

N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE  
SCIENTIFIQUE



UNIVERSITE DJILLALI LIABES  
FACULTE DES SCIENCES EXACTES  
SIDI BEL ABBES

# THESE DE DOCTORAT

*Présentée par : LARBI Abdelmadjid*

*Spécialité : Informatique*

*Option : Système d'Information et de connaissances*

*Intitulée*

**Conception des entrepôts de données :**

**Evaluation des besoins flexibles**

*Soutenue le 01/07/2018*

*Devant le jury composé de :*

*Président : SLAMA Zohra, Maitre de conference A ; UDL-Sidi Bel-Abbes.*

*Examineurs :*

*BENSLIMANE Sidi Mohamed, Professeur, ESI-Sidi Bel-Abbes.*

*AMAR BENSABER Djamel, Maitre de conference A, ESI-Sidi Bel-Abbes.*

*TOUMOUIH Adil, Maitre de conference A, UDL-Sidi Bel-Abbes*

*BERRABAH Djamel, Maitre de conference A, UDL-Sidi Bel-Abbes*

*Directeur de thèse : MALKI Mimoun, Professeur, ESI-Sidi Bel-Abbes.*

***Année universitaire 2017 / 2018***



# DEDICACE



*A ma Très chère maman et  
A mon cher Papa Hadj Abdelkrim*

*A ma chère femme*

*A ma fille et A mes enfants*

*Vous occupez une place particulière dans mon cœur.  
Je vous dédie ce travail en vous souhaitant un avenir radieux,  
plein de bonheur et de succès.*

*A mes très chères sœurs et frères*

*Je vous souhaite une vie pleine de bonheur et de prospérité*

*A mes amis, collègues et frères*

*En souvenir de nos éclats de rire et des bons moments.  
En souvenir de tout ce qu'on a vécu ensemble.*

*J'espère de tout mon cœur que  
Notre amitié durera éternellement.*

*Abdesmadjid*



# Remerciements



*Je remercie ALLAH de m'avoir aidé à réaliser ce travail.*

*Un très grand remerciement à mon directeur de thèse Pr. Malki Mimoun qui a dirigé cette thèse d'une main de maître. Tout au long de ces années, il a su orienter mes recherches aux bons moments, toujours dans les bonnes directions. Malgré l'éloignement, il a toujours été disponible pour prodiguer des conseils et des orientations ô combien pertinentes. Pour tout cela, pour m'avoir offert la chance d'en être là aujourd'hui, pour sa confiance et pour sa précieuse aide technique, aussi pour sa patience avec moi surtout dans les moments difficile de ce long chemin de préparation de thèse. Pour tous cela, je le remercie du fond du cœur.*

*Mes plus vifs remerciements vont également au Pr. Boukhalifa Kamel qui tout au long de ces années de thèse, malgré l'éloignement, a toujours été là par ses conseils, par son suivi minutieux de toutes mes propositions dans le cadre de cette thèse, par ses corrections, par ses orientations et suggestions, par de riches et longues discussions.*

*Je remercie Mme SLAMA Zohra, Maître de Conférences Classe A à l'ESI-Sidi Bel-Abbes, pour m'avoir fait l'honneur d'accepter d'examiner mes travaux de recherche et de présider le jury de ma soutenance de thèse.*

*Je remercie Monsieur BENSLIMANE Sidi Mohamed, Professeur à l'ESI-Sidi Bel-Abbes, pour m'avoir fait l'honneur d'accepter d'examiner mes travaux de thèse. Je remercie Monsieur AMAR BENSABER Djamel, Maître de Conférences Classe A à l'ESI-Sidi Bel-Abbes, pour m'avoir fait l'honneur d'accepter d'examiner mes travaux de recherche et d'être membre du jury de ma soutenance de thèse. Je remercie Monsieur BERRABAH Djamel, Maître de Conférences Classe A à l'UDL-Sidi Bel-Abbes pour m'avoir fait l'honneur d'accepter d'examiner mes travaux de recherche et d'être membre du jury de ma soutenance de thèse. Je remercie aussi Monsieur TOUMOUH Adil Maître de Conférences Classe A à l'UDL-Sidi Bel-Abbes pour m'avoir fait l'honneur d'accepter d'examiner mes travaux de recherche.*

