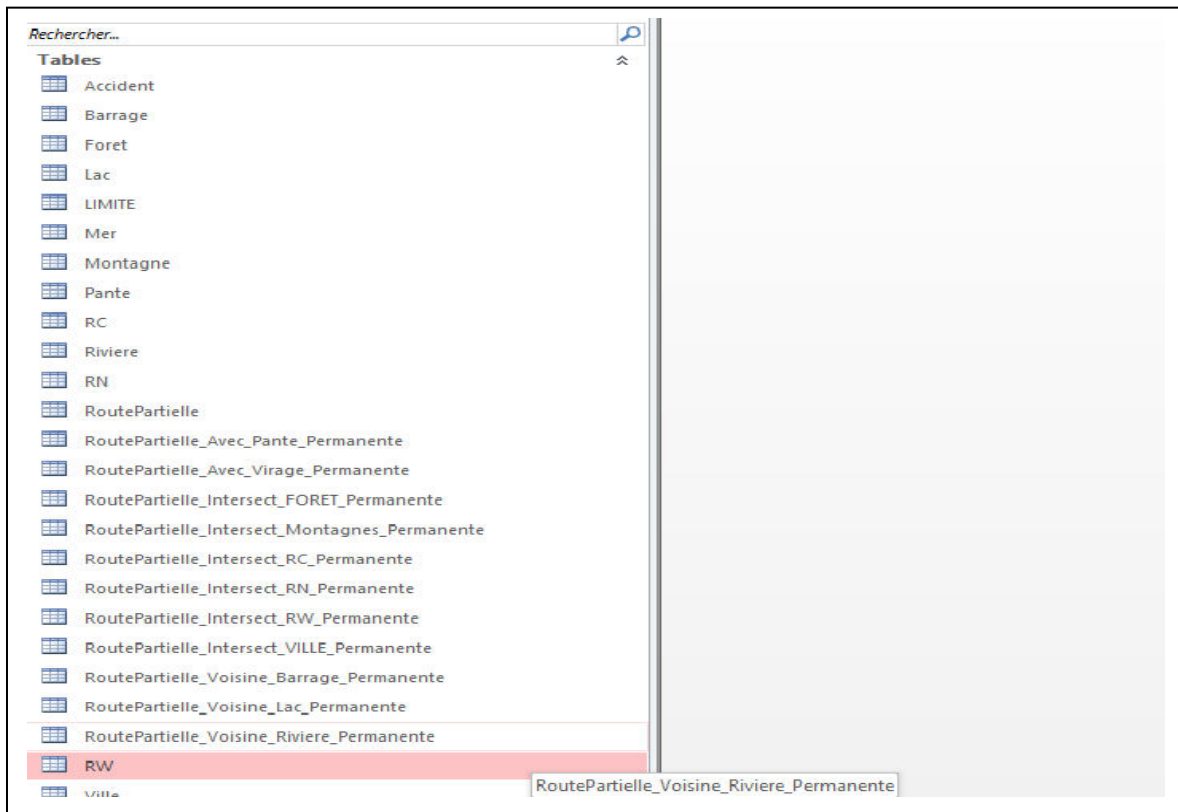


Figure4.18 L'affichage des différentes tables jointées

Voici le détail de la table special Route_Partielle_Intersect_foret

RoutePartielle_Intersect_FORET_Permanente									
OBJECTID	SHAPE	FID_RoutePartie	IdRoutePartielle	NomRoutePartiel	FID_Foret	IdForet	NomForet	SHAPE	
1	lignée binaire	69	41	R41	6	3	FOR3	13,171	
2	lignée binaire	156	115	R115	9	5	FOR5	11,449	
3	lignée binaire	174	133	R133	3	6	FOR6	17,080	
4	lignée binaire	175	134	R134	4	7	FOR7	6,7686	
5	lignée binaire	14	3	R3	6	3	FOR3	22,68	
6	lignée binaire	48	21	R21	7	1	FOR1	26,12	
7	lignée binaire	56	28	R28	8	2	FOR2	5,7667	
8	lignée binaire	57	29	R29	8	2	FOR2	7,4649	
9	lignée binaire	61	33	R33	9	5	FOR5	22,262	
10	lignée binaire	62	34	R34	9	5	FOR5	37,948	
11	lignée binaire	84	53	R53	2	9	FOR9	33,082	
12	lignée binaire	90	57	R57	1	8	FOR8	28,246	
13	lignée binaire	94	61	R61	10	10	FOR10	25,631	

Table4.6 La table spatiale route partielle intersection forêt

8- La construction de la table spatiale et attributaires :

Cette étape consiste à collecter les différents contextes de l'accident qui a eu lieu une route partielle donnée tels que le type de véhicule, le permis du chauffeur nouveau ou ancien et les reliefs de cette route partielle à savoir les forêts, les montagne, etc....

Ses statistiques peuvent être ramener des sources officielles telles que les services sécuritaires, la protection civile, les services de la santé.

On appris les Routes Partielles de la wilaya de M'Sila qu' on a préparé précédemment dans Arc Gis (statistiques non officielles les données sont remplies arbitrairement) .

Les autres tables c'est des tables spatiales ramenées d'Arc Gis comme exemple la Table4.7

Table4.7: La table attributaire accident

Accident										
IdAccident	IdRoutePart	NomRouteP	Permis	TempsAccid	TypeVihecu	Vitesse	EtatVihecu	Pluit	verglas	Cliquer pour ajo
1	1 R1		Nouveau	Jour	Leger	Elevee	Neuf	Oui	Oui	
2	1 R1		Ancien	Nuit	Lourd	Normale	Neuf	Non	Non	
3	1 R1		Nouveau	Nuit	Lourd	Normale	Neuf	Non	Non	
4	1 R1		Nouveau	Nuit	Lourd	Normale	Neuf	Non	Non	
5	1 R1		Nouveau	Nuit	Lourd	Normale	Neuf	Non	Non	
6	7 R7		Nouveau	Nuit	Leger	Elevee	Vieu	Non	Non	
7	8 R8		Nouveau	Jour	Leger	Normale	Vieu	Non	Non	
8	8 R8		Nouveau	Jour	Leger	Normale	Vieu	Non	Non	
9	6 R6		Nouveau	Nuit	Leger	Elevee	Neuf	Oui	Oui	
10	6 R6		Ancien	Nuit	Leger	Elevee	Neuf	Oui	Oui	
11	6 R6		Ancien	Nuit	Lourd	Elevee	Vieu	Oui	Oui	
12	6 R6		Ancien	Nuit	Leger	Elevee	Neuf	Oui	Oui	
13	6 R6		Ancien	Jour	Lourd	Elevee	Neuf	Oui	Non	
14	9 R9		Ancien	Jour	Lourd	Elevee	Neuf	Oui	Non	
15	9 R9		Nouveau	Jour	Lourd	Elevee	Neuf	Non	Oui	

