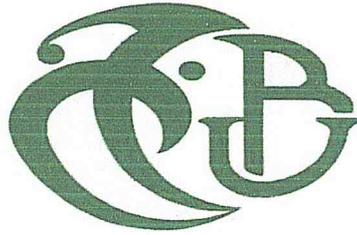


République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahlab Blida

N° D'ordre :



Faculté des sciences

Département d'informatique

Mémoire Présenté par :

Hentabli Chouaib Tarabet Yassine

En vue d'obtenir le diplôme de master

Domaine : Mathématique et informatique

Filière : Informatique
Spécialité : Informatique
Option : Ingénierie de logiciel

Sujet :

Thème

**Conception et réalisation d'un outil de collecte et
d'analyse des accidents de la circulation routière**

Soutenu le : 20/09/2015

M. Abed.H	Président
M. Chikhi.N	Examineur
M.	Examineur
Mme BENATHBA KARIMA	Promotrice
Mme BENATHBA KARIMA	Encadrante

Promotion : 2014 / 2015

Dédicaces

Je dédie ce modeste travail ;

À la mémoire de ma grand-mère et ma tante ;

À vous très cher père et très chère mère, en témoignage de mon amour et de ma reconnaissance ;

À mes frères : Sofiane, Reda, et leur épouses sans oublier Hamza;

À mes très chères sœurs : Nassima, Faiza, Amina, Chafia;

À mes nièces : Mahdia, Maïssa, Sarah, Amani;

À mes neveux : Samy, Zakaria, Ali et Youcef;

À Amina et Faiza qui n'ont pas cessé de me supporter durant toute cette période, avec le témoignage de toute ma reconnaissance ;

À mon encadreur M^{er} Mouloud Koudil et sa très chère famille ;

À mon promoteur M^{mer} BENATCHBA;

À M^{mer} Abed et Mr Mechalikh

À tous mes collègues, plus particulièrement : Khaled, Chouaib, Hamid, Sidali, Youcef en témoignage de mon amitié sincère;

À tous mes amis, plus particulièrement : Amine, Mehdi, Abdou, Tahar, Mohamed, Rafik, Ahmed, et en témoignage de mon amitié sincère;

À tous ceux qui m'ont soutenu, qu'ils trouvent ici l'expression de mon amour et ma profonde gratitude ;

Yassine

Je dédie ce modeste travail à tous ceux que j'aime ... sans vous, ma vie n'aura aucun sens.

*A mon père et ma mère, à qui je dois tout,
A ma sœur Soumaya, qui est toujours dans mon cœur,
A mes frères AMINE, ABDE RAHMANE et leur épouse, à qui je dois beaucoup,
...que dieu vous garde pour moi*

*A la famille HENTABLI, la famille NAIMI,
Entre vous, j'ai grandi entre de bonne main,
Mes succès sont à vous.*

*A tous mes amis de l'université de BLIDA,
Avec qui j'ai passé des moments agréables,
Leurs souvenirs sont éternels,*

*A mes chères amies de BLIDA,
Avec vous, j'ai passé des moments inoubliables,
Merci pour tout.*

*A Yassine mon binôme pour sa modestie, son courage, son ponctualité, son assiduité sa
motivation pour la réalisation de ce travail.*

Chouaib.

Remerciements

Nous tenons à remercier très vivement nos encadreur M^{mer} BENATCHBA et Mr Mouloud KOUDIL, Professeur et Directeur de l'ESI, pour leur entière disponibilité, leur amabilité et simplicité, leur conseils et encouragements. D'avoir tout mis en œuvre pour que nous puissions donner le meilleur de nous-même et pour tous les moyens qu'ils aient mis à notre disposition durant notre période de stage, nous le remercions sincèrement. Qu'ils trouvent ici l'expression de notre profonde gratitude d'avoir acceptés de diriger et de corriger ce travail.

Nos remerciements à Madame ABED, qui s'est toujours montrée à l'écoute et très disponible tout au long de la réalisation de ce projet.

Nous tenons à remercier aussi tout le cadre professoral et administratif du LMCS et de L'INCC pour leur aides et leur soutien notamment Mr Mechalikh.

Que les membres du jury trouvent ici le témoignage de notre reconnaissance pour avoir bien voulu juger notre travail.

Qu'il nous soit permis de remercier toute personne ayant contribué à la réussite de ce travail.

ملخص

وضع سياسة أمن الطرقات يتطلب منا الفهم الجيد لظاهرة حوادث المرور و ذلك لن يكون إلا عن طريق إستغلال الجيد لقاعدة بيانات خاصة بحوادث المرور.

هذه القاعدة تمثل مصدر بيانات هائل بالنسبة ل " مجال حوادث المرور " - نظام يركز على دراسة - و تحليل حوادث الطرق.

بالإضافة إلى ذلك، في السنوات الأخيرة هذه الأخيرة تعرف ازدياد ملحوظ في استخدام معلومات حول الموقع الجغرافي.

أساليب طورت لاستخلاص المعرفة من مجموعات البيانات الكبيرة. مع الأخذ في عين الاعتبار البيانات الجغرافية. هذه الأساليب هي جزء من نظام يسمى "التنقيب في البيانات الفضائية".

يعتبر التنقيب في البيانات الفضائية أحد أحدث المجالات التطبيقية المختصة في تحليل البيانات الجغرافية، تحتوي على مجموعة من التقنيات المختصة في تحليل البيانات الجغرافي.

في مشروعنا، نحن نركز على مجال حوادث المرور. خصوصا حول أساليب جمع البيانات لأقتراح نموذج اثبات. ثم حول مساهمة التنقيب في البيانات الفضائية في المجال المذكور سابقا؛ من أجل وضع أداة لجمع وتحليل حوادث الطرق.

الكلمات الدلالية

التنقيب في البيانات الفضائية , مجال حوادث المرور , ملف BAAC

Résumé

La définition d'une politique de sécurité routière impose de bien comprendre les mécanismes de l'accidentalité et donc de disposer d'une base de données fiable sur les accidents de la circulation routière.

Cette base représente une énorme source de données en « Accidentologie » - discipline qui s'intéresse à l'analyse des accidents et du risque routier -. De plus, cette dernière connaît depuis quelques années une utilisation accrue de l'information sur la localisation géographique. Des méthodes ont vu le jour, servant à extraire des connaissances à partir de grands jeux de données ; tout en tenant compte de la donnée géographique. Ces méthodes s'inscrivent dans une discipline nommée « Fouille de données spatiales » (*Ang. Spatial Data Mining*).

La fouille de données spatiales (*Ang. Data Mining Spatial*) est une discipline émergente, qui offre des techniques dédiées à l'analyse des données géographiques. Elle permet d'extraire des connaissances dans des bases de données géographiques.

Dans notre projet, nous nous intéressons au domaine d'accidentologie. Particulièrement aux méthodes de collecte de données afin de proposer un formulaire de constations et par la suite, l'apport du *Data Mining Spatial* au dit domaine ; afin de mettre en œuvre un outil de collecte et d'analyse des accidents routiers.

Mots clés : *Data Mining Spatial*, Accidentologie, fichier BAAC.

Abstract

The definition of a road safety policy requires to understand well the mechanisms of accidentality and therefore to dispose a reliable database on road traffic accidents.

This base represents a huge source of data in accidentology - discipline that focuses on the analysis of accidents and road risk -. Moreover from some years this last knows an increased use of the geographical location information. Methods have been developed, for extracting knowledge from large data sets;

These methods enroll in a discipline called « Data Mining Spatial ». Spatial data mining is an emerging discipline which offers a set of dedicated techniques to geographical data analysis. It allows extracting knowledge in geographic databases.

In our project, we interest to the accidentology domain. Especially to the methods of collecting data in order to suggest a finding form and subsequently, the contribution of the Spatial Data Mining in said field; so as to implement a gathering and analysis tool of road accidents.

Key words: Data Mining Spatial, Accidentology, BAAC file.

TABLES DES MATIÈRES

ملخص	V
RESUME	VI
ABSTRACT	VII
INTRODUCTION GENERALE	1
PREMIER CHAPITRE : ACCIDENTOLOGIE	
INTRODUCTION	4
1. GENERALITES	4
1.1. ACCIDENTS DANS LE MONDE	4
1.2. ACCIDENTS EN ALGERIE	5
2. ACCIDENTOLOGIE	6
3. SECURITE ROUTIERE	6
3.1. DEROULEMENT D'UN ACCIDENT DE LA CIRCULATION	7
3.2. FACTEURS DE RISQUE	8
3.2.1. <i>Facteurs de risques liés à l'homme</i>	8
3.2.2. <i>Facteurs de risque liés au matériel roulant</i>	10
3.2.3. <i>Facteurs de risque liés à l'état des infrastructures routières</i>	10
4. LES THEORIES DES CAUSES DES ACCIDENTS	11
4.1. LA THEORIE DES DOMINOS	12
4.2. LA THEORIE DU PUR HASARD	12
4.3. LA THEORIE DE LA PROBABILITE FAUSSEE	13
4.4. LA THEORIE DE LA PREDISPOSITION AUX ACCIDENTS	13
4.5. LA THEORIE DU TRANSFERT D'ENERGIE	13
4.6. LA THEORIE DES SYMPTOMES	14
4.7. THEORIE MONO-CAUSALE	14
4.8. LA THEORIE DES CAUSES MULTIPLES	14
5. MODELISATION D'UN ACCIDENT DE LA ROUTE	14
5.1. MODELE CVE (CONDUCTEUR, VEHICULE, ENVIRONNEMENT)	15
5.2. MODELE SEQUENTIEL	15
5.3. MODELE DE FONCTIONNEMENT DE L'OPERATEUR HUMAIN	16
6. INFORMATIONS BRUTES SUR LES ACCIDENTS	17
6.1. SOURCES DES INFORMATIONS	17
6.2. FORMAT DES INFORMATIONS	17
6.2.1. FORMULAIRES DE CONSTATIONS EN EUROPE (PAYS DEVELOPPES)	18
6.2.2. LE BAAC	18
LE CONTENU DU FICHIER BAAC	21
7. ANALYSE ET EXPLOITATION DES DONNEES RECUEILLIES SUR LES ACCIDENTS :	22
7.1. L'APPROCHE CLINIQUE	22
7.2. ÉPIDEMIOLOGIE QUANTITATIVE	23
7.3. CLASSIFICATION EN SCENARIOS-TYPES	23
8. ÉTUDES DÉTAILLÉES D'ACCIDENTS (EDA)	23

CONCLUSION	24
------------------	----

DEUXIÈME CHAPITRE : GENERALITES SUR LE DATA MINING SPATIAL

INTRODUCTION	26
--------------------	----

1. DATA MINING SPATIAL	26
------------------------------	----

2. DONNEES SPATIALES	26
----------------------------	----

2.1. LES DONNEES DESCRIPTIVES	27
-------------------------------------	----

2.2. LES DONNEES SPATIALES.....	27
---------------------------------	----

2.2.1. LES COORDONNEES CARTESIENNES.....	28
--	----

2.2.2. LES COORDONNEES GEOGRAPHIQUES.....	28
---	----

2.3. CARACTERISTIQUES DES DONNEES SPATIALES	29
---	----

2.3.1. INTERDEPENDANCE.....	29
-----------------------------	----

2.3.2. HETEROGENEITE	29
----------------------------	----

2.3.3. COMPLEXITE	30
-------------------------	----

2.4. LES RELATIONS SPATIALES	30
------------------------------------	----

3. OBJECTIFS DU DATA MINING SPATIAL.....	31
--	----

4. QUELQUES DOMAINES D'APPLICATION.....	31
---	----

4.1. L'ENVIRONNEMENT.....	32
---------------------------	----

4.2. LA SEISMOLOGIE	32
---------------------------	----

4.3. GEOMARKETING.....	32
------------------------	----

4.4. ASTRONOMIE.....	33
----------------------	----

4.5. CRIMINOLOGIE	33
-------------------------	----

4.6. ACCIDENTOLOGIE	34
---------------------------	----

CONCLUSION	34
------------------	----

TROISIÈME CHAPITRE : LES TACHES DU DATA MINING SPATIAL

INTRODUCTION	37
--------------------	----

1. CLASSIFICATION SPATIALE SUPERVISE.....	37
---	----

2. CLASSIFICATION NON SUPERVISEE (CLUSTERING).....	38
--	----

2.1. CLUSTERING PAR PARTITIONNEMENT.....	39
--	----

2.1.1. <i>k-Means</i>	39
-----------------------------	----

2.1.2. <i>K-Medoids</i>	40
-------------------------------	----

2.2. CLUSTERING HIERARCHIQUE	42
------------------------------------	----

2.2.1. <i>Approche agglomérative (ascendante)</i>	42
---	----

2.2.1.1. <i>BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)</i>	43
--	----

2.2.1.2. <i>CURE</i>	43
----------------------------	----

2.2.1.3. <i>CHAMELEON</i>	43
---------------------------------	----

2.2.2. <i>Approche divisive (descendante)</i>	44
---	----

2.3. CLUSTERING BASE DENSITE.....	44
-----------------------------------	----

2.3.1. <i>Algorithme DBSCAN (Density Based Spatial Clustering Application with Noise)</i>	45
---	----

2.3.2. <i>Algorithme DENCLUE (Density based Clustering)</i>	45
---	----

2.3.3. <i>Algorithme DBCLASD (Distribution Based Clustering of Large Spatial Databases)</i>	45
---	----

2.4. CLUSTERING BASE GRILLE.....	46
----------------------------------	----

2.4.1. <i>STING (Statistical Information Grid Based Method)</i>	46
---	----

2.4.2. <i>Clique (Clustering In Quest)</i>	47
--	----

2.4.3.	<i>Wave Cluster</i>	48
2.5.	ESTIMATEUR A NOYAU DE DENSITE.....	48
3.	ASSOCIATION SPATIALE	49
3.1.	REGLE D'ASSOCIATION.....	50
3.2.	EXTRACTION DES REGLES D'ASSOCIATION.....	51
4.	GEO-VISUALISATION	52
5.	APPLICATION DU DMS AU DOMAINE DE L'ACCIDENTOLOGIE	52
5.1.	PREMIERE ETUDE.....	53
5.2.	DEUXIEME ETUDE.....	54
5.3.	TROISIEME ETUDE.....	55
	CONCLUSION	56
QUATRIÈME CHAPITRE : ETUDE CONCEPTUELLE		
	INTRODUCTION	58
1.	PROBLEMATIQUE	58
2.	OBJECTIFS	59
3.	TRAVAUX EXISTANTS	59
3.1.	PREMIERE ETUDE.....	59
3.2.	DEUXIEME ETUDE.....	60
4.	SOLUTION PROPOSEE	60
4.1.	DIAGRAMME DE CONTEXTE.....	61
4.2.	DIAGRAMME DE CAS D'UTILISATION.....	62
4.2.1.	<i>Identification des acteurs</i>	63
4.2.2.	<i>Regroupement en paquetage</i>	66
4.2.3.	<i>Identification des cas d'utilisations</i>	66
4.2.3.1.	Paquetage « Gestion des accidents ».....	68
4.2.3.2.	Paquetage « Analyse ».....	74
4.2.3.3.	Paquetage « Statistiques ».....	76
4.2.3.4.	Paquetage « Administration ».....	77
4.3.	DIAGRAMME DE CLASSE.....	86
5.3.1.	<i>Liste des propriétés</i>	87
5.4.	SCHEMA RELATIONNEL.....	88
5.5.	DEVELOPPEMENT DU MODELE DYNAMIQUE.....	89
5.5.1.	<i>Diagramme de séquence associé au cas « Ajouter accident »</i>	90
5.5.2.	<i>Diagramme de séquence associé au cas « Consulter accident »</i>	91
5.5.3.	<i>Diagramme de séquence associé au cas « Valider accident »</i>	92
5.5.4.	<i>Diagramme de séquence associé au cas « Ajouter rubrique »</i>	93
5.5.5.	<i>Diagramme de séquence associé au cas « Ajouter variable »</i>	93
5.5.6.	<i>Diagramme de séquence associé au cas « Ajouter valeur »</i>	94
5.5.7.	<i>Diagramme de séquence associé au cas « Détection des hot-spots »</i>	95
5.5.8.	<i>Diagramme de séquence associés au cas « Classification »</i>	96
5.5.9.	<i>Diagramme de séquence associé au cas « Prévision »</i>	97
5.6.	Conception détaillé.....	97
5.6.1.	Architecture technique de l'application.....	97
5.6.1.1.	Module « Donnée ».....	98

a.	Module « Lecture »	98
b.	Module « Collecte de données »	98
c.	Module « Construction de la structure de la fiche d'accident ».....	99
5.6.1.2.	Module « Fouille »	99
a.	Module « Clustering ».....	99
b.	Module « Hot-spot »	102
c.	Module « Préviation »	103
5.6.2.	<i>Architecture réseau</i>	104
5.6.2.1.	Présentation de l'architecture à 2 niveaux (2 tiers).....	104
5.6.2.2.	Présentation de l'architecture à 3 niveaux	104
5.6.2.3.	La solution adoptée	105
CONCLUSION		105
 CINQUIÈME CHAPITRE : IMPLEMENTATION		
INTRODUCTION		107
1. TECHNOLOGIES UTILISEES		107
1.1.	ECLIPSE LUNA.....	107
1.2.	JAVA 8	107
1.3.	HTML ET CSS	107
1.4.	JAVASCRIPT	108
1.5.	BOOTSTRAP DE TWITTER	108
1.6.	WAMPSEVER	108
1.7.	MYSQL.....	109
1.8.	SPRING FRAMEWORK	109
1.10.	GLASSFISH.....	110
1.11.	WEKA	110
1.12.	OPENLAYERS	111
2. SCENARIO D'UTILISATION		111
2.1.	SCENARIO 1	111
2.2.	SCENARIO 2.....	116
2.3.	SCENARIO 3.....	118
CONCLUSION		123
CONCLUSION ET PERSPECTIVES		124
REFERENCES BIBLIOGRAPHIQUES		126

FIGURE 49. DESCRIPTION DU MODULE PREVISION.....	103
FIGURE 50. ARCHITECTURE A 2 NIVEAUX [WEB 6].....	104
FIGURE 51. ARCHITECTURE A 3 NIVEAUX [WEB 6].....	105
FIGURE 52.LES ETAPES DU SCENARIO 1.....	111
FIGURE 53.INTERFACE «AJOUTER ACCIDENT ».....	112
FIGURE 54.INTERFACE « FICHE BAAC ».....	113
FIGURE 55.INTERFACE DES RUBRIQUES.....	114
FIGURE 56.COMPLETER UNE RUBRIQUE.....	114
FIGURE 57.INTERFACE « CLOTURER ACCIDENT ».....	115
FIGURE 58.ETAPES DE SCENARIO 2.....	116
FIGURE 59.INTERFACE « CONFIGURER FICHE BAAC ».....	116
FIGURE 60.INTERFACE « LISTE DES VARIABLES ».....	117
FIGURE 61.INTERFACE « AJOUTER VARIABLE ».....	117
FIGURE 62.INTERFACE « LISTE DES VALEURS».....	118
FIGURE 63.INTERFACE « LISTES AJOUTER VALEUR».....	118
FIGURE 64.PRINCIPALES ETAPES DU 3E SCENARIO.....	119
FIGURE 65.INTERFACE « ALGORITHME D'ANALYSE ».....	119
FIGURE 66.INTERFACE « KDE ».....	120
FIGURE 67.INTERFACE « RESULTAT DE L'ANALYSE DES HOT-SPOTS ».....	120
FIGURE 68.INTERFACE « ALGORITHME K-MEANS ».....	121
FIGURE 69.INTERFACE « ALGORITHME K-MEANS ».....	121
FIGURE 70.INTERFACE « ALGORITHME APRIORI».....	122
FIGURE 71.INTERFACE « ALGORITHME APRIORI ».....	122

LISTE DES TABLEAUX

TABLEAU 1. ETAT DES ACCIDENTS DE LA CIRCULATION ROUTIERE ENREGISTRES AU COURS DES PREMIERS SEMESTRES DE 2007 ET 2008 [LEWHE.M ET ZEMMOUR.O, 2009].	6
TABLEAU 2. LES DIX WILAYAS, SIEGE DE 45% DES ACCIDENTS EN 2007 [LEWHE.M ET ZEMMOUR.O, 2009].	11
TABLEAU 3. LES DIFFERENTS FORMULAIRES DE CONSTATIONS DES ACCIDENTS EN EUROPE.....	18
TABLEAU 4. REPRESENTATION DES ACHATS DES CLIENTS SOUS FORME DE TABLEAU.....	50
TABLEAU 5.REGROUPEMENT EN PAQUETAGES DES CAS D'UTILISATIONS FONCTIONNELS.....	66
TABLEAU 6. IDENTIFICATION DES CAS D'UTILISATION.....	67
TABLEAU 7. LISTE DES PROPRIETES.....	88

LISTE DES ALGORITHMES

TABLEAU 1. ETAT DES ACCIDENTS DE LA CIRCULATION ROUTIERE ENREGISTRES AU COURS DES PREMIERS SEMESTRES DE 2007 ET 2008 [LEWHE.M ET ZEMMOUR.O, 2009].	6
TABLEAU 2. LES DIX WILAYAS, SIEGE DE 45% DES ACCIDENTS EN 2007 [LEWHE.M ET ZEMMOUR.O, 2009].	11
TABLEAU 3. LES DIFFERENTS FORMULAIRES DE CONSTATIONS DES ACCIDENTS EN EUROPE.....	18
TABLEAU 4. REPRESENTATION DES ACHATS DES CLIENTS SOUS FORME DE TABLEAU.....	50
TABLEAU 5.REGROUPEMENT EN PAQUETAGES DES CAS D'UTILISATIONS FONCTIONNELS.....	66
TABLEAU 6. IDENTIFICATION DES CAS D'UTILISATION.....	67
TABLEAU 7. LISTE DES PROPRIETES.....	88

Introduction Générale

Aujourd'hui, les accidents de la route sont très fréquents ; non seulement en Algérie, mais aussi dans le monde. Ils occupent les premiers rangs de causes de mortalité. Selon l'OMS¹, les accidents sont la huitième principale cause de décès, et ils sont la première cause de décès chez les personnes âgées de 15-29 ans.

D'après les statistiques faites par cette organisation en 2010, la circulation routière avait provoqué environ 1.24 million de décès dont 77% était des hommes. Ainsi, les pays à revenu intermédiaire ont eu la plus forte charge et les plus hauts taux de mortalité.

Pour les autorités, ce phénomène est devenu de plus en plus troublant. Non seulement, il cause des tragédies sociales en anéantissant des milliers de vies humaines, mais il engendre aussi des pertes économiques et financières fort importantes.

En Algérie, les statistiques fournies par le ministère des transports révèlent des chiffres de plus en plus alarmants concernant les accidents de la route. Ainsi, 40 000 accidents se produisent chaque année, entraînant plus de 4 000 morts et près de 60 000 blessés, dont des handicapés à vie (cf. Centre national de prévention et de sécurité routière - 2009). Ces drames routiers, qui imposent une contrainte majeure pour le développement du pays, ont incité la gendarmerie nationale à créer un laboratoire de recherche dans le domaine de l'accidentologie qui a pour mission l'étude et l'analyse détaillée des accidents de la circulation routière au sein de l'Institut national de criminalistique et de criminologie (INCC) afin d'améliorer la sécurité routière.

Cependant, il reste à savoir que les masses d'informations relevées concernant le phénomène d'accidentologie représentent une énorme source de données. L'exploitation efficace de cette dernière avec l'utilisation des techniques de la fouille de données spatiales, permet aux experts la découverte de connaissances afin d'améliorer la compréhension du dit phénomène, de souligner ses caractéristiques et de formuler des hypothèses sur son évolution.

La fouille de données spatiales (traduction de l'anglais « *Data Mining Spatial* ») est un domaine en pleine expansion. Elle est largement étudiée pour l'analyse des accidents de la route. Elle englobe un ensemble de méthodes, pour la découverte de connaissances dans les

¹ **Organisation mondiale de la santé (OMS)**, est une institution spécialisée de l'Organisation des Nations unies (ONU) pour la santé publique créée en 1948.

Accidentologie

Introduction

L'état actuel de la sécurité routière dans le monde, et plus particulièrement dans les pays en voie de développement, est digne d'attirer toutes les attentions, ce qui conduit à mettre l'accent sur le besoin de comprendre le phénomène des accidents routiers.

Une discipline a vu le jour avec pour objet l'étude du phénomène des accidents réels, dénommée « accidentologie ». Ce terme a été utilisé en 1968 par des chercheurs de l'ONSER. L'intention était d'individualiser une activité souvent confondue avec la traumatologie. Cette dernière s'intéresse à la phase finale de l'accident; celle qui provoque les lésions, alors que la compréhension de l'accident commence avant l'ouverture de la portière du véhicule par le conducteur qui va être l'auteur de l'événement et qui va en subir les conséquences.

1. Généralités

1.1. Accidents dans le monde

On rapporte que lors de l'enquête sur le premier décès dû à un accident de la circulation en 1896, le coroner aurait déclaré : « Cela ne doit plus jamais se reproduire »². Malheureusement et chaque année, 1,2 millions de personnes meurent sur les routes avec plus de 50 millions de blessés [WEB 1]. Les accidents de la route occupent le 11^{ème} rang des causes de décès dans le monde et 2,1% du taux de mortalité mondiale pour l'ensemble des tranches d'âges. Plus de 3000 personnes meurent chaque jour dans le monde des suites d'accidents de la circulation routière.

L'ensemble des décès (85%) survenus à la suite d'accidents de la circulation sont enregistrés dans des pays à revenu faible ou intermédiaire [WEB 1]. Selon les projections effectuées pour les années comprises entre 2000 et 2020, les cas de décès dus aux accidents de la circulation diminueront d'environ 30% dans les pays à haut revenu, mais augmenteront sensiblement dans les pays à revenu faible ou intermédiaire. Si l'on ne prend pas des mesures appropriées, les accidents de la circulation devraient représenter d'ici 2020 la troisième principale cause mondiale de maladies et de traumatismes [Lewhe.M et Zemmour.O, 2009]. En 1998 et selon une étude publiée par [KRUG .E, 1999], les accidents de la circulation ont causé la mort d'environ 1170694 personnes dans le monde et sur l'ensemble des décès (1029037), soit 87,9% se sont produits dans les pays à faible revenu et 12,1% dans les pays à revenu élevé. Ceci peut être mieux apprécié à travers la figure 1.

² Source: World's first road death. London, RoadPeace.

Chapitre 1 : Accidentologie

Désignation	1 ^{er} Semestre 2007	1 ^{er} Semestre 2008	Différence	Pourcentage (%)
Accidents	8232	8512	+ 280	+ 03.40
Blessés	9285	9804	+ 519	+ 05.58
Tués	343	368	+ 25	+ 07.28

Tableau 1. Etat des accidents de la circulation routière enregistrés au cours des premiers semestres de 2007 et 2008 [Lewhe.M et Zemmour.O, 2009].

2. Accidentologie

L'accidentologie est l'étude des accidents. C'est une discipline qui fait appel à un spectre de compétences étendu, qui va de l'ingénieur au médecin en passant par le psychologue, l'expert en mécanique ou en sciences cognitives [Ferrandez.F, 1995]. Il s'agit d'atteindre le plus haut degré d'entraide et d'échange d'informations pour obtenir une vision globale et exhaustive des accidents.

L'évolution de l'accidentologie est retracée dans [Fleury.D, 1998]. Cette discipline est très étroitement liée à la sécurité routière, son but est de fournir des données reflétant de façon objective la situation d'un accident. Après des débuts laborieux et empiriques dans la première moitié du 20^{ème} siècle, elle est restée quelques décennies sous une forme « primitive » [De Silva.H, 1951]. La démocratisation dans les pays riches, de l'usage de l'automobile a réellement eu lieu au début de la seconde moitié du 20^{ème} siècle. En conséquence, le nombre d'accidents et de victimes de la circulation a très vite augmenté.

L'accidentologie passe dans sa phase « moderne » dans les années 1970. Elle commence alors à se différencier de la traumatologie pure. Un accident est étudié de manière plus fine et plus nuancée. Les chercheurs et accidentologues de terrain tentent désormais de le comprendre dans sa globalité, en tenant compte de sa grande complexité [Ferrandez.F, 1995].

D'une autre manière, l'accidentologie représente « des études qui ont pour objectif de comprendre l'étiologie de l'accident et des lésions, afin de proposer des contre-mesures pour diminuer le nombre des accidents de la route et les dommages corporels conséquents et les évaluer » [Foret-Bruno et al, 2001].

3. Sécurité Routière

La sécurité (du latin securitas) désigne ce qui est hors de péril, dommage ou risque. Le concept de sécurité routière concerne donc la prévention d'accidents sur la route dans le but de protéger la vie des personnes [KRUG .E, 1999].

Chapitre 1 : Accidentologie

La sécurité routière a pour finalité spécifique d'assurer les déplacements routiers sans effets externes indésirables (sentiment d'insécurité, accidents de la route, blessures, etc). Elle doit être assurée tant dans l'organisation, qu'à l'occasion de chaque déplacement [Ferrandez.F, 1995].

Le besoin en sécurité routière est apparu avec l'apparition des accidents, dès la mise en circulation des premiers véhicules et de leur utilisation. Ce besoin s'est intensifié avec l'augmentation de l'insécurité routière.

3.1. Déroulement d'un accident de la circulation

A la fin des années 1960 aux Etats-Unis, William HADDON a tenté de modéliser les accidents de la circulation. De ses réflexions est née une méthode d'analyse appelée « matrice de Haddon ». Cette approche demeure jusqu'à aujourd'hui le principal cadre de réflexion théorique sur la sécurité routière.

La matrice de Haddon telle que représentée en figure 2, découpe l'accident selon trois phases chronologiques successives et leur associe, trois facteurs déterminants : le véhicule, l'infrastructure routière et le comportement humain. Les trois phases sont :

- **la phase avant, ou pré-crash**, constituée par les instants qui précèdent le choc,
- **la phase pendant**, démarre à l'instant du choc et dure jusqu'à ce que les véhicules et les corps impliqués dans l'accident atteignent une position d'équilibre - au sens physique -,
- **la phase après**, en fonction de l'accident, comprendra l'arrivée des secours, des forces de l'ordre . . . jusqu'à ce que les véhicules, les corps et/ou les débris aient été évacués ou dégagés et la circulation rétablie normalement.

La matrice de Haddon

Phases		Facteurs		
		Human	Vehicles and equipment	Environment
Pre-crash	Prévention primaire	Information Attitudes comportements Action de la police	Tenue de route phares Freins	Etat de la route Limitation de vitesse Protection des piétons
Crash	Prévention secondaire	Utilisation de la ceinture	Ceinture en bon état, airbags, coque, renforts, design protecteur	Barrière de sécurité Infrastructures adaptées
Post-crash	Prévention tertiaire	Premier secours Access aux soins	Facilité d'accès Risque d'incendie	Moyens des secours, embouteillages

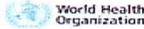
 World Health Organization
  INDIAN INSTITUTE OF TECHNOLOGY DELHI

Figure 2. La matrice de Haddon

Chapitre 1 : Accidentologie

L'objectif de la matrice de Haddon est de mettre en lumière, pour un accident donné, les facteurs clés qui à chacune des trois étapes successives, permettent de lutter contre les accidents routiers et leurs conséquences associées.

A partir de ce cadre d'analyse théorique, plusieurs autres concepts sont apparus, comme les notions de sécurité, selon [J.-Y. le Coz et Y. Page, 2003] et [T. Perron, 1997]. La sécurité routière peut être traitée sous trois angles (branches) :

- *la sécurité primaire*, a pour objectif de réduire les accidents par leur prévention et leur évitement,
- *la sécurité secondaire*, vise à réduire les conséquences des accidents (ex. protection des occupants),
- *la sécurité tertiaire*, vise à améliorer le secours aux personnes impliquées et diminuer les séquelles physiques et psychologiques après un accident.

Il existe aussi d'autres définitions qui font une distinction entre la sécurité active et la sécurité passive, différencier par le fait que:

- *la sécurité active*, fait référence aux systèmes nécessitant l'intervention du conducteur (ceinture de sécurité, ABS, etc.).
- *la sécurité passive*, fait référence aux contre-mesures indépendantes de l'intervention du conducteur (airbag, régulateur de vitesse, etc.).

Les différentes « branches » de la sécurité sont regroupées aujourd'hui dans ce qu'on appelle la « sécurité intégrée ». Ainsi, les objectifs de nos travaux s'intègrent dans les études en sécurité primaire. Rappelons que le but de la matrice de Haddon est d'associer aux différentes phases, trois facteurs déterminants : le comportement humain, le véhicule et l'infrastructure routière.

3.2. Facteurs de risque

3.2.1. Facteurs de risques liés à l'homme

Les statistiques mondiales accablent l'homme de la responsabilité de 80-95% des accidents de la voie publique. Le conducteur est sans doute l'élément primordial du phénomène. C'est lui qui, à tout moment, doit s'adapter si certains paramètres changent au niveau des deux autres facteurs (véhicule et infrastructure routière).

La Vitesse

Les Algériens roulent vite, trop vite à en croire les statistiques du groupement de la Gendarmerie Nationale d'Alger, qui a révélé, lors des journées portes ouvertes qui se sont

Chapitre 1 : Accidentologie

tenues du 18 au 20 juin 2008 à Maqam Echahid, que près de 18000 véhicules ont été «flashés» par les radars fixes et mobiles (Laser et Doppler) entre Janvier et fin Mai 2008. Pire, les gendarmes ont eu à vérifier cet excès de vitesse à travers une opération pilote, à savoir le respect des plaques de signalisation et de limitation de vitesse sur les autoroutes d'Alger, où la vitesse est limitée à 80 km/heure. Il est à noter que 91% des accidents de la route en Algérie sont dus à l'excès de vitesse, qui figure en première place dans les causes d'accidents. L'excès de vitesse, lié au dépassement dangereux notamment en zone rurale où l'on enregistre le taux le plus élevé d'accidents avec 54,66%, est dans tous les cas, la cause principale des accidents de la route en Algérie [G. Forbes et B. Malone, 2001].

L'alcool

Selon les estimations de l'IBSR, l'alcool est en cause dans 1 accident sur 4 avec des décès ou des blessés graves. Dans 85% des accidents dus à l'alcool, le conducteur est un buveur occasionnel et non un alcoolique. Nuits et weekends, le pourcentage des personnes admises aux urgences à la suite d'un accident passe de 28% à 50% [De Silva.H, 1951].

Drogue

Nous voyons notamment se développer la conduite sous l'emprise de la cocaïne, indique Charles Mercier-Guyon, médecin et secrétaire du comité médical de l'association Prévention routière. Les risques ne sont pas les mêmes : "conduire après avoir fumé du cannabis double le risque d'être responsable d'un accident mortel, il le multiplie par 8 après une prise de cocaïne" [Fleury.D, 1998].

Fatigue du conducteur

La fatigue ou la somnolence peuvent être associées à divers facteurs, certains concernent la circulation routière, comme la conduite sur de longues distances, le manque de sommeil et la perturbation des rythmes circadiens.

Le non-respect du code de la route

En Algérie, la majorité des conducteurs estime que le code de la route limite leur liberté individuelle. Les statistiques sont effarantes, ahurissantes. Une étude du département de sociologie et des sciences humaines de l'université de Bouzaréah révèle que 70% des Algériens ne respectent pas spontanément le nouveau code de la route.

La Gendarmerie a identifié cinq facteurs à l'origine de cette dramatique situation :

- Perte de contrôle du véhicule (1232 cas enregistrés).
- L'excès de vitesse (961 cas).
- L'imprudence des piétons (580 cas).

- Les dépassements dangereux (575 cas).
- Le non-respect de la distance de sécurité (318 cas).

Autrement dit, le non-respect du code de la route est la principale cause des accidents meurtriers [G. Forbes et B. Malone, 2001].

3.2.2. Facteurs de risque liés au matériel roulant

Ces causes occupent une place non négligeable dans la survenue des accidents. Des statistiques Nord-américaines (National Highway Traffic Safety Administration) et françaises (Prof SICARD) évaluent à 7% le nombre d'accidents de la voie publique imputables à des vices techniques du véhicule.

Le "National Safety Council des USA" estime à 2/5 le nombre des véhicules potentiellement dangereux. Si les progrès techniques ont réduit le nombre d'accidents imputables aux vices de fabrication et ont augmenté très notablement la sécurité des usagers des véhicules modernes, ils n'ont pas encore réussi à diminuer les vices imputables au vieillissement des organes.

Les principaux facteurs liés au matériel sont :

- Manque de protection anti-collision intégrée au véhicule.
- Utilisation de pièces de rechange non conformes aux normes.
- Visites techniques douteuses.
- Accroissement du trafic routier [Lewhe.M et Zemmour.O, 2009].

3.2.3. Facteurs de risque liés à l'état des infrastructures routières

Les accidents de la circulation ne sont pas répartis uniformément sur l'ensemble du réseau routier. Ils se produisent par grappes aux mêmes endroits, sur certains tronçons de routes ou un peu partout (notamment dans les quartiers socialement défavorisés). Les techniques routières peuvent sensiblement aider à réduire la fréquence et la gravité des accidents, mais elles peuvent aussi contribuer aux collisions. Le réseau routier influence le risque de collision parce qu'il détermine la façon dont les usagers de la route perçoivent leur environnement et leur indique au moyen de signalisations et de contrôles routiers ce qu'ils devraient faire. Quant aux travaux de Promising (2001) et de Kandela (1993), ils ont mis l'accent sur les défauts de conception des routes [Lewhe.M et Zemmour.O, 2009].

En Algérie, la saturation du trafic peut justifier des accidents de la route sur certaines sections routières. C'est le cas en particulier de la route nationale Constantine - Oran dont le niveau de saturation du trafic a été atteint et a été dépassé depuis longtemps.

Chapitre 1 : Accidentologie

La lenteur de la circulation, rendue plus dangereuse par la gêne des poids lourds sur de très grandes distances ne fait qu'augmenter l'impatience des conducteurs. Les risques qui deviennent très grands, sont concentrés sur dix wilayas et constituent la cause de 45% des accidents comme illustré en tableau 2.

Wilaya	Accidents	Parc x 1000
Alger	5 476	638
Sétif	2 096	52
Oran	1 351	142
Blida	1 316	190
Batna	1 263	87
Tlemcen	1 223	89
Mostaganem	1 216	75
Biskra	1 103	41
Mascara	1 035	64
Constantine	1 028	104

Tableau 2. Les dix wilayas, siège de 45% des accidents en 2007 [Lewhe.M et Zemmour.O, 2009].

L'inadaptation entre l'augmentation vertigineuse du parc roulant et les infrastructures routières existantes accroît de 10% par an le risque d'accident en Algérie. Avec un parc relativement modeste - d'environ 3 millions de véhicules - l'Algérie se retrouve avec un taux de risque 50 fois plus élevé que celui des pays scandinaves et 20 fois plus élevé que celui de la France [Lewhe.M et Zemmour.O, 2009].

Afin de diminuer le taux d'accidents, la sécurité routière est devenue de plus en plus nécessaire. Il est considéré comme impossible de prévenir un accident si l'on ne comprend pas ses causes.