*Je tiens à remercier du fond du cœur Pr Elhirèche A., Pr. Bellatrèche L., Dr. Djoudi M., Pr. A. Abdelmalek, Pr. Elberrichi Z., Pr. Feraoun et Dr. Bouachria A., pour les encouragements et les orientations.*

*Je tiens à remercier tous les amis du Laboratoire EEDIS de la faculté des sciences exactes de l'université Djillali Lyabes Sidi de Bel Abbes, je remercie également tous les collègues du laboratoire ENERGARID de l'université Tahri Mohamed de Béchar à leur tête Pr. Tamali pour le soutien, les encouragements et les orientations, Pr Draoui et surtout les membres de l'équipe SimulIA à leur tête Pr. Sellam et tous les collègues du département Mathématiques et Informatique.*

*J'ai sûrement oublié de remercier beaucoup d'autres personnes méritantes, des personnes qui m'ont offert leur amitié, qui m'ont ouvert leur cœur, qui m'ont ouvert leur porte, qu'elles trouvent ici l'expression de ma profonde gratitude et de mon amitié la plus sincère.*



# Résumé

La conception d'un entrepôt de données est très différente de celle d'une base de données. En plus, les besoins des utilisateurs de l'entrepôt ne sont aussi clairs que ceux des utilisateurs des systèmes OLTP. La réécriture des requêtes a attiré l'attention de nombreux chercheurs car elle est en relation avec plusieurs problèmes de gestion de données : optimisation de requêtes, intégration des données, conception des entrepôts de données.

Pour répondre aux problèmes d'évaluation et d'optimisation des requêtes en vue de spécifier les besoins des décideurs (requêtes classiques) surtout en cas où ces besoins sont vagues ou comprennent des parties imprécises ou indéterminées (requêtes flexibles), on peut faire appel à la logique floue ou aux systèmes experts ou même les ontologies pour pouvoir lever l'indétermination.

Les imprécisions des données sources (Base de production) des systèmes décisionnels peuvent conduire à des analyses OLAP erronées. Donc, la qualité des données extraites (erreurs, imprécision, etc.) et l'évaluation précise des besoins du décideur (Lever l'ambiguïté, l'incertitude et l'imprécision) avant de choisir la base de production est devenue dans ces dernières années un axe important de recherche en décisionnel en général et dans la conception ascendante du SAD en particulier. Les imperfections des données en systèmes décisionnels sont dues à de nombreuses causes comme la difficulté d'avoir des modèles proches de la réalité ou des limitations des instruments et des processus d'acquisition. De plus, une représentation inadaptée de l'information recueillie peut être aussi une source de dégradation de la qualité des données, la représentation devenant trop éloignée de l'objet du monde réel que l'on souhaite modéliser. L'imprécision peut être due à l'ambiguïté d'expression des besoins du décideur. En partant d'une mauvaise interprétation des besoins, toutes les conceptions du DW et les analyses OLAP qui s'en suivent seront erronées. Par conséquent, de nombreux travaux se sont intéressés à la prise en compte des données vagues dans les Systèmes d'Information mais peu de travaux qui ont étudié ce cas dans les entrepôts de données. Ce travail de thèse vient, avec d'autres travaux, combler le vide dans ce contexte et initier une réflexion autour de cette problématique, à travers quelques contributions permettant l'évaluation de l'imprécision des besoins décisionnels à base de requêtes et menant à l'amélioration de la qualité de conception de l'entrepôt de données et la pertinence de l'analyse OLAP par la suite.