9-La jointure spatiale des tables

On remarque que toutes les tables ont comme champs commun Id_ Route_ Partielle reliant ces tables avec elles, en faisant une requête SQL de jointure entre les table avec le champs commun Id_ Route_ Partielle :

```

SELECT Accident.Permis, Accident.TempsAccident, Accident.TypeVihecule,
Accident.Vitesse,
Accident.EtatVihecule, Accident.Pluit, Accident.Verglas,

iif(isnull(RoutePartielle_Avec_Pante_Permanente.NomRoutePartielle) , 'Non', 'Oui') AS
RavecPente,
iif(isnull(RoutePartielle_Avec_Virage_Permanente.IdVirage) , 'Non', 'Oui') as Virage,
iif(isnull(RoutePartielle_Intersect_FORET_Permanente.IdForet) , 'Non', 'Oui') as
Foret,
iif(isnull(RoutePartielle_Intersect_RW_Permanente.IdRW) , 'Non', 'Oui')as Rwilaya,
iif(isnull(RoutePartielle_Voisine_Lac_Permanente.IdLac) , 'Non', 'Oui')as Lac,
iif(isnull(RoutePartielle_Voisine_Barrage_Permanente.IdBarrage) , 'Non', 'Oui') as
Barrage,
iif(isnull(RoutePartielle_Intersect_Montagnes_Permanente.IdMontagne) , 'Non', 'Oui')as
Montagne,
iif(isnull(RoutePartielle_Voisine_Riviere_Permanente.IdRiviere ) , 'Non', 'Oui')as
Riviere,
iif(isnull(RoutePartielle_Intersect_RN_Permanente.IdRN) , 'Non', 'Oui')as RNationale,
iif(isnull(RoutePartielle_Intersect_VILLE_Permanente.IdVille) , 'Non', 'Oui')as
Ville,
iif(isnull(RoutePartielle_Intersect_RC_Permanente.IdRC) , 'Non', 'Oui') as Rcomunale

FROM ((((((((((Accident LEFT JOIN RoutePartielle_Avec_Pante_Permanente ON
Accident.IdRoutePartielle =
RoutePartielle_Avec_Pante_Permanente.IdRoutePartielle) LEFT JOIN
RoutePartielle_Avec_Virage_Permanente ON
Accident.IdRoutePartielle = RoutePartielle_Avec_Virage_Permanente.IdRoutePartielle)
LEFT JOIN RoutePartielle_Intersect_FORET_Permanente ON
Accident.IdRoutePartielle =
RoutePartielle_Intersect_FORET_Permanente.IdRoutePartielle) LEFT JOIN
RoutePartielle_Intersect_Montagnes_Permanente ON Accident.IdRoutePartielle =
RoutePartielle_Intersect_Montagnes_Permanente.IdRoutePartielle)
LEFT JOIN RoutePartielle_Intersect_RC_Permanente ON Accident.IdRoutePartielle =
RoutePartielle_Intersect_RC_Permanente.IdRoutePartielle)
LEFT JOIN RoutePartielle_Intersect_RN_Permanente ON Accident.IdRoutePartielle =
RoutePartielle_Intersect_RN_Permanente.IdRoutePartielle)
LEFT JOIN RoutePartielle_Intersect_RW_Permanente ON Accident.IdRoutePartielle =
RoutePartielle_Intersect_RW_Permanente.IdRoutePartielle)
LEFT JOIN RoutePartielle_Intersect_VILLE_Permanente ON Accident.IdRoutePartielle =
RoutePartielle_Intersect_VILLE_Permanente.IdRoutePartielle)
LEFT JOIN RoutePartielle_Voisine_Barrage_Permanente ON Accident.IdRoutePartielle =
RoutePartielle_Voisine_Barrage_Permanente.IdRoutePartielle)
LEFT JOIN RoutePartielle_Voisine_Lac_Permanente ON Accident.IdRoutePartielle =
RoutePartielle_Voisine_Lac_Permanente.IdRoutePartielle)
LEFT JOIN RoutePartielle_Voisine_Riviere_Permanente ON Accident.IdRoutePartielle =
RoutePartielle_Voisine_Riviere_Permanente.IdRoutePartielle;

```

”;

On obtient une table globale contenant tout les accidents avec leurs attributs statistiques et spatiaux comme le montre la Table 4.8