4. Les théories des causes des accidents

Les accidents sont définis comme des événements imprévus qui occasionnent des traumatismes, des décès, une perte de production ou des dommages aux personnes, aux biens et aux avoirs. Il est extrêmement difficile de les prévenir si l'on ne comprend pas leurs causes. De nombreuses tentatives ont été faites par des chercheurs de différentes disciplines pour élaborer une théorie des causes des accidents afin d'identifier, d'isoler et en fin de compte, de

Chapitre 1 : Accidentologie

supprimer les facteurs proches ou lointains de ces derniers; mais jusqu'ici, aucune ne s'est universellement imposée. Nous présentons brièvement ci-après diverses théories des causes des accidents citées par [J.M .Stellman, 2000], mais avant, nous avons estimé qu'il est utile de présenter l'évolution de celles-ci dans le temps (figure 3).

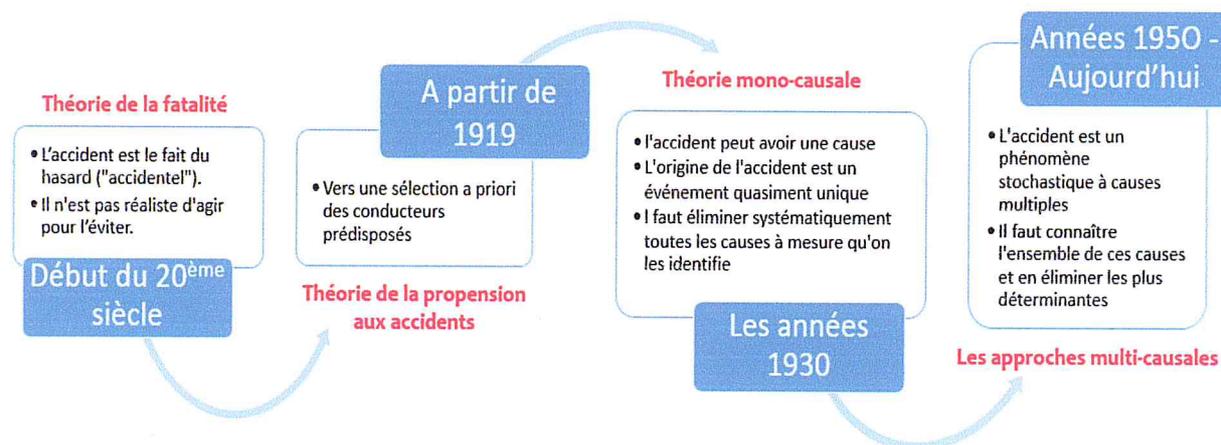


Figure 3. Historique des apparitions des théories de causes d'accidents.

4.1. La théorie des dominos

Selon [W.H. Heinrich, 1931], l'initiateur de la théorie dite des dominos, 88% des accidents sont provoqués par des gestes humains dangereux, 10% par des actes dangereux et 2% par le hasard. Heinrich a proposé une «séquence accidentelle à cinq facteurs» dans laquelle chaque facteur déclenche le suivant, de la même manière que, dans une rangée de dominos, le déséquilibre d'un domino entraîne la chute de tous les autres. Cette séquence est la suivante:

1. antécédents et environnement social,
2. faute du travailleur,
3. geste dangereux associé à un risque mécanique ou physique,
4. accident,
5. dommages matériels ou corporels.

Selon Heinrich, de même qu'il suffit d'enlever un seul domino de la rangée pour interrompre la succession de chutes, la suppression de l'un des cinq facteurs empêcherait l'accident et ses conséquences, le domino clé à enlever étant le troisième. Bien qu'Heinrich n'ait présenté aucune statistique à l'appui de sa théorie, celle-ci n'en constitue pas moins un point de départ utile pour la discussion et les recherches futures.

4.2. La théorie du pur hasard

Selon la théorie du pur hasard, la probabilité, dans un ensemble donné de travailleurs, d'être victime d'un accident est la même pour tous, et il est impossible de discerner un

schéma unique d'événements conduisant à un accident. Tous les accidents sont considérés comme dans la théorie de Heinrich - comme relevant du pur hasard, en partant du principe qu'aucune intervention ne saurait les empêcher.

4.3. La théorie de la probabilité faussée

L'idée sur laquelle repose cette théorie est que lorsqu'un travailleur a été victime d'un accident, la probabilité qu'il le soit de nouveau est augmentée ou diminuée par rapport à celle des autres travailleurs. Cette théorie n'aide guère, voire pas du tout, à trouver des mesures de prévention.

4.4. La théorie de la prédisposition aux accidents

On considère ici que, dans un ensemble donné de travailleurs, il existe un sous-ensemble dont les éléments sont plus enclins que les autres à être victimes d'accidents. Les chercheurs n'ont pas été en mesure d'apporter des preuves convaincantes, car la plupart des travaux n'ont pas été conduits de manière satisfaisante et les résultats sont le plus souvent contradictoires et peu concluants. Cette théorie n'est pas acceptée par tous, même si les données d'expérience viennent l'appuyer. Elle ne rend probablement compte que d'une très faible proportion d'accidents, sans signification statistique.

4.5. La théorie du transfert d'énergie

Pour les tenants de cette théorie, c'est un transfert d'énergie qui provoque des dommages corporels ou matériels et tout transfert d'énergie implique une source, une voie de transfert et un récepteur. Cette théorie est utile pour la détermination des causes des lésions et l'évaluation des risques liés au transfert d'énergie ainsi que des méthodes de contrôle. On peut mettre au point des stratégies de prévention, de limitation ou d'amélioration.

On peut agir sur le transfert d'énergie à la source par les moyens suivants:

- élimination de la source,
- modification de la conception ou des spécifications des éléments du poste de travail,
- maintenance préventive.

On peut modifier la voie de transfert par les moyens suivants:

- isolement de la voie de transfert,
- installation de barrières,
- installation d'absorbants,
- mise en place d'isolants.

On peut aider le récepteur du transfert d'énergie en adoptant les mesures suivantes:

- limitation de l'exposition,

- utilisation d'un équipement de protection individuelle.

4.6. La théorie des symptômes

La théorie des symptômes n'est pas tant une théorie qu'un avertissement dont il faut tenir compte si l'on veut comprendre les causes des accidents. Le plus souvent, lorsqu'on analyse un accident, on tend à privilégier les causes les plus évidentes et à négliger les causes profondes. Or, les conditions ou les actes dangereux sont les causes proches et non profondes de l'accident.

4.7. Théorie mono-causale

D'après les travaux de recherche de [VAN ELSLANDE, 2010], la théorie mono-causale se traduit par :

- Une unique cause d'accident.
- Une élimination systématique de toutes les causes identifiées.

Cependant, le travail de l'accidentologue ressemblait, par beaucoup d'aspects, à la mission d'un détective qui recherche un coupable. Par conséquent, on cherche plus à accuser l'un des protagonistes qu'à trouver des solutions.

4.8. La théorie des causes multiples

La théorie des causes multiples est un dérivé de la théorie des dominos, mais elle part du principe que de nombreux facteurs, causes et causes secondaires peuvent être à l'origine d'un accident qui résulte de certaines de leurs combinaisons. Elle distingue deux catégories de facteurs contributifs:

- *Les facteurs liés au comportement de l'opérateur:* attitude inadéquate, manque de connaissances, insuffisance de qualifications ou état physique ou mental inadapté.
- *Les facteurs liés à l'environnement:* les insuffisances de la protection contre des éléments dangereux existant sur le lieu de travail et la dégradation de l'équipement par l'usage ou du fait de méthodes dangereuses.

Le principal apport de cette théorie est qu'elle met l'accent sur le fait qu'un accident est rarement - sinon jamais - le résultat d'une seule cause ou d'un seul acte.

5. Modélisation d'un accident de la route

Dès la fin des années cinquante, de nombreuses études ont été menées dans le cadre de la recherche pour tenir compte de la complexité des accidents.

Il a fallu créer des modèles de représentation des accidents permettant leur analyse en détail tout en respectant leur aspect dynamique séquentiel et causal, ainsi que la représentation

Chapitre 1 : Accidentologie

des différents composants. Dans ce qui suit, nous allons présenter les modèles les plus utilisés, actuellement.

5.1. Modèle CVE (Conducteur, véhicule, Environnement)

Le modèle CVE permet une analyse des accidents selon les 3 points suivants :

- Véhicule.
- Conducteur.
- Environnement.

Ainsi que les interactions entre ces trois composants, les éléments de chacun des trois sous-systèmes sont identifiés et caractérisés par des attributs.

La Figure 4 représente une version simplifiée du modèle CVE avec quelques exemples d'interaction entre ces différents composants.

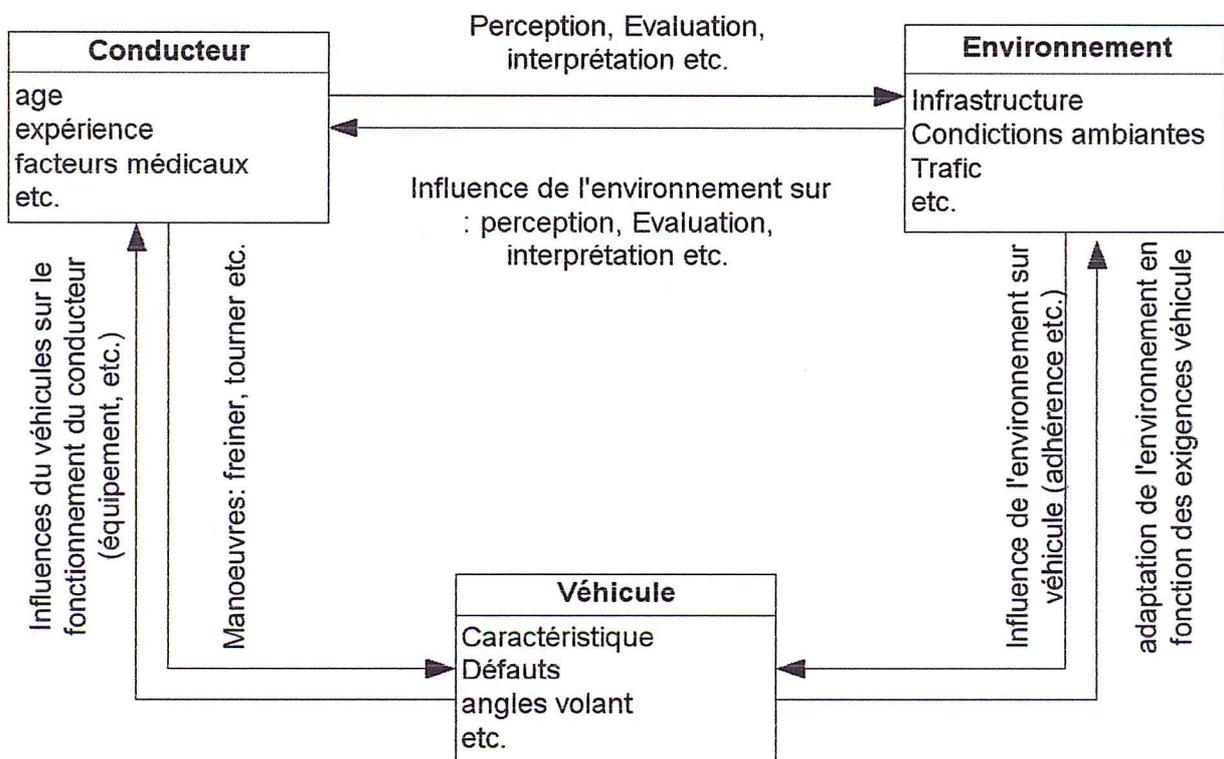


Figure 4. Modèle CVE [W.Ben Ahmed, 2005].

5.2. Modèle Séquentiel

L'approche séquentielle est parmi les approches les plus utilisées pour l'analyse de l'accident car elle représente en même temps l'aspect causal, dynamique et spatio-temporel de l'événement. Nous citons celui de Baker (1960) qui représente un accident selon une

Chapitre 1 : Accidentologie

séquence de phase, la figure 5 décrivant le déroulement de celui-ci à travers les cinq phases (ou situations) suivantes:

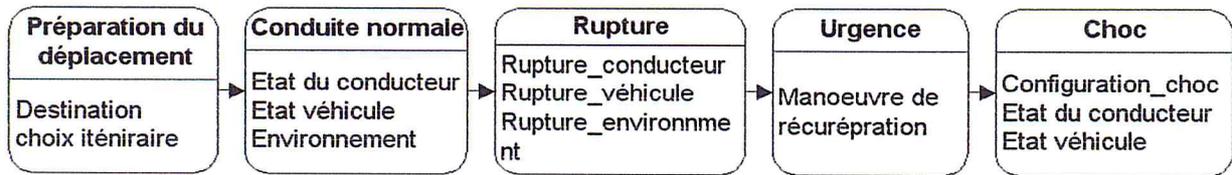


Figure 5. Modèle Séquentiel [W.Ben Ahmed, 2005]

Préparation du déplacement, correspond à l'itinéraire choisi et à la stratégie de conduite.

Une situation de conduite normale, correspond au comportement du système CVE. Elle est caractérisée par l'objectif du conducteur, sa stratégie de conduite et son mode de déplacement. Cette situation intègre aussi les cadres spatio-temporels de la fonction de conduite avant l'apparition du facteur initiateur de l'accident.

La situation de rupture, est instantanée et courte en durée. Il s'agit d'une rupture qui se produit par rapport à la situation précédente. Elle est caractérisée par un événement (manœuvre inattendue d'un autre usager, etc.) qui fait basculer le système CVE vers une situation d'urgence.

Situation d'urgence, c'est une situation dégradée dans laquelle le conducteur doit mettre en œuvre une tentative de récupération sous forme d'une manœuvre d'urgence pour faire face aux contraintes spatiales et temporelles de la situation.

Situation de Choc, marque l'échec des manœuvres d'évitement entreprises dans la situation d'urgence. Elle englobe le choc, son équivalent (chute, renversement,...) et leurs conséquences.

5.3. Modèle de fonctionnement de l'opérateur humain

Le conducteur occupe une position centrale dans la régulation du fonctionnement du modèle CVE dont il fait partie. Des études ont donc porté sur le conducteur, sur ses caractéristiques, son comportement d'une part en situation normale de conduite et d'autre part en situation d'accident [J.-Y. le Coz et Y. Page, 2003]. Elles se sont focalisées sur des études du processus de traitement de l'information par l'opérateur humain et sur les mécanismes de production de la défaillance.

6. Informations brutes sur les accidents

Les informations disponibles pour l'étude des accidents sont des données brutes relevées directement sur les lieux des accidents, ou de leurs exploitations et des analyses de ces données.

6.1. Sources des informations

Les informations sont recueillies sur les lieux des accidents et auprès des personnes impliquées ou des témoins éventuels. Dans les pays à revenus élevés, les informations proviennent essentiellement de trois sources. Dans le contexte français, ces sources d'informations sur les accidents corporels de la circulation sont :

- Les données collectées par les autorités compétentes, forces de l'ordre : Gendarmerie et Police Nationales. Malgré une uniformisation, les informations disponibles sont variables et parfois subjectives. Cet accès à l'information sur les accidents n'est pas toujours aisé (Secret d'instruction, respect de la vie privée). Les données se présentent sous forme de:
 - procès-verbaux d'accidents mortels (PVM),
 - fichiers de procès-verbaux d'accidents corporels (PVC) non mortels, au 1/50e (2% des cas sont prélevés et rapportés dans ce fichier),
 - bulletins d'analyse d'accidents corporels de la circulation (BAAC), qui synthétisent un ensemble d'accidents.
- Les données recueillies directement par des accidentologues (deux ou trois experts) sur le terrain selon la démarche des études détaillées d'accidents (EDA) [Ferrandez.F, 1995; DTLR, 2000]. Chacun des experts est responsable du recueil sur le terrain d'un type d'information bien défini:
 - recueil des témoignages et interrogation des personnes impliquées et des témoins éventuels, généralement par un psychologue,
 - collecte des informations sur le contexte, l'infrastructure et l'environnement,
 - relevé des données sur le ou les véhicules impliqués dans l'accident.
- La majorité des accidents matériels fait uniquement l'objet de constats pour les compagnies d'assurances. L'information contenue n'est pas exploitable à cause de la simplification.

6.2. Format des informations

Un des problèmes majeurs pour étudier l'accidentologie routière des pays en émergence est celui de l'inexistence ou le manque de fiabilité des données recueillies, et c'est le cas en

Chapitre 1 : Accidentologie

Algérie. En effet, les méthodologies d'analyse complètes de l'accident utilisées dans les pays développés permettent de comprendre son origine ainsi que son déroulement, afin de définir des actions efficaces et adaptées. Elles se basent sur des données d'accidentologie fiables et relativement détaillées.

Parmi nos objectifs, concevoir un fichier national des accidents de la circulation routière riche et fiable correctement renseigné à partir de formulaires de constations des accidents (questionnaire statistique national) offrant l'exhaustivité et la fiabilité des informations collectées. Pour cela, nous allons faire un tour sur les formulaires de constations utilisés dans différents pays du monde.

6.2.1. Formulaires de constations en Europe (pays développés)

Le tableau 3 liste les différents formulaires de constations des accidents en Europe.

Pays	Formulaires de constations des accidents
Belgique	Formulaire d'analyse des Accidents de la Circulation avec tués ou blessés (FAC)
Espagne	Cuestionario estadístico de Accidentes con Victimias
France	Bulletin d'Analyse d'Accident Corporel de la Circulation (BAAC)
Grèce	Road Accident Data Collection Form - DOTA (ΔΕΛΤΙΟ ΟΔΙΚΟΥ ΤΡΟΧΑΙΟΥ ΑΤΥΧΗΜΑΤΟΣ)
Hongrie	Questionnaire No. 1009: Road traffic accidents involving personal injury (Személy sérüléssel közúti közlekedési balesetek)
Italie	Incidenti Stradali
Luxembourg	Procès-verbaux
Norvège	Road Accident Report Form (Rapport om veitrafikkuhell)
Pologne	Karta zdarzenia drogowego
Royaume-Uni	STATS19 (Great Britain)
République tchèque	Statistic form on road accident registration (Statistický formulář pro registraci dopravních nehod)
Suède	Informationsunderlag Vägtrafikolycka

Tableau 3. Les différents formulaires de constations des accidents en Europe.

6.2.2. Le BAAC

Tout accident corporel de la circulation routière doit normalement faire l'objet d'une fiche, dit, le Bulletin d'analyse des accidents corporels de la circulation ou fiche BAAC (figure 6)

Chapitre 1 : Accidentologie

que les forces de l'ordre renseignent pour chaque accident corporel ayant provoqué une blessure. Ils établissent le procès-verbal décrivant l'accident qui constitue la base de la procédure aboutissant éventuellement à des sanctions pénales et à l'indemnisation des victimes.

Ce fichier constitue une base de données d'information indispensable pour connaître l'état et l'évolution de l'insécurité routière. Ses exploitations statistiques orientent les politiques nationales et locales de sécurité routière [WEB 5]. Il s'agit d'une base de données qui regroupe les variables décrivant de façon précise les caractéristiques, les lieux, les véhicules et les usagers impliqués dans un accident. Chaque variable est codifiée à l'aide d'un ensemble de valeurs. Par exemple, la variable « catégorie administrative du véhicule » est décrite avec 21 valeurs qui vont de la bicyclette au tracteur agricole. Compte tenu de la nature très différente de ces variables, celles-ci sont découpées en 4 rubriques :

- *la rubrique des caractéristiques* qui décrit les circonstances générales de l'accident (date, heure, lieu, conditions atmosphériques, type de collision),
- *la rubrique des lieux* (type de voie, profil, état, aménagements, signalisation, environnement) qui peuvent être multiples dans le cas des intersections,
- *la rubrique des véhicules impliqués* (catégorie, chargement, état, point de choc, obstacle heurté, manœuvres avant l'accident).
- *la rubrique des usagers impliqués* (place, sexe, âge, catégorie socio- professionnelle, blessures, éventuelle alcoolisation, caractéristiques du permis de conduire, nature du trajet).

Chapitre 1 : Accidentologie

IDENTIFIANT										BULLETIN D'ANALYSE D'ACCIDENT CORPOREL DE LA CIRCULATION																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
CODE UNITE		NUMERO DE PV N° FEUILLE		ETABLI PAR		1. Département		2. Police de la route		3. G.R.		4. P.A.		5. Gendarmerie		6. Autre		7. Sans catégorie		8. Sans catégorie																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
1 - CARACTERISTIQUES										2 - LIEUX																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
LUMIERE					LOCALISATION					INTERSECTION					CONDITION ATMOSPHERIQUE					TYPE DE COLLISION																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
DATE										HEURE										CODE INSEE LIEU ACCIDENT																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
10										11										12										13										14										15										16										17										18										19										20										21										22										23										24										25										26										27										28										29										30										31										32										33										34										35										36										37										38										39										40										41										42										43										44										45										46										47										48										49										50										51										52										53										54										55										56										57										58										59										60										61										62										63										64										65										66										67										68										69										70										71										72										73										74										75										76										77										78										79										80									
3 - VEHICULES										4 - USAGERS										5 - MANOEUVRE PRINCIPALE AVANT L'ACCIDENT										6 - INFRACTION																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
CATEGORIE ADMINISTRATIVE										CATEGORIE CIVILE / PROFESSIONNELLE										CATEGORIE ALCOOLIQUE										PERMIS DE CONDUIRE										TRAJET										EQUIPEMENT DE SECURITE										MANGIURE DU PIETON										DROQUE																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
LETTRE CONVENTIONNELLE										PLACE DANS LE VEHICULE										SEXE										TAUX D'ALCOOL										DATE D'OBTECTION DU PERMIS										INFRACTION										UTILISATION										LOCALISATION DU PIETON										DROQUE																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					

Figure 6. Bulletin d'analyse d'accident corporel de la circulation en France.

Ces quatre rubriques sont précédées d'un identifiant ; qui précise plus particulièrement l'unité des forces de l'ordre à l'origine de la fiche BAAC, qui est rempli par le service de police ou de gendarmerie compétent.

Chacune des variables contenues sur une rubrique doit pouvoir être reliée aux variables des autres. Par exemple, il peut être nécessaire de connaître les véhicules dans lesquels se trouvaient les usagers lors de l'accident et sur quel type de voie circulaient ces véhicules. C'est ce que l'on appelle établir un lien entre les variables qui décrivent un accident. Ce lien est établi par le biais de deux variables :

- Le code route, affecté à chaque voie un numéro d'ordre : 1, 2, 3, etc. Ce numéro est systématiquement repris pour chaque véhicule impliqué et circulant sur la voie précédemment décrite.
 - La lettre conventionnelle, affecté à chaque véhicule impliqué une lettre : A, B, C, etc. Cette lettre conventionnelle est systématiquement reprise pour toutes les personnes à l'intérieur du véhicule et pour les piétons ayant été heurtés par le véhicule.
- Cette codification particulière permet de relier entre elles les variables des trois rubriques : lieux, véhicules et usagers.

Introduction

Les données spatiales sont de plus en plus utilisées grâce à l'évolution des outils d'acquisition de l'information géo-spatiale. Les méthodes de Data Mining Classique (DM) ne sont pas adaptées pour traiter ce type de données à cause de l'interdépendance spatiale implicite entre ces dernières (différentes relations spatiales) qui n'existent pas dans les autres données. Cette particularité rend le processus d'extraction de connaissance classique sur ces données plus compliqué. Le Data Mining Spatial (le DMS) est apparu afin de résoudre ce problème, il permet d'analyser des observations d'un jeu de données stockées dans des bases de données volumineuses à références spatiales afin d'extraire de la connaissance. Les méthodes de cette technique ne sont qu'une extension de celle du Data Mining Classique en tenant compte de l'interdépendance qui existe entre les données géo-référencées.

Dans ce chapitre, nous allons parcourir le domaine du Data Mining Spatial et présenter quelques notions générales le concernant.

1. Data Mining Spatial

« Le Data Mining spatial ou la découverte de la connaissance dans les bases de données spatiales, est défini comme l'extraction de connaissances implicites, relations spatiales, ou d'autres patterns non stockés explicitement dans les bases de données spatiales » [Koperski et Han, 1995].

En d'autres termes, le Data Mining Spatial (DMS), fouille de données spatiale, est l'exploration de données ou encore l'extraction de connaissances implicite, de relations spatiales ou autre propriétés non explicitement stockées dans les bases de données spatiales, à partir d'une grande masse de données géographiques.

Avant de nous intéresser aux principales approches et méthodes du Data Mining Spatial, nous allons nous familiariser avec la notion de données spatiales.

2. Données spatiales

Les données spatiales (géo-référencées) sont des données statistiques caractérisées par une position géographique (localisation) [WEB 2].

Autrement dit, une donnée spatiale est une traduction d'une entité du monde réel. Elle est représentée dans une base de données spatiale, sous une structure qui contient la donnée descriptive (sémantiques, tabulaires ou attributaires) et la donnée spatiale définie par sa position (coordonnées spatiales).

2.1. Les données descriptives

Ce sont des informations textuelles, qualitatives ou quantitatives. Elles décrivent les caractéristiques de l'objet géographique, par exemple le nom d'une ville, nombre d'habitants, numéro de la commune, nombre de places à l'école... etc.

2.2. Les données spatiales

Les données spatiales sont des objets situés à la surface terrestre (route, bâtiment...), Elles doivent être référencés et positionnés les uns par rapport aux autres dans l'espace, structuré sous la forme de couches grâce à deux types de coordonnées, cartésiennes et géographiques.

Les données spatiales se manifestent sous trois formes (voir Figure 7):

- Le point (x,y) ou ponctuel, permet de définir le positionnement d'éléments géographiques très petits qui n'ont pas de surface réelle.
- La ligne $((x_1, y_1), \dots, (x_n, y_n))$ ou linéaire, permet de représenter les objets géographiques qui ont des formes trop étroites (l'axe d'une route, rue, rivières ..., etc.) ou qui n'ont pas de surface, mais ont une longueur (les courbes de niveau, les réseaux de communication..., etc.).
- Le polygone, pour représenter la forme et le positionnement des objets géographiques homogènes ayant une existence dans la réalité comme les pays, villes, type de sols, forêt...etc.

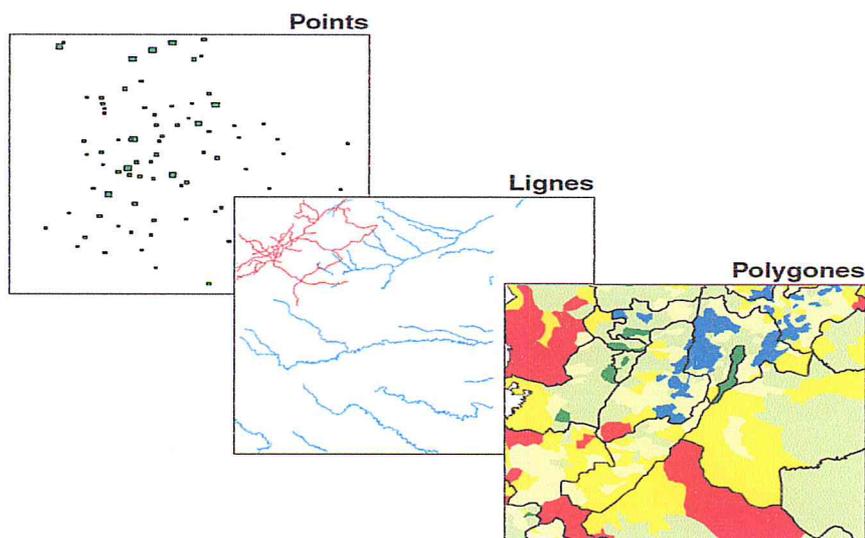


Figure 7. Présentation des objets géométrique [HAB 00].

2.2.1. Les coordonnées cartésiennes

Les coordonnées d'un point dans un espace de n dimensions ou dans un plan peuvent être exprimées sous la forme de coordonnées géocentriques (X, Y, Z) liées respectivement aux trois axes orthogonaux Ox, Oy, Oz qui passent par le centre de la Terre. Les axes Ox, Oy sont dans le plan de l'équateur, l'axe Oz est confondu avec l'axe de rotation de la Terre.

2.2.2. Les coordonnées géographiques

Les coordonnées géographiques d'un objet dans un espace donné sont des coordonnées angulaires, définies par rapport à deux plans de référence. Le premier est par rapport à l'axe de rotation de la Terre, le second est défini par le plan de l'équateur. Ils donnent généralement la localisation la plus précise par rapport à d'autres moyens de définition de la localisation, la (Figure 8) ci-dessous montre un exemple de ces coordonnées.

- **Latitude** : est une mesure angulaire permettant d'exprimer un positionnement nord-sud d'un point sur Terre, qui se situe au nord ou au sud de l'équateur, avec une étendue de 0° à l'équateur à 90° aux pôles.
- **Longitude** : est une valeur angulaire qui permet d'exprimer un positionnement est-ouest d'un point sur Terre. La longitude est donc une mesure angulaire sur 360° par rapport à un méridien de référence qui est le méridien de Greenwich, s'étendant de $+180^\circ$ à -180° et à 180° ouest.

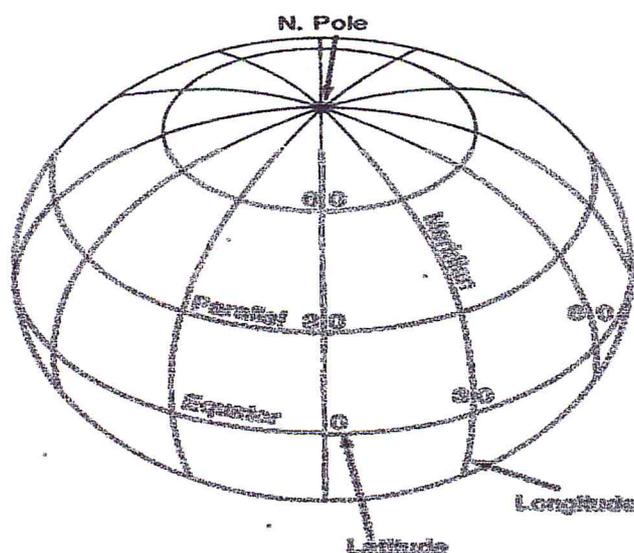


Figure 8. Les coordonnées géographiques.

2.3. Caractéristiques des données spatiales

Les données spatiales, contrairement aux données classiques (numériques ou chaînes de caractères) traitées par le Data Mining, contiennent des informations utiles à la description des relations spatiales [Chelghoum et Zeitouni, 2004].

L'une des principales caractéristiques qu'engendrent les relations spatiales par rapport aux données spatiales est la dépendance ou l'influence des voisinages entre les entités spatiales [Zeitouni et al. 2001].

2.3.1. Interdépendance

Les données spatiales se caractérisent par leur nature interdépendante. Selon la première loi de la géographie, connue aussi comme loi de Tobler : « ... les entités spatiales sont inter-reliées. Celles plus proches s'influencent plus comparé à celles plus distantes » [Tobler, 1970]. Cette définition explique l'influence d'une entité sur son voisinage.

Les données spatiales sont aussi caractérisées par une forte hétérogénéité.

2.3.2. Hétérogénéité

Selon [Jayet, 2001], « Toute analyse statistique d'une population suppose que les éléments de cette population ont des points communs, sur lesquels on peut fonder des comparaisons et asseoir des régularités ». Néanmoins, cette hétérogénéité des données spatiales complique leur analyse. L'hétérogénéité repose sur le fait que les valeurs de variation des données dépendent de leurs emplacements [Ouattara, 2010].

Il existe trois formes d'hétérogénéité [Jayet, 2001] :

- **Hétérogénéité de taille:** se présente dans le cas des données avec attributs d'échelles de grandeur de variation remarquable. Pour remédier à ce type, les analystes utilisent un indicateur de dimension, et appliquent le taux (%) aux calculs au lieu des agrégats (Exemple d'une route à grande circulation et une route secondaire).
- **Hétérogénéité de forme:** relève de l'aspect graphique des objets spatiaux. Elle est donc très compliquée à contourner, dans la mesure où la construction des indicateurs de correction est très difficile. Toutefois, il faut la prendre en considération pour cerner d'éventuels problèmes (exemple d'une maison et d'un bâtiment).
- **Hétérogénéité de structure:** concerne la structure de la population. On ne peut élaborer une comparaison entre des individus obtenus de localisations différentes, à cause de leurs structures variables. Un indicateur de référence est utilisé pour le calcul d'effets structurels, en représentant les performances de chaque zone.

La dépendance et l'hétérogénéité suscitent une troisième caractéristique, qui est la complexité.

2.3.3. Complexité

Le fait que les entités spatiales soient dépendantes les unes des autres et qu'elles soient hétérogènes, rend le traitement beaucoup plus complexe et long. Cela affecte le temps d'exécution.

2.4. Les relations spatiales

Le Data Mining spatial est fondé sur les relations spatiales. Cela vient du fait qu'il s'intéresse non seulement à la localisation des objets spatiaux et la description de ses individus mais aussi à l'étude des relations existantes qui les lient.

Une relation spatiale est définie comme étant une ou plusieurs descriptions d'objets. Par exemple, la localisation d'un objet ou sa forme est liée à la localisation ou la forme d'un autre objet [Schneuwly et al, 2010].

Les relations spatiales sont définies de manière implicite ou explicite dans les bases de données spatiales. On distingue trois types de relations [Zeitouni, 2006]:

- **Les relations topologiques (figure 9):** Une relation topologique est définie entre deux entités géographiques. Elle décrit l'interaction entre les frontières, l'intérieur et l'extérieur de deux entités. On peut citer comme relation : la disjonction, l'intersection, l'adjacence et l'inclusion. Si on pivote ou on translate les deux objets en même temps, ces relations restent préservées [Lagarde, 2009].

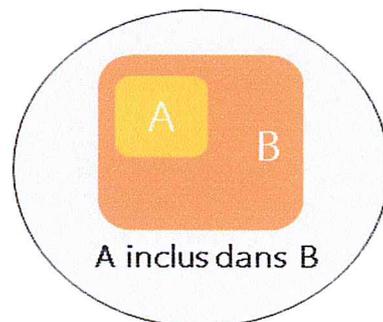
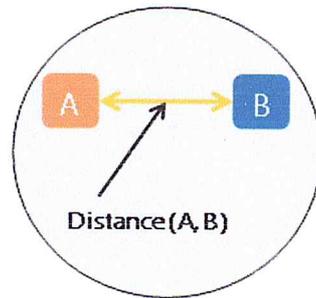


Figure 9. Exemple d'une

relation topologique

- **Les relations métriques (figure 10):** Une relation métrique entre deux objets spatiaux repose sur le calcul de distance réelle (euclidienne) qui sépare deux données spatiales pour déterminer leur proximité. Son existence revient à comparer la mesure à un certain seuil, en utilisant les opérateurs arithmétiques, en général le « < ». Son but est d'exprimer la séparation et l'écart entre les objets [Bogorny, 2005].

Figure 10. Exemple d'une



relation euclidienne

- Les relations

directionnelles permettent en général de définir avec précision le positionnement géographique d'un objet par rapport à un autre (Sud, Nord-Est, ...etc.) [Chelghoum et Zeitouni, 2004].

directionnelles (figure 11): Les

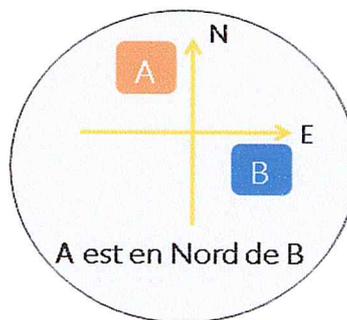


Figure 11. Exemple d'une

relation directionnelle

3. Objectifs du Data Mining spatial

Les deux objectifs majeurs de la fouille de données spatiales sont la prédiction et la description [Yubin, 2004]:

- **La prédiction:** Prédire le développement de l'évènement spatial. Elle consiste à étudier les interactions existantes entre les données spatiales et leurs associer des modèles formels afin d'obtenir des hypothèses. La prédiction s'appuie principalement sur la classification supervisée, la régression et l'association.
- **La description :** Extraire les propriétés communes entre les données étudiées ce qui permet de percevoir l'impact de la position géographique sur l'apparition et le développement de phénomènes spatiaux.

4. Quelques Domaines d'application

De nos jours, le Data Mining Spatial est utilisé dans de nombreux domaines et secteurs d'activités. Avec la diversité des méthodes qu'elle englobe, la fouille de données spatiales,

Chapitre 2 : Généralités sur le Data Mining spatial

garantit la réalisation de plusieurs tâches, qui répondent aux attentes des analystes et décideurs dans ces domaines. A titre d'exemple on peut citer:

4.1. L'environnement

Le DMS est de plus en plus utilisé pour prévoir les risques de dépassement du seuil de concentration d'ozone dans les agglomérations et l'impact qu'aurait ce dépassement sur notre planète. Il sert également à la production des cartes de prédiction des niveaux de pollution (pour pouvoir déterminer les fontaines contaminées par exemple). Ces dernières années, il est de plus en plus utilisé pour l'élaboration de nouvelles stratégies d'adaptation au changement climatique.

4.2. La séismologie

Le Data Mining Spatial est utilisé en séismologie pour prédire l'activité sismique et volcanique, qui s'étend sur une zone donnée. Il sert aussi à prédire les conséquences et les répercussions que peuvent avoir les séismes et les éruptions sur les régions concernées.

4.3. Géomarketing

C'est une branche du Marketing dont le principal objectif est la segmentation du marché, en utilisant les propriétés socio-économiques des clients dans des régions géographiques différentes.

Pour établir des stratégies du Marketing plus efficace, les activités économiques avec leur dimension spatiale ont eu recours à la fouille de données spatiales. Elle est ainsi utilisée dans diverses applications, par exemple [WEB 4]:

- Zones de chalandises: Zone d'attraction commerciale d'un point de vente tout en minimisant leur distance à la clientèle (zone d'où provient la majorité des clients).
- Localisation commerciale: Etude du potentiels, la sectorisation, l'optimisation des moyens du marketing direct.
- L'aide à l'implantation d'un nouveau point de vente
- L'orientation des stratégies de développement : Etude d'implantation, rénovation...etc.

Par exemple, les sociétés de vente par correspondance analysent, avec les techniques du DMS, le comportement des consommateurs. Ceci afin de dégager des similarités de comportement, accorder des cartes de fidélité, ou établir des listes de produits à proposer en vente additionnelle.

4.4. Astronomie

L'astronomie est l'une des sciences les plus vieilles qui s'intéresse à l'acquisition, la systématisation et l'interprétation de grandes quantités de données relatives aux galaxies. La nouvelle génération d'instruments astronomiques digitaux (télescopes, détecteurs, et caméras) sert à l'observation de régions célestes très vastes et la mesure de plusieurs paramètres, concernant des centaines de millions d'objets extragalactiques [Brunner et al, 2001].

Les méthodes traditionnelles d'identification de la population astronomique se fiaient à l'œil nu, en examinant les paramètres des parcelles de l'espace. Cependant, ces méthodes ont atteint leurs limites pour la visualisation et l'interprétation des données astronomiques. Actuellement, ces données sont stockées dans d'immenses bases de données, facilitant ainsi l'exploitation efficaces à l'aide des techniques du Data Mining Spatial. Ces dernières sont le fondement d'une nouvelle ère pour la recherche astronomique [Brunner et al, 2001].