**Mots clés :** Imprécision, besoin décisionnel, data warehouse, design, système expert, ontologie floue,

# Abstract

The data warehouse design is very different from database one. In addition, the needs of warehouse users are only as clear as those of OLTP users. Rewriting Requête has attracted the attention of many researchers because it is related to several data management issues: query optimization, data integration, data warehouse design.

To answer the evaluation and optimization problems of requests in order to specify the needs of the decision-makers (traditional requests) especially in cases where these needs are vague or include inaccurate or indeterminate parts (flexible requests). Fuzzy logic or expert systems or even ontologies can be used to remove indeterminacy.

The data source inaccuracies (production base) of decision-making systems can lead to erroneous OLAP analysis. Therefore, the extracted data quality (errors, inaccuracy, etc.) and the precise assessment of the decision-maker needs (Unambiguity, uncertainty and inaccuracy) before choosing the production base has become in the latter years a major focus of decision-making research in general and in the decisional system design in particular. The data imperfections in decision-making systems are due to many causes such as the difficulty of having models close to reality or instruments limitations and acquisition processes. In addition, an inadequate representation of the collected information can also be a source of data quality degradation, the representation becoming too remote from the real world object that we wish to model. The inaccuracy may be due to the needs expression ambiguity of the decision maker. Starting from a misinterpretation of requirements, all DW designs and subsequent OLAP analyzes will be erroneous. As a result, many studies have focused on taking into account vague data in Information Systems but few studies have studied this case in data warehouses. This thesis work comes, along with other works, to fill the void in this context and to initiate a reflection around this problematic, through some contributions allowing the vagueness evaluation of the decisional needs based on requests and leading to the improvement of the data warehouse design quality and the relevance of the OLAP analysis thereafter.

**Keywords:** Imprecision, decisional requirement, data warehouse, conception, expert system, fuzzy ontology.

# ملخص

تصميم مستودع البيانات يختلف كثيرا عن تصميم قاعدة البيانات. وبالإضافة إلى ذلك، فإن احتياجات مستعملي المستودعات ليست واضحة إلا مقارنة بحاجيات مستخدمي OLTP. وقد جذبت إعادة كتابة الاستفسارات انتباه العديد من الباحثين لأنه يرتبط بعدد قضايا إدارة البيانات: الاستعلام الأمثل، تكامل البيانات، وتصميم مستودع البيانات.

للإجابة عن مسائل التقييم والاستفادة المثلى من الطلبات من أجل تحديد احتياجات صانعي القرار (الطلبات الكلاسيكية) خاصة في الحالات التي تكون فيها هذه الاحتياجات غامضة أو تتضمن أجزاء غير دقيقة أو غير محددة (طلبات مرنة). ويمكن استخدام منطق غامض أو نظم الخبراء أو حتى أو الأنتولوجيا لإزالة عدم التحديد.

ويمكن أن يؤدي عدم دقة بيانات المصدر (قاعدة الإنتاج) في نظم صنع القرار إلى تحليل خاطئ للتحليل OLAP. ولذلك، فإن جودة البيانات المستخرجة (أخطاء، عدم دقة، الخ) والتقييم الدقيق لاحتياجات صانع القرار (إزالة الغموض وعدم اليقين وعدم الدقة) قبل اختيار قاعدة الإنتاج قد أصبح في السنوات الأخيرة من التركيز الرئيسي لبحوث صنع القرار بشكل عام وفي التصميم من أسفل إلى أعلى على وجه الخصوص. وتعزى عيوب البيانات في نظم صنع القرار إلى أسباب كثيرة مثل صعوبة وجود نماذج قريبة من الواقع أو قيود الأدوات وعمليات الاقتناء. وبالإضافة إلى ذلك، فإن التمثيل غير الكافي للمعلومات التي تم جمعها يمكن أن يكون مصدرا لتدهور نوعية البيانات، بحيث يصبح التمثيل بعيد المنال عن كائن العالم الحقيقي الذي نود أن نمذجه. وقد يرجع عدم الدقة إلى الغموض في التعبير عن احتياجات صانع القرار. بدءا من سوء تفسير المتطلبات، فإن جميع التصاميم والتحليلات اللاحقة OLAP تكون خاطئة. ونتيجة لذلك، ركزت العديد من الدراسات على مراعاة البيانات المبهمة في نظم المعلومات ولكن القليل من الدراسات درس هذه الحالة في مستودعات البيانات. ويأتي هذا العمل "أطروحة"، جنبا إلى جنب مع غيرها من الأعمال، لملء الفراغ في هذا السياق، والشروع في التفكير حول هذه المشكلة، من خلال بعض المساهمات مما يسمح بتقييم غموض احتياجات القرار على أساس الطلبات ويؤدي إلى تحسين نوعية تصميم مستودع البيانات وأهمية تحليل OLAP بعد ذلك.