N Accident	description
<input type="checkbox"/> 1	Permis Nouveau # TempsAccident Jour # TypeVehicule Leger # Vitesse Elevee # EtatVehicule Neuf # Pluit Oui # Verglas Oui # RavecPente Oui # Virage Oui # Barrage Oui
<input type="checkbox"/> 2	Permis Ancien # TempsAccident Nuit # TypeVehicule Lourd # Vitesse Normale # EtatVehicule Neuf # RavecPente Oui # Virage Oui # Barrage Oui # RNationale Oui #
<input type="checkbox"/> 3	Permis Nouveau # TempsAccident Nuit # TypeVehicule Lourd # Vitesse Normale # EtatVehicule Neuf # RavecPente Oui # Virage Oui # Barrage Oui # RNationale Oui #
<input type="checkbox"/> 4	Permis Nouveau # TempsAccident Nuit # TypeVehicule Lourd # Vitesse Normale # EtatVehicule Neuf # RavecPente Oui # Virage Oui # Barrage Oui # RNationale Oui #
<input type="checkbox"/> 5	Permis Nouveau # TempsAccident Nuit # TypeVehicule Lourd # Vitesse Normale # EtatVehicule Neuf # RavecPente Oui # Virage Oui # Barrage Oui # RNationale Oui #
<input type="checkbox"/> 6	Permis Nouveau # TempsAccident Nuit # TypeVehicule Leger # Vitesse Elevee # EtatVehicule Vieu # RNationale Oui #
<input type="checkbox"/> 7	Permis Nouveau # TempsAccident Jour # TypeVehicule Leger # Vitesse Normale # EtatVehicule Vieu # Riviere Oui # RNationale Oui # Ville Oui #
<input type="checkbox"/> 8	Permis Nouveau # TempsAccident Jour # TypeVehicule Leger # Vitesse Normale # EtatVehicule Vieu # Riviere Oui # RNationale Oui # Ville Oui #
<input type="checkbox"/> 9	Permis Nouveau # TempsAccident Nuit # TypeVehicule Leger # Vitesse Elevee # EtatVehicule Neuf # Pluit Oui # Verglas Oui # RNationale Oui # Ville Oui #
<input type="checkbox"/> 10	Permis Ancien # TempsAccident Nuit # TypeVehicule Leger # Vitesse Elevee # EtatVehicule Neuf # Pluit Oui # Verglas Oui # RNationale Oui # Ville Oui #
<input type="checkbox"/> 11	Permis Ancien # TempsAccident Nuit # TypeVehicule Lourd # Vitesse Elevee # EtatVehicule Vieu # Pluit Oui # Verglas Oui # RNationale Oui # Ville Oui #
<input type="checkbox"/> 12	Permis Ancien # TempsAccident Nuit # TypeVehicule Leger # Vitesse Elevee # EtatVehicule Neuf # Pluit Oui # Verglas Oui # RNationale Oui # Ville Oui #
<input type="checkbox"/> 13	Permis Ancien # TempsAccident Jour # TypeVehicule Lourd # Vitesse Elevee # EtatVehicule Neuf # Pluit Oui # RNationale Oui # Ville Oui #
<input type="checkbox"/> 14	Permis Ancien # TempsAccident Jour # TypeVehicule Lourd # Vitesse Elevee # EtatVehicule Neuf # Pluit Oui # Riviere Oui # RNationale Oui #
<input type="checkbox"/> 15	Permis Nouveau # TempsAccident Jour # TypeVehicule Lourd # Vitesse Elevee # EtatVehicule Neuf # Verglas Oui # Riviere Oui # RNationale Oui #

Table4.8La table attributaire et spatiale accident

Par exemple pour le premier accident le chauffeur a un nouveau permis, se fait la nuit, le type de véhicule Leger, la vitesse été élevée ,l'état de véhicule neuf, il y'avait de la Pluie et le verglas plus les attributs spatiaux la route est avec pente, il y'a des virages, il y'a un barrage a cote de cette route ,c'est une route nationale etc.

La même chose pour les autre accidents ses attributs on les considère comme items pour extraire les règles d'association.

10-L'exportation des bases spatiales vers C Sharp pour extraire les règles d'association.

Passant maintenant à l'extraction des règles d'association, comme on a vu précédemment près l'exécution de notre requête qui a comme table (Route _ Partielle)et les autres tables à savoir :

(Route Partielle_ Avec _Virage_ Permanente .Id Virage)

(Route Partielle _Intersect_ FORET _Permanente. IdForêt)

(Route Partielle _Intersect_ RW _Permanente .IdRW)

(Route Partielle _Voisine _Lac _Permanente .IdLac)

(Route Partielle _Voisine _Barrage _Permanente .IdBarrage)

(Route Partielle _Intersect _Montagnes _Permanente .IdMontagne)

(Route Partielle _Voisine_Riviere_Permanente .IdRivière)

(Route Partielle _Intersect _RN_Permanente .IdRN)

(Route Partielle _Intersect _VILLE_Permanente .IdVille)

(Route Partielle _Intersect _RC_Permanente .IdRC)

Les résultats de cette requête nous fournissent une table riche d'information sur les circonstances de chaque accident à savoir les capacités du chauffeur l'état de véhicule les différents reliefs du lieu de l'accident etc. Chaque attribut d'un accident on le prend comme un item et la réunion de ces attributs comme un n-Items.

Pour cela on a utilisé le langage c-Sharp pour développer une application faisant ce traitement bien sûr avec prise en charge des différentes optimisations fournies par l'algorithme Apriori.

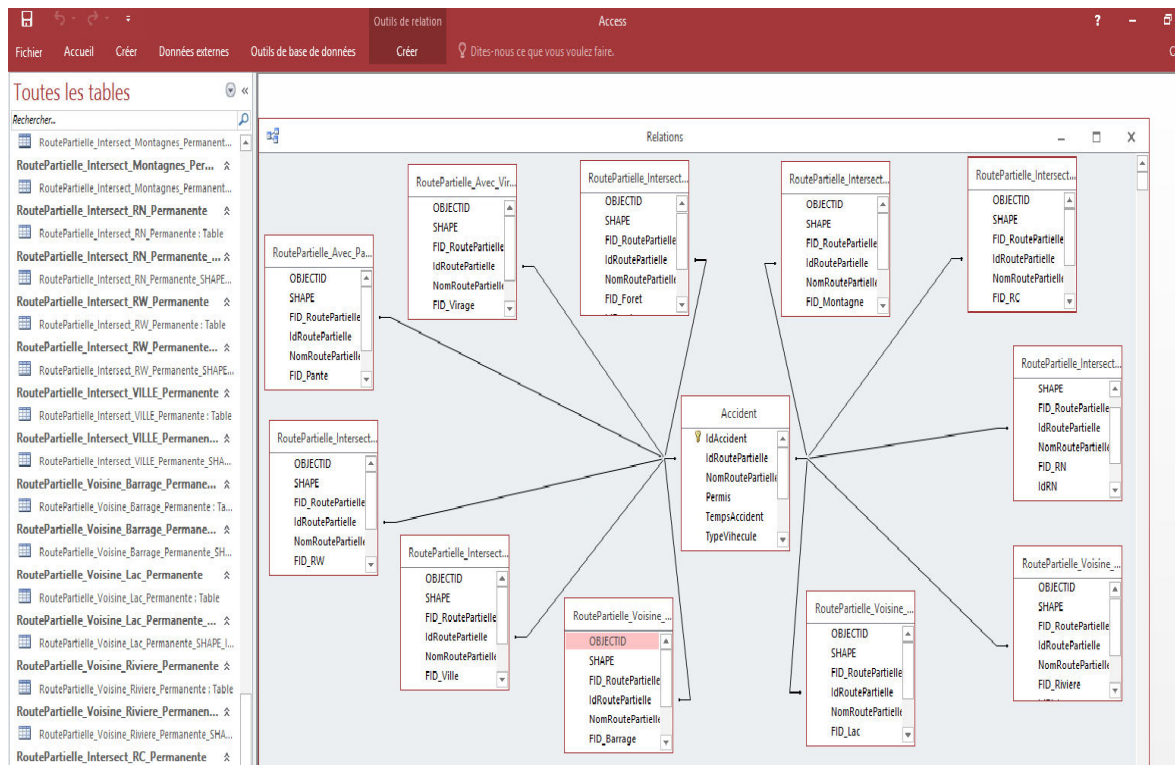


Figure4.19L'affichage des différentes tables jointée avec la table Accident sur Access