Cette nouvelle discipline fait face à plusieurs défis dont:

- Reconnaissance des galaxies, étoiles et d'autres objets stellaires, en se basant sur les propriétés telles que les magnitudes et l'intensité.
- La fouille et test des images de télescopes pour la découverte de nouveaux objets astronomiques.
- Découverte de nouvelles agglomérations astronomiques.
- Établissement d'une carte systématique de l'univers.
- Découverte de nouveaux phénomènes entre les objets astronomiques.
- Nouvelle technique de visualisation et présentation des données astronomiques.

4.5. Criminologie

Les techniques de fouille de données spatiale permettent aux spécialistes de cerner les régions à haute concentrations de crimes, pour renforcer les mesures de sécurité.

D'après [Keefe et al, 2011], les techniques de la fouille de données spatiales offrent deux types d'analyse de crimes:

- Rétrospectives : basée sur l'examen des phénomènes criminels et les facteurs connexes ayant déjà eu lieu.
- Prédictives : identification précoce des signes de crimes pour définir les stratégies de préventions.

De nombreux travaux ont été effectués ou sont encore en phase de recherche grâce à l'introduction du DMS au domaine de la criminologie, par exemple :

- Identification des liens entre les différentes causes de crimes.

- Étude de l'influence de l'environnement socio-économique sur les criminels.
- Prédiction des zones à risque.

4.6. Accidentologie

Les accidents de la route constituent l'une des plus importantes causes de mortalité au monde et la raison principale de la congestion routière. Les administrations ou les services de police et de gendarmerie, enregistrent chaque année un taux d'accidents, de plus en plus élevé. Réduire ce taux, augmenter la mobilité sur les routes et améliorer la sécurité routière, est devenue une des priorités des autorités concernées.

Les informations d'accidentologie enregistrées, représentent une énorme source de données. L'exploitation efficace de cette source permet aux experts de la sécurité routière, d'établir des stratégies préventives, et des plans d'action, ou encore des campagnes de sensibilisation.

L'analyse des accidents de trafic routier, agit sur différents types d'informations, qui concernent l'environnement de l'accident, le conducteur, ainsi que le véhicule lui-même. Cette analyse porte également sur des informations de localisation de l'accident, ce qui a motivé l'utilisation de techniques de la fouille de données spatiale. Cette discipline permet de tenir compte des liens existants, entre l'emplacement de l'évènement et son voisinage.

L'aspect descriptif de cette analyse, vise généralement à identifier des zones à risque selon la répartition des accidents dans un espace géographique, ainsi que leur classification selon les facteurs de risque. Quant à l'aspect prédictif, il permet d'estimer des relations existantes entre les causes dans le but est de prédire de futures zones à risque d'accidentologie.

Conclusion

Le Data Mining Spatial est un domaine en plein expansion, il est de plus en plus utilisé dans différents domaines. Ce dernier est une variante de l'analyse et l'exploitation des données, qui a émergé pour prendre en considération les spécificités des données spatiales.

Il permet principalement l'analyse des interactions entre les entités spatiales. Cette discipline utilise un éventail de techniques dédiées, ayant pour objectif la description et la compréhension de phénomènes spatiaux actuels, et la prédiction d'éventuelles évolutions. Le choix de ces méthodes dépend des propriétés du domaine d'étude.

Dans ce chapitre, nous avons fait le tour sur les différentes notions liées à cette nouvelle discipline (le DMS) en vue d'avoir une idée sur celle-ci.

Les techniques du DMS se divisent suivant leurs objectifs, en des méthodes descriptives, et d'autres prédictives. Dans le prochain chapitre, nous examinons de près les différentes tâches

Chapitre 2 : Généralités sur le Data Mining spatial

de cette discipline, et nous détaillons les méthodes les plus fréquemment employées pour chaque objectif. Ainsi, nous allons voir leur application au domaine de l'accidentologie.

Les Tâches du Data Mining Spatial

Introduction

Le Data Mining spatial se présente comme un catalogue de méthodes, réalisant des tâches fonctionnelles qui peuvent être classées en deux familles. La première est orientée vers la découverte et l'exploration (description synthétique, classification, recherche de tendances) et la seconde réunit des analyses dans un objectif décisionnel ou prédictif (recherche de règles, modélisation). En outre, le processus de Data Mining n'est pas une exploitation simple des données mais un processus complexe faisant appel à des techniques de visualisation, de gestion de bases de données, d'analyse de données statistiques et de l'intelligence artificielle.

Dans ce chapitre, nous allons étudier les principales tâches du DMS telles que la classification (supervisée et non supervisée), l'association et la géo-visualisation. Ainsi, nous aborderons les techniques les plus répandues pour chaque tâche.

1. Classification spatiale supervisée

La recherche de règles de classement vise à structurer un ensemble d'objets en classes d'objets ayant des propriétés communes. Elle permet d'affecter les objets qui ont des propriétés communes à une classe prédéfinie. Contrairement au clustering qui identifie des classes, 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. La classification spatiale supervisée est souvent exprimée par les arbres de décision.

Un arbre de décision spatial est une structure de connaissances hiérarchique, correspondant à une séquence de règles de décision. Le principe de la classification par arbre de décision spatial est de diviser récursivement la population d'apprentissage en des populations plus homogènes, en appliquant un critère de subdivision [CHELGHOUM et al, 2002]. Les résultats de cette classification sont présentés sous forme d'arbres de connaissances. Ce dernier est facilement interprétable par un utilisateur qui n'est pas familier avec le domaine d'analyse de données. La figure 12 montre un exemple d'arbre de décision spatial.

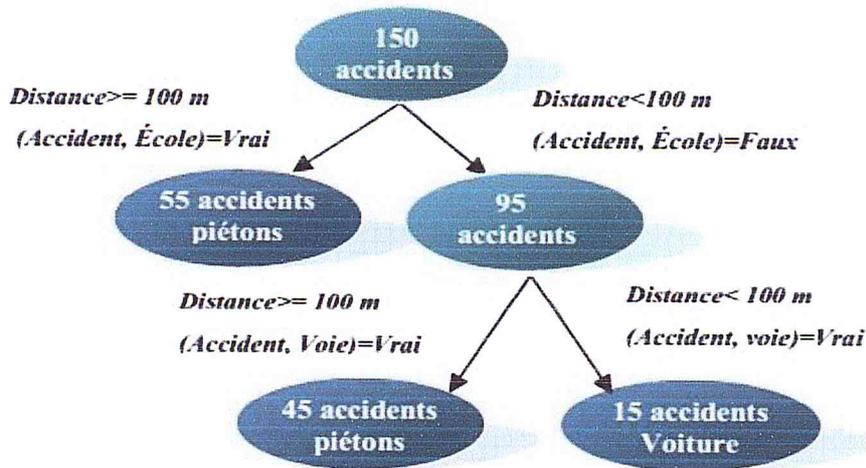


Figure 12. Arbre de décision spatial.

2. Classification non supervisée (clustering)

C'est la tâche descriptive la plus connue de la fouille de données spatiales. Le clustering est une méthode de classification automatique non supervisée qui regroupe des objets dans des classes. Son but est de maximiser la similarité intraclasse⁵ et de minimiser la similarité interclasses⁶.

La transposition au domaine spatial des méthodes de clustering s'appuie sur une mesure de similarité d'objets localisés suivant leur distance métrique. Néanmoins, la finalité du clustering spatial n'est pas tant de former des classes que de détecter des concentrations anormales (par exemple, détecter un point chaud dans l'étude de criminalité, ou des zones à risque en accidentologie). Avant de présenter quelques méthodes dédiées au clustering, nous présentons ci-dessous un diagramme représentant une taxonomie des méthodes de clustering (Figure 13).

⁵Intraclasse: les objets appartenant à la même classe.

⁶Interclasse: les objets appartenant aux classes différents.

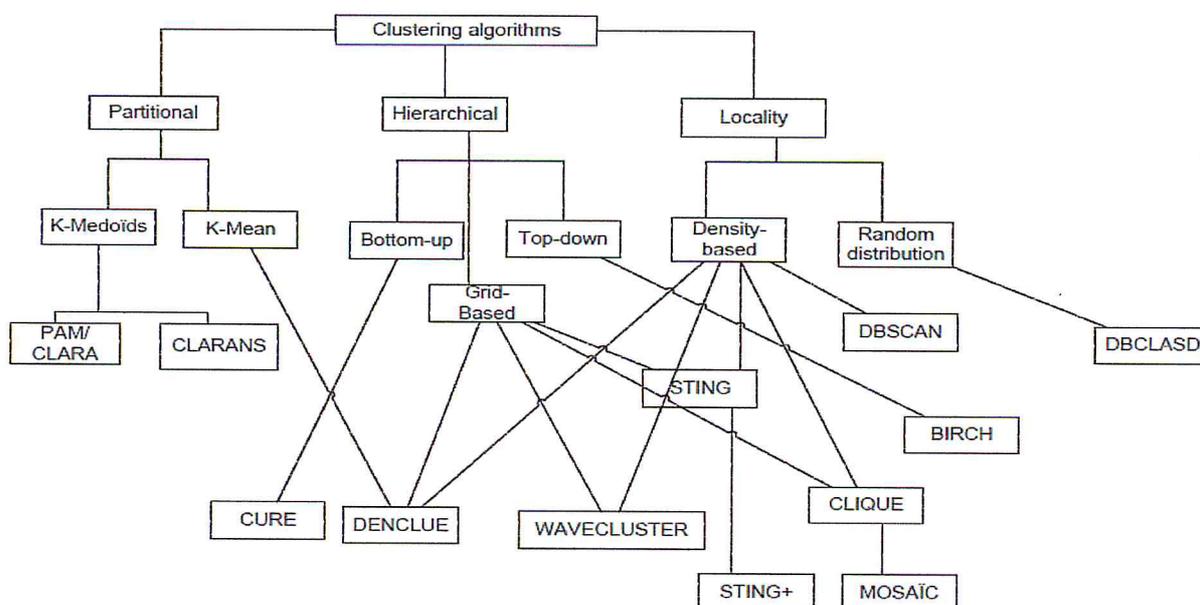


Figure 13. Digramme des méthodes et algorithmes de clustering spatial [Mamadou, 2010].

2.1. Clustering par partitionnement

Consiste à découper l'ensemble des objets en des classes homogènes (ayant les mêmes caractéristiques) avec un nombre de classes prédéfini. Chaque objet appartient seulement à une seule classe.

Le principe de l'approche par partitionnement est de grouper les données autour du point de référence dans le but de construire les classes en cherchant à trouver les k meilleures partitions d'un ensemble de n objets (données) [Kumar et al, 2012].

On distingue deux importantes familles d'algorithmes de clustering par partitionnement : K-means et K-medoids.

2.1.1. k-Means

C'est le représentant le plus répandu de cette famille. Dans cet algorithme, le point de référence est le centre de chaque cluster. Son objectif est de partitionner un ensemble de n objets (x_1, x_2, \dots, x_n) en k groupes (G_1, G_2, \dots, G_k) tout en minimisant la distance entre les objets d'un cluster et leurs centroïdes⁷ respectifs. La distance utilisée est la spatiale pour les emplacements géographiques ou l'eulidienne pour les points numériques. Chaque cluster est représenté par la moyenne (mean) ou la moyenne pondérée (centroïde) [Roussel, 2012].

Dans ce qui suit, nous donnons les étapes de l'algorithme K-means :

⁷ Le centroïde (centre de gravité) d'une classe C , est un vecteur dont chaque composante correspond à la moyenne arithmétique de tous les objets de cette classe.

Chapitre 3 : Les Tâches du Data Mining Spatial

- **Initialisation:** A partir d'un jeu de données, choisir les centroides initiaux de manière aléatoire.
- **Assignment :** Associer les données au plus proche centre.
- **Mise à jour :** A chaque itération, calculer les nouveaux centres de gravité pour chaque classe et réaffecter les données de manière à ce que la distance entre les données de chaque groupe soit minimale.
- **Stabilité :** l'algorithme s'achève lorsqu'aucune réaffectation n'est possible.

La figure 14 ci-dessous montre le déplacement du centre de gravité après chaque itération.

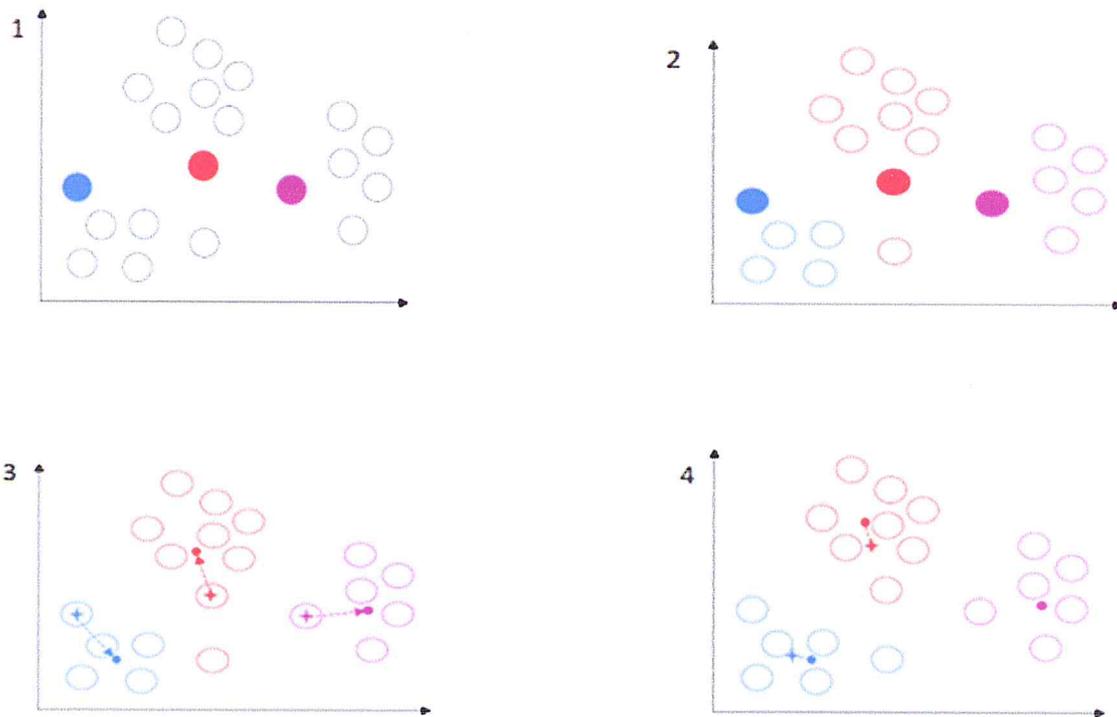


Figure 14. Partitionnement basé sur k-means.

2.1.2. K-Medoids

L'algorithme de K-medoids est l'un des dérivés de l'algorithme K-means. Sa seule différence est que l'algorithme K-medoids choisit ses centres de gravités parmi les objets à classer que nous appelons « médoïde⁸ ». Il est moins sensible aux données incohérentes et outliers⁹, il offre plus de stabilité par rapport au K-Means car il se base sur la notion de médoïds pour la création des classes [Ouattara, 2010]. Son inconvénient réside dans son

⁸ Un médoïde est l'élément d'un ensemble qui minimise la somme des distances entre lui et chacun des autres éléments de cet ensemble, Medoids sont les plus couramment utilisés sur les données quand un centre de gravité ne peut pas être défini.

⁹ (Johnson, 1992) définit un outlier comme une observation dans un ensemble de données qui semble être incompatible avec le reste de l'ensemble de données.

Chapitre 3 : Les Tâches du Data Mining Spatial

exigeance d'un temps élevé pour le traitement par rapport au K-Means, Il existe trois algorithmes très populaires dans cette famille : PAM, CLARA, CLARANS.

- **PAM (Partitionning Around Medoids):** PAM est une technique de clustering utilisant les K-medoids pour l'identification des clusters. Il permet de déterminer K-clusters en trouvant pour chacun d'eux l'objet le plus centralement localisé dans le cluster : « le Médoïd ». La démarche de PAM est la suivante : après avoir sélectionné arbitrairement K medoid au départ, les objets non sélectionnés sont assignés à chaque cluster en fonction de la distance séparant chaque objet de chaque cluster. Un objet sera assigné à un cluster si la distance ou fonction de dissimilarité est minimale comparée aux autres. Les étapes suivantes consistent pour PAM à effectuer des itérations afin d'améliorer la qualité des clusters trouvés. Ainsi pour chaque cluster, PAM essaie de retrouver un médoïd de meilleure qualité. Si cela est le cas, le médoïd en cours est remplacé et les distances entre non-médoïds et le nouveau cluster sont évaluées afin de redéfinir l'appartenance des objets aux clusters. L'itération continue jusqu'à ce que les K-clusters trouvés soient de meilleure qualité [Ouattara, 2010].

On note que PAM est performant dans le traitement des ensembles de données moyennement volumineux. Plus les données sont larges, plus on note une dégradation des performances de PAM. Ce qui a d'ailleurs motivé la mise en œuvre d'un nouvel algorithme dénommé CLARA.

- **CLARA (Clustering LARge Applications):** Le problème avec PAM est sa complexité croissante en fonction du nombre d'objets n à traiter (le calcul de la matrice de dissimilarité $O(n^2)$). Pour cela, CLARA calcule cette matrice de dissimilarité pour un échantillon du jeu de donnée [Kumar et al, 2012]. Il construit un ensemble d'échantillons de même taille, puis applique PAM sur chacun d'eux. Les medoids retenus sont ceux de l'échantillon pour lequel PAM minimise le plus la dissimilarité intra groupe. Ainsi CLARA peut traiter des jeux de données plus volumineux.
- **CLARANS (Clustering Large Applications based on RANdomized Search):** C'est l'algorithme le plus adaptée aux jeux de données spatiales. Il utilise une abstraction de graphe pour représenter le problème de recherche de l'ensemble de medoids optimal. Il procède d'abord par la construction d'un graphe où chacun des nœuds contient k points du jeu de données. Deux nœuds sont voisins s'ils ne diffèrent que d'un seul point. Il applique PAM pour un nœud initial, puis parcourt un sous ensemble des nœuds voisins et leur applique PAM, pour trouver celui qui minimise la dissimilarité

générale [Kumar et al, 2012]. Pour l'application de CLARANS, deux paramètres doivent être spécifiés en entrée: le nombre de nœuds voisins à visiter à chaque étape et le nœud initial.

2.2. Clustering hiérarchique

Les méthodes hiérarchiques permettent l'accès aux différents niveaux de granularité. Leur objectif est d'avoir des résultats sous forme d'arbre hiérarchique appelé dendrogramme où chaque sous arbre représente un cluster [Elghazel, 2007].

Ces méthodes présentent plusieurs avantages tels que leur flexibilité concernant le niveau de granularité (le choix d'atteindre une classe fine ou épaisse). Elles présentent également quelques inconvénients comme la difficulté de fixer un critère d'arrêt, le besoin d'accéder plusieurs fois aux données et leur coût de calcul élevé [Cleuziou, 2004].

Il existe deux approches de clustering hiérarchiques. L'approche agglomérative (approche ascendante), et l'approche divisive (descendante).

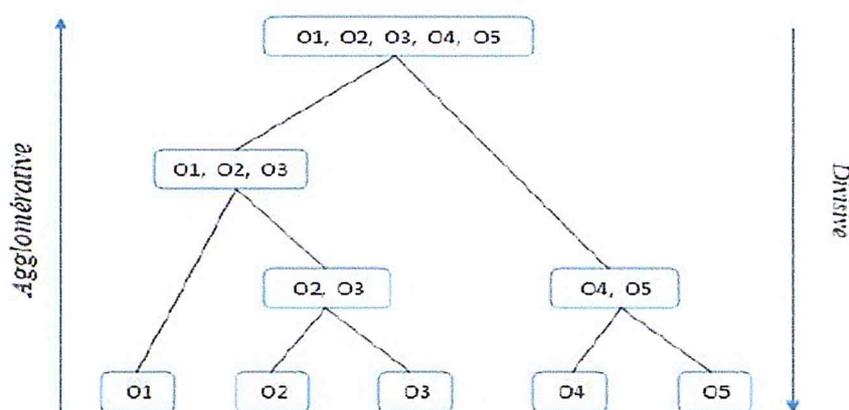


Figure 15. Méthodes divisives et agglomératives pour le clustering.

2.2.1. Approche agglomérative (ascendante)

Les méthodes agglomératives commencent toujours avec n clusters (n étant le nombre de données en entrée), et au fur et à mesure, deux clusters sont fusionnés jusqu'à l'obtention d'un seul cluster au final (voir Figure 15).

Le processus de fusionnement des clusters s'effectue en maximisant, à chaque étape, la similarité entre les deux clusters à fusionner. Les mesures de similarités les plus connues sont: saut minimal «*Single Linkage*», saut maximal «*Complete Linkage*» ou diamètre, saut moyen «*Average Linkage*» et la distance aux barycentres [Cabanes, 2010].

Parmi les algorithmes les plus utilisés dans le clustering hiérarchique ascendant, on trouve l'algorithme BIRCH, CURE et CHAMELEON.

2.2.1.1. BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)

Il a été développé par [Zhang et al, 1996]. Son principe s'appuie sur les approches ascendantes. Il n'agit pas sur tout l'ensemble de données, contrairement aux autres, mais sur un résumé compact de cet ensemble. Il peut, donc, être appliqué sur de gros jeux de données en utilisant une mémoire limitée.

Cet algorithme utilise une structure de donnée hiérarchique CF-tree (Clustering Feature-tree) dans le but de classifier dynamiquement les données selon l'ordre de leurs arrivées. Il réalise la classification en deux grandes étapes [Kolatch, 2001].

- **Première étape:** Parcourir la BDD en entier pour transformer toutes les données sous forme d'une structure CF-tree dont chaque feuille représente un résumé d'une donnée.
- **Deuxième étape:** Appliquer une technique de clustering classique, tel que K-means ou K-medoids, sur ces résumés.

2.2.1.2. CURE

L'algorithme considère tous les points de données comme des classes indépendantes. Ensuite et dans chaque classe, il choisit un nombre C de données bien réparties. Ces données sont utilisées pour identifier la forme et la taille de la classe. Par la suite, en utilisant une fonction prédéterminée, il réduit le nombre de données sélectionnées autour du centre de gravité de la classe. Puis, il fusionne successivement les paires de classes les plus proches jusqu'à atteindre le nombre k de classes désirées. Il calcule le représentant de chaque classe obtenue en choisissant des points bien éparpillés afin de diminuer leurs distances par rapport au centre de la classe. Ce traitement se fait en passant par plusieurs itérations.

En premier lieu, le premier point éparpillé est choisi. Ensuite, les points les plus éloignés de ce point sont choisis. Enfin, après avoir constitué l'ensemble C des points éparpillés, les représentants sont obtenus par rétrécissement de ces points par une fraction vers le centre de la classe [Zhao et Karypis, 2002].

2.2.1.3. CHAMELEON

CHAMELEON (cf. [Karypis et al, 1999] pour plus de détail) est une technique combinant le partitionnement et un clustering hiérarchique. Similaire à CURE, il améliore la qualité des clusters par l'application d'un critère élaboré dans la fusion de deux clusters. En effet, deux clusters seront fusionnés si l'inter-connectivité et la proximité du cluster fusionné est assez similaire à l'inter-connectivité et à la proximité des clusters à fusionner. L'inter-connectivité et la proximité sont déterminées en prenant en compte les caractéristiques propres à chaque cluster.

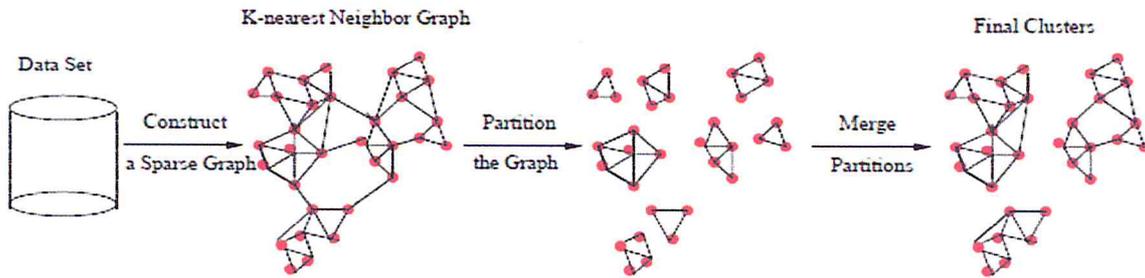


Figure 16. Un aperçu de CHAMELEON [Karypis et al, 1999].

2.2.2. Approche divisive (descendante)

Contrairement aux méthodes agglomératives, les méthodes divisives débutent par un cluster contenant toutes les n données. Au fur et à mesure, un cluster est choisi et scindé en deux clusters descendants suivant une mesure de dissimilarité, de telle sorte à obtenir n clusters en fin de traitement (voir Figure 15).

2.3. Clustering basé densité

C'est une famille d'algorithmes de la classification non supervisée basée sur le calcul d'une mesure de densité. Les clusters sont considérés comme des régions à haute densité séparées par des régions à faible densité. La densité comme le montre la figure 17 est une mesure qui évalue le nombre d'objets dans le voisinage (ou encore le ratio nombre d'objets / espace étudié) [Ouattara, 2010].

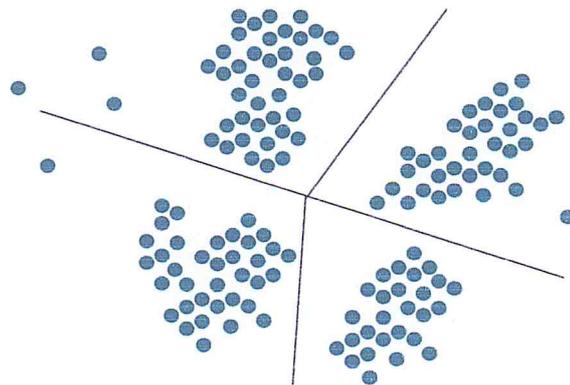


Figure 17. Variations de densité dans un espace d'étude.

Il existe deux types d'algorithmes basés densité : ceux basés sur la connectivité de densité (ex: DBSCAN), et ceux basés sur la fonction de densité (ex : DENCLUE). L'algorithme le plus appliqué est donné pour chaque famille et détaillé dans ce qui suit.

2.3.1. Algorithme DBSCAN (*Density Based Spatial Clustering Application with Noise*)

À la différence de CLARANS, DBSCAN effectue un clustering basé sur la densité. Cette technique se base sur le principe selon lequel la densité est plus élevée à l'intérieur d'un cluster qu'à l'extérieur. L'algorithme prend deux(2) paramètres en entrée : Eps et MinPts. MinPts représente le nombre de points maximaux à l'intérieur d'un cluster. Eps quant à lui représente la distance maximale qui devrait séparer un point quelconque du centre du cluster [Kolatch, 2001].

Ces deux paramètres doivent être déterminés avant chaque exécution de DBSCAN. Ce qui constitue un des inconvénients majeurs de cet algorithme en plus de l'incapacité de traiter les données à grande dimensionnalité. En termes d'avantages, on note que DBSCAN est performant dans le traitement des larges volumes de données ainsi que des données déviées et inconsistantes.

2.3.2. Algorithme DENCLUE (*Density based Clustering*)

DENCLUE est un algorithme qui hérite tout à la fois des méthodes basées sur le partitionnement, la densité et la hiérarchie. Cet algorithme s'appuie sur l'hypothèse selon laquelle l'influence d'un point sur ses voisins peut être modélisée mathématiquement sous forme d'une fonction dénommée fonction d'impact. Cette fonction est alors appliquée à chaque point de l'espace de données pour ensuite dériver la densité qui n'est autre que la somme des fonctions d'impact [Kolatch, 2001].

On note que cet algorithme est efficace dans le traitement des données à large dimensionnalité ainsi que dans le traitement de clusters de formes différentes.

2.3.3. Algorithme DBCLASD (*Distribution Based Clustering of Large Spatial Databases*)

Il s'agit d'un algorithme de clustering basé sur la densité qui contrairement à DBSCAN suppose que les points sont uniformément repartis à l'intérieur d'un cluster. En effet, on note que la distance d'un point quelconque à son voisin est plus petite à l'intérieur d'un cluster qu'à l'extérieur.

Selon [Kolatch, 2001] et [Xu et al, 1998], DBCLASD est un algorithme incrémental c'est-à-dire que les points sont traités en fonction des points précédemment traités. Cela fait de DBCLASD un algorithme dépendant de l'ordre d'entrée des données. Cependant, pour remédier à cette faiblesse, l'algorithme retraits les points qui n'ont pas été assignés avec succès à un cluster. Ce qui peut entraîner une réassignation des points à d'autres clusters.

En termes d'avantages, DBCLASD est capable de traiter des larges volumes de données ainsi que des données à haute dimensionnalité. En outre, à la différence d'autres algorithmes, il ne requiert pas de paramètres en entrée et traite efficacement les données inconsistantes car fondé sur une probabilité basée sur le facteur distance. A son désavantage, DBCLASD, comparé à d'autres algorithmes, consomme beaucoup de ressources. En plus, l'hypothèse selon laquelle les points sont uniformément répartis est un facteur limitatif quant à son efficacité.

2.4. Clustering basé grille

C'est une approche permettant de réduire la complexité du clustering. L'idée clé de ces méthodes est de subdiviser l'espace d'étude en un nombre fini de cellules homogènes formant ainsi une grille [Ouattara, 2010], et les cellules denses forment des clusters (figure 18).

Cependant, deux paramètres doivent être définis en entrée: le nombre de cellules dans toute la grille et le nombre d'objets dans chaque cellule. Globalement, ces algorithmes peuvent être classés comme algorithmes par partitionnement ou hiérarchique [Hanoi et Tuan, 2004], les algorithmes les plus connus sont Sting, Clique et Wave Cluster.

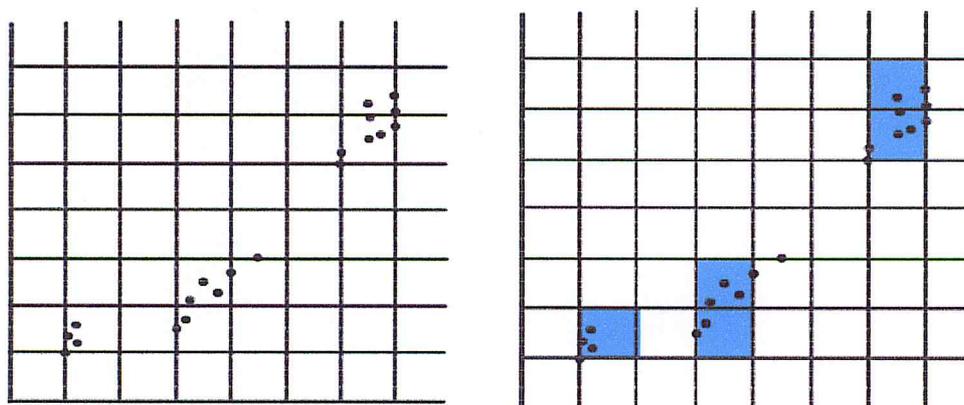


Figure 18. Clustering basé sur les grilles.

2.4.1. STING (Statistical Information Grid Based Method)

STING divise la zone spatiale en différentes cellules régulières suivant par exemple la latitude et la longitude. Les cellules sont structurées sous forme hiérarchique [Kolatch, 2001], dans le sens où chaque cellule est récursivement divisée en deux sous-cellules de niveau inférieur.

A chaque niveau, la pertinence des cellules est évaluée et seules les cellules les plus significatives sont gardées. Le parcours de ces niveaux se fait de manière ascendante.

En termes d'avantages, cet algorithme n'est pas sensible à l'ordre des données ni aux données inconsistantes et déviées. Il s'en sort efficacement dans le traitement des larges volumes de données et fournit d'ailleurs des clusters de qualité tant que la granularité reste fine [Ouattara, 2010].

2.4.2. Clique¹⁰ (Clustering In Quest)

CLIQUE est un algorithme de clustering basé sur la densité et les grilles qui a comme particularité la possibilité de traiter les données à grandes dimensions c'est-à-dire avec un nombre élevé d'attributs.

Afin d'obtenir une approximation de la densité, chaque dimension est partitionnée en intervalles de taille égales en utilisant une approche ascendante. Chaque partition ayant alors le même volume de données, les densités sont ensuite dérivées selon le nombre de points contenus dans chaque partition. Ces densités aident à l'identification automatique des sous-espaces dans lesquels les clusters sont déterminés en séparant les points selon une fonction de densité et un regroupement des partitions à haute densité connectées dans le sous-espace.

L'identification des clusters au niveau de CLIQUE est réalisée en trois (3) grandes phases :

- Identification des sous espaces contenant les clusters : cette opération est réalisée en utilisant un algorithme ascendant de recherche d'unités denses.
- Identification des clusters : consiste à trouver les composants connectés en utilisant les sommets des unités denses. L'identification des clusters est bien entendue fonction du nombre d'unités denses.
- Génération d'une description minimale des clusters grâce aux composants déterminés dans la phase précédente.

CLIQUE est efficace dans le traitement de données contenant un certain taux de données déviées et inconsistantes. Il requiert de l'utilisateur, deux paramètres représentant le seuil de densité et le nombre d'intervalles d'égale longueur. On note toutefois que la qualité des clusters déterminés peut être mise en cause lorsque le seuil de densité est moins élevé.

¹⁰ Cf. (Agrawal et al, 1998) pour plus de détails.

2.4.3. Wave Cluster

Transformation de l'espace d'étude en un domaine de fréquence, où les signaux à haute fréquence correspondent aux limites des clusters, et les faibles représentent le contenu de ces derniers. Le regroupement des objets dans les clusters tient compte de deux critères: le nombre minimum d'éléments dans chacun des clusters et la distance entre deux éléments supérieurs à un certain seuil [Hannou et Medani, 2013].

2.5. Estimateur à noyau de densité

L'estimateur à noyau de densité «*Kernel Density Estimation*» est une estimation de la densité. Elle combine le principe général du clustering basé grille et basé densité pour cerner les zones de concentration de la population étudiée [Di Salvo et al, 2005]. Cette technique a pour but de comprendre les phénomènes géographiques liés entre eux.

L'espace d'étude est représentée comme une grille, divisée en cellules de tailles égales (taille de la cellule à spécifier). Le KDE associe à chaque point de cet espace un voisinage circulaire, puis estime sa densité.

Les voisinages denses (dont la mesure de densité dépasse un seuil prédéfini) sont regroupés pour former des zones fortement peuplées.

La fonction d'estimation est dite à noyau de densité car elle considère un noyau autour de chaque point.

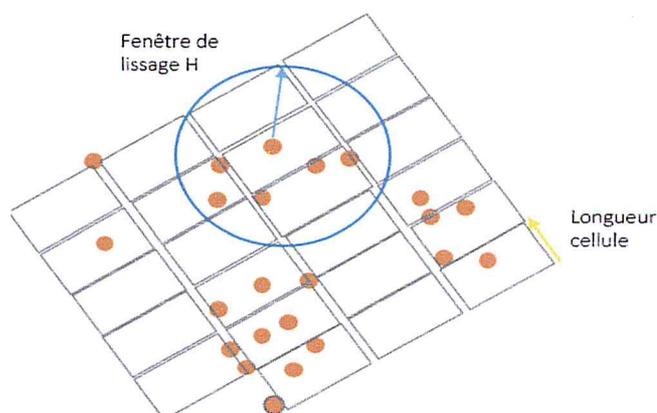
Ce noyau est évalué par une fonction dans la formule de calcul de densité qui est donnée dans ce qui suit [Di Salvo et al, 2005] :

$$F(c) = \frac{1}{n \cdot h^2} \sum_{i=1}^n k\left(\frac{x_i - c}{h}\right)$$

Avec :

- c : un point de l'espace d'étude ayant une position géographique
- x_i : un point du voisinage construit autour de c .
- h : rayon de recherche (ou fenêtre de lissage quand la fonction noyau est la gaussienne [Di Salvo et al, 2005]).
- n : le nombre de points dans le voisinage

Figure 19.illustration de quelques paramètres du KDE



Chapitre 3 : Les Tâches du Data Mining Spatial

- $k(x)$: fonction noyau (Kernel) exprime la distance d'un point du voisinage du centre.

Le choix de ces paramètres influe considérablement sur les résultats de la technique KDE.

La fenêtre de lissage fixe la précision désirée dans l'estimation. Plusieurs fonctions sont utilisées comme noyau : l'uniforme, la quartique, la triangulaire, l'exponentielle et la normale (gaussienne). Cette dernière est la plus couramment employée [Hannou et Meddani, 2013].

L'estimation à noyau de densité s'applique pour filtrer les régions à haute densité d'évènement, dans l'espace étudié. Par exemple en médecine, les stratégies de prévention contre la propagation d'épidémies sont établies suivant la distribution des décès. Le KDE peut être hybridé avec une méthode de clustering, tel que K-Means, pour identifier les zones à haut risque de décès où peuvent être organisés des campagnes de vaccination.

La popularité de cette approche est due à sa simplicité et à la qualité de visualisation qu'elle fournit. Cela permet une meilleure présentation des résultats [Andersson, 2009].

3. Association spatiale

L'association consiste à rechercher les règles de dépendance entre certaines caractéristiques des données, le plus souvent sous forme de règles d'association [Koperski and Han, 1995]. Le but des règles d'association est la découverte de relations existantes entre les objets d'un phénomène étudié. Elles permettent de prédire la probabilité qu'un événement futur se produise, en se basant sur les événements actuels et leurs relations avec cet événement futur.

L'ensemble des objets et événements étudiés en association spatiale doit avoir au moins une dimension spatiale (liée à la position géographique).

L'association peut être appliquée à plusieurs domaines tels que l'accidentologie, le marketing, la bio-informatique, les diagnostics médicaux, la criminologie... etc.

Par exemple, les grands magasins collectent énormément de données sur les achats des consommateurs via les tickets (Tableau 4). Chaque rang correspond à une transaction et reporte le numéro de ticket ainsi qu'une liste de produits achetés. Les commerçants sont intéressés par l'analyse de ce type de données afin de mieux comprendre le comportement des achats de leurs clients.

Transaction	Items
1	{pain, lait}
2	{pain, couches, bière, œufs}

Chapitre 3 : Les Tâches du Data Mining Spatial

3	{lait, couches, bière, coca}
4	{pain, lait, couches, bière}
5	{pain, lait, couches, coca}

Tableau 4. Représentation des achats des clients sous forme de tableau.

Des méthodes d'analyse d'associations sont utilisées pour découvrir des relations dans de grandes bases de données. Par exemple, on peut tirer la transaction suivante du tableau 4:

{Couches} → {Bière}.

Chaque transaction contient un sous ensemble d'items. On appelle Item, tout objet dans le domaine d'étude. Il est dit spatial si sa position géographique est connue ou peut être déterminée. L'utilisation des règles d'association peut être confrontée à deux problèmes:

- L'extraction des items peut être coûteuse en présence de base de données de taille importante.
- Certaines associations peuvent être fausses, sans importances ou apparaissent seulement par hasard.

3.1. Règle d'association

Une règle d'association est une relation de la forme si X alors Y, à laquelle on associe une valeur de probabilité de réalisation, où X et Y sont des prédicats relatifs au domaine étudié. Pour évaluer la probabilité de la règle d'association, deux métriques sont employées: le support %s et la confiance %c.