الكلمات المفاتيح: عدم الدقة، حاجة صاحب القرار، مستودع البيانات، التصميم، نظام الخبراء، الأنتولوجيا الغامضة.



# Table des matières

## **CHAPITRE I**

### Introduction générale

1.1 Introduction	01
1.2 Contexte & Problématique	01
1.3 Objectifs	05
1.4 Méthodologie suivie	05
1.5 Contributions	06
1.6 Organisation du manuscrit	07

## **CHAPITRE II**

### Conception d'entrepôts de données & Modélisation des besoins décisionnels

2.1 Conception d'entrepôts de données	09
2.1.1 Concepts de base	09
2.1.2 Architecture d'un SID	12
2.1.3 Méthodologie de développement de DW	15
2.1.4 Modélisation multidimensionnelle	15
2.2 Modélisation des besoins décisionnels	16
2.2.1 Ingénierie des besoins décisionnels	16
2.2.2 Ingénierie des besoins & Ingénierie des SIDs	18
2.2.3 Classification des besoins	20
2.2.4 Langages de modélisation des besoins	21
2.2.5 Cycle de vie de l'ingénierie des besoins	22
2.2.6 Conclusion	23
2.3 Conclusion du chapitre	23

# CHAPITRE III

## EVALUATION DES BESOINS DECISIONNELS

3.1 Evaluation des BDFs utilisant une solution structurelle	25
3.1.1 Modèle multidimensionnel	25
3.1.2 Modélisation d'un entrepôt de données	25
3.1.3 Métadonnée et méta modèle d'entrepôts de données	27
3.1.3 Conclusion	27
3.2 Evaluation des BDFs utilisant un système de connaissance	28
3.2.1 Evaluation des BDFs utilisant un système expert	28
3.2.1.1 Logique classique & logique floue	28
3.2.1.2 Extension de la logique floue (CWP) .	33
3.2.1.3 Evaluation de requêtes flexibles à base de système expert	34
3.2.1.4 STDS : Schéma de translation dirigée par la syntaxe	37
3.2.1.5 Conclusion	37
3.2.2 Evaluation des BDFs utilisant une ontologie floue	38
3.2.2.1 Définition d'Ontologie	38
3.2.2.2 Composantes de l'ontologie	39
3.2.2.3 Cycle de vie d'une ontologie	41
3.2.2.4 Les langages de description des ontologies	42
3.2.2.5 Ontologies floues	47
3.2.2.6 Conclusion (Evaluation de BDFs à base d'ontologie)	52
3.3 Conclusion du chapitre	52