Voici les différentes tables reliées avec elles via le lien Id_Route_Partielle, après la requête de jointure on obtient une table comme suit :

Liste Des Accidents			
N Accident	Items	dic	description
<input type="checkbox"/> 1	a,e,t,u,o,q,s,d,f,g,j	aetuoqsdgjf	Permis Nouveau # TempsAccident Jour # TypeVihecule Leger # Vitesse Elevee # EtatVihecule Neuf # Pluit
<input type="checkbox"/> 2	z,r,y,i,o,d,f,g,j	zryiodfgj	Permis Ancien # TempsAccident Nuit # TypeVihecule Lourd # Vitesse Normale # EtatVihecule Neuf # Rave
<input type="checkbox"/> 3	a,r,y,i,o,d,f,g,j	aryiodfgj	Permis Nouveau # TempsAccident Nuit # TypeVihecule Lourd # Vitesse Normale # EtatVihecule Neuf # Rav
<input type="checkbox"/> 4	a,r,y,i,o,d,f,g,j	aryiodfgj	Permis Nouveau # TempsAccident Nuit # TypeVihecule Lourd # Vitesse Normale # EtatVihecule Neuf # Rav
<input type="checkbox"/> 5	a,r,y,i,o,d,f,g,j	aryiodfgj	Permis Nouveau # TempsAccident Nuit # TypeVihecule Lourd # Vitesse Normale # EtatVihecule Neuf # Rav
<input type="checkbox"/> 6	a,r,t,u,p,j	artupj	Permis Nouveau # TempsAccident Nuit # TypeVihecule Leger # Vitesse Elevee # EtatVihecule Vieu # RNati
<input type="checkbox"/> 7	a,e,t,i,p,h,j,k	aetiphjk	Permis Nouveau # TempsAccident Jour # TypeVihecule Leger # Vitesse Normale # EtatVihecule Vieu # Riv
<input type="checkbox"/> 8	a,e,t,i,p,h,j,k	aetiphjk	Permis Nouveau # TempsAccident Jour # TypeVihecule Leger # Vitesse Normale # EtatVihecule Vieu # Riv
<input type="checkbox"/> 9	a,r,t,u,o,q,s,j,k	artuoqsjk	Permis Nouveau # TempsAccident Nuit # TypeVihecule Leger # Vitesse Elevee # EtatVihecule Neuf # Pluit
<input type="checkbox"/> 10	z,r,t,u,o,q,s,j,k	zrtuoqsjk	Permis Ancien # TempsAccident Nuit # TypeVihecule Leger # Vitesse Elevee # EtatVihecule Neuf # Pluit O
<input type="checkbox"/> 11	z,r,y,u,p,q,s,j,k	zryupqsjk	Permis Ancien # TempsAccident Nuit # TypeVihecule Lourd # Vitesse Elevee # EtatVihecule Vieu # Pluit O
<input type="checkbox"/> 12	z,r,t,u,o,q,s,j,k	zrtuoqsjk	Permis Ancien # TempsAccident Nuit # TypeVihecule Leger # Vitesse Elevee # EtatVihecule Neuf # Pluit O
<input type="checkbox"/> 13	z,e,y,u,o,q,h,j	zeyuoqjh	Permis Ancien # TempsAccident Jour # TypeVihecule Lourd # Vitesse Elevee # EtatVihecule Neuf # Pluit O
<input type="checkbox"/> 14	z,e,y,u,o,q,h,j	zeyuoqjh	Permis Ancien # TempsAccident Jour # TypeVihecule Lourd # Vitesse Elevee # EtatVihecule Neuf # Pluit O
<input type="checkbox"/> 15	a,e,y,u,o,s,h,j	aeyuoshj	Permis Nouveau # TempsAccident Jour # TypeVihecule Lourd # Vitesse Elevee # EtatVihecule Neuf # Vergl

Figure4.20 ListView de l'ensemble des accidents avec attributs

Voici l'interface principal de notre application qui contient :