Les deux notions prédicat et items sont définis comme suit:

- **Prédicat:** Descripteur portant sur un ou plusieurs items. Il peut caractériser une opération sur l'item ou des interactions entre items (contient, intersection..). Ces prédicats sont dit spatiaux s'ils décrivent la position géographique relative entre items (derrière, à gauche, à droite ...).
- **Item:** Objet ou évènement dans le domaine de l'étude. Il est dit spatial si sa position géographique est connue ou peut être déterminée.

Une règle d'association est représentée par la relation $X \rightarrow Y$ (%c, %s) où X et Y sont des prédicats du domaine d'étude et satisfaisants:

- X et Y n'ont pas d'éléments en commun $X \cap Y = \emptyset$
- X ou Y doivent avoir au moins une dimension spatiale (item spatial ou relation spatiale) : $X \cup Y$ a au moins une composante spatiale.

- $\%s$: métrique de support, évaluer la probabilité d'existence de X, Y en même temps dans l'ensemble de données: $P(X \cup Y)$.
- $\%c$: métrique de confiance. Evaluer la probabilité d'existence de Y, sachant que X existe dans l'ensemble de données: $P(Y | X)$. La confiance mesure la pertinence de l'inférence faite par une règle (plus la confiance de $X \rightarrow Y$ est élevée, plus la probabilité d'observer Y avec X est forte).

3.2. Extraction des règles d'association

L'extraction des règles d'association spatiales d'un jeu de données revient à définir les règles les plus significatives dans chacun des niveaux de granularité définis, et pour évaluer l'importance d'une règle d'association on se base sur les métriques de probabilités qui lui sont associées. Ces métriques acceptent des seuils qui filtrent les items fréquents et les règles fortes. On définit $\text{minSup}(L)$, et $\text{minConf}(L)$ comme les seuils respectifs de support et de confiance pour le niveau de granularité L.

- **Item fréquent** : Un item est dit fréquent dans un niveau de granularité L si sa probabilité $P(\text{item}) \geq \text{minSup}(L)$ et que tous les ancêtres de l'item (items correspondants dans les niveaux de granularité précédents) sont fréquents dans leurs niveaux respectifs.
- **Règle d'association forte** : Une règle d'association est forte dans un niveau de granularité L si $\%s \geq \text{minSup}(L)$ et $\%c \geq \text{minConf}(L)$.

L'approche générale pour extraction des règles d'associations forte est la suivante:

Dans chaque niveau de granularité L:

- Définir l'ensemble de tous les items de l'étude.
- Extraire le sous-ensemble des items fréquents du niveau L.
- Elaborer les règles d'association entre tous les éléments de l'ensemble des items fréquents.
- Evaluer le support et la confiance de chacune des règles établies.
- Définir l'ensemble des règles d'associations fortes dans le niveau L.

Entre les niveaux de granularité :

- Elaborer les règles d'association entre les ensembles des items fréquents de tous les niveaux [Hannou et Medani, 2013].
-

4. Géo-visualisation

C'est la visualisation de l'information géographique (Figure 20). Elle traite des données ayant une dimension géographique pour visualiser les connaissances et les relations entre ces données [Kraak, 1997].

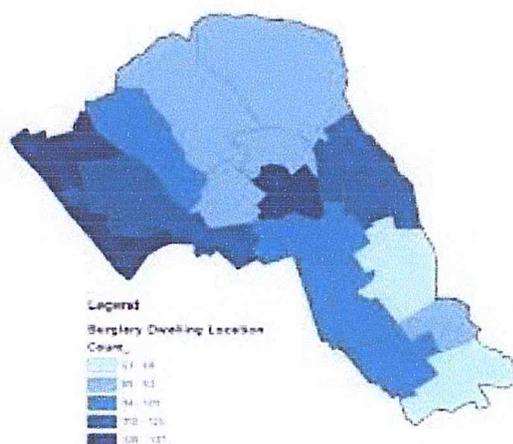


Figure 20. La visualisation des données géographiques

Cette technique manipule des données ayant une dimension géographique, afin de visualiser les connaissances ainsi que les relations implicites entre ces données. Elle intègre aussi des méthodes de visualisation scientifique, cartographie, analyse d'images, et d'analyse exploratoire des données.

5. Application du DMS au domaine de l'accidentologie

Dans la littérature, trois axes suivants, dans l'application du DMS à l'accidentologie, ressortent :

- Identification des zones chaudes d'accidentologie pour cerner les régions à haut risque.
- Classification des accidents selon les caractéristiques communes relatives à l'environnement d'accident (temps, route), au conducteur et au véhicule lui-même.
- Prédiction des futurs accidents, par segmentation des routes, et analyse des facteurs d'accidents.

Ces objectifs sont les plus étudiés dans les travaux de recherches, pour l'analyse d'accidentologie, La première permet à la police de mieux organiser ses patrouilles pour surveiller ces régions. Elle permet aussi aux institutions d'intervention médicales de préparer

des plans d'action pour qu'elles puissent organiser leurs ressources de manière optimale. Quand à La classification, elle permet d'étudier la répartition des accidents sur un espace géographique, et d'expliquer les différents facteurs de risques. Elle fournit ainsi aux spécialistes, les informations nécessaires pour mieux comprendre les différents scénarios d'accidents. Détailler ces scénarios permet d'élaborer une mesure préventive propre à chaque scénario. La troisième tâche a un impact préventif. Elle permet d'extraire la connaissance à partir d'un grand jeu de données, pour être exploitée pour la sensibilisation des conducteurs ou pour prendre des mesures de sécurité afin de réduire le taux des accidents.

Dans ce qui nous allons présenter des études de l'application de ces deux tâches dans le domaine de l'accidentologie.

5.1. Première étude

Cette étude porte sur l'article « Estimateur à noyau de densité pour les accidents de la route dans un réseau spatial » [Xie et al, 2008]. L'objectif de ce travail est la détection des régions à haute fréquence d'accidents en utilisant une technique de clustering basée densité (estimateur à noyau de densité-KDE-). La méthode a été appliquée sur la base de données du réseau de transport et des accidents de la circulation, survenus l'année 2005 à Bowling Green, état de Kentucky, USA.

Le principe général de l'étude repose sur la méthode de l'estimateur à noyau de densité. Pour chaque point accident de l'espace d'étude, on associe un voisinage circulaire. Ces voisinages sont, ensuite évalués par une mesure de densité, pour déterminer les plus significatifs pour l'étude. Les voisinages denses se regroupent pour former des zones chaudes (hot spots). La fonction noyau choisie pour le calcul de la densité est la fonction gaussienne.

Un exemple de série de tests réalisée : Six tailles pour la fenêtre de lissage ont été prises : 20m, 100m, 250m, 500m, 1000m, 2000m, avec une taille de cellule fixe de 10m a permis d'avoir les résultats que montre la figure ci-dessous.

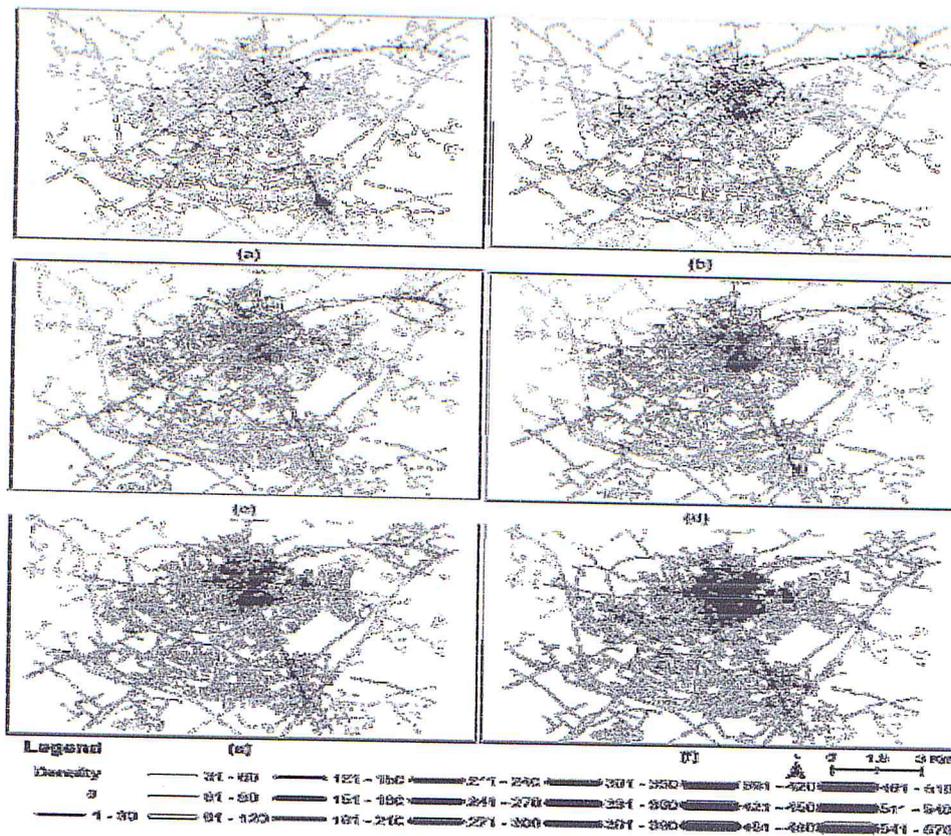


Figure 21. L'impact de la taille de la fenêtre de lissage sur l'identification des zones chaudes.

L'étude s'est basée sur le principe général de l'algorithme de l'estimation à noyau de densité. Cette approche offre une grande précision de calcul, vu qu'elle considère la mesure de densité pour chaque point d'accident. D'autres parts, elle nécessite beaucoup de calculs donc augmente la complexité.

5.2. Deuxième étude

Cette étude intitulée « Estimateur à noyau de densité et la classification K-means pour l'analyse des zones chaudes d'accidentologie » [Andersson, 2009], vise l'identification des zones à haut risque d'accidentologie.

Elle utilise une technique de classification pour déterminer les indicateurs fréquents dans chaque groupe de zones. La base de données considérée est celle des accidents de la circulation pendant 5 ans (1999-2003) à Londres-Angleterre.

Cette étude vise l'analyse des accidents de la route, et comporte deux étapes principales :

- L'identification des régions qui enregistrent le plus grand taux d'accidentologie, en définissant les zones chaudes (hot spots) en appliquant l'algorithme KDE.
- L'identification des facteurs de risques les plus fréquents, en regroupant les zones chaudes suivant les facteurs les plus influents en appliquant l'algorithme K-means.

basent sur la théorie Mono-causale et s'appuient sur les causes fréquentes d'accidents (excès de vitesse, ébriété, ...).

Vu les inconvénients que cette approche présente, cette institution a opté pour une stratégie multi-causale qui se base sur l'étude des interactions existantes entre ces causes, Cependant, cette stratégie nécessite de recourir à des techniques d'analyse de données plus élaborées.

D'où est né ce besoin de se doter d'une plateforme performante qui va non seulement va leur permettre d'automatiser la collecte de données mais aussi fournir des informations de synthèse utiles pour mieux visualiser la répartition des accidents sur le territoire national et leur évolutions dans le temps et explorer ces données pour tirer les liens entre les différentes causes fréquentes d'accidents.

2. Objectifs

Pour répondre aux besoins exprimés dans la problématique, nous réalisons un outil de collecte et d'analyse des accidents de la route, permettant de:

- Décrire un accident de façon plus claire, détaillée et qui répond à des normes internationales.
- Gérer les informations sur les accidents.
- Comprendre les facteurs de risques d'accidents, et leur corrélation.
- Identifier les endroits à haute fréquence d'accidents.
- Prédire l'éventuelle croissance du taux d'accidentologie dans des régions précises.

3. Travaux existants

Dans ce qui suit nous allons présenter les deux premières versions de cette plateforme réalisées au sein de la gendarmerie nationale. En effet, les travaux concernés sont ceux de [HANOU et MEDANI, 2013] portant sur la « Conception et réalisation d'un outil d'analyse d'accidents de trafic routier » et de [Aghiles, 2013] « Conception et réalisation d'un outil de collecte et d'analyse d'accidents de trafic routier ».

3.1. Première étude

Cette étude porte sur le projet de fin d'étude (PFE) « Conception et réalisation d'un outil d'analyse d'accidents de trafic routier » [HANNOU et MEDANI, 2013]. Ce travail a pour but d'employer des techniques de fouille de données spatiales, pour mettre en œuvre un outil d'analyse et d'exploration de bases de données d'accidentologie.

Le travail présenté est structuré en trois principales étapes :

- 1) L'identification des zones chaudes, en implémentant une méthode de "clustering basée densité".
- 2) Le regroupement des données dans des classes sur plusieurs échelles spatiales, en implémentant une méthode de "clustering hiérarchique", dans chaque échelle spatiale.
- 3) La prédiction de futures régions à haut risque d'accidents, en implémentant des techniques de "règles d'association et d'arbres de décision".

Cette solution n'a pas été déployée par la gendarmerie nationale, car la fiche accident proposée est statique contrairement à leur besoin. Deuxièmement, cette solution est de type desktop ce qui fait qu'elle ne peut pas être déployé sur un réseau interne. Notons également, l'absence d'un module de collecte de donnée. Pour pallier à ces problèmes, une deuxième version de la plateforme a vu le jour et est présentée ci-dessous.

3.2. Deuxième étude

Suite aux insuffisances détectées dans la première version, une deuxième version à été réalisée [Y.Aghiles, 2013].

Se travail a permis d'automatiser la collecte de données et donc répondre à certains besoins manifestés par les enquêteurs de la gendarmerie nationale afin de combattre les accidents de la route en Algérie. Parmi les objectifs du projet, nous pouvons citer :

- Décrire un accident de façon plus claire, détaillée et qui réponde à des normes internationales.
- Automatiser la collecte des informations
- Identifier les zones à risque, possédant une fréquence importante d'accidents.
- Comprendre les facteurs de risques d'accidents, et leur corrélation.
- Fournir des statistiques concernant l'évolution « d'accidentologie ».

Toutefois cette version présente des insuffisances au niveau de la gestion des utilisateurs et ne s'adapte pas à la hiérarchie générale de la gendarmerie nationale (ex : un intervenant a le privilège de valider un accident). De plus, le système ne garde pas trace de l'historique des modifications effectuées, et la partie analyse n'a pas était complètement abordés

4. Solution proposée

Nous proposons de concevoir un système qui permette de gérer les différentes informations sur les accidents de la route, y compris leurs localisations. Ensuite nous allons utiliser différentes techniques de la fouille de données spatiales, pour la réalisation des différentes

tâches de l'outil. Ce dernier représente un processus complet d'analyse d'accidentologie, structuré en quatre principales étapes :

1. Automatisation de la collecte de données.
2. Identification des zones chaudes, en implémentant la méthode de clustering basée densité (KDE) (Voir chapitre 3).
3. Compréhension des facteurs de risques en utilisant l'algorithme K-means.
4. Prédiction de futures régions à haut risque d'accidents, en implémentant des techniques d'extraction des règles d'association.

La plateforme doit aussi respecter certaines contraintes comme :

- Gestion des utilisateurs.
- Gestion des accidents.
- Gestion de la fiche accident (fiche BAAC).
- Fournir des statistiques concernant les indicateurs de la fiche accident.
- Historique de toute modification effectuée sur les fiches accident.
- Déploiement sur le territoire national.

4.1. Diagramme de contexte

Le diagramme de contexte est une modélisation des échanges de messages et d'informations, qui se font entre le système d'Information et les acteurs qui interagissent avec ce dernier (figure 23).

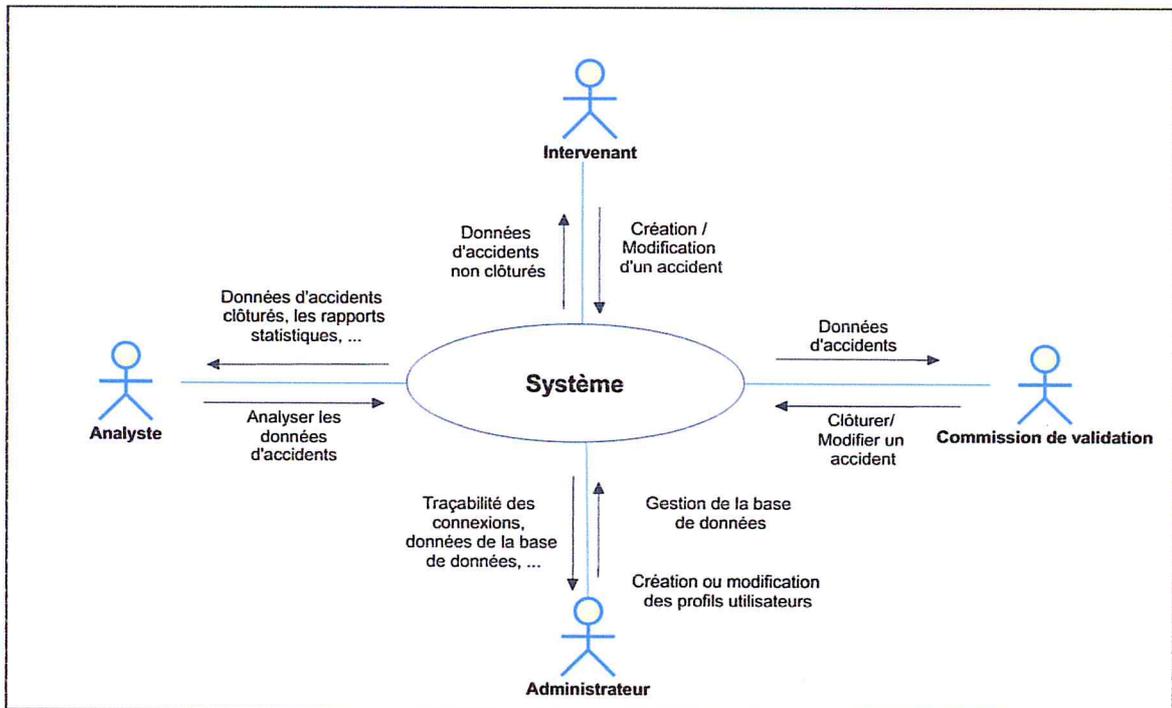


Figure 23. Diagramme de contexte dynamique.

4.2. Diagramme de cas d'utilisation

Pour assurer un bon fonctionnement du système et satisfaire les besoins et les attentes, il faut prendre en considération les besoins fonctionnels - comme un repère de saisie - et les analyser à travers les principaux cas d'utilisations.

Ce repère doit décrire le fonctionnement du système pour garantir un meilleur service aux différents utilisateurs de notre outil en faisant interagir un ensemble d'entités qui sont:

- **Le système:** Le système définit l'application informatique. Il ne contient donc pas les acteurs, mais les cas d'utilisation et leurs associations.
- **Les acteurs:** Un acteur représente une personne ou un périphérique qui joue un rôle (interagit) avec le système.
- **Les cas d'utilisations:** Chaque cas d'utilisation décrit un ensemble d'interactions successives d'une entité en dehors du système (utilisateur) avec le système lui-même pour réaliser une fonctionnalité.

Et pour affiner le diagramme de cas d'utilisation, UML définit trois types de relations standardisées entre les cas d'utilisation :

- **Une relation d'inclusion**, formalisée par le mot-clé `<<include>>` : le cas d'utilisation de base en incorporant explicitement un autre, de façon obligatoire.

- **Une relation d'extension**, formalisée par le mot-clé <<extend>> : le cas d'utilisation de base en incorporant implicitement un autre, de façon optionnelle.
- **Une relation de généralisation/spécialisation**: les cas d'utilisation descendants héritent de la description de leur parent commun. Chacun d'entre eux peut néanmoins comprendre des interactions spécifiques supplémentaires.

Bien que de nombreux diagrammes d'UML permettent de décrire un cas, il est recommandé de rédiger une description textuelle, car c'est une forme souple qui convient dans bien des situations.

Une description textuelle des cas d'utilisation est une description très utile pour décrire la communication des utilisateurs avec le système. Dans notre cas, cette description contiendra une première partie permettant d'identifier le cas d'utilisation et une deuxième partie contient la description du fonctionnement du cas sous la forme d'une séquence de messages échangés entre les acteurs et le système. Cette partie contient toujours une séquence nominale qui correspond au fonctionnement nominal du cas. Parfois, il peut avoir des comportements imprévisibles. La séquence nominale ne suffit donc pas pour décrire tous les comportements possibles. À la séquence nominale, s'ajoutent fréquemment des séquences alternatives et des séquences d'exceptions.

Dans ce qui suit nous allons procéder de la manière suivante :

- Identifier les acteurs de notre système.
- Identifier les différents cas d'utilisations et leurs objectifs.
- Regrouper les cas d'utilisation dans des paquetages par domaine fonctionnel.
- Décrire chaque cas d'utilisation par une description textuelle.

4.2.1. Identification des acteurs

« Un acteur représente un rôle joué par une entité externe (utilisateur humain, dispositif matériel ou autre système) qui interagit directement avec le système étudié. Un acteur peut consulter et/ou modifier directement l'état du système, en émettant et/ou en recevant des messages susceptibles d'être porteurs de données.»[ROQUES, 2008].

Les acteurs que nous avons identifié, qui interagiront avec le nouveau système sont les suivants :

1. **Intervenant**: c'est un utilisateur chargé d'ajouter un nouvel accident ou de le modifier tant qu'il n'a pas été clôturé (validé).
2. **Commission de validation**: Elle est chargée de la clôture des accidents. Cet utilisateur peut aussi modifier les données de l'accident.

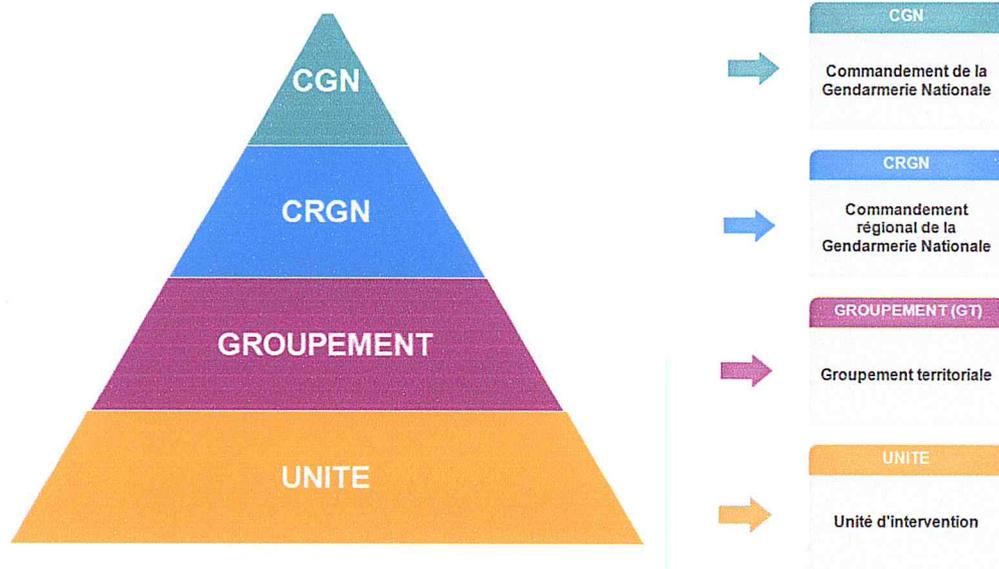
Chapitre 4 : Etude conceptuelle

3. **Analyste:** cet acteur a accès aux informations des accidents clôturés et applique des analyses sur ces derniers. Il peut aussi consulter les rapports statistiques, et interpréter les résultats.
4. **Administrateur:** il est chargé de gérer les utilisateurs, il définit les profils, attribue les privilèges et droits d'accès. Il s'occupe aussi de la gestion de la base de données.

Avant de passer à l'étape suivante, nous voyons qu'il est nécessaire de présenter la hiérarchie de la gendarmerie nationale détaillée en figure 24 et ce afin de voir comment les acteurs de notre système sont répartis.

Chapitre 4 : Etude conceptuelle

Gendarmerie Nationale



Localisation des acteurs

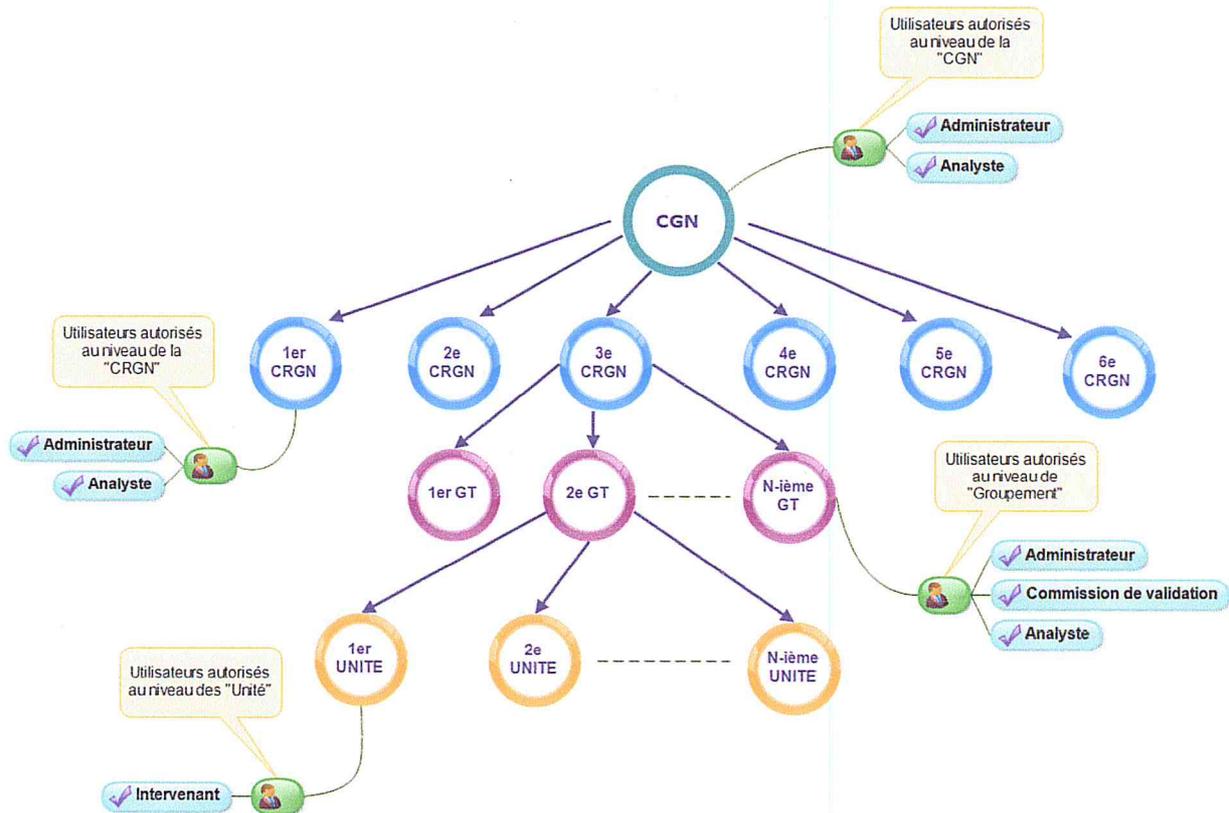


Figure 24. Hiérarchie de la gendarmerie nationale.

Chapitre 4 : Etude conceptuelle

4.2.2. Regroupement en paquetage

Dans ce qui suit, nous allons donner le regroupement en paquetages de différents cas d'utilisation identifiés précédemment. Quatre paquetages sont à distinguer:

Paquetage	Nom du paquetage
Paquetage 1	Gestion des accidents
Paquetage 2	Administration
Paquetage 3	Analyse
Paquetage 4	Statistique

Tableau 5. Regroupement en paquetages des cas d'utilisations fonctionnels.

4.2.3. Identification des cas d'utilisations

N	Cas d'utilisation	Acteurs	Objectif
1	Ajouter un accident	Intervenant	décrire les étapes permettant d'ajouter un accident
2	Consulter la liste des accidents	Administrateur, intervenant, analyste et commission de validation.	Consulter les données des accidents.
3	Etablir une fiche BAAC	Intervenant	Permet de recenser un ensemble d'information concernant l'accident.
4	Modifier un accident	Intervenant, commission de validation	Permet de modifier les différentes données concernant un accident.
5	Rechercher un accident	Intervenant, commission de validation	Faire une recherche sur accident quelconque.
6	Classification	Analyste	Permet une classification des données et une visualisation des résultats.
7	Détection des hot-spots	Analyste	Permet de détecter les zones chaudes.
8	Prévision	Analyste	Prévenir d'éventuelles zones à haute concentration de risque

Chapitre 4 : Etude conceptuelle

			d'accidents.
9	Exécuter des requêtes statiques	Analyste	Permet à l'analyste d'exécuter des requêtes prédéfinies.
10	Analyse des variables	Analyste	Permet à l'analyste de mettre le point sur l'impact de cette variable sur l'ensemble des accidents
11	Gestion des utilisateurs	Administrateur	Gérer les utilisateurs en tout ce qui concerne l'ajout, modification, suppression...
12	Gestion de la fiche BAAC	Administrateur	Permet à l'administrateur de gérer la fiche BAAC en tout ce qui concerne l'ajout la modification, la suppression... d'une rubrique, variable et valeur.
13	Gestion des entités	Administrateur	Permet à l'administrateur de gérer une entité (cause, groupement, unité...) en tout ce qui concerne l'ajout la modification, la suppression...
14	Valider un accident	Commission de validation	la clôture de l'accident.

Tableau 6. Identification des cas d'utilisation.

4.2.3.1. Paquetage « Gestion des accidents »

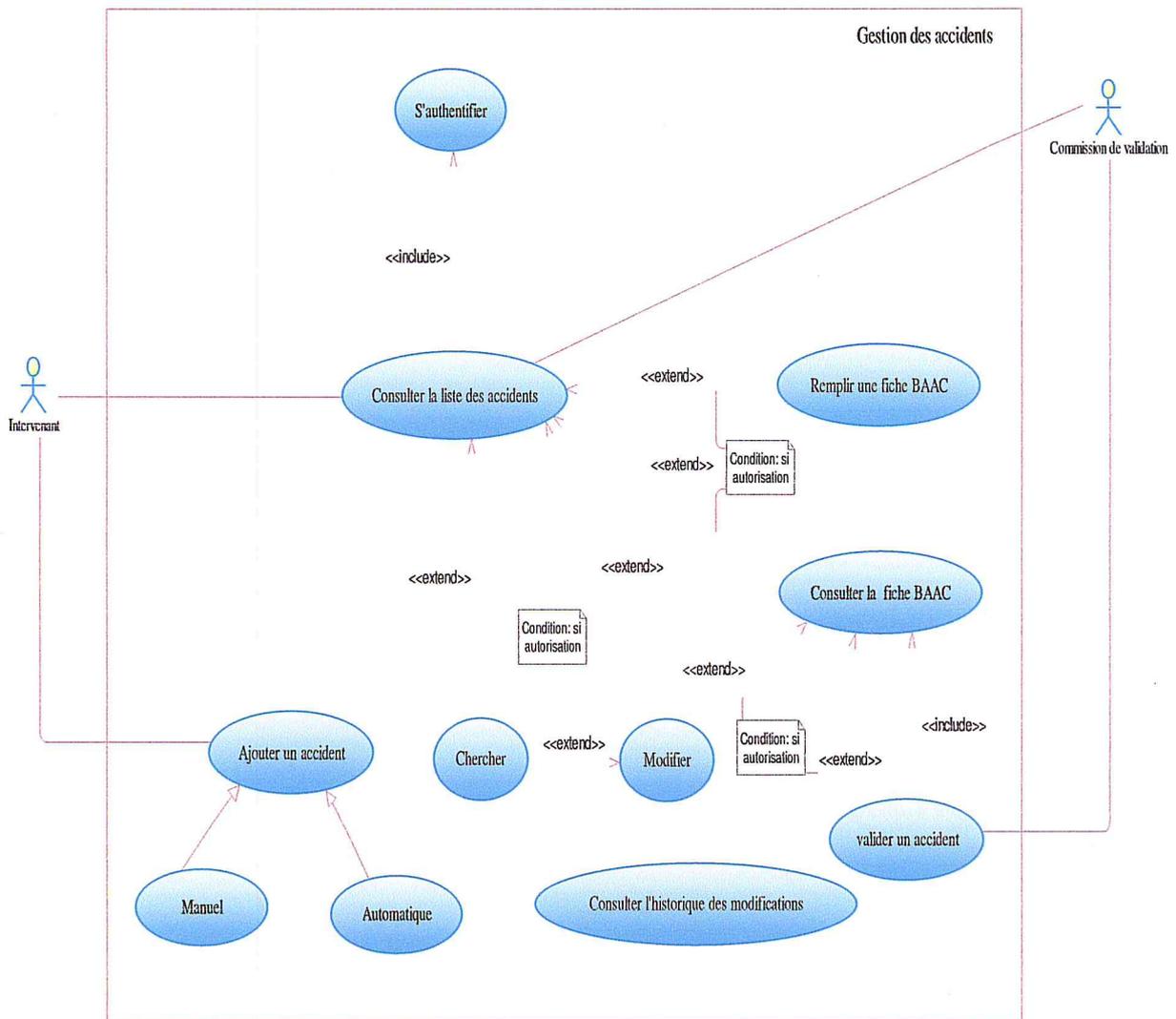


Figure 25. Détails du paquetage « Gestion des accidents ».

Description textuelle

a. Ajouter un accident

Description du cas « Ajouter un accident »
<p>Identification :</p> <p>Nom du cas : « Ajouter un accident »</p> <p>Objectif : décrire les étapes permettant d'ajouter un accident.</p> <p>Acteur: Intervenant.</p>
<p>Pré-conditions :</p> <ul style="list-style-type: none">– L'utilisateur s'est identifié correctement à l'application. <p>N1 : Ajout manuel</p> <ol style="list-style-type: none">1. Le système affiche un menu comprenant les différentes fonctionnalités possibles pour cet utilisateur.2. L'utilisateur choisit un ajout d'un accident.3. Le système affiche une fenêtre composée de plusieurs rubriques (caractéristique, usager, véhicule, lieu, bilan général).4. L'utilisateur remplit les différentes informations concernant chaque rubrique et valide ces dernières.5. Le système enregistre les données saisies et affiche un message de confirmation. <p>N2 : Ajout automatique</p> <ol style="list-style-type: none">1. L'utilisateur choisit une insertion automatique.2. L'utilisateur joint le fichier et lance l'exécution.3. Le système analyse et ajoute l'accident.4. L'utilisateur effectue une vérification et valide.5. Le système enregistre l'accident. <p>Enchaînement d'exception :</p> <p>E1 : le système détecte que des champs obligatoires sont vides</p> <p>Le système affiche un message pour demander à l'utilisateur de remplir tous les champs obligatoires.</p> <p>E2 : le système détecte que le fichier joint est incompatible avec les formats pris en charge</p> <p>Le système affiche un message pour avertir l'utilisateur.</p> <p>Post-conditions :</p> <ul style="list-style-type: none">– Accident ajouté.– Date de création de l'accident enregistré.

b. Consulter la liste des accidents

Description du cas « Consulter la liste des accidents »
<p><u>Identification :</u></p> <p>Nom du cas : « Consulter la liste des accidents »</p> <p>Objectif : avoir accès à la liste des accidents.</p> <p>Acteurs: Intervenant, Commission de validation.</p>
<p><u>Pré-conditions :</u></p> <ul style="list-style-type: none">– L'utilisateur s'est identifié correctement à l'application.
<p><u>Enchaînement nominal :</u></p> <ol style="list-style-type: none">1. Le système affiche le menu.2. L'utilisateur choisit de consulter la liste des accidents.3. Le système affiche la liste des accidents auxquels l'utilisateur a accès.
<p><u>Enchaînement alternatif :</u></p> <p>A1 : l'enchaînement démarre après le point 3 de l'enchaînement nominal.</p> <ol style="list-style-type: none">4. Si l'utilisateur veut chercher un accident particulier ; appel du cas : « Chercher un accident ». <p>La séquence nominale reprend après le point 3.</p>
<p><u>Post-condition :</u></p> <ul style="list-style-type: none">– Aucune.

c. Remplir une fiche accident

Description du cas « Remplir une fiche accident »
<p><u>Identification :</u></p> <p>Nom du cas : « Remplir une fiche accident »</p> <p>Objectif : ajouter les détails de l'accident sous les différentes rubriques (caractéristiques, lieux, ...).</p> <p>Acteurs: Intervenant.</p>
<p><u>Séquencement :</u></p> <p>Le cas démarre au point 3 de l'enchaînement nominal de la description du cas « Consulter la liste des accidents ».</p>

Enchaînement nominal :

N1 : Ajouter une fiche accident

4. L'utilisateur choisit de remplir une fiche accident.
5. Le système affiche le formulaire d'ajout.
6. L'utilisateur saisit les informations de chaque rubrique.
7. L'utilisateur valide l'ajout.
8. Le system affiche un message de confirmation.

N2 : Consulter une fiche accident

4. L'utilisateur choisit de consulter la fiche d'un accident.
5. Le système affiche les informations.

N3 : Modifier une fiche accident

4. L'utilisateur choisit de modifier la fiche d'un accident.
5. Le système affiche la fenêtre de modification.
6. L'utilisateur saisit les modifications et valide.
7. Le système enregistre la modification et affiche un message de confirmation.

Post-conditions :

- Fiche accident ajouté.
- Fiche accident modifié.

d. Consulter l'historique des modifications

Description du cas « Consulter l'historique des modifications »

Identification :

Nom du cas : « Consulter l'historique des modifications »

Objectif : afficher l'historique des modifications apportées à une fiche accident.

Acteurs: Commission de validation.

Séquencement :

Le cas démarre après le point 5 de l'enchaînement nominal de la description du cas « Remplir une fiche accident » - N2: Consulter la fiche accident.

Enchaînement nominal :

1. L'utilisateur choisit de voir l'historique des modifications d'une fiche accident.
2. Le system affiche une fenêtre; composer de deux champs, la date de modification et l'utilisateur qui a fait la modification.

Post-condition :

- Aucun.

e. Modifier un accident

Description du cas « Modifier un accident »
<p><u>Identification :</u></p> <p>Nom du cas : « Modifier un accident »</p> <p>Objectif : permet de modifier les différentes données concernant un accident.</p> <p>Acteurs: Intervenant, commission de validation.</p>
<p><u>Pré-condition</u></p> <ul style="list-style-type: none">– L'accident n'est pas encore validé. <p><u>Séquencement :</u></p> <p>Le cas démarre au point 3 de la description du cas « Consulter la liste des accidents ».</p> <p><u>Enchaînement nominal :</u></p> <ol style="list-style-type: none">1. L'utilisateur choisit de modifier un accident.2. Le système affiche la fenêtre de modification.3. Il saisit les modifications et valide.4. Le système enregistre la modification et affiche un message de confirmation. <p><u>Post-conditions :</u></p> <ul style="list-style-type: none">– Accident modifié.– Date de modification enregistrée.

f. Chercher un accident

Description du cas « Chercher un accident »
<p><u>Identification :</u></p> <p>Nom du cas : « Chercher un accident »</p> <p>Objectif : permet de décrire les étapes pour faire une recherche sur un accident.</p> <p>Acteurs: Commission, Intervenant, Analyste.</p>

Séquencement :

Le cas démarre au point 4 de l'enchaînement alternatif de la description du cas « Consulter la liste des accidents ».

Enchaînement nominal :

1. L'utilisateur choisit de rechercher un accident
2. L'utilisateur saisit le critère de la recherche.
3. Le système affiche les résultats de la recherche.