# CHAPITRE IV

## ETAT D'ART

4.1 Evaluation des BDFs : Imprécision dans les besoins décisionnels .....	54
4.1.2 Approches ascendantes .....	54
4.1.3 Approches descendantes .....	54
4.1.4 Approches mixtes .....	55
4.1.5 Récapitulatif .....	55
4.1.6 Conclusion .....	56
4.2 Revue de littérature : Evaluation des besoins décisionnels flexibles & Solution structurelle	58
4.2.1 Analyse des besoins décisionnels et traitement de l'imprécision .....	59
4.2.2 Imprécision des données dans l'entrepôt de données .....	59
4.2.3 Revue de littérature .....	59
4.2.4 Synthèse .....	59
4.2.5 Solution structurelle (modèle FASEL) .....	63
4.2.6 Conclusion .....	63
4.3 Revue de la littérature : Evaluation des besoins décisionnels flexibles & Système expert	67
4.3.1 Outils décisionnels utilisant les approches flous .....	68
4.3.2 Systèmes experts, gestion des connaissances, prise de décision et ingénierie des besoins	69
4.3.3 Revue de littérature .....	69
4.3.4 Requêtage d'une base de données flexible .....	70
4.3.5 Etude Comparative .....	71
4.3.6 Conclusion .....	74
4.4 Revue de la littérature : Evaluation des besoins décisionnels flexibles & Ontologie	75
4.4.1 Conception de DW basée sur les ontologies .....	76
4.4.2 Récapitulatif .....	78
4.4.3 Conclusion .....	79
4.5 Conclusion du chapitre .....	79

# CHAPITRE V

## Evaluation de besoins décisionnels flexibles : CONTRIBUTIONS & APPLICATION

5.1 Contributions et approches proposées .....	80
5.2 Approche 1 : Evaluation des BDFs utilisant le changement de structure .....	82
5.2.1 Présentation de l'approche 1 (Solution Structurale) .....	82
5.2.2 Exemple illustratif : Evaluation de risque dans le transport routier .....	86
5.2.3 Expérimentation .....	88
5.2.4 Conclusion .....	91
5.3 Approche 2 : Evaluation des BDFs utilisant un système expert .....	92
5.3.1 Présentation de l'approche 2 (Solution à base de système expert) .....	92
5.3.2 Exemple illustratif : Evaluation de risque de gestion intelligente de projet .....	96
5.3.3 Expérimentation .....	98
5.3.4 Conclusion .....	100
5.4 Approche 3 : Evaluation des BDFs utilisant une ontologie floue .....	101
5.4.1 Présentation de l'approche 3 (Solution Ontologique) .....	101
5.4.2 Exemple illustratif : Evaluation de risque dans le transport routier .....	110
5.4.3 Expérimentation .....	111
5.4.4 Conclusion .....	116
5.5 Conclusion du chapitre : Récapitulatif des trois solutions (Structure, Système expert, Ontologie)	116

# CHAPITRE VI

## CONCLUSION GENERALE

6.3 Conclusion générale .....	117
6.2 Perspectives .....	120

Références bibliographiques .....	119
Annexe .....	131
Annexe A .....	131
Annexe B .....	135
Annexe C .....	140
Annexe D .....	146