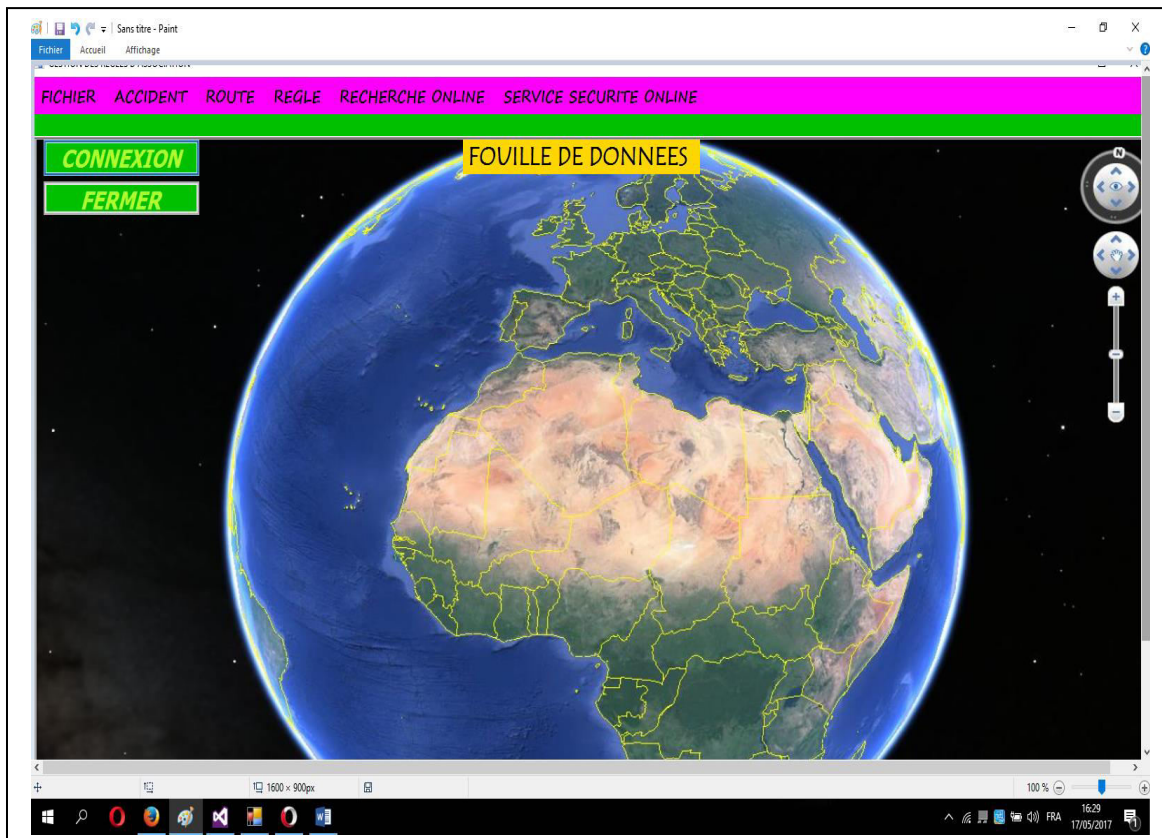


Figure4.21Forme principale de l'application

Le bouton Connexion :

- 1- Pour choisir la base de données à analyser
- 2- Pour s'identifier

Information de connexion

Base Donnees : Parcourir..

Administrateur :

Mot de passe :



Connexion  Annuler 

Figure4.22Forme d'identification

Le bouton Accident pour :

- 1- afficher la liste des accidents.
- 2- ajouter, supprimer, modifier un accident.

IdAccident	IdRoutePartielle	NomRoutePartielle	Permis	TempsAccident	TypeVehicule	Vitesse	EtatVehicule	Pluit	Verglas
1	1	R1	Nouveau	Jour	Leger	Elevee	Neuf	Oui	Oui
2	1	R1	Ancien	Nuit	Lourd	Normale	Neuf	Non	Non
3	1	R1	Nouveau	Nuit	Lourd	Normale	Neuf	Non	Non
4	1	R1	Nouveau	Nuit	Lourd	Normale	Neuf	Non	Non
5	1	R1	Nouveau	Nuit	Lourd	Normale	Neuf	Non	Non
6	7	R7	Nouveau	Nuit	Leger	Elevee	Vieu	Non	Non
7	8	R8	Nouveau	Jour	Leger	Normale	Vieu	Non	Non
8	8	R8	Nouveau	Jour	Leger	Normale	Vieu	Non	Non
9	6	R6	Nouveau	Nuit	Leger	Elevee	Neuf	Oui	Oui
10	6	R6	Ancien	Nuit	Leger	Elevee	Neuf	Oui	Oui
11	6	R6	Ancien	Nuit	Lourd	Elevee	Vieu	Oui	Oui
12	6	R6	Ancien	Nuit	Leger	Elevee	Neuf	Oui	Oui
13	6	R6	Ancien	Jour	Lourd	Elevee	Neuf	Oui	Non
14	9	R9	Ancien	Jour	Lourd	Elevee	Neuf	Oui	Non
15	9	R9	Nouveau	Jour	Lourd	Elevee	Neuf	Non	Oui
*									

Figure4.23Forme ajouter accident

Le bouton ROUTE pour afficher les listes des différentes couches ou les tables de la base spatiale

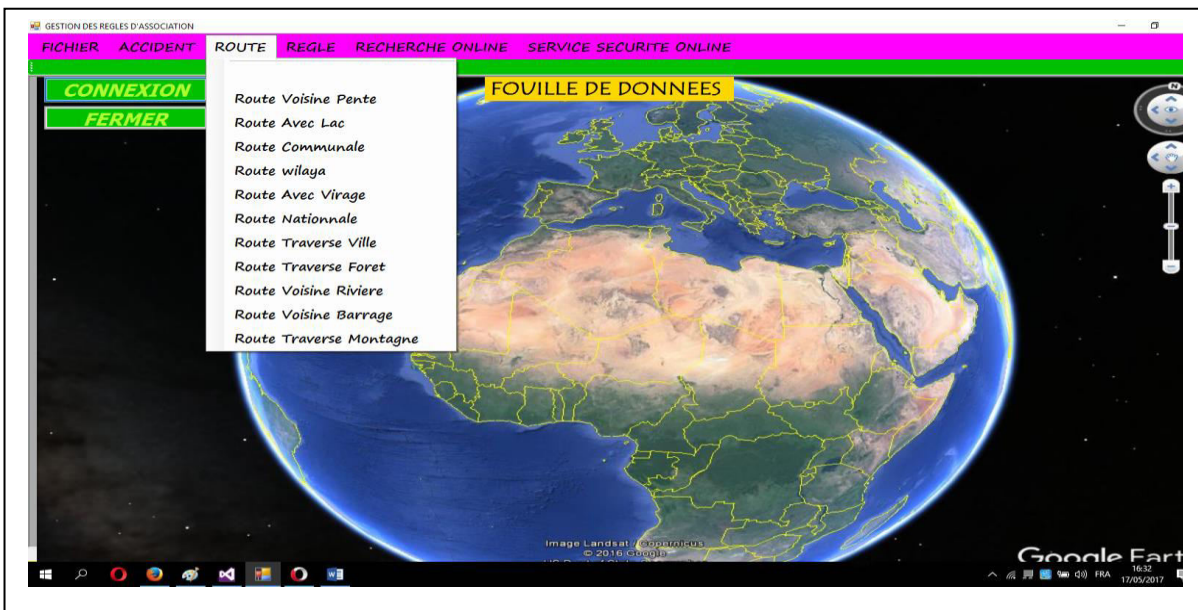


Figure4.24Forme de toutes les routes partielles avec leurs reliefs

Le bouton REGLE pour avoir la fenêtre d'extraction des règles d'association :