Post-condition :

- Aucune.

g. Valider un accident

Description du cas « Valider un accident »

Identification :

Nom du cas : « Valider un accident »

Objectif : la clôture de l'accident.

Acteur: Commission de validation.

Séquencement :

Le cas démarre après le point 5 de l'enchaînement nominal de la description du cas « Remplir une fiche accident » - N2 : Consulter une fiche accident.

Enchaînement nominal :

1. L'utilisateur vérifie les informations de l'accident et complète la fiche d'accident, si c'est nécessaire.
2. L'utilisateur clique sur valider l'accident.
3. Le système valide l'accident.

Post-condition :

- Accident validé.

4.2.3.2. Paquetage « Analyse »

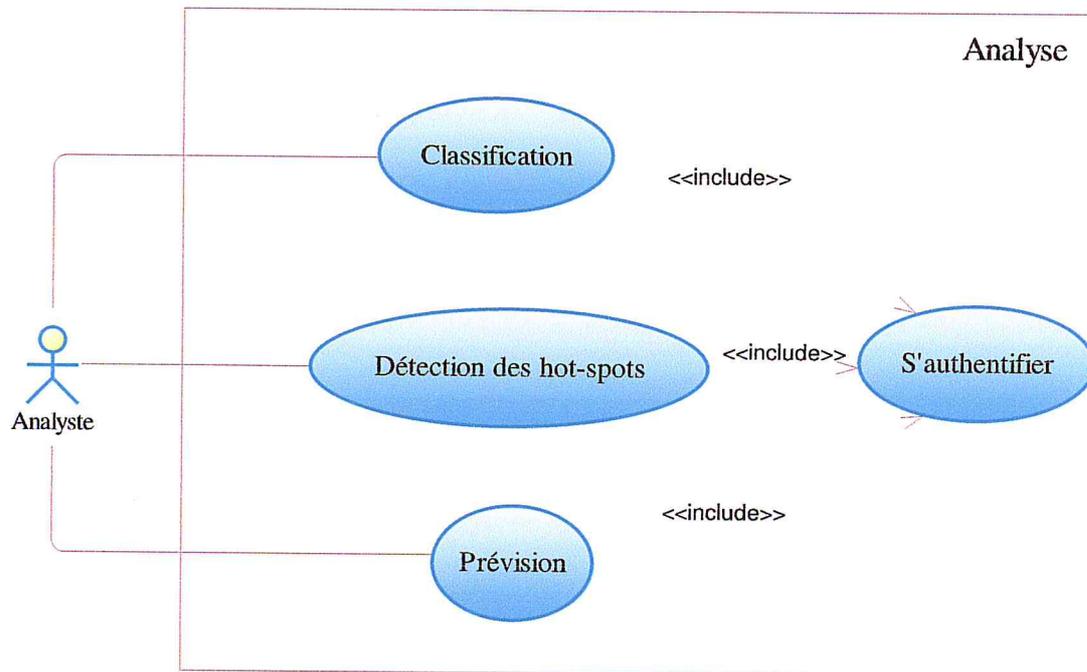


Figure 26. Détails du paquetage « Analyse ».

Description textuelle

a. Classification

Description du cas « Classification »
<p>Identification :</p> <p>Nom du cas : « Classification »</p> <p>Objectif : obtenir une classification des données à l'aide de l'algorithme K-means, et visualiser les résultats.</p> <p>Acteur: Analyste.</p>
<p>Pré-condition :</p> <ul style="list-style-type: none">– L'utilisateur s'est identifié correctement à l'application. <p>Enchaînement nominal :</p> <ol style="list-style-type: none">1. L'utilisateur choisit d'appliquer une classification.2. L'utilisateur saisit les paramètres.3. Lancer le traitement de l'analyse.4. Le système affiche le résultat. <p>Si l'utilisateur le souhaite, il peut appliquer les différentes tâches d'analyse qu'offre l'outil (classification, hot spots,...etc.), sur le résultat obtenu.</p>

b. Détection des hot-spots

Description du cas « Détection des hot-spots »
<p><u>Identification :</u></p> <p>Nom du cas : « Détection des hot-spots »</p> <p>Objectif : détecter les zones chaudes (concentration des accidents).</p> <p>Acteur: Analyste.</p>
<p><u>Pré-conditions :</u></p> <ul style="list-style-type: none">– L'utilisateur s'est identifié correctement à l'application. <p><u>Enchaînement nominal :</u></p> <ol style="list-style-type: none">1. L'utilisateur saisit les paramètres.2. Lancer l'identification des zones chaudes.3. Le système affiche le résultat. <p>Si l'utilisateur le souhaite, il peut appliquer les différentes tâches d'analyse qu'offre l'outil (classification, hot-spots,...etc.), sur le résultat obtenu.</p>

c. Prévision

Description du cas « Prévision »
<p><u>Identification :</u></p> <p>Nom du cas : « prévision »</p> <p>Objectif : prévenir d'éventuelles zones à haute concentration de risques d'accidents.</p> <p>Acteur: Analyste.</p>
<p><u>Pré-condition :</u></p> <ul style="list-style-type: none">– L'utilisateur s'est identifié correctement à l'application. <p><u>Enchaînement nominal :</u></p> <ol style="list-style-type: none">1. L'utilisateur choisit d'extraire les règles d'association.2. L'utilisateur saisit les paramètres.3. L'utilisateur Lance l'analyse.4. Le système affiche le résultat. <p>Si l'utilisateur le souhaite, il peut appliquer les différentes tâches d'analyse qu'offre l'outil (classification, hots spots,...etc.), sur le résultat obtenu.</p>

4.2.3.3. Paquetage « Statistiques »

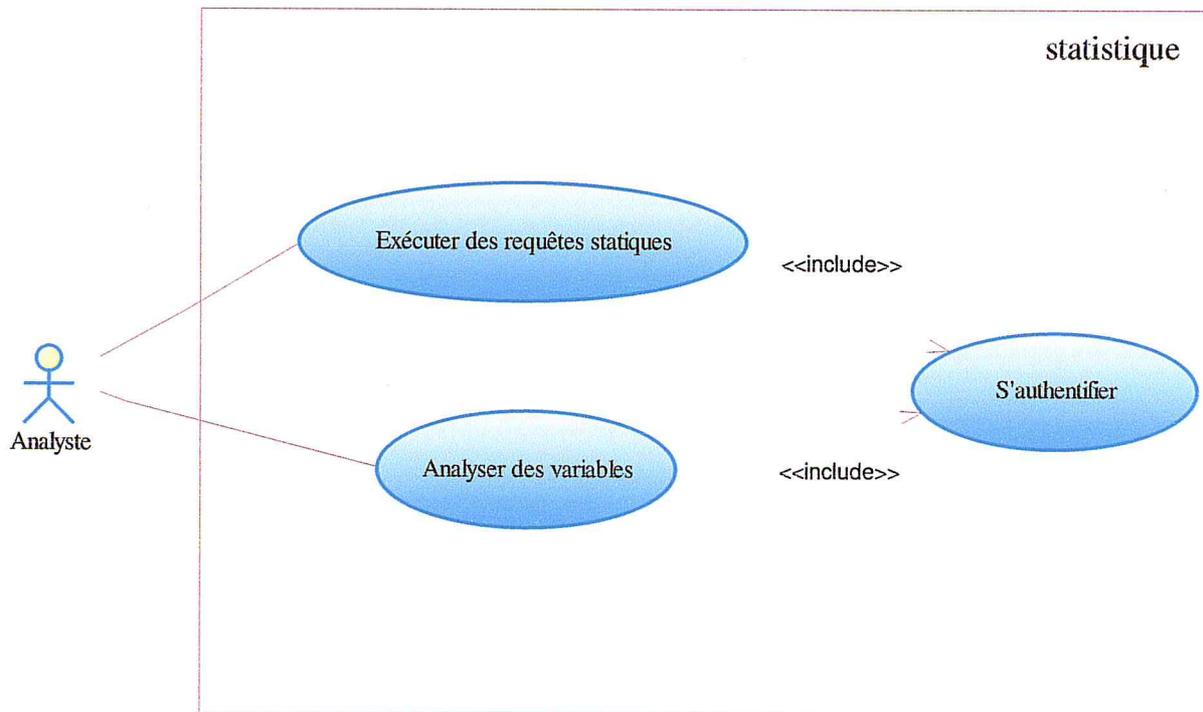


Figure 27. Détails du paquetage « Statistiques ».

Description textuelle

a. Exécuter des requêtes statiques

Description du cas « Exécuter des requêtes statiques »
<p>Identification :</p> <p>Nom du cas : « Exécuter des requêtes statiques »</p> <p>Objectif : exécuter des requêtes prédéfinies.</p> <p>Acteur: Analyste.</p>
<p>Pré-condition :</p> <ul style="list-style-type: none">– L'utilisateur s'est identifié correctement à l'application. <p>Enchaînement nominal :</p> <ol style="list-style-type: none">1. L'Analyste choisit « requêtes statiques ».2. Le système affiche la liste des requêtes.3. L'analyste choisit une requête et lance l'exécution.4. Le système affiche le résultat.

b. Analyser des variables

Description du cas « Analyser des variables »
<p>Identification :</p> <p>Nom du cas : « Analyser des variables »</p> <p>Objectif : l'analyse des variables consiste à montrer l'impact des différentes valeurs qu'elle peut prendre sur l'ensemble des accidents enregistrés.</p> <p>Acteur: Analyste.</p>
<p>Pré-condition :</p> <ul style="list-style-type: none">- L'utilisateur s'est identifié correctement à l'application.
<p>Enchaînement nominal :</p> <ol style="list-style-type: none">1. L'analyste choisit « analyser les variables ».2. Le système affiche la liste des variables3. L'analyse choisit la variable.4. Le système affiche les résultats.
<p>Enchaînement alternative :</p> <ol style="list-style-type: none">1. L'analyste choisit d'appliquer une des techniques du data mining spatial (hot spot, règle d'association) sur le résultat obtenu et lance l'exécution.2. Le Système affiche les résultats.

4.2.3.4. Paquetage « Administration »

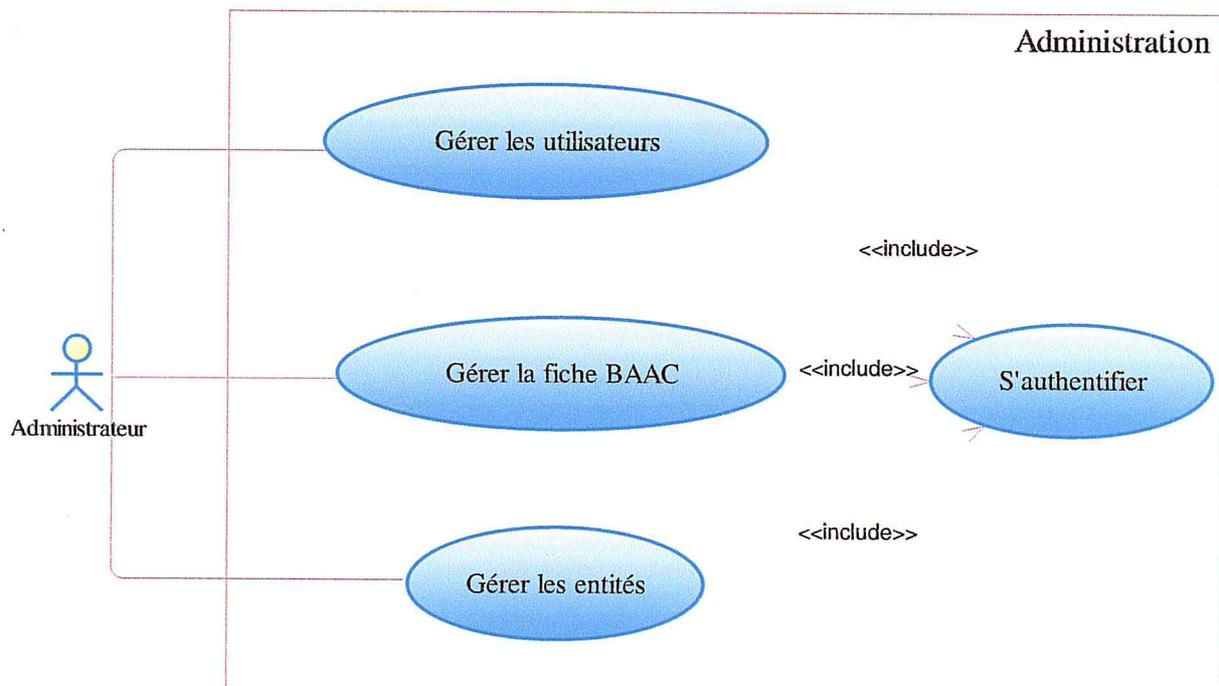


Figure 28. Détail de paquetage administration.

Chapitre 4 : Etude conceptuelle

Afin de mieux comprendre en quoi consistent les cas d'utilisation de ce paquetage, on a préféré passer au 2^{ème} niveau pour avoir plus de détails sur ces derniers.

▪ Gestion des utilisateurs

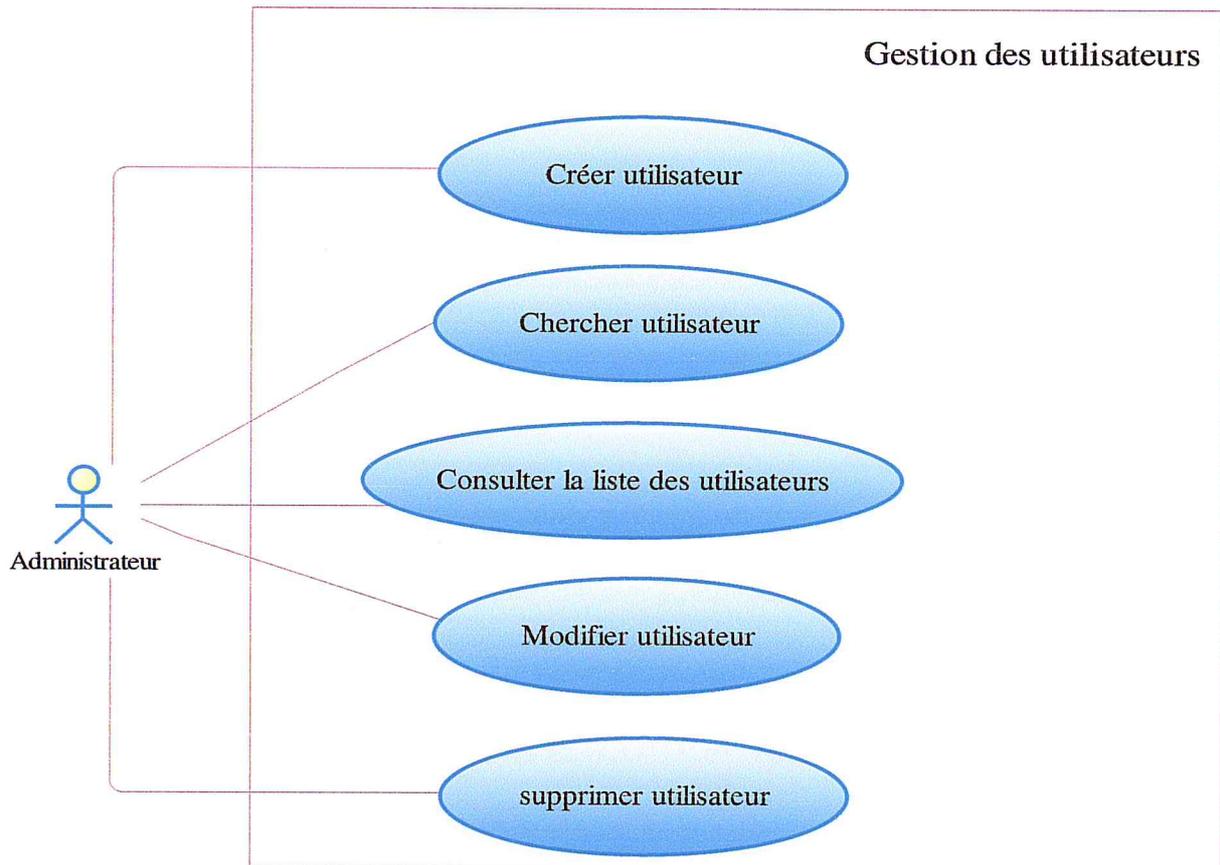


Figure 29. Diagramme du cas d'utilisation « Gestion des utilisateurs ».

Description textuelle

a. Gestion des utilisateurs

Description du cas « Gestion des utilisateurs »
Identification : Nom du cas : « Gestion des utilisateurs » Objectif : gérer les utilisateurs sur tout ce qui concerne l'ajout, la modification, la suppression... Acteur: Administrateur.
Pré-condition : – L'utilisateur s'est identifié correctement à l'application.
Enchaînement nominal : 1. L'utilisateur se rend dans l'espace gestion des comptes utilisateurs.

N1 : Créer utilisateur

2. L'utilisateur choisit d'ajouter un nouveau compte utilisateur.
3. Le système lui retourne la fenêtre d'ajout.
4. L'utilisateur saisit les informations et les valide.
5. Si un des champs obligatoires est vide, le système demande à l'utilisateur de terminer la saisie avant de valider.

N2 : Consulter utilisateur

2. Le système affiche la liste des utilisateurs.
3. L'utilisateur clique sur le compte à consulter.
4. Le système affiche les informations concernant ce compte.

N3 : Modifier utilisateur

2. Le système affiche la liste des utilisateurs.
3. L'utilisateur sélectionne un compte à modifier.
4. Le système affiche la fenêtre de modification avec les différentes informations concernant le compte.
5. L'utilisateur modifie les informations du compte et valide.
6. Le système modifie les informations du compte et affiche un message de confirmation.

N4 : Supprimer utilisateur

2. Le système affiche la liste des utilisateurs.
3. L'utilisateur choisit de supprimer un compte utilisateur.
4. Le système supprime le compte et affiche un message de confirmation.

N5 : Chercher utilisateur

Le cas démarre après le point 3 de l'enchaînement alternatif « A1 ».

3. L'utilisateur choisit de rechercher un utilisateur.
4. L'utilisateur saisit le critère de la recherche.
5. Le système affiche les résultats de la recherche.

Enchaînement alternatif :

A1 : l'enchaînement démarre après le point 2 de N2, N3, N4.

3. Si l'utilisateur veut chercher un compte particulier ; appel du « N5 : Chercher utilisateur ».

La séquence nominale reprend après le point 3 de N2, N3, N5.

Post-conditions :

- Utilisateur ajouté.
- Utilisateur modifié.
- Utilisateur supprimé.

a. Gérer la fiche BAAC

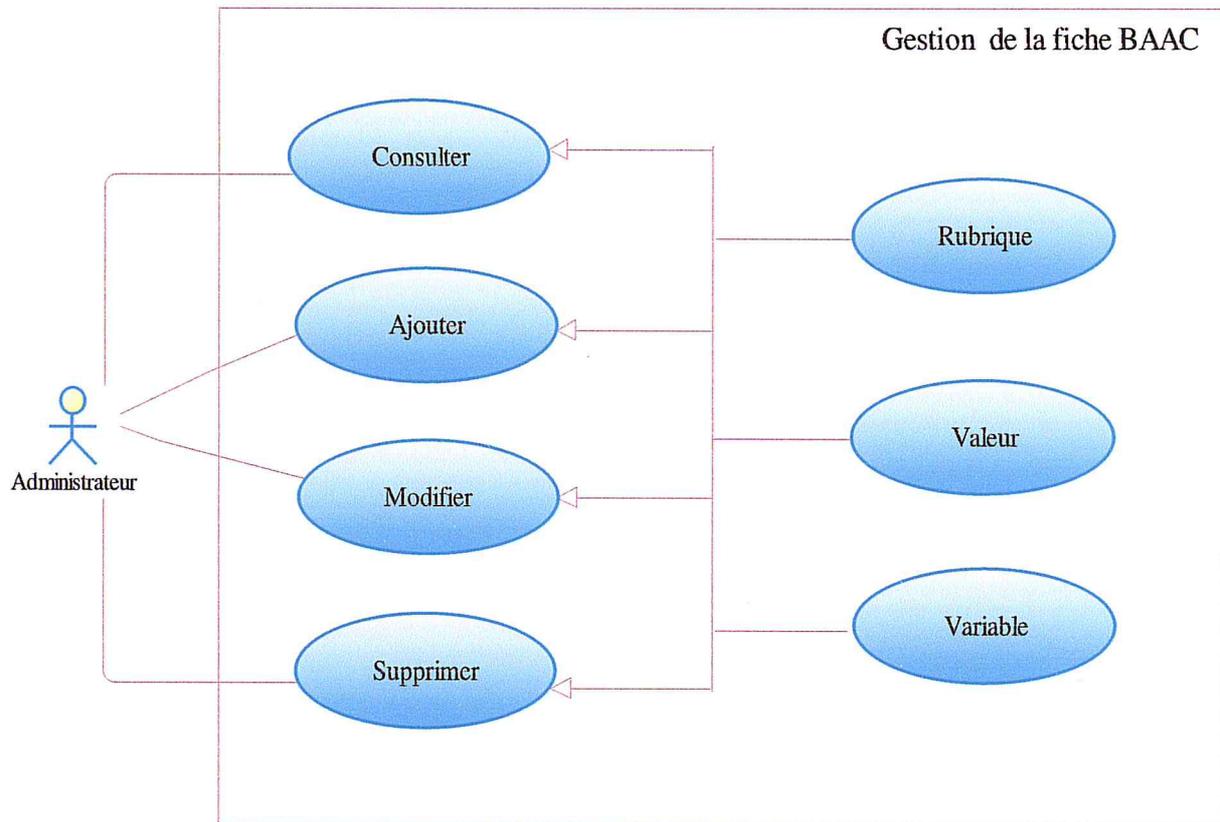


Figure 30. Diagramme du cas d'utilisation « Gérer la fiche BAAC ».

Description textuelle

a. Gestion des rubriques

Description du cas « Gestion des rubriques »
<p>Identification :</p> <p>Nom du cas : « Gestion des rubriques »</p> <p>Objectif : gérer les rubriques en tout ce qui concerne l'ajout, la modification, la suppression...</p> <p>Acteur: Administrateur.</p>
<p>Pré-condition :</p> <ul style="list-style-type: none">- L'utilisateur s'est identifié correctement à l'application.
<p>Enchaînement nominal :</p> <ol style="list-style-type: none">1. L'utilisateur se rend dans l'espace gestion des rubriques.
<p>N1 : Créer une rubrique</p> <ol style="list-style-type: none">2. L'utilisateur choisit d'ajouter une nouvelle rubrique.3. Le système affiche le formulaire d'ajout.

4. L'utilisateur saisit les informations et valide l'ajout.

N2 : Modifier une rubrique

2. L'utilisateur choisit de modifier une rubrique.
3. Le système affiche la liste des rubriques.
4. L'utilisateur choisit une rubrique.
5. Le système affiche les informations de la rubrique.
6. L'utilisateur modifie les informations.

N3 : Consulter rubrique

2. L'utilisateur choisit de consulter une rubrique.
3. Le système affiche la liste des rubriques.
4. L'utilisateur clique pour voir plus de détails.
5. Le système affiche les différentes informations concernant la rubrique ainsi que leurs variables.
6. Appel de cas gestion des variables « N3 : Consulter variable ».

N4 : Supprimer rubrique

2. Le système affiche la liste des rubriques.
3. L'utilisateur sélectionne la rubrique à supprimer.
4. Le système supprime la rubrique et affiche un message de confirmation.

Post-conditions :

- Rubrique ajoutée.
- Rubrique modifiée.
- Rubrique supprimée.

b. Gestion des variables

Description du cas « Gestion des variables »

Identification :

Nom du cas : « Gestion des variables »

Objectif : gérer les variables en tout ce qui concerne l'ajout, la modification, la suppression...

Acteur: Administrateur.

Pré-condition :

- L'utilisateur s'est identifié correctement à l'application.

Enchaînement nominal :

1. L'utilisateur se rend dans la gestion des variables

N1 : Créer une variable

2. L'utilisateur choisit d'ajouter une variable.
3. Le système affiche le formulaire d'ajout.
4. L'utilisateur saisit les informations.
5. Le système enregistre l'ajout et affiche un message de confirmation.

N2 : Modifier une variable

2. L'utilisateur choisit la rubrique ou se situe la variable qu'il souhaite modifier.
3. Le système affiche la liste des variables.
4. L'utilisateur clique sur la variable à modifier.
5. Le système affiche les informations de la variable.
6. L'utilisateur saisit les nouvelles informations.
7. Le système enregistre les modifications et affiche un message de confirmation.

N3 : Consulter une variable

2. L'utilisateur choisit de consulter une variable.
3. Le système affiche la liste des variables.
4. L'utilisateur sélectionne une variable pour voir plus de détails.
5. Le système affiche les différentes informations concernant la variable.

N4 : Supprimer une variable

2. L'utilisateur choisit la rubrique ou se situe la variable qu'il souhaite supprimer.
3. Le système affiche la liste des variables.
4. L'utilisateur sélectionne la variable à supprimer.
5. Le système supprime la variable et affiche un message de confirmation.

N5 : Chercher variable

Le cas démarre après le point 3 de N2, N3 ou N4.

4. L'utilisateur choisit de rechercher une variable.
5. L'utilisateur saisit le critère de la recherche.
6. Le système affiche les résultats de la recherche.

La séquence nominale reprend au point 4 de N2, N3 ou N4

Post-conditions :

- Variable ajoutée.
- Variable modifiée.
- Variable supprimée.

c. Gestion des valeurs

Description du cas « Gestion des valeurs »
<p>Identification :</p> <p>Nom du cas : « Gestion des valeurs ».</p> <p>Objectif : gérer les variables en tout ce qui concerne l'ajout, la modification, la suppression...</p> <p>Acteur: Administrateur.</p>
<p>Pré-condition :</p> <ul style="list-style-type: none">– L'utilisateur s'est identifié correctement à l'application. <p>Enchaînement nominal :</p> <ol style="list-style-type: none">1. L'utilisateur se rend dans l'espace gestion des valeurs.2. L'utilisateur choisit de consulter la liste des variables. <p>N1 : Ajouter une valeur :</p> <ol style="list-style-type: none">3. Appel du cas « Gestion des variables » - N3 : consulter variable.4. L'utilisateur choisit d'ajouter une nouvelle valeur à une variable.5. Le système affiche le formulaire d'ajout.6. L'utilisateur saisit la valeur et la valide.7. Le système enregistre l'ajout et affiche un message de confirmation. <p>N2 : Modifier une valeur :</p> <ol style="list-style-type: none">2. Appel du cas « Gestion des variables » - N3 : consulter variable.3. L'utilisateur choisit de modifier une valeur.4. Le système donne la main à l'utilisateur de modifier.5. L'utilisateur modifie la valeur et valide.6. Le système enregistre la modification et affiche un message de confirmation. <p>N3 : Supprimer une valeur :</p> <ol style="list-style-type: none">2. Appel du cas « Gestion des variables » - N3 : consulter variable.3. L'utilisateur choisit de supprimer une valeur.4. Le système supprime la valeur et affiche un message de confirmation. <p>Enchaînement d'exception :</p> <ul style="list-style-type: none">– Si le système détecte que la valeur existe, il affiche un message d'erreur à l'administrateur pour l'informer que la valeur existe déjà.– Si la saisie n'est pas complète, le système affiche un message d'erreur à l'utilisateur pour lui demander de terminer la saisie. <p>Post-conditions :</p>

- Valeur ajoutée.
- Valeur modifiée.
- Valeur supprimée.

b. Gestion des entités

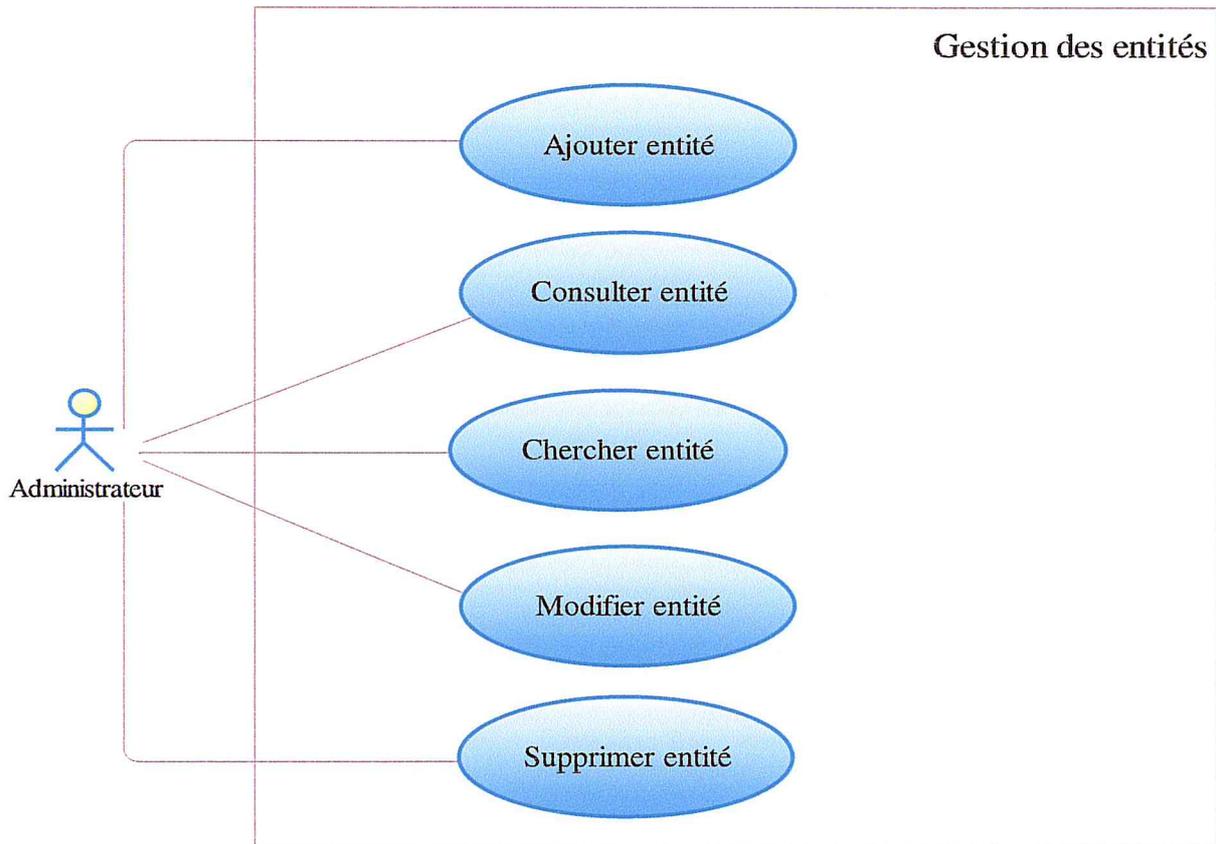


Figure 31. Diagramme de cas d'utilisation « Gestion des entités ».

Description textuelle

a. Gestion des entités

Description du cas « Gestion des entités »
<p>Identification :</p> <p>Nom du cas : « Gestion des entités »</p> <p>Objectif : gérer les entités en tout ce qui concerne l'ajout, la modification, la suppression...</p> <p>Acteur: Administrateur.</p>
<p>Pré-condition :</p>

- L'utilisateur s'est identifié correctement à l'application.

Enchaînement nominal :

1. L'utilisateur se rend dans l'espace gestion des entités.

N1 : Ajouter entité

2. L'utilisateur choisit d'ajouter une entité.
3. Le système affiche le formulaire d'ajout.
4. L'utilisateur saisit les informations et les valide.
5. Le système enregistre l'ajout et affiche un message de confirmation.

N2 : Consulter entité :

2. L'utilisateur choisit de consulter une entité.
3. Le système affiche la liste des entités.
4. L'utilisateur sélectionne une entité pour voir plus de détails.
5. Le système affiche les différentes informations concernant l'entité.

N3 : Modifier entité

2. L'utilisateur choisit de consulter la liste des entités.
3. Le système affiche la liste des entités.
4. L'utilisateur choisit l'entité qu'il souhaite modifier.
5. Le système affiche les informations correspondant.
6. L'utilisateur saisit les nouvelles informations et les valide.
7. Le système enregistre les modifications et affiche un message de confirmation.

N4 : Supprimer entité

2. L'utilisateur choisit de consulter la liste des entités.
3. Le système affiche la liste des entités.
4. L'utilisateur sélectionne l'entité qu'il souhaite supprimer.
5. L'utilisateur clique sur supprimer.
6. Le système supprime l'entité et affiche un message de confirmation.

N5 : Chercher entité

Le cas démarre après le point 3 de N2, N3.

4. L'utilisateur choisit de rechercher une entité.
5. L'utilisateur saisit le critère de la recherche.
6. Le système affiche les résultats de la recherche.

La séquence nominale reprend au point 4 de N3 ou N4.

Enchaînement d'exception :

- Si l'entité existe déjà le système informe l'utilisateur.

4.3. Diagramme de classe

La base de données d'accidentologie doit contenir les informations décrivant les accidents, de manière complète. Pour identifier les caractéristiques des accidents qui doivent figurer dans cette base, nous nous sommes rapprochés d'un expert de la gendarmerie nationale. Nous nous sommes également appuyés sur le contenu de la fiche BAAC (Bulletins d'Analyse des Accidents Corporels), qui répond aux normes internationales.

La figure ci-après représente le schéma de la base de données conçue :

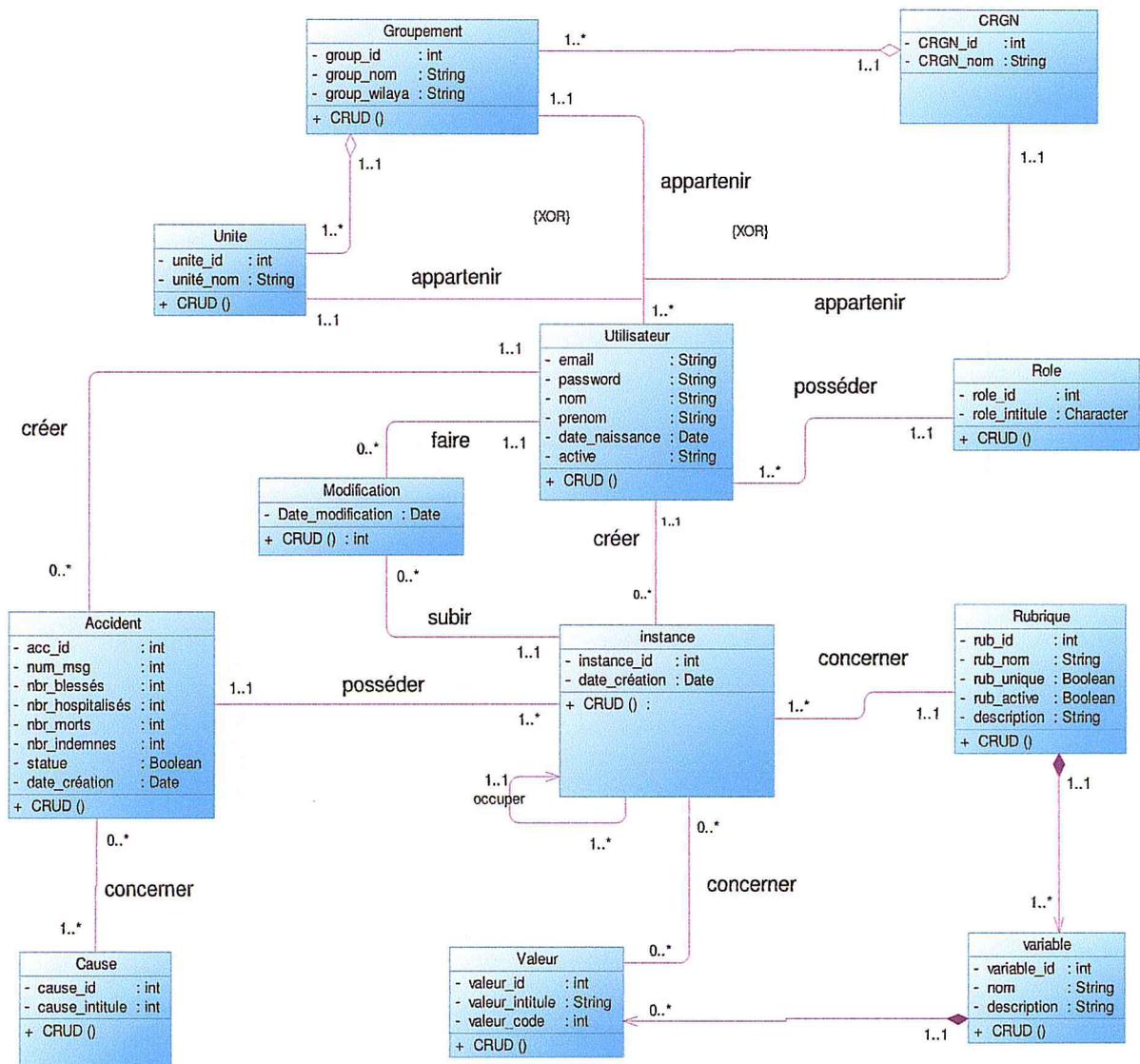


Figure 32. Diagramme de classe.

5.3.1. Liste des propriétés

	Attributs	Désignation	Type
Accident	Acc_id	Identifiant	Entier long
	Num_msg	Un numéro utilisé par la gendarmerie	Entier
	Nbr_blessés	Nombre de blessés	Entier
	Nbr_morts	Nombre de morts	Entier
	Nbr_indemnes	Nombre d'indemnes.	Entier
	Nbr_hospitalisés	Nombre d'hospitalisés	Entier
	Statut	Indique l'état de l'accident (en cours ou validé)	Booléen
	Date_création	La date de création	Date
	Date_modification	La date de modification	Date
Instance	Instance_id	Identifiant	Entier long
	Date_création	La date de création	Date
Rubrique	Rub_id	Identifiant	Entier
	Rub_nom	Nom de la rubrique	Chaine de caractères
	Rub_unique	Indique si l'accident peut avoir une ou plusieurs instances d'une rubrique donnée	Booléen
	Rub_active	Indique si la rubrique est en état active ou non	booléen
	Description	Description	Chaine de caractères
Variable	Var_id	Identifiant	Entier
	Var_nom	Le nom de la variable	Chaine de caractères
	Description	Description	Chaine de caractères
Valeur	Val_id	Identifiant	Entier
	Val_intitule	L'intitulé de la valeur	Chaine de caractères
Cause	Cause_id	Identifiant	Entier
	Cause_intitule	L'intitule de la cause	Chaine de caractères
CRGN	CRGN_id	Identifiant	Entier
	CRGN_nom	Nom	Chaine de caractères

Groupement	Group_id	Identifiant	Entier
	Group_nom	Nom	Chaine de caractères
	Group_wilaya	La wilaya	Chaine de caractères
Unité	Unité_id	Identifiant	Entier
	Unité_nom	Nom	Chaine de caractères
Utilisateur	Username	Identifiant	Chaine de caractères
	Password	Le mot de passe	Chaine de caractères
	Nom	Nom	Chaine de caractères
	Prénom	Prénom	Chaine de caractères
	Date_naissance	La date de naissance	Date
	Active	Etat de l'utilisateur	Booléen
Rôle	Role_id	Identifiant	Entier
	Role_intitule	L'intitulé du rôle	Chaine de caractères
Modification	Date_modification	Identifiant	Date

Tableau 7. Liste des propriétés.