# Liste des Figures

<b>Chapitre II</b>	
Figure 2.1 Les trois grands types de systèmes d'information	10
Figure 2.2 Métriques d'aide à la décision	11
Figure 2.3: Architecture d'un système décisionnel	12
Figure 2.4 Les différentes composantes du décisionnel	13
Figure 2.5 Evolution des bases de données décisionnelles	14
Figure 2.6 Processus d'ingénierie des besoins.	19
Figure 2.7 Processus de collecte et de formalisation des besoins utilisateurs	20
Figure 2.8 Hétérogénéité des vocabulaires et des langages de modélisation des besoins	22
<b>Chapitre III</b>	
Figure 3.1 Exemple d'un schéma en étoile	28
Figure 3.2 Exemple d'un schéma en flocon de neige	28
Figure 3.3 Exemple d'un schéma en constellation	29
Figure 3.4 Deux sous-ensembles selon la logique classique	30
Figure 3.5 Deux sous-ensembles selon la logique floue	30
Figure 3.6 Variables linguistiques	31
Figure 3.7 Fonction d'appartenance	32
Figure 3.8 Formes de fonctions d'appartenance	33
Figure 3.9 Valeurs d'appartenance	33
Figure 3.10 structure d'un système expert	37
Figure 3.11 Cycle de vie d'une ontologie	45
Figure 3.12 Exemple d'une ontologie floue avec des concepts flous	49
Figure 3.14 Composants de l'ontologie floue	51
Figure 3.15 Le Modèle en oignon	51
<b>Chapitre IV</b>	
Figure 4.1 Nouveau schéma d'ED flou incluant une hiérarchie floue	59
Figure 4.2 Schéma conceptuel IFDW	60
Figure 4.3 Schéma logique IFDW	61
Figure 4.4 Modèle conceptuel multidimensionnel pour Fait Events	61
Figure 4.5 Correspondance entre le DW et schéma intégré	62
Figure 4.6 Les tables ajoutées FCT & FMT	62
Figure 4.7 Propriétés des tables FCT & FMT	62
Figure 4.8 Modèle FDW	64
Figure 4.9 Membership Function of Customer_age	66
Figure 4.10 Les tables FCT et FMT additionnelles dans une solution	66
Figure 4.10a Les tables FCT et FMT additionnelles avec nombre important de TA	67
Figure 4.12 Les caractéristiques DSS	68
Figure 4.13 Architecture d'analyseur programme statistique (SPA)	72
Figure 4.14 Schéma général de l'Application (2007)	72
Figure 4.15 Architecture de base du prototype développé	73
Figure 4.16 Architecture générale de l'évaluation de requêtes SQL Floues (bypass)	74

<b>Chapitre V</b>	
Figure 5.1 Démarche générale de notre approche	81
Figure 5.2 Métabase & Entrepôts de données	83
Figure 5.3 Solution proposée	84
Figure 5.4 Tables additives	84
Figure 5.5 Entrepôts de données à base d'Ontologie	85
Figure 5.6 Nouvelle structure issue de notre solution	85
Figure 5.7 Classes de DW existante	87
Figure 5.8 Règles de risque appliqué pour le cas de BRType=1	90
Figure 5.9 Tables FCT et FMT ajoutées pour l'attribut cible TA=BRType	90
Figure 5.10 Classe d'équivalence	91
Figure 5.11 Evaluation de risque de transport routier (Cas d'accident)	93
Figure 5.12 Architecture de notre interface – Niveau d'analyse de requête SQL Floue	93
Figure 5.13 Une partie de la grammaire utilisée	94
Figure 5.14 Evaluation de requête floue utilisant notre framework Seuil=0.7	98
Figure 5.15 Evaluation de requête floue utilisant notre framework Seuil=0.3	99
Figure 5.16 Influence du seuil sur la réponse de requête floue	99
Figure 5.17 Approche ontologique proposée GLMR	102
Figure 5.18 Architecture de notre système	103
Figure 5.19 Le Schéma global de notre Ontologie GLMR	105
Figure 5.20 Marge cible initial de l'opérateur NEAR Négatif	109
Figure 5.21 Algorithme Operateur Flou NEAR-	110
Figure 5.22 Interface Risque-Ontologie	110
Figure 5.23 Schéma de spécification de la requête floue	111
Figure 5.24 Application Fuzzy Ontology	112
Figure 5.25 Solution GLMR pour exemples de Bus Risk Type	115
Figure 5.26 Résultat expérimental	115