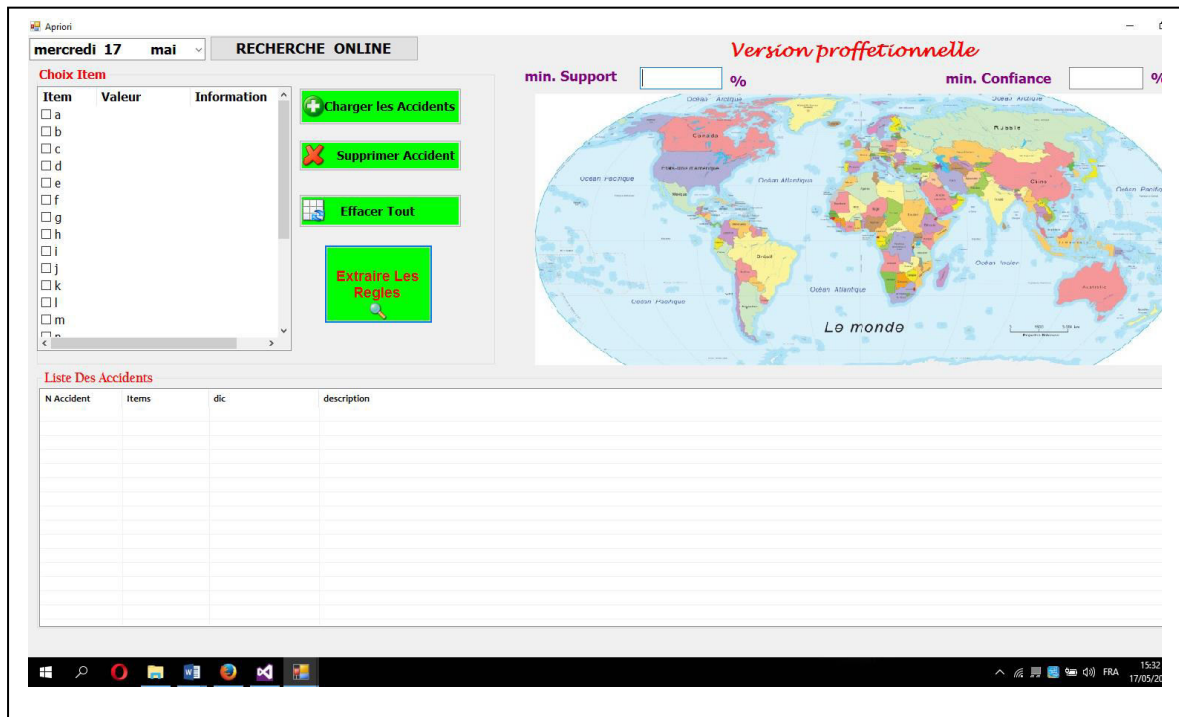


Figure4.25Forme charger tous les accidents de la Base pour le traitement

avec un cliques sur Charger les accidents toutes les accidents de la base seront chargés sur la liste View .

On associe à chaque attribut ou item une lettre correspondante comme exemple :

Permis Nouveau → a

Permis Ancien → z

Temps Accident Jour → e

Donc pour le traitement on va utiliser les lettres alphabétiques comme items comme il est montré sur la figure 4.26

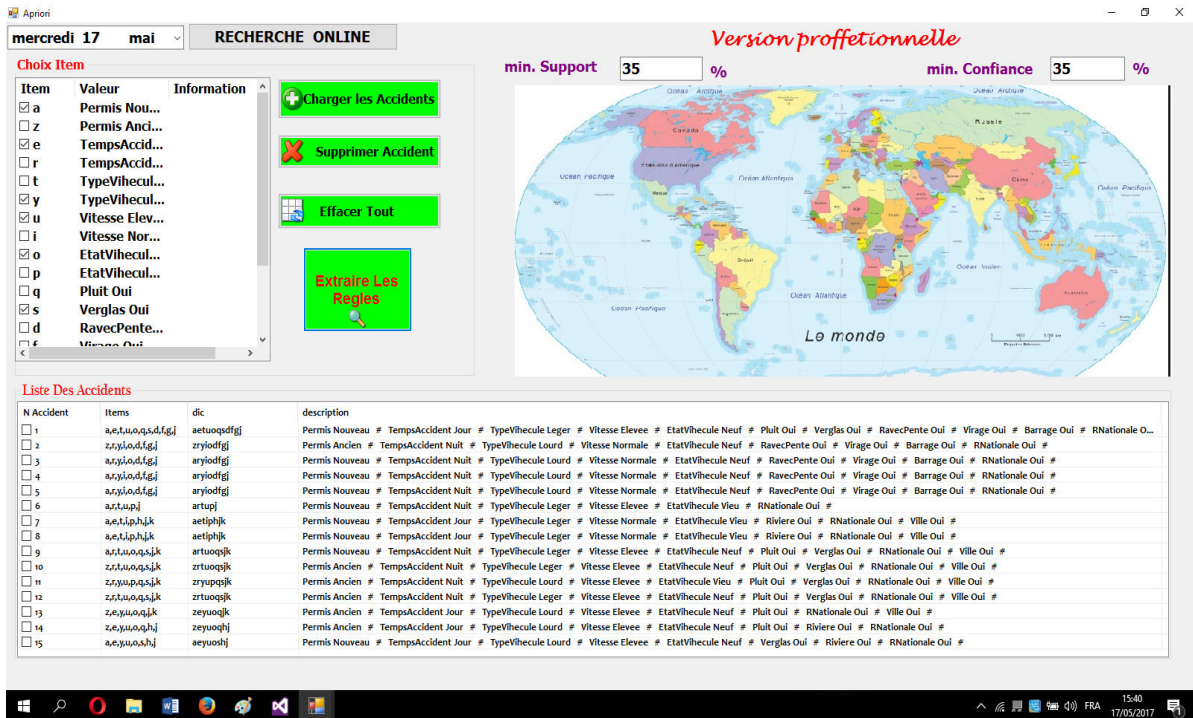


Figure 4.26 Affichage de tous les accidents de la base sur la ListView pour le traitement

On choisit le min Support et le min Confiance et on exécute par le bouton Extraire Les Règles, on obtient le résultat comme suit :