5.4. Schéma relationnel

Le modèle relationnel a pour but de décrire et manipuler les données d'une base de données. Pour créer et interroger une base de données relationnelle, nous choisissons le langage SQL.

Le langage SQL utilise le modèle relationnel. De ce fait, nous devons faire le passage du modèle UML (classes-associations) au modèle relationnel. Les principaux changements s'effectuant en passant du modèle UML au modèle relationnel sont:

- Passage d'un modèle disposant de deux structures (classes-associations) à un modèle ne disposant que d'une seule structure (la relation ou la table).
- Application d'un ensemble de règles qui garantissent la cohérence sémantique entre le modèle UML et le modèle relationnel.

Chapitre 4 : Etude conceptuelle

Ce passage est régi par les règles suivantes :

- **Règle 01** : toute classe du schéma UML se traduit par une table avec les mêmes attributs. Si aucun identifiant n'est défini pour l'entité, un attribut est rajouté pour jouer le rôle de clé primaire.
- **Règle 02** : une relation 1 vers 1 se traduit en portant la clé primaire de l'une de ces entités dans la table de l'autre.
- **Règle 03** : une relation N vers 1 se traduit en portant la clé primaire de la table du côté 1, vers la table du côté N.
- **Règle 04** : une relation N vers N se traduit, en créant une nouvelle table ayant pour clé primaire la combinaison des clés primaires des deux tables.

Ci-après le schéma relationnel correspondant :

CRGN (CRGN_id, CRGN_nom).

Groupement (group_id, group_nom, group_wilaya, CRGN_id*).

Unite (unite_id, unite_nom, group_id*).

Rôle (role_id, intitule_role).

Utilisateur_Role (username, role_id*)

Utilisateur (username, password, nom, prénom, date_naissance, email, active, role_id*, unite_id*, group_id*, CRGN_id*).

Accident (acc_id, num_msg, nbr_blessés, nbr_hospitalisés, nbr_morts, nbr_indemnes, statue, date_creation, util_email*).

Cause (cause_id, cause_intitule).

Accident_Cause (acc_id*, cause_id*)

Modification (date_heure_modification, util_email*, instance_id*).

Instance (instance_id-usager, instance_id-véhicule*, acc_id*, rub_id*, username*).

Rubrique (rub_id, rub_nom, rub_unique, rub_active, description).

Variable (variable_id, var_nom, description, rub_id*).

Valeur (valeur_id, valeur_intitule, valeur_code, variable_id*).

Instance_Valeur (valeur_id*, instance_id*).

5.5. Développement du modèle dynamique

Cette étape se focalise sur le comportement des classes du système à partir des diagrammes de séquence et les diagrammes d'état des cas d'utilisation.

Chapitre 4 : Etude conceptuelle

Le diagramme de séquence permet de représenter les messages échangés entre les différents objets du système, avec un ordre chronologique représenté de haut en bas [AUDIBERT, 2009]. Le diagramme de séquence n'est pas la seule modélisation possible des échanges de messages entre objets; le diagramme de communication le permet aussi. La différence entre les deux diagrammes réside dans le fait que le diagramme de séquence met l'accent sur l'aspect chronologique des communications [ROQUES et al, 2004] alors que le diagramme de communication met l'accent sur les relations structurelles des participants qui interagissent [ROQUES et al, 2004]. Cette différence justifie le choix du diagramme de séquence, qui reste meilleur pour notre système.

5.5.1. Diagramme de séquence associé au cas « Ajouter accident »

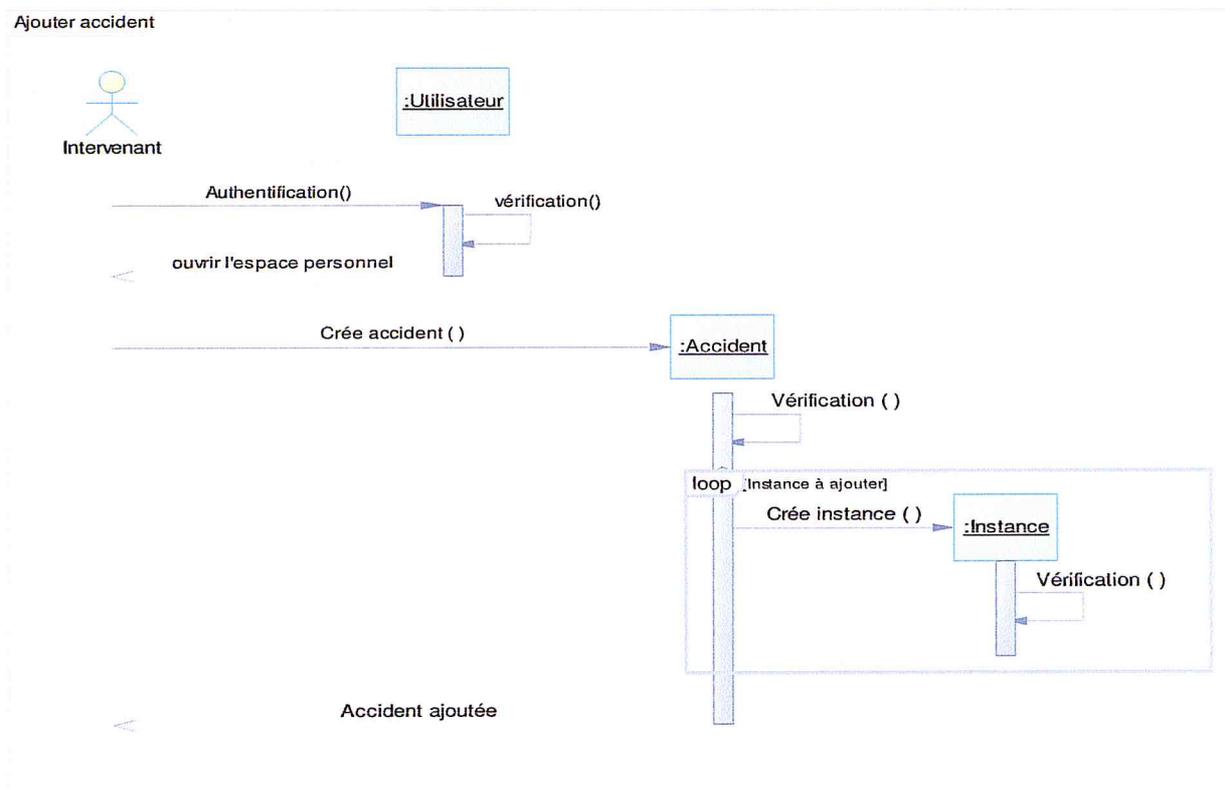


Figure 33. Diagramme de séquence « Ajouter accident ».

5.5.3. Diagramme de séquence associé au cas « Valider accident »

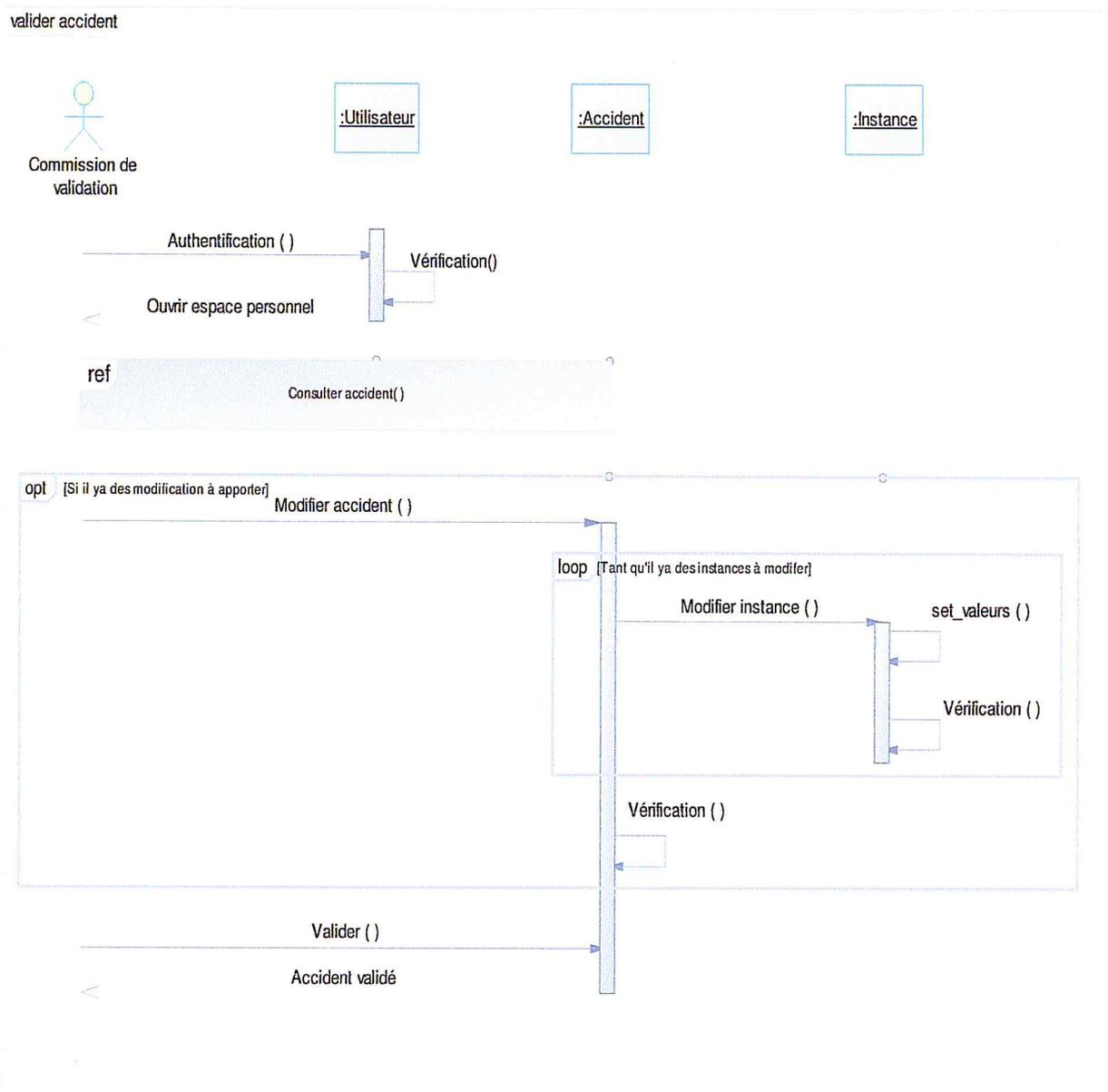


Figure 35. Diagramme de séquence « Valider accident ».

5.5.4. Diagramme de séquence associé au cas « Ajouter rubrique »

DiagrammeSequence_1

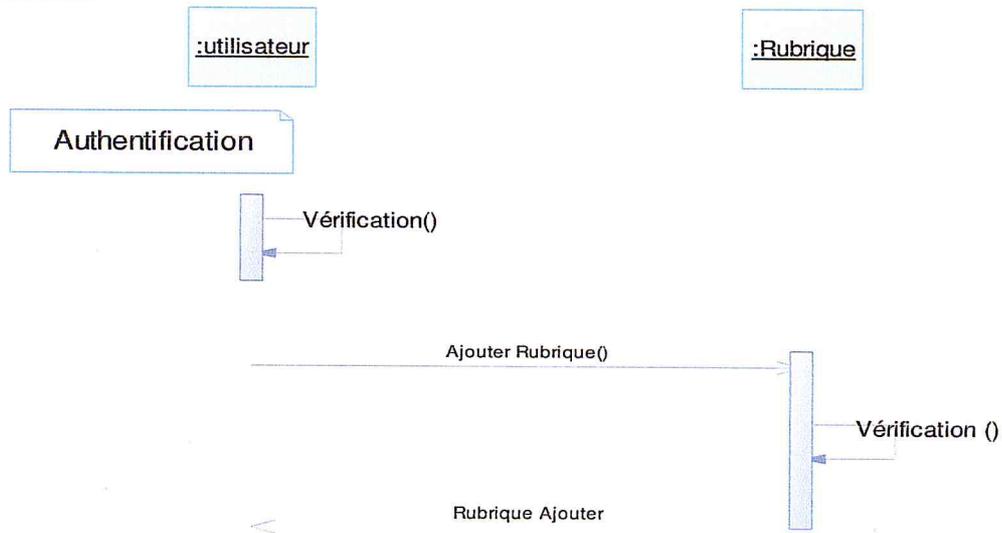


Figure 36. Diagramme de séquence « Ajouter rubrique ».

5.5.5. Diagramme de séquence associé au cas « Ajouter variable »

Ajouter variable

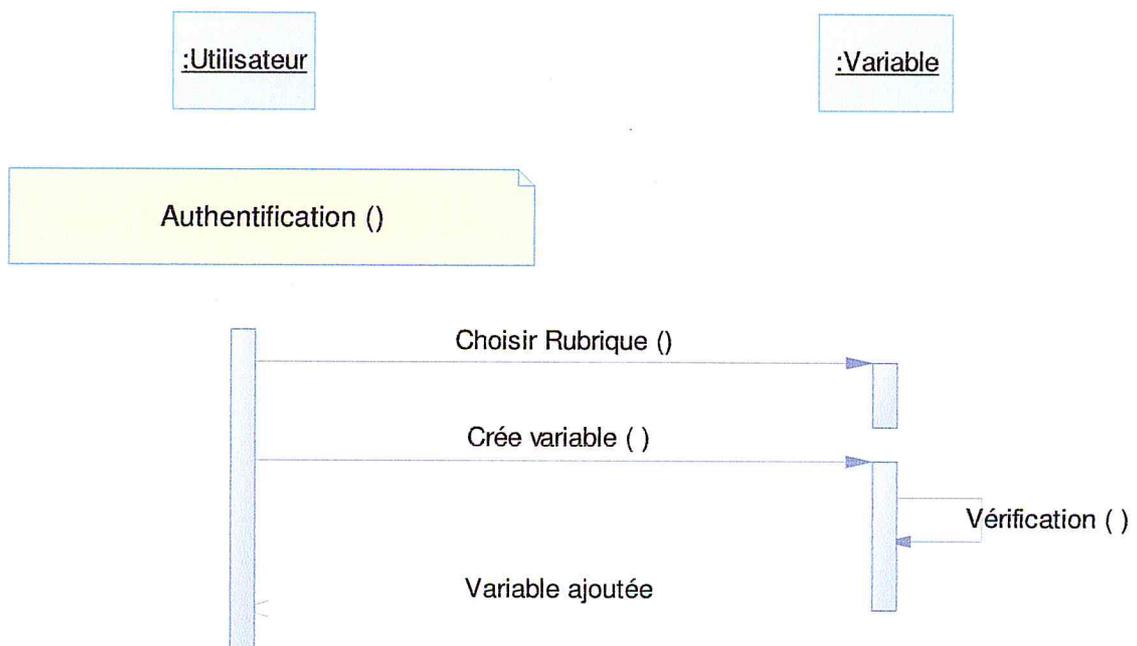


Figure 37. Diagramme de séquence « Ajouter variable ».

5.5.6. Diagramme de séquence associé au cas « Ajouter valeur »

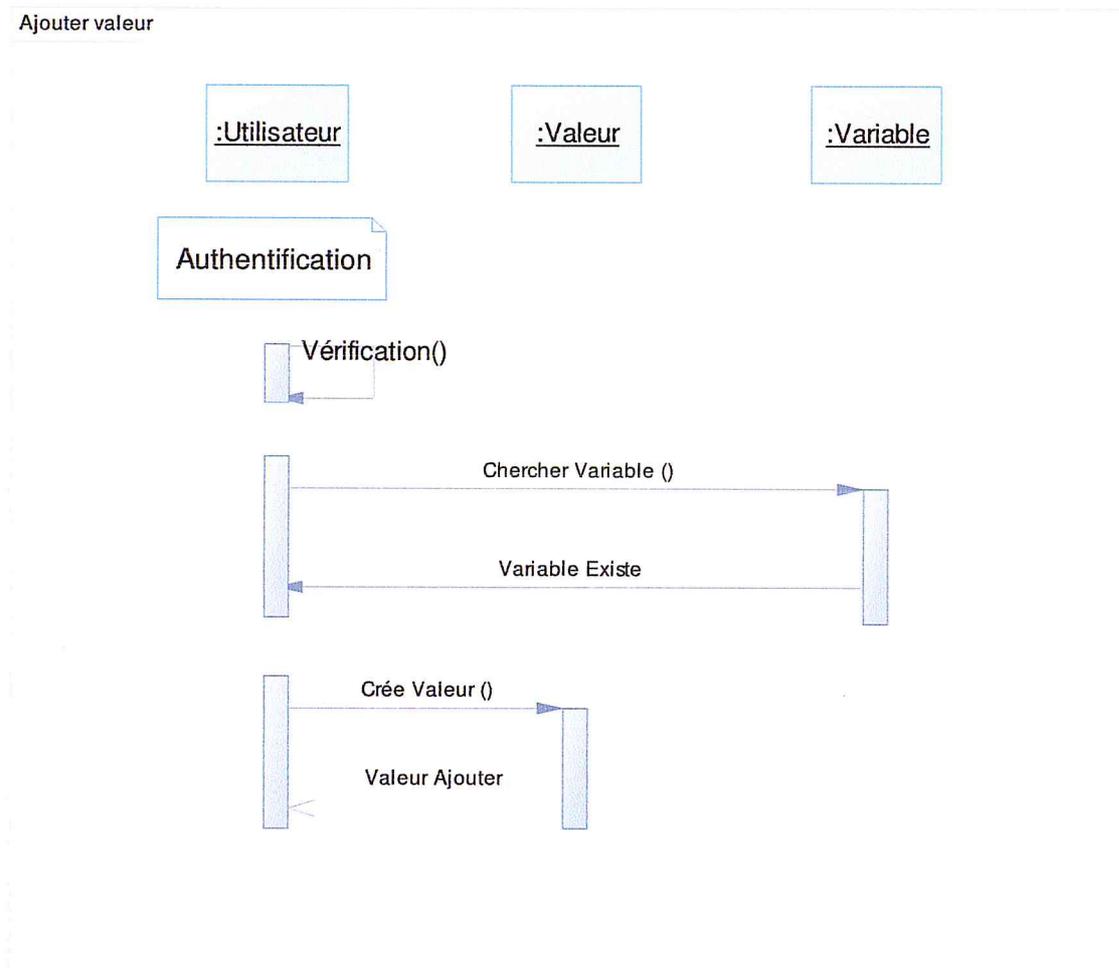


Figure 38. Diagramme de séquence « Ajouter valeur ».

5.5.7. Diagramme de séquence associé au cas « Détection des hot-spots »

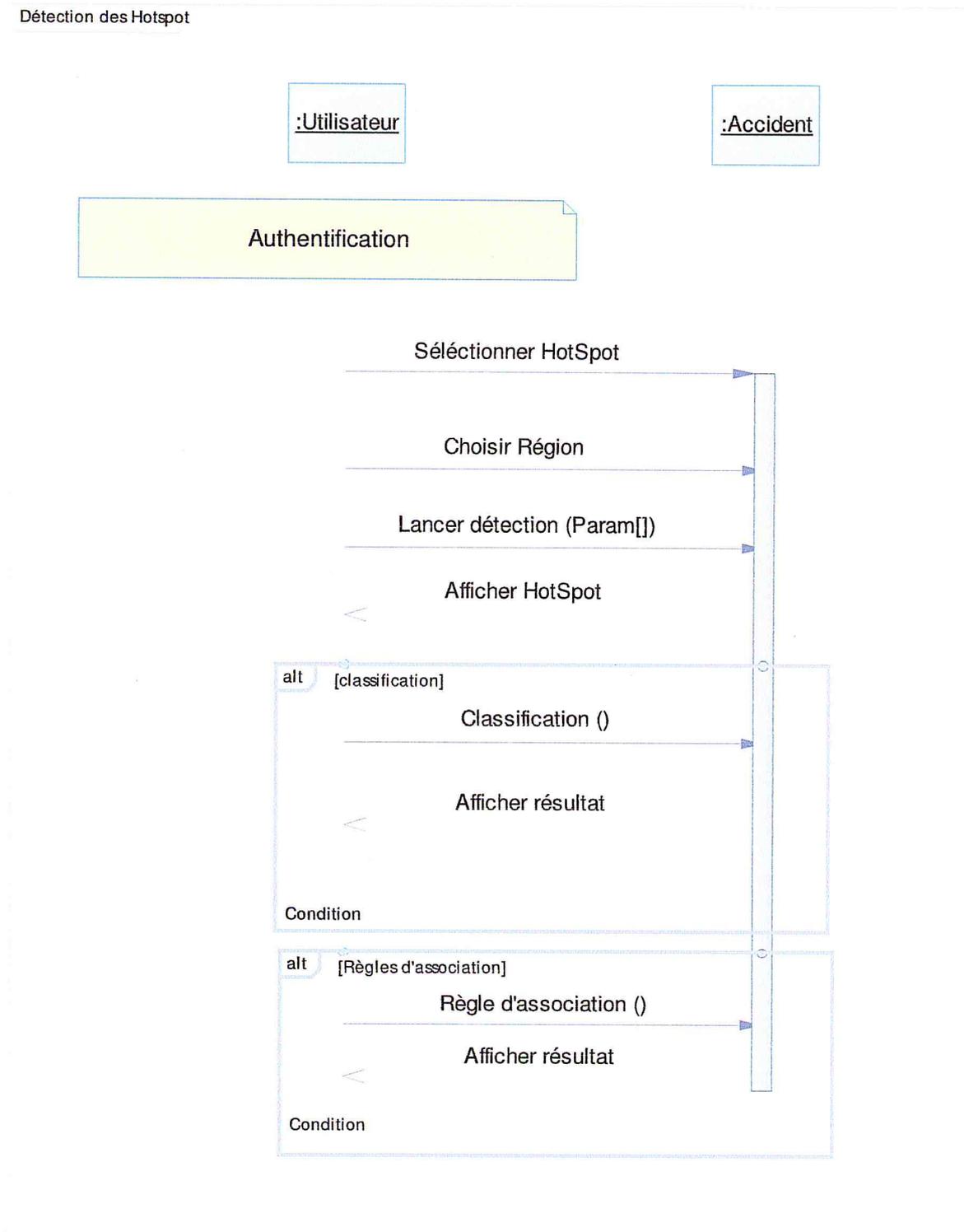


Figure 39. Diagramme de séquence « Détection des hot-spots ».

5.5.8. Diagramme de séquence associés au cas « Classification »

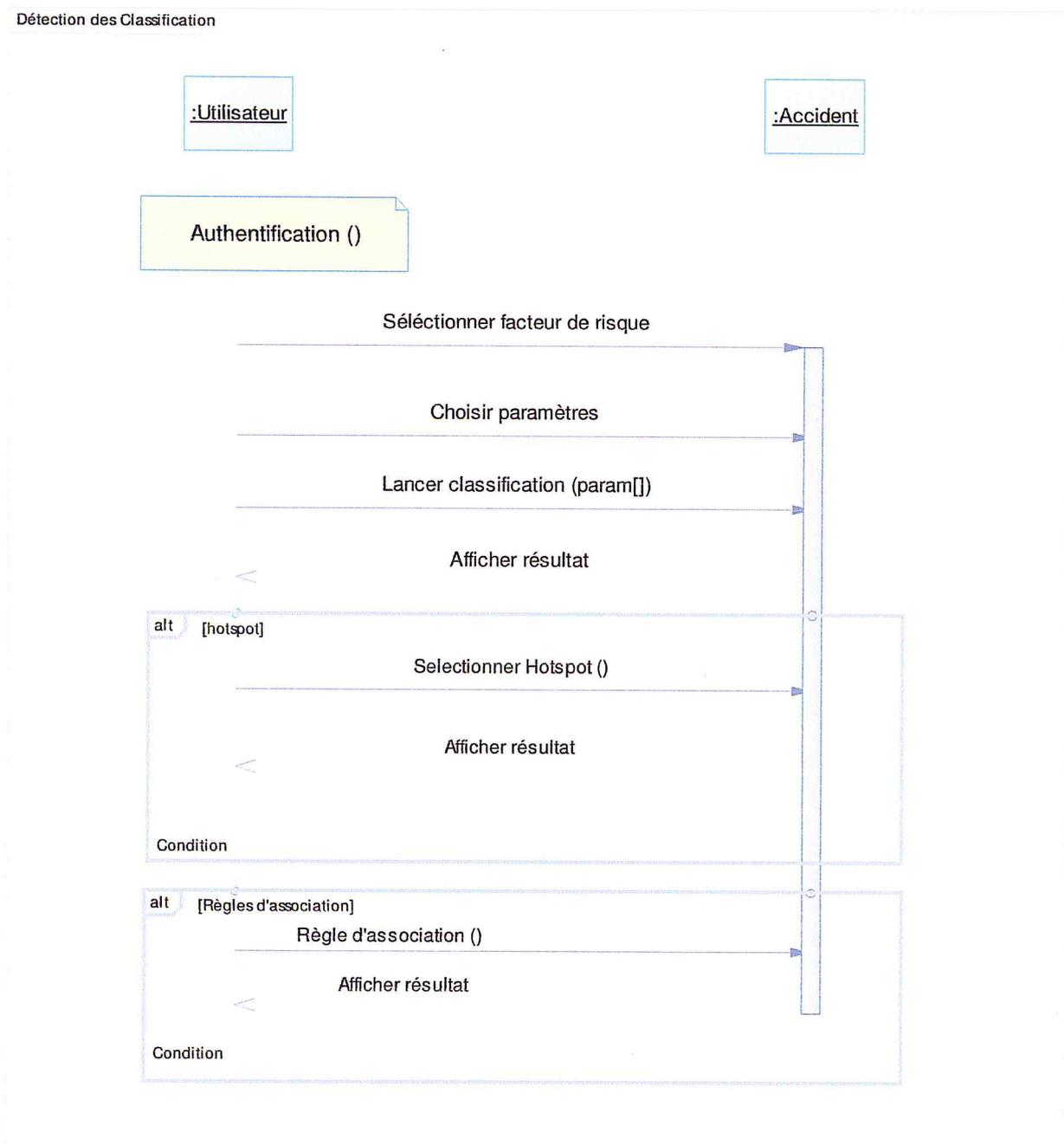


Figure 40. Diagramme de séquence « Classification ».

5.5.9. Diagramme de séquence associé au cas « Prévission »

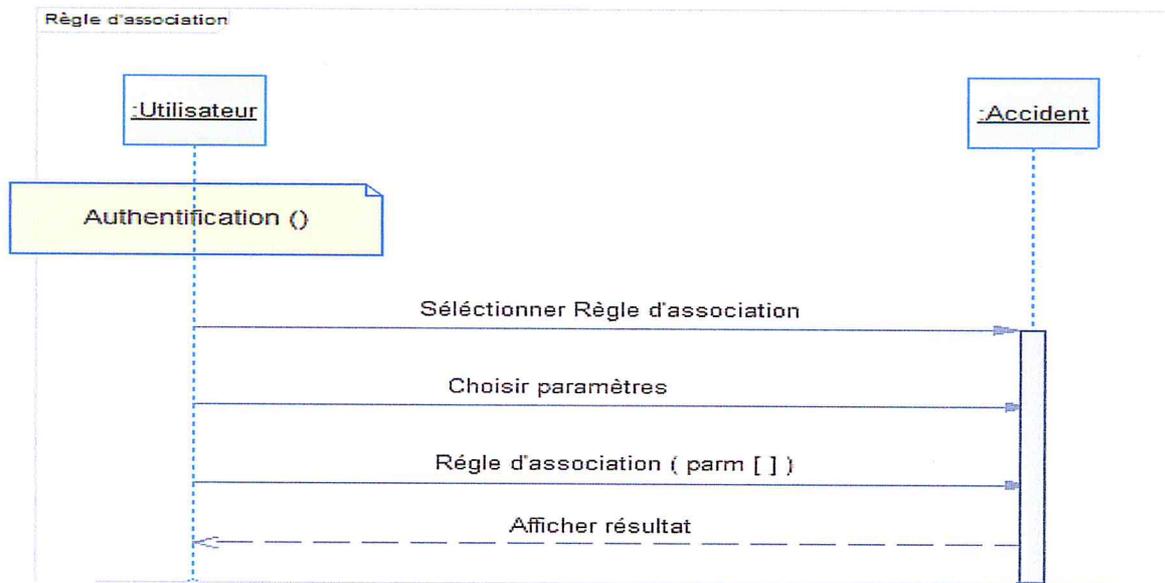
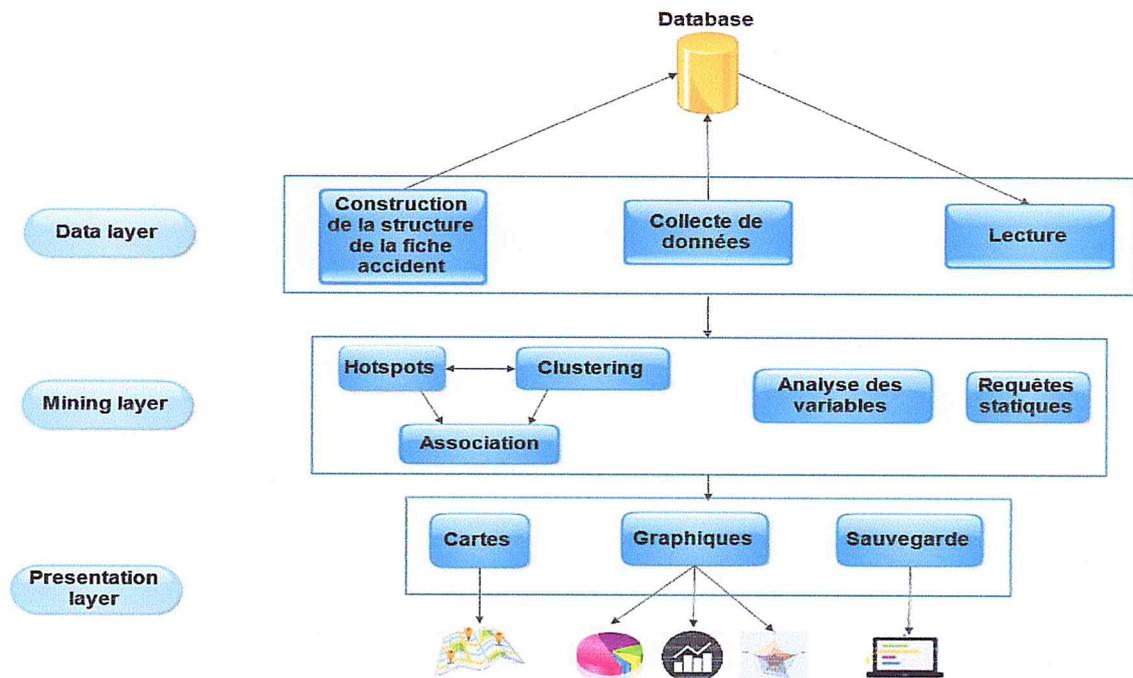


Figure 41. Diagramme de séquence « Prévission »

5.6. Conception détaillé

5.6.1. Architecture technique de l'application

L'application que nous avons conçue, dans le but d'atteindre les différents objectifs cités précédemment, se compose de différents modules. Chaque module est amené à réaliser une tâche précise. La relation entre ces modules est illustrée dans la figure 42.



L'outil « RS-DATA »

Figure 42. Architecture technique de l'outil.

5.6.1.1. Module « Donnée »

Ce module comporte toutes les opérations d'interaction de l'outil avec la base de données. Ces opérations sont réparties sur trois modules : lecture, collecte de données et construction de la structure de la fiche accident (fiche BAAC).

a. Module « Lecture »

Ce module permet de récupérer le contenu de la base de données. Il affiche un sous ensemble de données, selon le souhait de l'utilisateur représenté par une requête SQL, qui délimite le jeu de données.

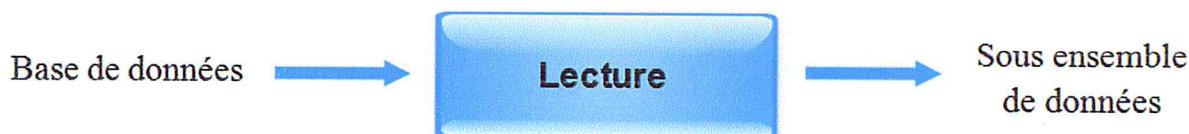


Figure 43. Description du module « Lecture ».

b. Module « Collecte de données »

Ce module peut réaliser la création ou la modification des informations de la base de données.

Ce dernier enveloppe trois sous modules :

- Module « création ».
- Module « modification ».
- Module « suppression ».

▪ Sous Module « Création »

Ce module permet d'insérer de nouvelles données dans la base de données. Ceci permet en pratique d'enrichir la base à analyser avec des informations récoltées en temps réel.



Figure 44. Description du sous module « Création ».

▪ Sous Module « Modification »

Il permet de mettre à jour des données déjà existantes dans la base de données.

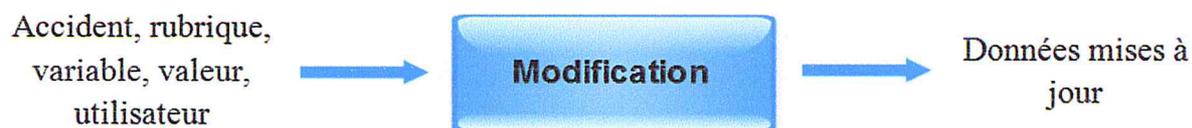


Figure 45. Description du module « Modification ».

▪ Sous Module « Suppression »

Ce module permet de supprimer des informations de la base de données.

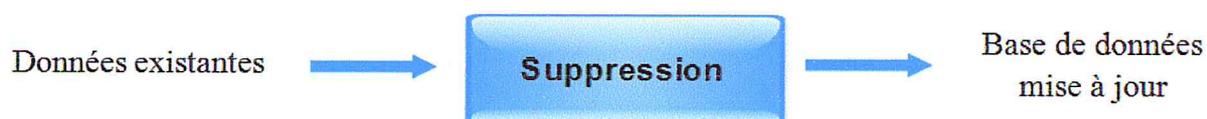


Figure 46. Description du module « Suppression ».

c. Module « Construction de la structure de la fiche d'accident »

Ce module permet de définir la structure de la fiche accident (fiche BAAC). Il consiste à ajouter les rubriques, chacune avec ses variables et les valeurs de ces dernières. En d'autres termes, il permet de définir les différents composants de la fiche accident. Ce dernier regroupe les mêmes sous modules définis au-dessus (module « Collecte de données »).

5.6.1.2. Module « Fouille »

Il regroupe l'ensemble des méthodes de la fouille de données. Dans notre plateforme, nous avons implémenté trois méthodes, structurées en trois modules :

- Module « Clustering ».
- Module « Hot-spots ».
- Module « Prévision ».

a. Module « Clustering »

Ce module permet d'identifier les facteurs de risque les plus dominants. Il permet d'extraire et de comprendre les causes des accidents. Pour ce dernier, on a utilisé la méthode de partitionnement qui est l'une des méthodes de classification.

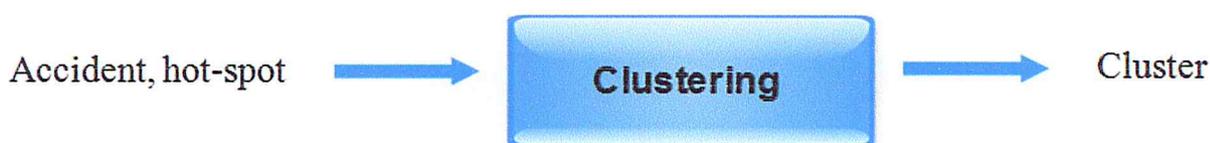


Figure 47. Description du module « Clustering ».

Parmi les algorithmes de partitionnement, nous avons choisi l'algorithme « k-means ».

Chapitre 4 : Etude conceptuelle

Comme le note [Kanungo, et al, 2000] [Larose, 2005] [Matteo, 2008], K-Means est une technique de clustering très populaire dont la principale force réside dans sa simplicité. En effet, il est l'une des techniques de clustering les plus utilisées et son algorithme est simple d'implémentation. De plus, il a prouvé son efficacité dans la production de clusters de bonne qualité dans bien des domaines [Alsabti, et al, 1998].

Le k-means implémenté, il agit soit sur les accidents ou sur les hot-spots pour créer des partitions de données, ayant des caractéristiques communes.

L'algorithme général du k-means est donné ci-après (K étant le nombre de classes à créer):

Algorithme K-means
1. Initialisation : choisir k points centroïdes initiaux.
2. Former les K clusters initiaux en affectant tous les points du jeu de données aux centroïdes les plus proches (Distance minimale)
3. Calculer les centroïdes des clusters formés
4. Si la distance générale entre tous les points et les centroïdes, est minimisée, alors arrêter, sinon aller a 2.

Algorithme 1. Algorithme K-Means [JIEPING, 2007].

Pour l'initialisation du k means, nous nous sommes basées sur l'algorithme ELAGHA INITIALIZATION présenté dans [El AGHA et al, 2012]. Cette technique d'initialisation, récente, a démontré son efficacité par rapport aux méthodes courantes. Elle permet de fournir un ensemble optimal de points références en un temps réduit.

Son principe est de créer une grille sur l'espace d'étude, composée de k colonnes et de k lignes (k étant le nombre de clusters à former).en considérant l'espace d'étude sur deux dimensions x et y, x représentent la longitude et y la latitude. Ces choix peuvent être étendus à plusieurs dimensions (prendre en considération d'autre attributs autre que la localisation spatiale).

L'algorithme calcule un tableau de centres candidats, les centroïdes à retenir sont au niveau de la diagonale de ce tableau (voir Algorithme 3).

ELAghaInitialization

```
1. Xc = Xmin, Yc=Ymin
2. Xw = largeur de la grille / k
3. Yw = hauteur de la grille / k
4. Pour chaque ligne de la grille n<k
   {
     Pour chaque cellule de la ligne i<k
     {
       CentresCandidats[n][i].X = Xc + random(Xw) - random(Xw/2)
       CentresCandidats[n][i].Y = Yc + random(Yw) - random(Yw/2)
       Xc = Xc + Xw
     }
     Xc = Xmin
     Yc = Yc + Yw
   }
```

Algorithme 2. ELAghaInitialization [El AGHA et al, 2012].

Pour le calcul de la distance entre deux point, nous avons utilisé la distance géographique donner comme suit :

Le calcul de la distance géographique S_{A-B} entre deux points A et B, ne peut se faire de la manière traditionnelle, du moment où ces points sont positionnés sur une sphère (globe terrestre). En effet, calculer cette distance revient à calculer l'abscisse curviligne S (AB) sur le grand cercle passant par A et B [Garnir, 2004].

Étant donnés:

\varnothing_A : latitude du point A, λ_A : longitude du point A.

\varnothing_B : latitude du point B, λ_B : longitude du point B.

S_{A-B} : distance géographique entre A et B.

Le calcul de S_{A-B} se fait comme suit :

Calcul de la distance géographique

- Initialiser S_{A-B} à 0.
- Si \varnothing_A , λ_A , \varnothing_B , λ_B sont en degrés, alors transformer ces coordonnées en radian.
- Calculer $S_{A-B} = \arccos(\sin \varnothing_A * \sin \varnothing_B + \cos \varnothing_A * \cos \varnothing_B * \cos(\lambda_B - \lambda_A))$.