## Liste des tableaux

<b>Chapitre II</b>	
Tableau 2.1 Différences entre les SIO et les SID	11
<b>Chapitre III</b>	
Tableau 3.1 Opérations sur les ensembles flous	34
Tableau 3.2 quelques règles du processus de conduite	34
<b>Chapitre IV</b>	
Tableau 4.1 Synthèse des approches	57
Tableau 4.2 Etude Comparative	63
Tableau 4.3 Représentation des prédicats flous	73
Tableau 4.4 Comparaison des solutions proposées	74
Tableau 4.5 Approches basées sur les ontologies	78
<b>Chapitre V</b>	
Tableau 5.1 Comparaison entre le modèle FASEL et notre proposition	82
Tableau 5.2 Statistiques des accidents en Algérie	86
Tableau 5.3 Fonction (Count BRTYPE)	91
Tableau 5.4 Influence du seuil sur la réponse de requête floue	99
Tableau 5.5 Comparaison entre GLMP & GLMR	104
Tableau 5.6 Comparaison entre les concepts de GLMP & GLMR	107
Tableau 5.7 CPRI Calcul des bus de l'entreprise pour la classe de bus Inacceptable	114
Tableau 5.8 Calcul du CPRI pour les bus de la compagnie - classe «unwanted bus»	114
Tableau 5.9 Résultat expérimental	114
Tableau 5.10 Issue des résultats de l'expérimentation	115

## Liste des Algorithmes

Algorithme	Page
Algorithme 5.1 Algorithme Principal	95
Algorithme 5.2 Algorithme SELECTVALIDE	95
Algorithme 5.3 Algorithme INSERTVALIDE	96
Algorithme 5.4 Algorithme DELETEVALIDE	96
Algorithme 5.5 Algorithme UPDATEVALIDE	96

## **LISTE DES ACRONYMES**

<b>BDF</b>	<b>Besoins Décisionnels Flous</b>
<b>DW</b>	<b>Datawarehouse</b>
<b>SID</b>	<b>Système d'information décisionnel</b>
<b>SIO</b>	<b>Système d'information Opérationnel</b>
<b>BDF</b>	<b>Besoin décisionnel flou</b>
<b>FCT</b>	<b>Fuzzy Classification Table</b>
<b>FMT</b>	<b>Fuzzy Membership Table</b>
<b>MDA</b>	<b>Membership Degree Attribute</b>
<b>CMA</b>	<b>Class Membership Attribute</b>
<b>FDW</b>	<b>Fuzzy Data warehouse</b>
<b>DSS</b>	<b>Decisional System</b>
<b>OLTP</b>	<b>On-Line Transaction Processing</b>
<b>OLAP</b>	<b>On-Line Analysis Processing</b>
<b>ROLAP</b>	<b>OLAP Relationnel</b>
<b>MOLAP</b>	<b>OLAP Multidimensionnel</b>
<b>BNF</b>	<b>Backus Naur Form</b>
<b>GLMP</b>	<b>Granular Linguistic Model of Perception</b>
<b>GLMR</b>	<b>Granular Linguistic Model of Requirement</b>
<b>CWP</b>	<b>Computational Word Perception</b>
<b>EIS</b>	<b>Executive Information System</b>
<b>ETL</b>	<b>Extract Transform Load</b>
<b>IBE</b>	<b>Ingénierie des besoins et besoins</b>
<b>IB</b>	<b>Ingénierie des besoins</b>
<b>IBD</b>	<b>Ingénierie des besoins décisionnels</b>
<b>IC</b>	<b>Ingénierie des Connaissances</b>
<b>IA</b>	<b>Intelligence Artificielle</b>
<b>SE</b>	<b>Système Expert</b>
<b>SBC</b>	<b>Systèmes à Base de Connaissances</b>
<b>RDFS</b>	<b>Resource Description Framework Schéma</b>
<b>OWL</b>	<b>Ontology Web Langage</b>
<b>MKT</b>	<b>Meta Knowledge Table</b>
<b>SAD</b>	<b>Systèmes d'Aide à la Décision</b>
<b>MCF</b>	<b>Meta Connaissances</b>



# CHAPITRE I

---

---

## Introduction générale

<b>1.1 Introduction .....</b>	<b>01</b>
<b>1.2 Contexte &amp; Problématique .....</b>	<b>01</b>
<b>1.3 Objectifs .....</b>	<b>05</b>
<b>1.4 Méthodologie suivie .....</b>	<b>05</b>
<b>1.5 Contributions .....</b>	<b>06</b>
<b>1.6 Organisation du manuscrit .....</b>	<b>07</b>