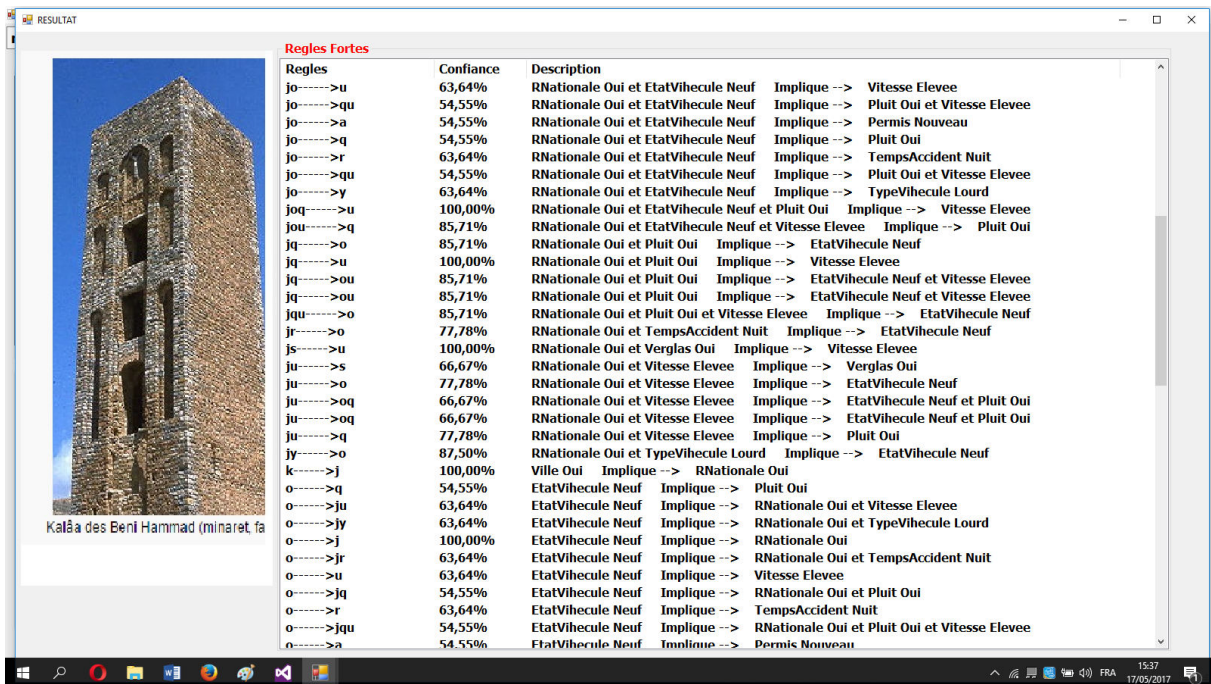


Figure 4.27: Affichage de toutes les règles d'Association après le traitement

11-Conclusion :

La méthode hybride qu'on a utilisé a permis d'une communication topologique entre les entités spatiales via Arc Gis d'une part et la communication relationnelle entre les entités attributaires d'une autre part ainsi la jointure effectuée entre les deux, cette dernière nous a donné la possibilité d'avoir des connaissances à partir d'un ensemble de bases de données hétérogènes, le premier est spatiale son traitement topologique est réalisé via Arc Gis, le deuxième est relationnel son traitement et relationnel via Access, ce qui a préparé une plateforme sur la quelle on a entamé l'analyse de datamining comme si on est sur une base de données ordinaire.

CONCLUSION GENERALE :

L'exploitation des bases de données multiples à référence spatiale correspondant au réseau routier avec l'ajout des bases normales contenant l'ensemble des circonstances des différents accidents pour l'extraction des connaissances via la fouille de données est un projet important pour les services concernés. Nous avons étudié la mise en œuvre de ce projet a deux étapes, le premier consiste représenter et découper le réseau routier de la wilaya de Msila en plusieurs tranches en lui associant les différents reliefs via le logiciel Arc Gis, avant d'avoir un ensemble de bases de données spatiales, ces dernières sont transférées avec les données fournies par les services de sécurité contenant l'ensemble des accidents ont eu lieu la région de Msila vers le logiciel C Sharp.

Dans le but d'atteindre les objectifs du projet, on a réalisé une application permettant à extraire les règles d'association à partir des bases spatiales et autre normales reliées. Au terme de ce projet, nous avons appris à mieux développer un projet à travers le logiciel C Sharp. En outre nous avons appris les contours du concept spatial. Grace à ce concept nous avons pu étudier la mise en œuvre de la localisation spatiale des objets tels que les routes, les forêts, les montagne etc. et leur représentation map sur carte.

Par ailleurs, Nous avons appris les différents modes d'interactions entre les logiciels SIG et les autres technologies de développement classiques comme C Sharp. L'étude nous a permis également d'élaborer le processus de réalisation du système et de mettre en œuvre le système à travers une méthode hybride qui mixte entre les entités spatiales et attributaires afin de générer des connaissances malgré l'hétérogénéité des entités.

Le projet, en dehors de l'aspect régional pourrait connaître une extension nationale car le problème des accidents routiers touche tout le territoire sans exception, et comme dit le proverbe

Il vaut mieux prévenir que guérir
ou un homme avertit en vaut deux.

BIBLIOGRAPHIES

[1] ABDICHE Fethi, ATMANI Baghdad Equipe de recherche « Simulation, Intégration et Fouille de données (SIF) »Département d'Informatique, Faculté des Sciences, Université d'Oran Es-Senia, BP 1524, El-M'Naouer, 31000, Oran, Algérie.

fethi_abdiche_box@yahoo.fr, atmani.baghdad@univ-oran.dz

[2] Agrawal R., Imielinski T. & Swami A., Mining Association Rules between sets of items in large databases. Proceedings of the ACM SIGMOD. Washington, DC, pp. 207-216 (1993).

[3] Marie-Aude Aufaure, Laurent Yeh, Karine Zeitouni

Laboratoire d'Ingénierie des Systèmes d'Information - INSA de Lyon

Bâtiment 501 - F-69621 Villeurbanne Cedex

Marie-Aude.Aufaure@lisi.insa-lyon.fr

Laboratoire Prism - Université de Versailles

45, avenue des Etats-Unis - F-78 035 Versailles Cedex

{Prénom.Nom@prism.uvsq.fr}

[4] BENNIS-ZEITOUNI K., DAVID B., MORIZE-QUILIO I., VIEMONT Y., "Géotropics: Database Support Alternatives for Geographic Applications", *Proc. of the 4th Int. Symposium on Spatial Data Handling*, Zurich, Suisse, 1990.

[5] Breiman L., J.H. Friedman, R.A. Olshen, C.J. Stone, (1984),

Classification and Regression Trees Ed. Wadsworth & Brooks. Monterey, California.