Chapitre 4 : Etude conceptuelle

Notre implémentation permet de faire la classification selon 1 seule critère qui est la localisation des accidents (Longitude et latitude)

Toute fois la puissance du K-means n'a pas été vraiment exploité pour la détection des facteurs de risque et ceci est causé par l'absence d'un SIG, ce dernière fournit les données géographique nécessaire pour faire l'analyse et donc fournir les facteurs de risque des accidents en relations avec les caractéristique des lieux

Une autre alternative est posé, c'est d'appliqué le Kmeans classique sur les résultats de notre implémentation, ceci permettra d'avoir les facteurs de risque commun des accidents qui ont des caractéristique spéciale commune, sauf que le résultat concerne les données des accidents et non pas du lieu commun (donnée spatiale).

b. Module « Hot-spot »

Ce module permet de définir les régions de concertations des accidents à partir de la base de données. Le but de l'identification des hots-spots, est de cerner les zones qui nécessitent une attention particulière des forces de l'ordre.

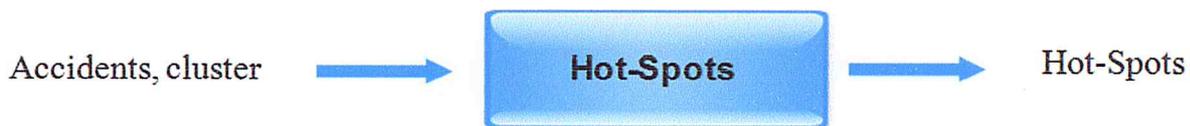


Figure 48. Description du module « Hot-spot ».

Pour la détection des hot-spots nous avons utilisé l'algorithme « KDE ». C'est une méthode de classification non supervisée.

Notre choix a porté sur cet algorithme vu sa popularité qui est dû à sa simplicité et à la qualité de visualisation qu'il fournit. Cela permet une meilleure présentation des résultats et aussi le fait qu'il combine le principe général du clustering basé grille et basé densité pour cerner les zones de concentration de la population étudiée. Cela a fait de lui l'algorithme le plus populaire.

L'algorithme général du KDE est donné ci-après (algorithme 4).

Algorithme Kernel Density Estimation

1. Diviser l'espace d'étude en cellules de même taille W de sorte à créer une grille.
2. Calculer les centres c de cellules.
3. Tracer un voisinage circulaire de rayon h autour de chaque cellule (fenêtre de lissage).
4. Calculer la densité de chaque cellule F , définie par la densité du centre de la cellule :

$$F(c) = \frac{1}{n.h^2} \sum_{i=1}^n K\left(\frac{X_i - c}{h}\right)$$

5. Éliminer les cellules ayant une faible densité $F < S$, où s est le seuil de densité .et n : le nombre de points dans le voisinage, h : fenêtre de lissage, X_i : le i ème point du voisinage circulaire de c , K : fonction noyau (Kernel).

Algorithme 3. Algorithme Kernel Density Estimation [Di Salvo et al, 2005].

c. Module « Préviation »

Ce module permet de prévenir d'éventuelles zones à haute concentration de risques d'accidents, et cela, par l'extraction d'un ensemble de règles d'association régissant les interactions entre les facteurs d'accidents, avec des probabilités calculées.

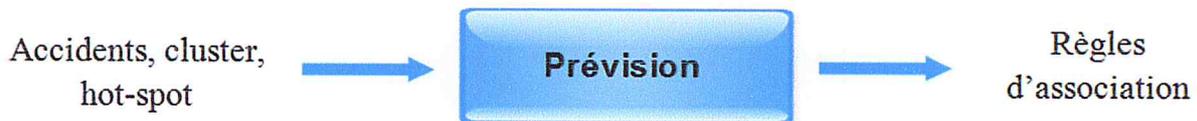


Figure 49. Description du module préviation.

On considère chaque valeur d'attribut de la base comme un item dans l'étude (voir algorithme 5).

Algorithme APRIORI

1. Définition de l'ensemble E des items fréquents i ($P(i) > \text{minsup}$)
2. Construire toutes les paires (i_1, i_2) possibles à partir de l'ensemble E
3. Pour chaque paire (i_1, i_2) , calculer la confiance $\%c$ et le support $\%s$ pour la règle $i_1 \rightarrow i_2$.
4. Retenir les règles d'associations fortes : $\%c > \text{minConf}$ et $\%s > \text{minSup}$.

Pour la découverte des règles d'associations dans les classes hiérarchiques, nous avons besoin de définir des seuils de confiance et support dans chaque niveau.

On considère l'étape sans regroupement (tous les éléments singleton), comme le niveau 0, et on progresse de L niveaux jusqu'à atteindre les k classes créées par l'algorithme hiérarchique.

Pour chaque niveau de granularité l faire :

1. Définition de l'ensemble E des items fréquents i dans le niveau l ($P(i) > \text{minsup}(l)$ et i est fréquent dans tous les niveaux inférieurs $< l$).

2. Construire toutes les paires (i_1, i_2) possibles à partir de l'ensemble E.
3. Pour chaque paire (i_1, i_2) , calculer la confiance $\%c$ et le support $\%s$ pour la règle $i_1 \rightarrow i_2$.
4. Retenir les règles d'associations fortes : $\%c > \text{minConf}(l)$ et $\%s > \text{minSup}(l)$.

Algorithme 4. Algorithme Apriori [Pascal, 2007].

5.6.2. Architecture réseau

Avant de définir le choix d'architecture adoptée pour notre système, nous allons d'abord présenter les différentes architectures réseau existantes:

5.6.2.1. Présentation de l'architecture à 2 niveaux (2 tiers)

L'architecture à deux niveaux (aussi appelée architecture 2-tiers, tiers signifiant rangée en anglais) caractérise les systèmes clients/serveurs pour lesquels le client demande une ressource et le serveur la lui fournit directement, en utilisant ses propres ressources. Cela signifie que le serveur ne fait pas appel à une autre application afin de fournir une partie du service.

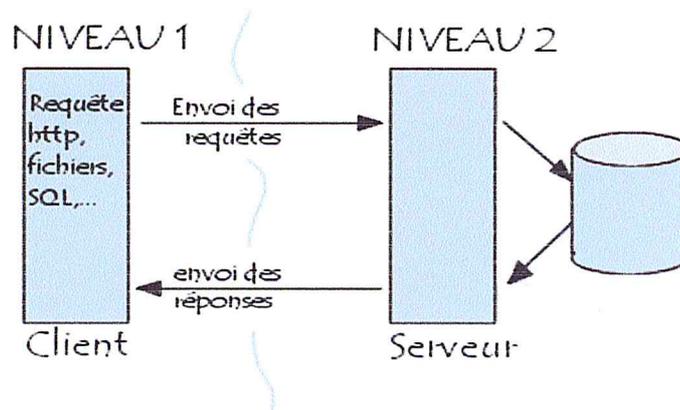


Figure 50. Architecture à 2 niveaux [WEB 6].

5.6.2.2. Présentation de l'architecture à 3 niveaux

Dans l'architecture à 3 niveaux (appelée architecture 3-tiers), existe un niveau intermédiaire (figure 48), c'est-à-dire que l'on a généralement une architecture partagée entre:

1. Un client, l'ordinateur demandeur de ressources, équipé d'une interface utilisateur (généralement un navigateur web) chargé de la présentation.
2. Le serveur d'application (appelé également middleware), chargé de la ressource mais faisant appel à un autre serveur.
3. Le serveur de données, fournissant au serveur d'application les données dont il a besoin.

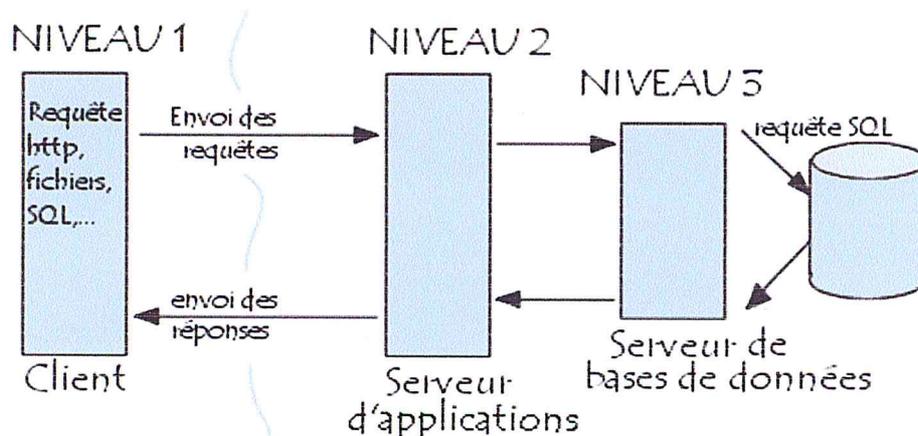


Figure 51. Architecture à 3 niveaux [WEB 6].

5.6.2.3. La solution adoptée

La solution choisie pour notre système est l'architecture « trois tiers ». De point de vue développement, cette architecture permet une meilleure répartition des tâches et fonctions (développeur de modèle/designer, programmeur, administrateur de bases de données...), et enfin le facteur géographique car les utilisateurs sont répartis sur le territoire national. Avec cette architecture, il suffit d'installer un navigateur sur les postes clients pour accéder au système.

Conclusion

La conception de l'outil est une étape primordiale avant la phase de réalisation. Elle a pour objectif de formaliser les étapes de développement du système, pour répondre aux besoins soulevés par l'utilisateur.

Tout d'abord, nous avons commencé par la conception de la base de données. Ensuite, nous avons décrit l'architecture globale de notre application, en énumérant et expliquant chacun de ses modules. Puis, nous avons exposé les différentes interactions entre ces derniers. L'explication de l'utilisation de ces outils, l'implémentation de notre application et la réalisation de BDD feront l'objet du chapitre suivant.

Dans le chapitre suivant, nous allons résumer les différents outils et approches déployés pour la réalisation de notre application ainsi que quelques scénarios d'utilisation possible.

Implémentation

Introduction

Après une étude approfondie des concepts théoriques des techniques à implémenter (Voir chapitres 1, 2 et 3), nous nous intéressons dans ce chapitre à la mise en place de notre outil « RS-DATA », décrit et conçu dans le chapitre de conception.

Ce chapitre est structuré en trois principales parties, la première concerne la présentation des outils de développement, que nous avons utilisés pour la réalisation de cet outil. La deuxième partie, présente quelques scénarios d'utilisation en illustrant les différentes fonctionnalités de l'outil réalisé.

1. Technologies utilisées

1.1. Eclipse Luna

Eclipse IDE, Integrated Development Environment (EDI environnement de développement intégré en français) est un environnement de développement intégré libre (le terme Eclipse désigne également le projet correspondant, lancé par IBM) extensible, universel et polyvalent, permettant potentiellement de créer des projets de développement mettant en œuvre n'importe quel langage de programmation. Eclipse IDE est principalement écrit en Java (à l'aide de la bibliothèque graphique SWT, d'IBM), et ce langage, grâce à des bibliothèques spécifiques, est également utilisé pour écrire des extensions.



Pour notre projet nous avons choisi de travailler avec extension Eclipse Luna pour le développement d'applications web Java EE.

1.2. JAVA 8

Java 8 est la dernière version de Java et offre de nouvelles fonctionnalités, des performances accrues et des corrections de bug pour améliorer l'efficacité de développement et d'exécution des programmes Java [WEB 2].



1.3. HTML et CSS

HTML est un langage informatique utilisé pour créer des pages web, L'acronyme signifie "HyperText Markup Language", ce qui



signifie en français "langage de balisage d'hypertexte".

Cette signification porte bien son nom puisqu' effectivement ce langage permet de réaliser de l'hypertexte à base d'une structure de balisage. Ce n'est pas à proprement parlé un langage de programmation, mais plutôt un langage qui permet de mettre en forme du contenu. Les balises permettent de mettre en forme le texte et de placer des éléments interactifs, tel des liens, des images ou bien encore des animations.

Quand à CSS est l'acronyme anglais de "Cascading Style Sheets" qui peut se traduire par "feuilles de style en cascade". Le CSS est un langage informatique utilisé sur internet pour mettre en forme les fichiers HTML ou XML. Ainsi, les feuilles de style, aussi appelées, les fichiers CSS, comprennent du code qui permet de gérer le design d'une page en HTML.

1.4. JavaScript

JavaScript un langage de programmation directement interprété par le navigateur, crée par Netscape, utilisé pour les pages interactives, il permet de les dynamiser grâce aux animations, les effets sur le texte, contrôle de saisie ... etc.



1.5. Bootstrap de Twitter

Twitter Bootstrap est une collection d'outils, intuitif et puissant pour la création de site web rapide. Il est composé de codes HTML, CSS et Javascript qui en constituent des composants tels que les formulaires, les boutons, des éléments interactifs de navigation ainsi que des extensions javascript facile à intégrer.



Twitter Bootstrap a été conçu par Mark Otto et Jacob Thomson qui en ont fait l'un des projets les plus populaires du web. Il est devenu un outil incontournable du web développeur et du web designer pour la réalisation simple et rapide d'un site web responsive.

1.6. WAMPSEVER

C'est une plateforme de développement web, permettant de faire fonctionner localement (sans se connecter à un serveur externe) des scripts PHP. WampServer n'est pas en soi un logiciel, mais un environnement comprenant deux serveurs (Apache et MySQL), un interpréteur de script (PHP), ainsi que **phpMyAdmin** pour



l'administration Web des bases MySQL.

Il dispose d'une interface d'administration permettant de gérer et d'administrer ses serveurs au travers d'un tray icon (icône près de l'horloge de Windows).

1.7. MySQL

MySQL, le plus populaire des serveurs de bases de données SQL Open Source, c'est un système de gestion de bases de données relationnelles (SGBDR). Il est distribué sous une double licence GPL et propriétaire. Il fait partie des logiciels de gestion de base de données les plus utilisés au monde, autant par le grand public (applications web principalement) que par des professionnels, en concurrence avec Oracle, Informix et Microsoft SQL Server.



1.8. Spring Framework

Spring est un framework libre pour construire et définir l'infrastructure d'une application java, dont il facilite le développement et les tests.

Spring est ainsi un des frameworks les plus répandus dans le monde Java ; sa popularité a grandi au profit de la complexité de Java EE notamment pour ses versions antérieures à la version 5 mais aussi grâce à la qualité et la richesse des fonctionnalités qu'il propose :



- ✓ son cœur reposant sur un conteneur de type IoC assure la gestion du cycle de vies des beans et l'injection des dépendances.
- ✓ l'utilisation de l'AOP.
- ✓ des projets pour faciliter l'intégration avec de nombreux projets open source ou API de Java EE.

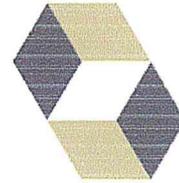
Spring était un framework applicatif à ses débuts mais maintenant c'est une véritable plateforme composée du framework Spring, de projets qui couvrent de nombreux besoins et de middlewares.

Spring est associé à la notion de conteneur léger (lightweight container) par opposition aux conteneurs lourds que sont LES serveurs d'applications Java EE.

1.9. Hibernate

Hibernate est un framework de mapping Objet/Relationnel pour applications JAVA (et .NET avec Nhibernate). Supporter par Jboss/RedHat Hibernate vous permet de créer une couche d'accès aux données (DAO) plus modulaire, plus maintenable, plus performante qu'une couche d'accès aux données 'classique' reposant sur l'API JDBC.

La promesse d'hibernate : libérer le développeur de 95% des tâches de programmation classique d'une couche d'accès aux données via JDBC.



HIBERNATE

1.10. GlassFish

GlassFish est le nom du serveur d'applications Open Source Java EE 5 et qui sert de fondation au produit Sun Java System Application Server de Sun Microsystems. Sa partie Toplink persistance provient d'Oracle. C'est la réponse aux développeurs Java désireux d'accéder aux sources et de contribuer au développement des serveurs d'applications de nouvelle génération de Sun.



1.11. WEKA

Weka est une plateforme d'algorithmes data mining écrite en java implémentant les principaux algorithmes de data mining. Il est disponible gratuitement à l'adresse www.cs.waikato.ac.nz/ml/weka, dans des versions pour Unix et Windows. Ce logiciel est développé en parallèle avec un livre : Data Mining par I. Witten et E. Frank (éditions Morgan Kaufmann), Weka peut s'utiliser de plusieurs façons :

- Par l'intermédiaire d'une interface utilisateur.
- Sur la ligne de commande.
- Par l'utilisation des classes fournies à l'intérieur de programmes Java.



1.12. OpenLayers

OpenLayers est une bibliothèque JavaScript pur pour afficher des données de carte dans la plupart des navigateurs web modernes, sans dépendances côté serveur. OpenLayers implémente une API JavaScript pour la construction de riches applications géographiques basés sur le web, similaires à Google Maps et MSN API Virtual Earth, avec une différence importante - OpenLayers est un logiciel libre, développé par et pour la communauté du logiciel Open Source.



2. Scénario d'utilisation

Nous présentons dans cette section trois scénarios d'utilisation de notre outil, dans le but d'illustrer ses fonctionnalités. Afin de procéder à l'analyse décrite par ces scénarios, l'utilisateur doit être connecté à notre plateforme.

2.1. Scénario 1

Ce scénario décrit les principales étapes de création jusqu'à la validation de l'accident. Le schéma ci-dessous décrit la procédure à suivre :

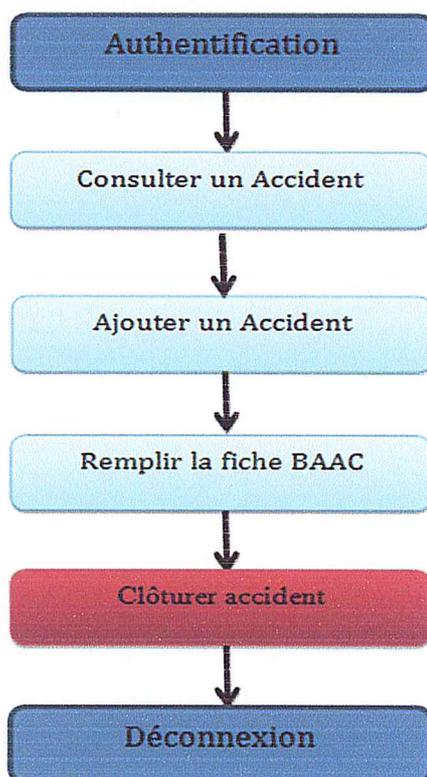


Figure 52. Les étapes du Scénario 1.

Chapitre 5 : Implémentation

Dans ce qui suit nous allons décrire les étapes précédentes avec des captures d'écran depuis l'interface utilisateur (Intervenant).



Figure 53.Interface «Ajouter Accident ».

Une fois que l'utilisateur ajoute l'accident, il peut cliquer sur « Fiche Baac » pour la remplir.

Chapitre 5 : Implémentation

The image displays two screenshots of a web application interface for accident management. The top screenshot shows the 'Liste des Accidents' page, which includes a search bar, a table of accident records, and a sidebar with navigation options. A red box highlights the first row of the table, and a red arrow points to the 'Fiche BAAC' link in the 'Actions' column. The bottom screenshot shows the 'Fiche Accident' page for accident number 20150003654, featuring a sidebar with navigation options and a main content area with 'Informations générales de l'accident:'.

Liste des Accidents

✓ L'accident a été créé avec succès.

Rechercher: + Nouvel accident Recherche avancée

#	Date	Heure	Adresse	Latitude	Longitude	Date d'ajout	Statut	Actions
20150003654	2012-04-20		Cergy Pontoise 95100	50.355685254225	3.2545265444	29-08-2015	En cours	✓ Fiche BAAC
2012000119	2012-12-30	15:18:40	Rte du Faubg Soyeres	50.13555	3.67933	30-12-2012	En cours	✓ Fiche BAAC
20120001062	2012-12-01	04:09:51	88 rte de Kergrist	48.77907	-3.07563	01-12-2012	En cours	✓ Fiche BAAC
20120003892	2012-11-21	03:30:08	D53/ave des pyrénées	43.51175	1.17418	21-11-2012	En cours	✓ Fiche BAAC
20120002387	2012-11-12		Rue le Gorgeu	0	0	12-11-2012	En cours	✓ Fiche BAAC
20120003560	2012-10-15	14:17:39		0	0	15-10-2012	En cours	✓ Fiche BAAC
20120000059	2012-10-11			50.73293	2.6639154	11-10-2012	En cours	✓ Fiche BAAC
20120000170	2012-10-09	11:15:44	rue de Tournai	50.48741	3.26077	09-10-2012	En cours	✓ Fiche BAAC
20120002339	2012-09-13	08:48:04	L'OUCHE A GUERRY	44.757	2.192	13-09-2012	En cours	✓ Fiche BAAC
20120001330	2012-07-01	16:14:44	Rond-point de la Resist	45.2683	6.36769	01-07-2012	En cours	✓ Fiche BAAC

Fiche Accident

Informations générales de l'accident:

Accident N°:	20150003654
Message N°:	1632
Date de l'accident:	2012-04-20
Heure de l'accident:	
Date création de l'accident:	2015-08-29 18:16:46.0
Date dernière modification de l'accident:	2015-08-29 18:16:46.0
Statut:	En cours

Figure 54. Interface « Fiche Baac ».

L'utilisateur peut naviguer sur les différentes rubriques existantes, afin de les compléter.

Chapitre 5 : Implémentation

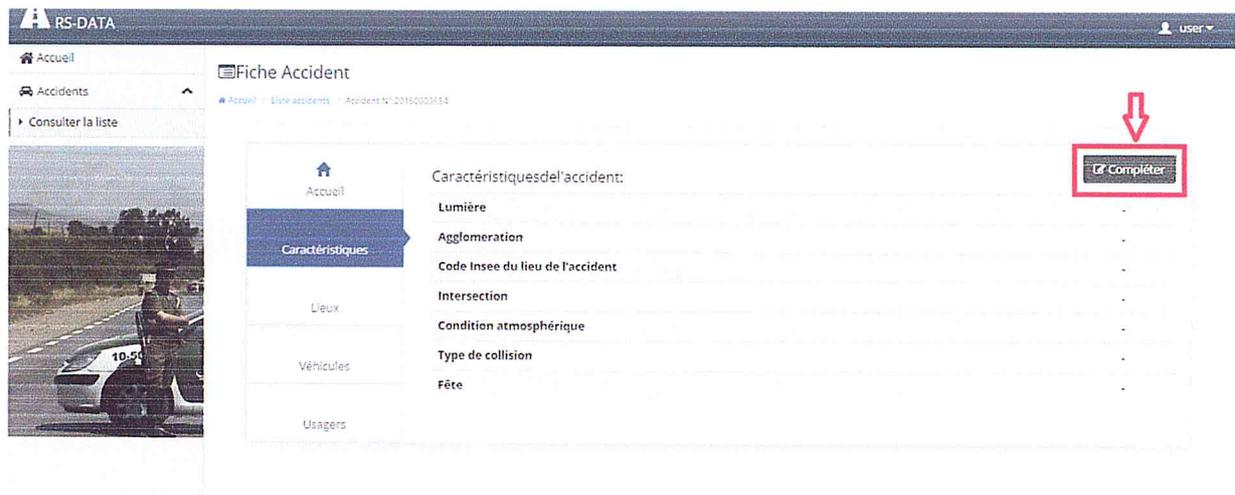


Figure 55. interface des rubriques

D'après la figure ci-dessus, l'utilisateur a choisi de compléter la rubrique « Caractéristiques ».

Un formulaire d'ajout apparaîtra, comprenant les différentes variables de cette dernière.

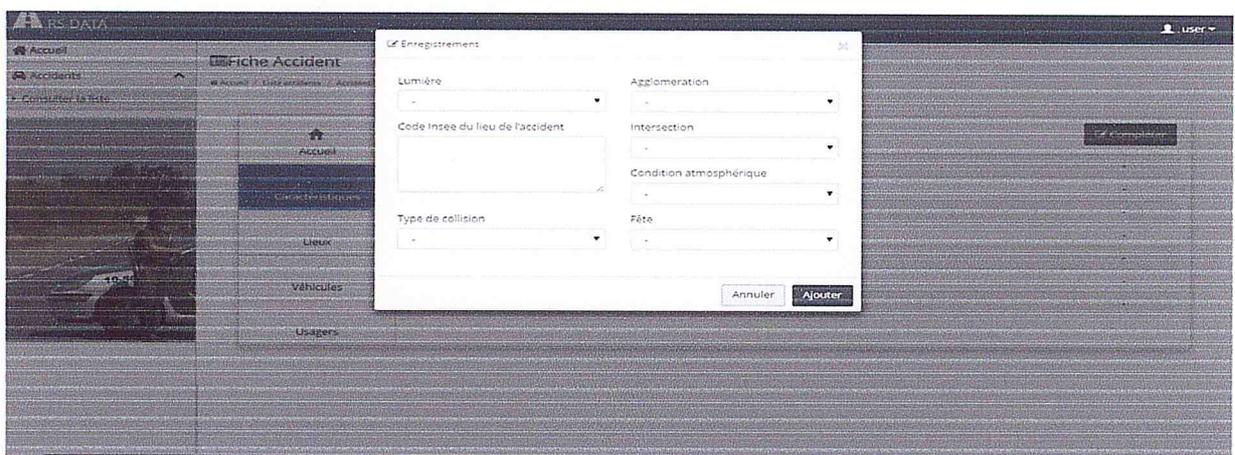


Figure 56. compléter une rubrique

Chapitre 5 : Implémentation

Une fois que l'intervenant termine la saisie des informations sur l'accident, la commission de validation peut consulter les données le concernant afin de le valider.

Dans ce qui suit, nous allons connecter avec un compte utilisateur (commission de validation) ; pour voir les démarches de validation de l'accident.

D'après la figure ci-dessous, le système a vérifié le rôle (les rôles) de l'utilisateur connecté et parmi ce dernier (ces derniers), il doit trouver le rôle « commission de validation ». afin d'afficher la rubrique « clôturer ».

RS-DATA validator16

Accueil

Accidents

Consulter la liste

Fiche Accident

Accueil

Caractéristiques

Lieux

Véhicules

Usagers

Clôturer

Veuillez renseigner les champs suivants avant de clôturer l'accident:

Nombre de morts:
3

Nombre de blessés:
2

Nombre de victimes hospitalisées:
0

Nombre d'indemnes:
2

Clôturer

RS-DATA validator16

Accueil

Accidents

Consulter la liste

Liste des Accidents

Rechercher:

Recherche avancée

#	Date	Heure	Adresse	Latitude	Longitude	Date d'ajout	Statut	Actions
20150003654	2012-04-20			50.355685254225	3.2545265444	28-08-2015	Validé	Fiche BAAC
20120000119	2012-12-30	15:18:40	Rte du Faubg Soyeres	50.13555	3.67933	30-12-2012	En cours	Fiche BAAC
20120002741	2012-12-11	16:20:28		0	0	11-12-2012	Validé	Fiche BAAC
20120002626	2012-12-07	18:03:08	6 rue du broustey	44.941	-4.94	07-12-2012	Validé	Fiche BAAC
20120000538	2012-12-01	18:25:40		45.5212	5.21293	01-12-2012	Validé	Fiche BAAC
20120001062	2012-12-01	04:09:51	88 rte de Kergrist	48.77907	-3.07563	01-12-2012	En cours	Fiche BAAC
20120003844	2012-11-26	19:59:53	route de cépet.	43.73581	1.46254	26-11-2012	Validé	Fiche BAAC
20120003892	2012-11-21	03:30:08	D53/ave des pyrénées	43.51175	1.17418	21-11-2012	En cours	Fiche BAAC
20120000063	2012-11-16	04:47:14	23 RUE DU 8 MAI	50.63658	2.70717	16-11-2012	Validé	Fiche BAAC
20120002387	2012-11-12		Rue le Gorgeu	0	0	12-11-2012	En cours	Fiche BAAC

Showing 1 to 10 of 54 entries (filtered from 4.607 total entries)

Figure 57. Interface « Clôturer Accident ».

2.2. Scénario 2

Dans cette partie nous allons montrer comment un administrateur procède afin de configurer la « fiche Baac ». Les principales étapes sont :

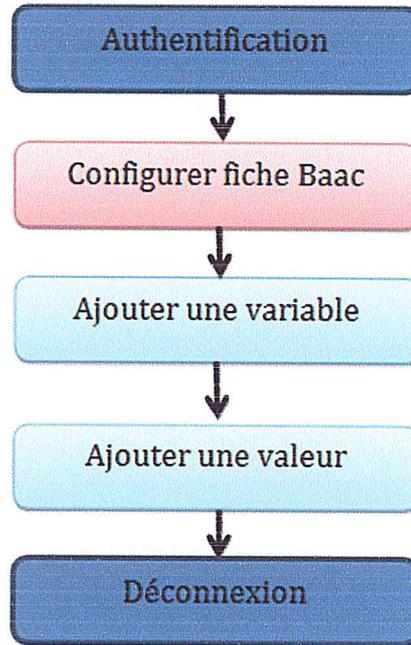


Figure 58. Etapes de scénario 2.

Dans ce qui suit, nous allons décrire l'étape précédente avec des captures d'écran depuis l'interface utilisateur (Administrateur).

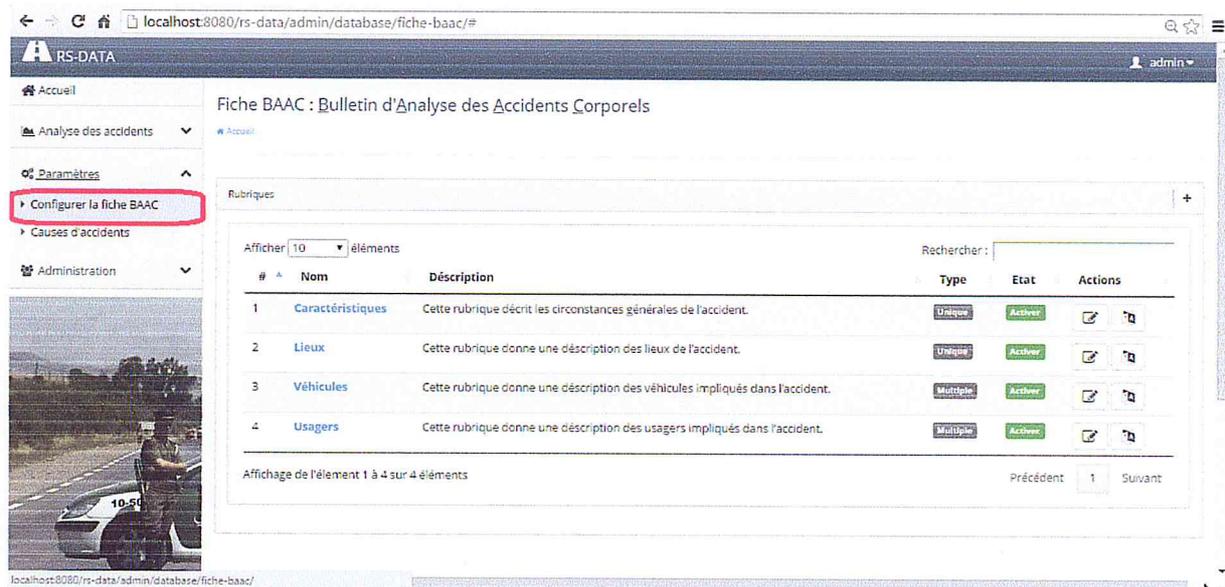


Figure 59. Interface « Configurer fiche Baac ».

Chapitre 5 : Implémentation

- L'administrateur choisit maintenant une rubrique et click sur le bouton « Ajouter Variable ».
- Le système affiche le formulaire correspondant.
- L'utilisateur doit terminer de remplir les champs et valide.

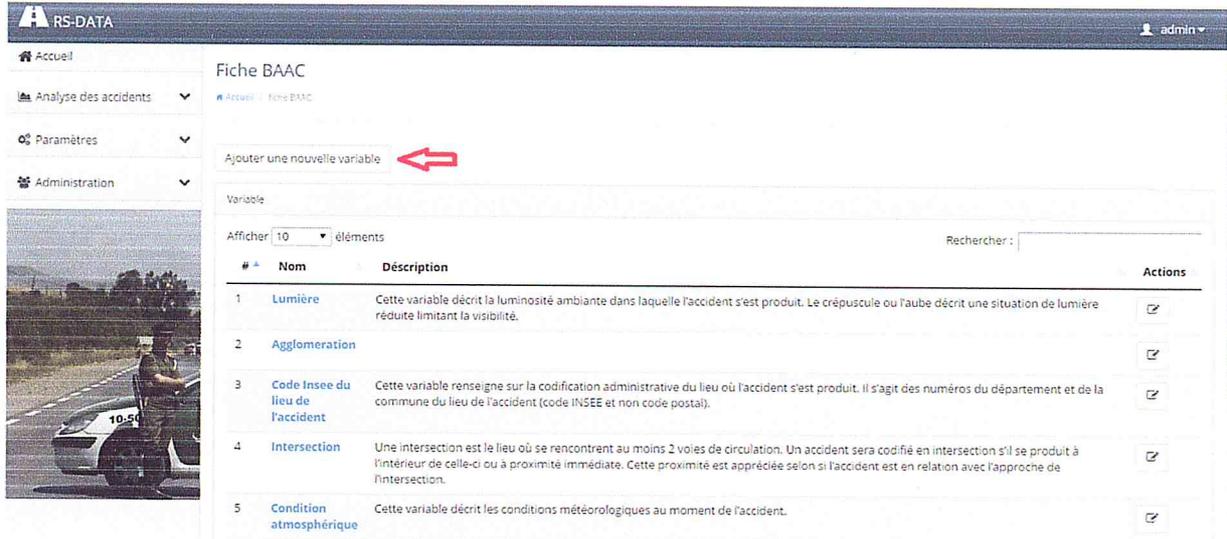


Figure 60. Interface « Liste des variables »

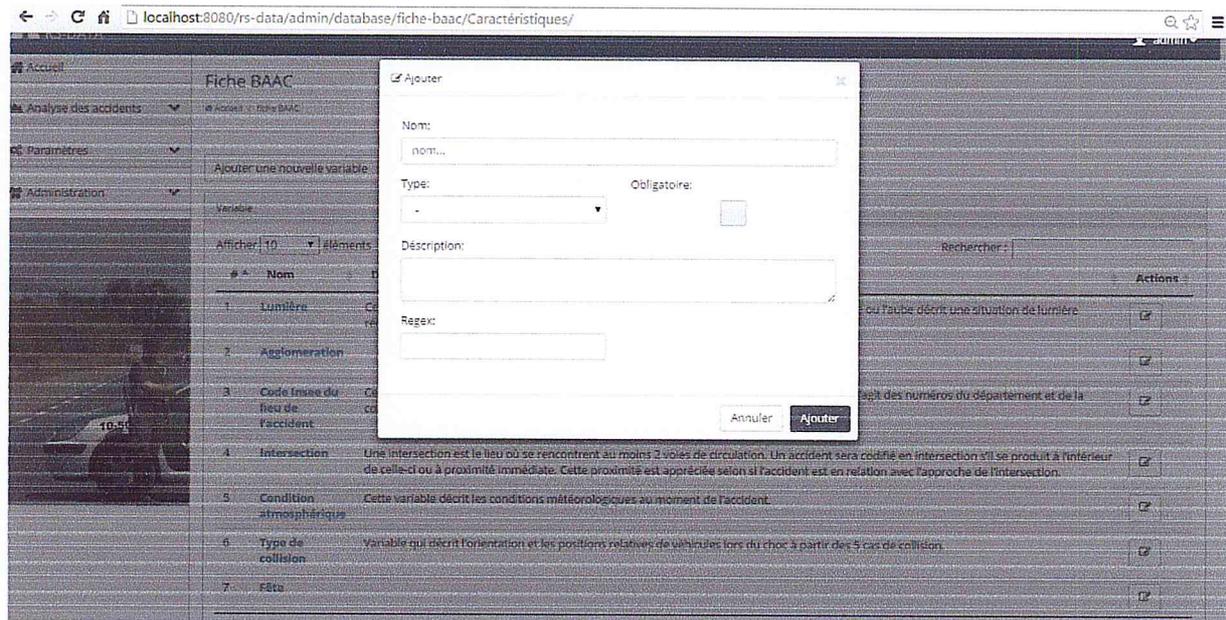


Figure 61. Interface « Ajouter variable »

Chapitre 5 : Implémentation

Si l'administrateur souhaite ajouter des valeurs à une variable, il clique sur le nom de la variable, ensuite il saisit la nouvelle valeur (les nouvelles valeurs) et valide. Ci-dessous les captures d'écran correspondantes.

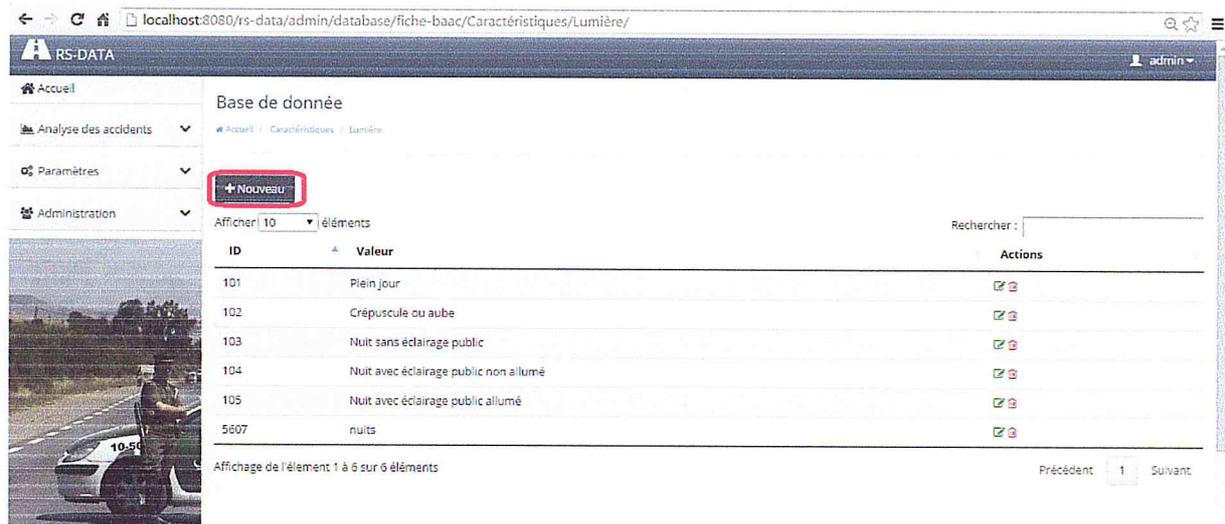


Figure 62. Interface « Liste des Valeurs ».

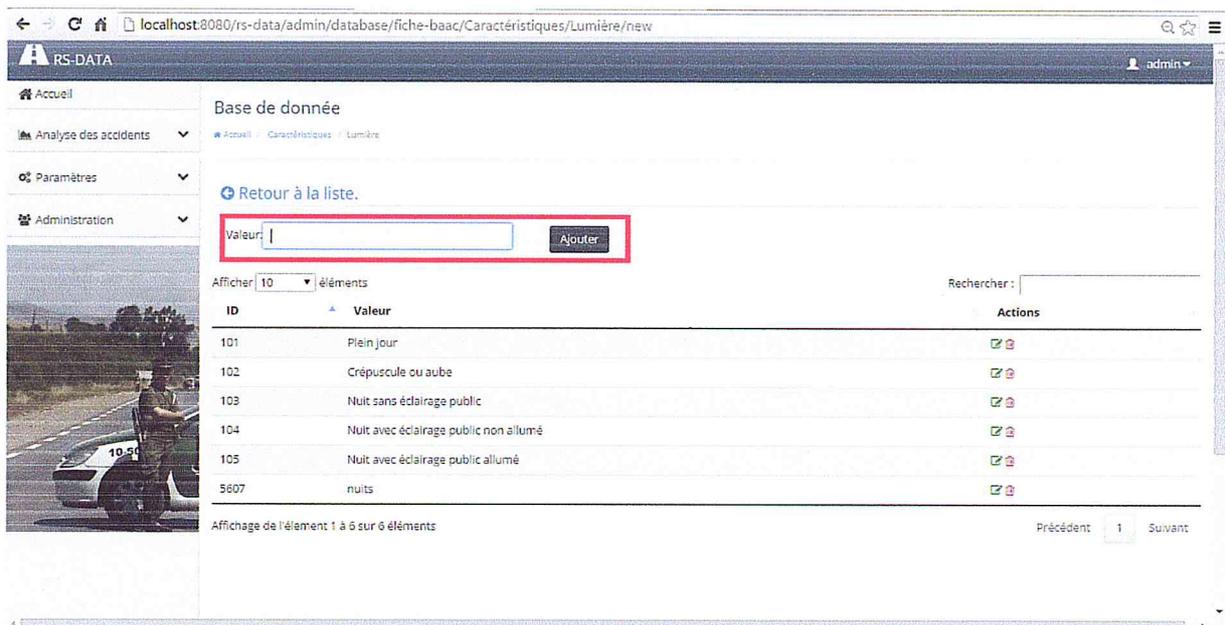


Figure 63. Interface « Listes Ajouter valeur »

2.3. Scénario 3

Dans cette partie, nous allons montrer comment un utilisateur (analyste) procède pour utiliser les algorithmes d'analyse. Les principales étapes sont :

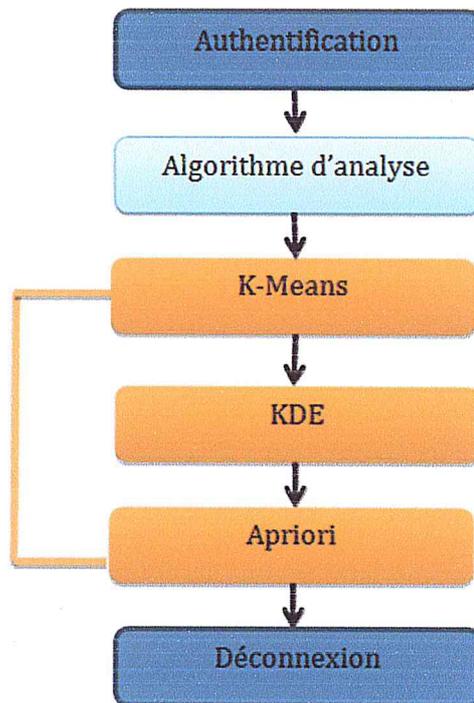


Figure 64.Principales étapes du 3e scénario.

Dans ce qui suit, nous allons décrire les étapes précédentes (voir figure 65) avec des captures d'écran depuis l'interface utilisateur (analyste).

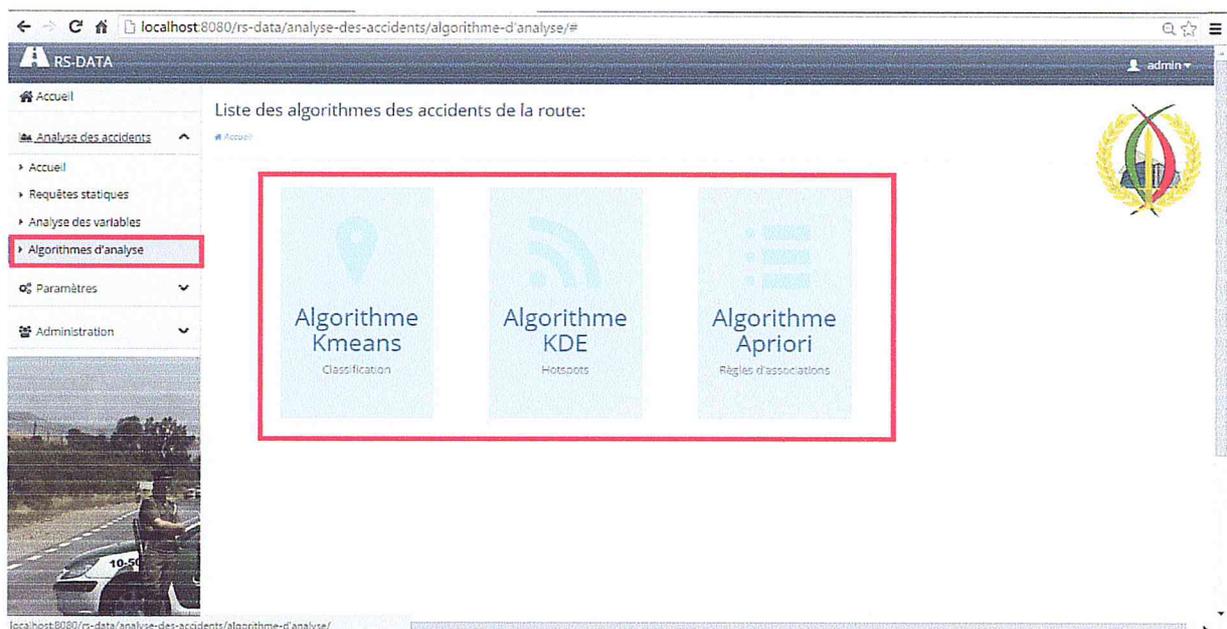


Figure 65.Interface « algorithme d'analyse ».

Si l'analyste veut utiliser l'algorithme KDE pour la détection des Hot-spots, il clique sur « Algorithme KDE ».

Chapitre 5 : Implémentation

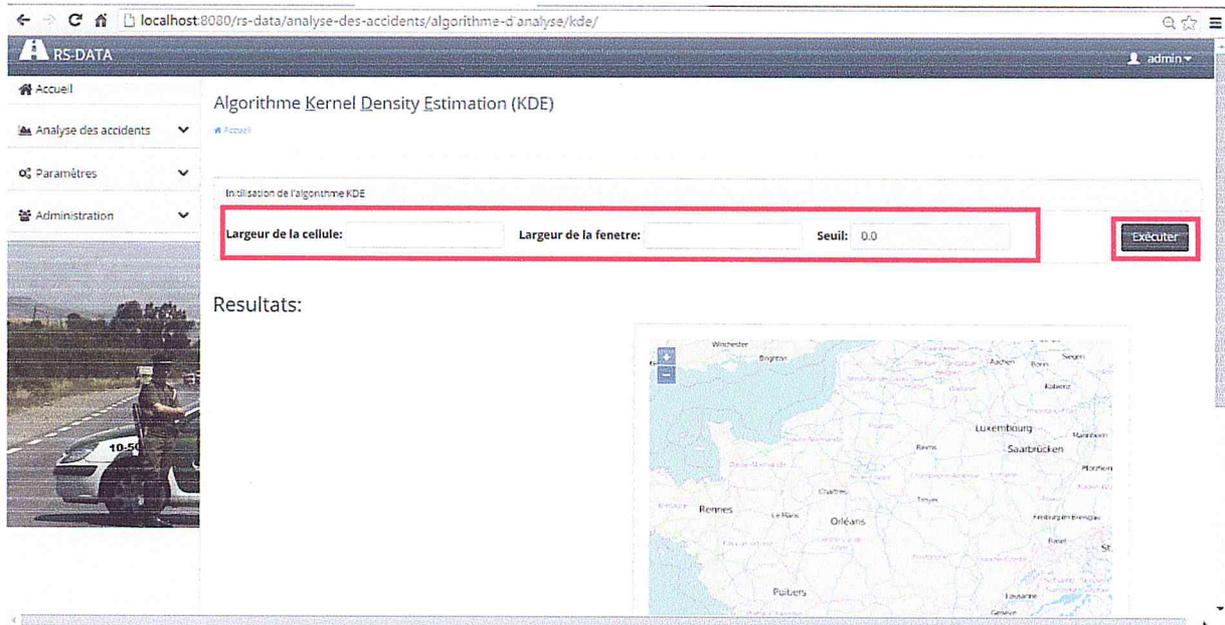


Figure 66. Interface « KDE ».

L'analyste renseigne les champs, puis clique sur « exécuter ». Le système affiche les résultats de l'analyse.

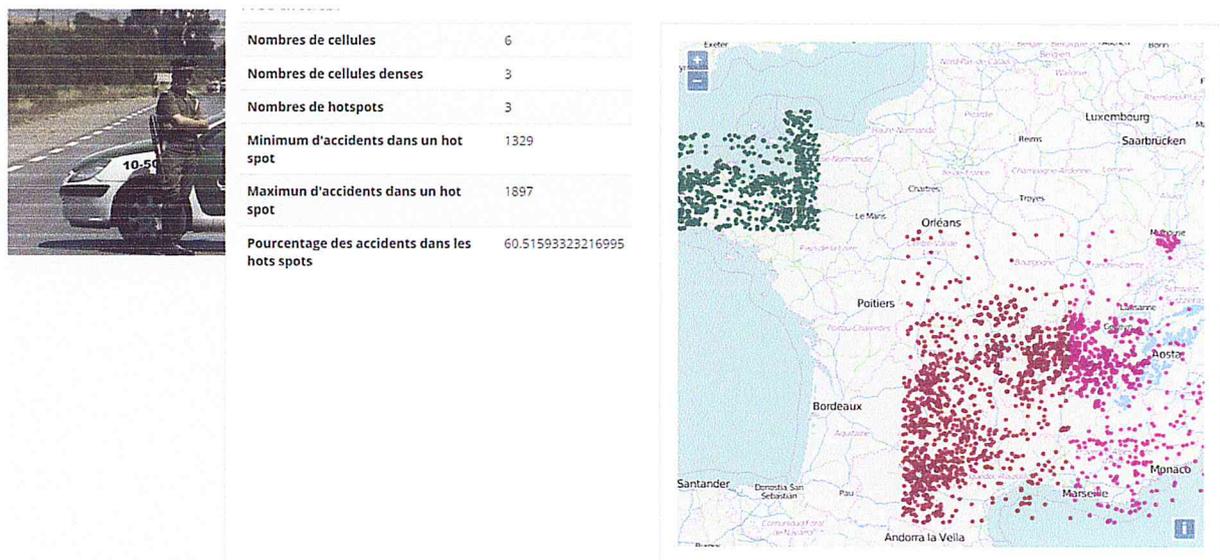


Figure 67. Interface « résultat de l'analyse des hot-spots ».

Un autre cas d'utilisation, l'analyste choisit l'algorithme K-means. Pour cela, il doit choisir « K-means » (voir figure 68), ci-dessous l'interface appropriée.

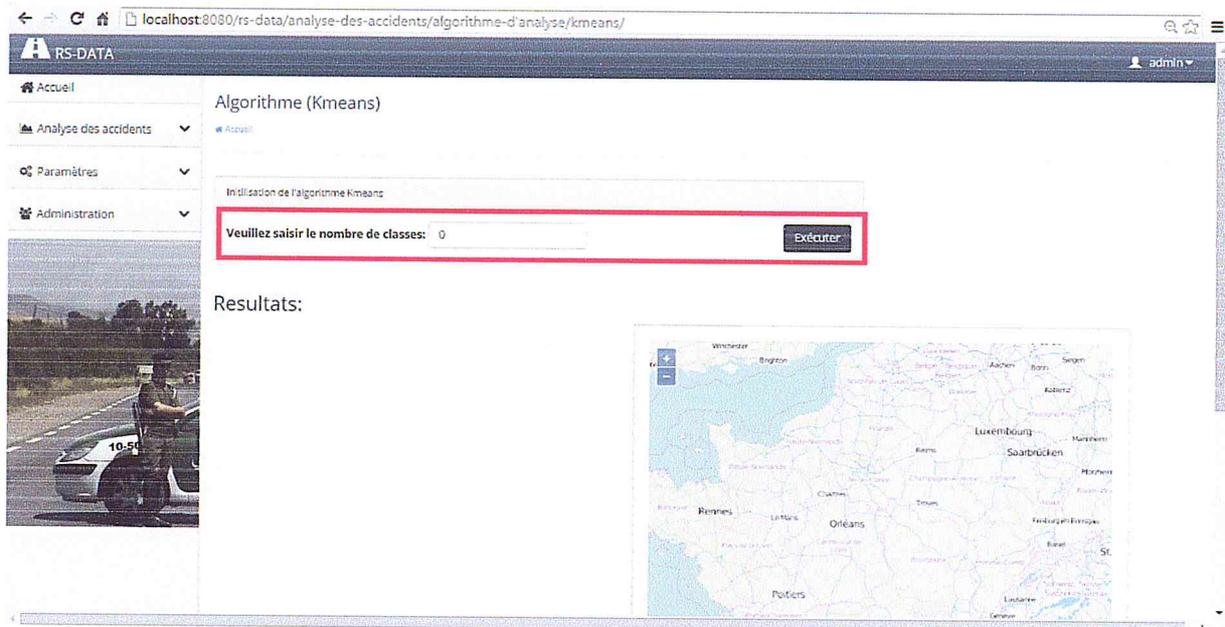


Figure 68. Interface « Algorithme K-means ».

Ensuite, il saisit le nombre de classes (ex: nombre de classe = 4) et lance l'exécution. Ci-dessous le résultat :

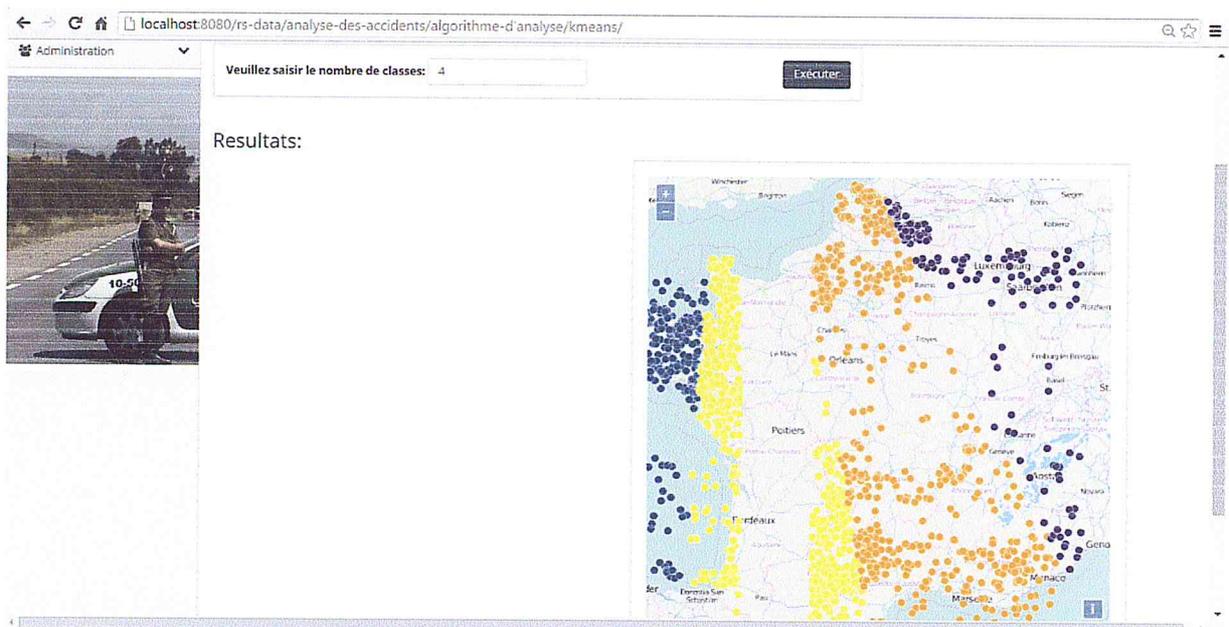


Figure 69. Interface « Algorithme K-means ».

Un autre cas d'utilisation, consiste à étudier l'interaction entre les facteurs de risques. Le décideur utilise donc notre plateforme pour extraire les règles d'association en exécutant l'algorithme « Apriori ». La figure ci-après décrit les étapes de réalisation de ce dernier :

Chapitre 5 : Implémentation

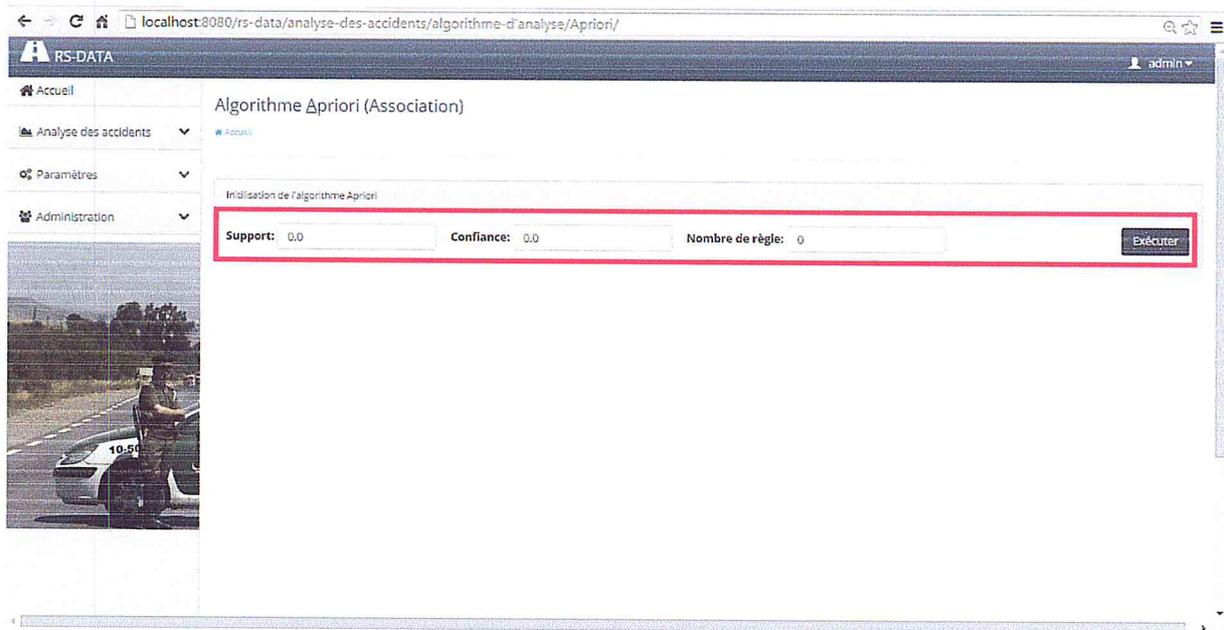


Figure 70. Interface « Algorithme Apriori ».

D'après la figure ci-dessus, l'utilisateur doit donner le support (ex : sup = 0,2), la confiance (ex : conf= 0,5) ainsi que le nombre de règles (ex : nombre de règles =1000) qu'il souhaite avoir après l'analyse, puis lance l'exécution.



Figure 71. Interface « Algorithme Apriori ».

Conclusion

Nous avons présenté dans ce chapitre les détails techniques liés à la mise en œuvre de notre plateforme « RS-DATA ». Nous avons commencé par présenter les technologies utilisées pour la réalisation de la plateforme. Quand la deuxième partie elle été consacré à la présentation des différents scénarios d'utilisations ainsi que les résultats obtenu.

Conclusion et Perspectives

Dans le cadre de notre projet de fin d'étude, nous avons conçu et réalisé une plateforme de collecte et d'analyse des accidents de la route, basé sur les techniques du DMS. Pour ce faire, nous avons divisé l'étude théorique en trois parties. La première représente une étude théorique à travers laquelle nous avons approfondi nos connaissances dans le domaine de l'accidentologie. La deuxième a permis de consulter de près les techniques du DMS en visualisant les généralités concernant la DMS. Quand à la troisième, elle a été consacrée à ses principales tâches ainsi que des études réalisées sur leur application dans domaine de l'accidentologie

Grâce aux notions vues dans la partie théorique, nous avons pu identifier les méthodes les plus adaptées pour l'analyse des accidents de la route. En nous appuyant sur ses acquis, nous avons mis en place la conception de notre solution.

D'abord, nous avons conçu une base de données des accidents de la route. Ensuite, nous avons entamé la conception et la réalisation de l'ensemble des fonctionnalités de l'outil. A la fin de la réalisation de l'outil, nous avons abouti aux fonctionnalités suivantes :

- Gestion des accidents.
- Gestion de la fiche BAAC.
- Gestion des utilisateurs.
- Détection des zones à haute concentration d'accidents.
- Extraction des facteurs de risque des accidents.
- Extraction des règles d'associations qui permettent la prévention des accidents de la route.
- Sauvegarde des résultats obtenus.
- Affichage des statistiques des accidents en Algérie.

Néanmoins, les améliorations à apporter à notre travail sont nombreuses. En effet, il serait intéressant de :

- Étendre la plateforme en plusieurs langues, en particulier la langue Arabe.
- Etendre notre solution à des plateformes mobiles (iPad, smartphone..) pour permettre principalement d'enrichir la base de données à temps réel.
- Analyser n'importe qu'elle zone sur la carte en cliquant dessus.
- La communication entre les différents algorithmes.

Conclusion Générale

- Permettre la communication avec d'autres SIG, afin de profiter pleinement des techniques de fouille de données spatiales.
- En cas de mot de passe oublié, envoyer un mail à l'administration.

Ce projet de fin d'étude, étant une première expérience personnelle et professionnelle, nous a permis d'une part, de concrétiser nos connaissances théoriques acquises pendant le cursus universitaire, et d'autre part, de nous familiariser avec un environnement dynamique et professionnel tout en s'adaptant à l'exigence informationnelle et technologique du domaine.

Références Bibliographiques

A

- [Andersson, 2009] ANDERSSON T.K. “Kernel density estimation and K-means clustering to profile road accident hotspots”. IN Accident Analysis and Prevention 4th. pp. 359–364. 2009.
- [Alsabti et al, 1998] Alsabti, Khaled, Ranka, Sanjay and Singh, Vineet. 1998. An Efficient K-Means Clustering Algorithm. 1998.
- [Agrawal et al, 1998] Agrawal, Rakesh, Johannes Gehrke, Dimitrios Gunopulos and Prahakar Raghavan, Automatic Subspace Clustering of High Dimensional Data for Data Mining Application. *Proceedings of the 1998 ACM-SIGMOD International Conference on Management of Data*. Juin 1998.
- [AUDIBERT, 2009] AUDIBERT, Laurent. UML 2 De L'apprentissage à La Pratique. Ellipses. Info +, 2009. [Online]. Available : <http://laurent-audibert.developpez.com/CoursUML/html/Cours-UML047.html>.

B

- [B. Megherbi, 1994] B. Megherbi, “L’insécurité d’une ville moyenne approchée par l’analyse fine des accidents et leur agrégation en scénarios-types,” Master’s thesis, Université Paris XII, 1994.
- [B.Idiri, 2012] BILAL IDIRI, ALDO NAPOLI. Découverte de règles d’association pour l’aide à la prévision des accidents maritimes. Yves Lechevallier, Guy Melançon, Bruno Pinaud. Conférence internationale Extraction et Gestion des Connaissances - EGC 2012, Jan 2012, Bordeaux, France. Hermann Editions, p. 243-248 - ISBN 978-2-7056-8310-8. <hal-00747588>

Références bibliographiques

- [Bogorny, 2005] Bogorny V., Martins E., Alvares L.O. "A Reuse-based Spatial Data Preparation Framework for Data Mining". In Proceedings of the 17th International Conference on Software Engineering and Knowledge Engineering. pp. 649-652. 2005.
- [Brunner et al., 2001] BRUNNER R, DJORGOVSKI G, PRINCE T, SZALAY A. Chapter 1: Massive datasets in astronomy. Maryland: Université de Johns-Hopkins, 2011.

C

- [Cabanes, 2010] CABANES G. Classification non supervisée à deux niveaux guidée par le voisinage et la densité. Thèse de doctorat. Paris: Université Paris 13 - École Doctorale Galilée, 2010.
- [Chelghoum et Zeitouni, 2004] Chelghoum .N. et Zeitouni K., "Data mining spatial un problème de data mining multi-tables", RIST 134 Vol.14 N°02, pp. 129-145. ISSN 1111-0015. 2004.
- [Chelghoum et al, 2002] CHELGHOUM N, ZEITOUNI K, BOULMAKOUL A. Fouille de données spatiales par arbre de décision multi-thème. EGC, pp. 281-286, 2002.
- [Cleuziou, 2004] Cleuziou.G., "Une méthode de classification non-supervisée pour l'apprentissage de règles et la recherche d'information", Thèse de doctorat, Université d'Orléans, 2004.

D

- [Di Salvo et al, 2005] Di Salvo M., GADAIS M., ROCHE G., WOILLEZ. "L'estimation de la densité par la méthode des noyaux ; méthode et outils". Rapport technique : Centre d'étude sur les réseaux et l'urbanisme, Lyon. 2005.
- [De Salvo.H, 1951] De Silva.H, « why we have automobile accidents », Wiley, ch. Pedestrians, pp 220–239, 1951.
- [D. Wood et O'Riordain, 1994] S. D. Wood and S. O'Riordain, "Monte Carlo simulation methods applied to accident reconstruction and avoidance analysis," in SAE ICE, paper 940720, 1994.

Références bibliographiques

[DTLR, 2000] DTLR, "OTS accident research on the spot," 2000.

E

[El AGHA et al, 2012] EL AGHA M., M.ASHOUR M. "Efficient and fast initializationalgorithm for kmeansclustering". IN Modern Education and Computer Sciences Pres. 2012.

[Elghazel, 2007] ELGHAZEL H. Classification et prévision des données hétérogènes : Application aux trajectoires et séjours hospitaliers. Thèse de doctorat. Lyon: Université Claude Bernard Lyon 1. 2007.

[Ester et al, 1997] ESTER M., KRIEGEL H.-P., SANDER J., "Spatial Data Mining: A Database Approach", *Proc. 5th Symp. on Spatial Databases*, Berlin, Germany, 1997.

F

[Foret-Bruno et al, 2001]. Foret-Bruno, J. Y., Huère, J. F., & Page, Y, *Projet de recherche E. LY. SA : Epidémiologie Routière et Analyse des Situations Accidentelles*, 2001

[Ferrandez..F, 1995] Ferrandez.F, *L'étude détaillée d'accidents orientée vers la sécurité primaire*. Presses ENPC, pp. 88-89, 1995.

[Ferrandez.F, et al, 1986] Ferrandez.F, Fleury.D, Malaterre.G *l'étude détaillée d'accidents (EDA), une nouvelle orientation de la recherche en sécurité routière, recherche transport sécurité*, 9-10, 7-33, 1986.

[Fleury.D, 1998] Fleury.D, *Sécurité et urbanisme - La prise en compte de la sécurité routière dans l'aménagement urbain*. Presses ENPC, 1998.

G

[G. Forbes et B. Malone, 2001] G. Forbes and B. Malone, "Are we speaking the same language? defining systemic safety problems," in *Proceedings of the Canadian multidisciplinary road safety*

Références bibliographiques

conference XII, June 10–13, 2001.

H

- [Hannou et Meddani, 2013] Hannou F, Meddani N "Conception et réalisation d'un outil d'analyse d'accidents de trafic routier ", Ecole National Supérieur D'informatique, Algérie, 2013.
- [Hanoi et Tuan, 2004] le Anh Tuan, IFI Hanoi" réduction de base de données par la classification automatique ", Institut de la francophonie pour l'informatique. Décembre 2004.

J

- [J. Schofer et al, 1995] J. Schofer, K. Kaufer Christoffel, M. Donovan, J. Lavigne, R. Tanz, and K. Wills, "Child pedestrian injury taxonomy based on visibility and action," *Accid. Anal. and Prev.*, vol. 27, no. 3, pp. 217–233, 1995.
- [Jayet, 2001] JAYET H. "Econométrie et données spatiales : Une introduction à la pratique". *Cahiers d'économie et sociologie rurales*, n° 58-59. 2001.
- [JIEPING, 2007] JIEPING Y. "Numerical linear algebra for data exploration-Clustering". 2007.
- [J.M .Stellman, 2000] Jeanne Mager Stellman, « théories des causes », *Encyclopédie de sécurité et de santé au travail*, vol 2, pp 56.6-56-7, 2000.
- [J.-Y. le Coz et Y.Page, 2003] J.-Y. le Coz et Y.Page, « la démarche accidentologique au service de l'évolution de la sécurité des véhicules », *revue de la gendarmerie nationale*, vol.207, 2003.
- [Johnson R, 1992] Johnson R, *Applied Multivariate Statistical Analysis*. Prentice Hall, 1992.

K

- [Karypis et al, 1999] Karypis, George, Eui-Hong Han, and Vipin Kumar,

Références bibliographiques

- [Keefe et al., 2011] Hierarchical Clustering Algorithm Using Dynamic Modeling. *IEEE Computer*. Août 1999, Vol. 8, 32, pp. 68-75.
- [Kanungo et al, 2000] KEEFE R., SULLIVAN T. "Resource-constrained Hot spot identification". RAND technical report: National Security Research Division. p 92. 2011.
- [Koperski et Han, 1995] Kanungo, Tapas, et al. 2000. The Analysis of a Simple k-Means Clustering Algorithm. 2000.
- [Koperski et Han, 1995] K. Koperski and J. Han. Discovery of spatial association rules in geographic information databases. In *Proc. 4th Int'l Symp. on Large Spatial Databases (SSD'95)*, pages 47--66, Portland, Maine, Aug. 1995.
- [Kolatch, 2001] Kolatch Erica, "Clustering Algorithms for Spatial Databases: A Survey". Department of Computer Science. University of Maryland, College Park. 2001.
- [Kolatch Erica. 2001] Clustering Algorithms for Spatial Databases: A Survey. 2001.
- [Koperski et Han, 1998] KOPERSKI K., HAN J., and STEFANOVIĆ N., "An Efficient Two-Step Method for Classification of Spatial Data", In *Proc. International Symposium on Spatial Data Handling (SDH'98)*, p. 45-54, Vancouver, Canada, July 1998.
- [Kraak, 1997] KRAAK M.J, VAN D.E.R. Principles of hypermaps. Computers and geosciences, pp. 457-464, 1997.
- [KRUG .E, 1999] KRUG .E « Injury À leading Cause of the global burden of disease: World Health Organization », violence and injury prevention, Geneva, 1999.
- [Kumar et al, 2012] KUMAR C.N., RAMULU V.S., REDDY K.S., KUTHA S., CUMAR C.M. "Spatial data mining using cluster analysis". IN *International Journal of Computer Science & Information Technology (IJCSIT)* Vol 4, No 4. 2012.

L

Références bibliographiques

- [Lagarde, 2009] Mélanie Lagarde. "La valorisation documentaire par les SIG : Le cas D'EPAMARNE". Mémoire de master II pour obtenir le Titre professionnel : Chef de projet en ingénierie documentaire. Institut National Des Techniques de La Documentation (INTD). 2009.
- [Larose, 2005] Larose, Daniel T. 2005. *Discovering Knowledge in Data: An introduction to data mining*. Hoboken, New Jersey : John Wiley & Sons, Inc., 2005.
- [L. Grösch et al,1987] L. Grösch, H. P. Kaiser, and W. Schmid, "Mathematical movement and load simulation for persons involved in an automobile accident," in SAE, paper 871109, 1987.
- [Lewhe M et Zemmour O, 2009] Lewhe Mahugnon Jorys, Mlle Zemmour Ouarda, « Accidents De La Route Et Identification des Facteurs De Risque », mémoire de fin d'étude en vue de l'obtention du grade de docteur en médecine, pp 11, 2009.
- [L. Evans, 1991] L. Evans, "Traffic safety and the driver", Van Nostrand Reinhold, 1991.

M

- [Mamadou, 2010] O.Mamadou. « Fouille de données : vers une nouvelle approche intégrant de façon cohérente et transparente la composante spatiale ». la Faculté des études supérieures de l'Université Laval dans le cadre du programme de Maîtrise en sciences géomatiques pour l'obtention du grade de maître science. QUÉBEC. 2010.
- [Matteo, 2008] Matteo, Matteucci. 2008. A Tutorial on Clustering Algorithms K-Means Clustering. [Online] 2008. http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html.
- [M. Keal, 1995] M. Keal, "Pedestrian exposure to risk of road accident in New Zealand," *Accid. Anal. and Prev.*, vol. 27, no. 5, pp. 729-740, 1995.

Références bibliographiques

- [M.Orlando, 2000] Metroplan Orlando, "A study of 617 motorist-versus-pedestrian crashes on state roads 50 & 436 and US Highways 17/92, 192 and 441 during the years 1993 through 1997," Orlando Urban Area Arterial Pedestrian Crash Study, Tech. Rep. 2000.

N

- [N. Metropolis et S. Ulam, 1949] N. Metropolis and S. Ulam, "The Monte Carlo method," *Journal of the American Statistical Association*, vol. (247), pp. 335–341, 1949.
- [N. Takubo et K. Mizuno, 2000] N. Takubo and K. Mizuno, "Accident analysis of sports utility vehicles: human factors from statistical analysis and case studies," *JSAE review*, vol. 21, pp. 103–108, 2000.

O

- [Ouattara, 2010] Ouattara M. "Fouille de données : vers une nouvelle approche intégrant de façon cohérente et transparente la composante spatiale". Québec : Faculté des études supérieures de l'Université Laval. 208 p. 2010.

P

- [Pascal, 2007] Pascal Poncelet. "Extraction de motifs : Règles d'association et motifs séquentiels". LIRMM Pascal. Poncelet@lirmm.fr. <http://www.lirmm.fr/~poncelet>. 2007.
- [Pisati, 2012] PISATI M. Spatial data analysis in stata: An overview. Support de cours. 2012.

Q

- [Quinlan, 86] QUINLAN J.R., "Induction of Decision Trees." *Machine Learning*, v.1 p.81-106, 1986.

R

- [ROQUES et al, 2004] Pascal Roques and Franck Vallée, UML 2 en action, de l'analyse des besoins à la conception J2EE, 3ème édition.

Références bibliographiques

EYROLLES, 2004.

[ROQUES, 2008] Pascal Roques, UML 2 Modéliser une application Web, 4ème édition. EYROLLES, 2008.

[Roussel, 2012] Guillaume Roussel. "Développement et évaluation de nouvelles méthodes de classification spatiales spectrale d'image hyper spectrales ". Université de Toulouse. page 14- 20. Le10 juillet 2012.

S

[Schneuwly et al., 2010] SCHNEUWLY D, CALOZ R. "Les concepts spatiaux fondamentaux". Geographic Information Technology Training Alliance GITTA. 2010.

T

[Talbi, 00] E-G Talbi. « Fouille de données (Data Mining) - Un tour d'horizon ». Présentation enligne. Laboratoire d'informatique fondamentale de Lille (LIFL).

[T. Brénac and D. Fleury, 1999] T. Brénac and D. Fleury, "Le concept de scénario type d'accident de la circulation et ses applications," RTS, no. 63, 1999.

[T. Brénac, 1997] T. Brénac, "L'analyse séquentielle de l'accident de la route (méthode INRETS), ser". Outils et méthodes. INRETS, no. 3, 1997.

[T.Brénac et B. Megherbi, 1996] T. Brénac and B. Megherbi, "Diagnostic de sécurité routière sur une ville : intérêt de l'analyse fine de procédures d'accidents tirées aléatoirement," RTS, no. 52, pp. 59-71, 1996.

[T. Brénac et al, 1996] T. Brénac, J. Delcamp, S. Pelat, and G. Teisseire, "Scénarios types d'accidents de la circulation dans le département des bouches-du-rhône," INRETS, Tech. Rep. MA 9611-2, 1996.

[Tobler, 1970] Tobler W. R. "A computer movie simulating urban growth in the Detroit region," Economic geography, pp. 234-240. 1970.

Références bibliographiques

[T.Perron, 1997] T.Perron, « méthode d'analyse de sécurité primaire automobile pour la spécification fonctionnelle et l'évaluation prévisionnelle d'efficacité de systèmes d'évitement d'accidents », Thèse de Doctorat, Laboratoire Génie Industriel, Ecole Centrale Paris, 1997.

[T, Perron, 2001] Perron, Thierry. « Spécification de systèmes automobiles de sécurité active ». Les Cahiers du numérique 2, no 1 (1 mars 2001) : 133-55

X

[Xu et al, 1998] Xu, Xiaowei, Martin Ester, Hans-Peter Kriegel and Jörg Sander, A Distribution-Based Clustering Algorithm for Mining in Large Spatial Databases. *Proceedings of the 14th International Conference on Data Engineering (ICDE 98)*. 1998.

Y

[Yubin, 2004] Yubin Y., "Spatial Data Mining", Joint Laboratory for Geo-information Science The Chinese University of Hong Kong. 2004.

Z

[Zeitouni, 2006] Zeitouni K., "Analyse et extraction de connaissances des bases de données spatiotemporelles", mémoire, Université de Versailles Saint-Quentin-en-Yvelines. Décembre 2006.

[Zeitouni et al., 2001] Zeitouni K., Yeh L., and Aufaure M.-A., "Join indices as a tool for spatial data mining," in *Temporal, Spatial, and Spatio-Temporal Data Mining*, ed: Springer, pp. 105-116. 2001.

[Zhang et al., 1996] ZHANG T, RAMAKRISHNAN R, LIVNY M. Birch: an efficient data clustering method for very large databases. *SIGMOD '96 Proceedings of the 1996 ACM SIGMOD international conference on Management of data*, pp. 103-114, 1996.

Références bibliographiques

- [Zhao et Karypis, 2002] Zhao Y. and Karypis G.. "Evaluation of hierarchical clustering algorithms for document datasets ".In CIKM. 2002.

W

- [W.Ben Ahmed, 2005] Walid Ben Ahmed, « Application à la construction et à l'interprétation de scénario d'accident de la route », Laboratoire 2005, génie industriel.
- [W.H. Heinrich, 1931] 1931: Industrial Accident Prevention (New York, McGraw-Hill).
- [WEB 1] Sécurité routière : un problème de santé publique, 29 mars 2004. Disponible sur le site : <
http://www.who.int/features/2004/road_safety/fr/ >
- [WEB 2] "Données géo-spatiales", 2011. Les bibliothèques/UdeM. Bibliothèque Université de Montréal [en ligne]. Disponible sur le site < <http://guides.bib.umontreal.ca/disciplines/147-Donnees-geospatiales> >.
- [WEB 3] <http://landtrustgis.org/images/GIS%20Technology/GIS%20data%20basics/5longitudelatitude.png/view>.
Site web: Land Trust GIS.
- [WEB 4] "Le géomarketing", 2013, Asterop. Disponible sur le site : <<http://www.asterop.com/fr/products/geomarketing.aspx>>.
- [WEB 5] http://www.insee.fr/fr/themes/document.asp?reg_id=0&id=924.
- [WEB 6] Réseaux - Architecture client/serveur à 3 niveaux,
<http://www.commentcamarche.net/contents/221-reseaux-architecture-client-serveur-a-3-niveaux>