[6] Cecil M., A. Appice, D. Malerba,

Spatial Associative Classification at Different Levels of Granularity: A Probabilistic

Approach, in J.-F. Boulicaut, F. Esposito, F. Giannotti, & D. Pedreschi (Eds.), Knowledge

Discovery in Databases: PKDD 2004, Lecture Notes in Artificial Intelligence, 3202, 99111, Springer, Berlin, Germany, 2004.

[7] Jean-Paul DONNAY

Professeur ordinaire, Université de Liège – Unité de Géomatique

[8] Ester M., Kriegel H.P., Sander J., "Spatial Data Mining: A Database Approach",

in proceedings of 5th Symposium on Spatial Databases, Berlin, Germany, 1997.

[9] google earth

[10] http://live.osgeo.org/fr/overview/geoserver_overview.html

(consulté le 20/02/2017)

[11] Mohamed Daoud Joumoua

<https://www.youtube.com/watch?v=GbZ3MQn-cOU>

(consulté le 10/02/2017)

<https://www.youtube.com/watch?v=udlmBpnsWAc>

(consulté le 10/02/2017)

<https://www.youtube.com/watch?v=TwOnw5TceKk>

[12] Koperski K., Han J., Stefanovic N., "An Efficient Two-Step Method for Classification of Spatial", In proceedings of International Symposium on Spatial Data Handling (SDH'98), p. 45-54, Vancouver, Canada, July 1998.

[13] Koperski K. and Han J., "Discovery of Spatial Association Rules in Geographic Information Databases", In Advances in Spatial Databases (SSD'95), p. 47-66, Portland, ME, August 1995.

[14] Koperski K., "A progressive refinement approach to spatial data mining", PhD Thesis, Simon Fraser University. April 1999.

[15] Laboratoire de cartographie appliquée –
Élisabeth HABERT - IRD - 2000

[16] LARUE T., PASTRE D., VIEMONT Y., "Strong Integration of Spatial Domains and Operators in a Relational Database System", *Proceedings of the 3rd Symposium on Spatial Data bases (SSD'93)*, Singapour, June 1993, Lecture Notes in Computer Science N°692, Springer Verlag (Ed.), p. 53-72.

[17] Lebart L. et al. "Statistique exploratoire multidimensionnelle",
Edition Dunod, Paris, 2 Edition, 1997.

[18] Longley, P.A.; Goodchild, M.F.; Maguire, D.J.; Rhind, D.W. *Geographic Information Systems and Science*. 2nd ed. Chichester: Wiley, 2005, 517 p.

Karine.Zeitouni@prism.uvsq.fr, Tsin-Shu.Yeh@prism.uvsq.fr

[19] Malerba D., F.A. Lisi, An ILP Method for Spatial Association Rule Mining.
In A. Knobbe and D.van der Wallen (Eds.), Notes of the ECML/PKDD 2001 Workshop on Multi-Relational Datamining, 18-29, Germany Freiburg, 2001.

[20] Marc Souris Les principes des systèmes d'information géographique
Principes, algorithmes et architecture du système Savane

[21] R. Marghoubi - A. Boulmakoul - K. Zeitouni

Utilisation des treillis de Galois pour l'extraction et la visualisation des règles d'association spatiales *Faculté des Sciences Technique de Mohammedia (FSTM), B.P. 146 Mohammedia, Maroc. 2*

Agence Nationale de Réglementation des Télécommunications (ANRT), Complexe d'Affaires aile sud Ar-Ryad, Hay Riad 10100, BP 2939 Rabat Maroc 3 Laboratoire PRiSM, Université de Versailles, 45 avenue des Etats-Unis, 78035 Versailles Cedex France.

E-mail : {marghoubi@anrt.net.ma ; boulmakoul2000@yahoo.fr,

Karine.Zeitouni@prism.uvsq.fr}

[22] Open Geospatial Consortium (OGC). OGC Référence

Model. OGC 03-040, version 0.1.3, 200

<http://www.opengeospatial.org/specs/?page=orm> (consulté le 21 décembre 2005)

[23] MAMADOU OUATTARA. Mémoire présenté à la Faculté des études supérieures de l'Université Laval dans le cadre du programme de Maîtrise en sciences géomatique pour l'obtention du grade de maître ès science (M.SC.

[24] Christine Plumejeaud UMR LIENSs 2 rue Olympe de Gouges 17000 La Rochelle
cplumejeaud@gmail.com

[25] Quinlan J.R., Induction of Decision Trees, *Machine Learning* (1), 82 - 106, 1986.

[26] S. Sagar, I. R. Farah et M. B. Ahmed

Laboratoire de Recherche en Informatique Arabisée et Documentique Intégrée (R.I.A.D.I), Ecole Nationale des Sciences de l'Informatique. Campus Universitaire de la Manouba, 2010 Manouba, Tunis, Tunisie.

[27] Shekhar, S., Zhang P., Yan H., Ranga V.: Trends in Spatial Data Mining: Next Generation Challenges and Future Directions, Hillol Kargupta and Anupam Joshi(editors), AAI/MIT Press (2003)

[28] SOURIS M., *Systèmes d'information géographique et bases de données*, Colloques et Séminaires sur le Traitement des données localisées, Paris, Editions de l'ORSTOM, p. 29-87

[29] TOBLER W. R., "Cellular geography", In Gale S. Olsson G. (eds) *Phylosophy in Geography*, Dordrecht, Reidel, p.379-86, 1979.

[30] M.Traori Ibrahim, mémoire de fin de cycle en vue de l'obtention du diplôme d'ingénieur de conception en informatique ; thème : Mise en place d'une base de données à référence spatiale pour la gestion du recensement de la population et des demandes de parcelles.

[31] Wikipédia

<http://fr.wikipedia.org/wiki/MapInfo>(consulté le 10/02/2017)

[32] world Algeria- algeria-road-map.jpg

[33] Karine Zeitouni — Laurent Yeh *Laboratoire PRISM - Université de Versailles - Saint-Quentin, 45, Avenue des Etats-Unis, F-78035 Versailles cedex*

[34] Karine Zeitouni. Analyse et extraction de connaissances des bases de données spatiotemporelles .Interface homme-machine [Cs. HC]. Université de Versailles-Saint Quentin en Yvelines, 2006. <tel-00325468>

[35] Zighed A., Ricco R., "Graphes d'induction - Apprentissage et Data Mining", Edition Hermès Sciences, 2000.