# CHAPITRE I

---

## Introduction générale

Hésiter, c'est déjà prendre une décision

*Stanislaw Jerzy Lec*  
Artiste, écrivain, Poète  
(1909 - 1966)

### 1.1 Introduction

Le sujet de cette thèse entre en général dans le cadre de recherche de la qualité de conception de l'entrepôt de données à tous les niveaux du processus décisionnel (conceptuel, logique et physique) et en particulier dans la recherche d'amélioration de la phase d'analyse des besoins décisionnel vu l'impact des résultats de cette phase sur la qualité du processus décisionnel.

En raison de la complexité croissante de Data Warehouse (DW), une attention continue doit être accordée à l'évaluation de sa qualité tout au long de sa conception et de son développement. Une bonne qualité du modèle de besoins peut mener à une bonne qualité du DW. Divers auteurs ont proposé des métriques pour assurer la qualité des modèles conceptuels, logiques et physiques pour le DW. À partir d'une mauvaise interprétation des besoins flous, toutes les conceptions d'entrepôt de données et l'analyse OLAP qui en découleront seront erronées. Pour concevoir un modèle DW, il est nécessaire de prendre en compte non seulement les sources de données, mais aussi les besoins décisionnels.

### 1.2 Contexte et problématique

Le Business Intelligence (BI), ou Informatique Décisionnelle [Gam et al., 2006], se définit comme étant l'ensemble des technologies permettant de traiter, valoriser et présenter les données à des fins de compréhension et de décision. L'intérêt du BI est multiple : visibilité sur une activité, reporting fiable et détaillé, conseil en matière de comportement clients, détection de faiblesses, prise de décisions stratégiques ou tactiques, réactivité face à un événement. Le BI donne aux dirigeants une visibilité sur la performance de leur entreprise afin d'améliorer la capacité de celle-ci à réagir plus rapidement que ses concurrents face à de nouvelles opportunités ou aux risques du marché. Le Business Intelligence s'appuie sur un système d'information spécifique appelé Système d'Information Décisionnel (SID), par opposition aux systèmes d'informations opérationnels (SIO). Cependant, un nombre important d'études a montré que [Boukhari et al., 2014] les échecs dans la mise en œuvre et l'exploitation des projets informatiques sont souvent dus à une mauvaise compréhension des besoins ainsi qu'à des incohérences, et des ambiguïtés de leurs définitions [Golfarelli et al., 2009] et [Lubars et al., 1993]]. Les besoins peuvent également contribuer à l'identification des activités redondantes. Rappelons que plusieurs approches de conception des applications avancées sont censées être centrées utilisateur. Cette approche consiste à considérer les utilisateurs et leurs besoins tout au long du processus de développement d'applications informatiques. Parmi les systèmes concernés par cette approche, nous pouvons citer les bases de données [Rolland et al., 1999],

les entrepôts de données ou approche orientée besoins [Khouri et al., 2011], les systèmes d'intégration des données [Goh et al., 1999], et les interfaces homme-machines [Ait Ameer et al., 2003]. Ainsi, l'ingénierie des besoins et des exigences (IBE) [Amokrane et al., 2016], qui est une discipline commune à toutes les ingénierie métiers (génie logiciel, mécanique, électronique, etc.) comme à l'Ingénierie Système, la phase décisive la plus en amont dans tout projet d'ingénierie d'un système, est primordial à tout projet de développement de système d'information décisionnel. Ainsi, sous-estimer cette importance [Gam et al., 2006] peut mener à l'inachèvement des projets, au dépassement de leurs budgets, ou à la défaillance des systèmes développés ne rendant pas les services attendus par les décideurs et la communauté des usagers du SID. Cependant, quelle que soit la méthode proposée [Amokrane et al., 2016], l'objectif des travaux en IBE est d'améliorer la qualité des besoins en atténuant les défauts, omissions ou erreurs souvent rencontrés dans la phase d'expression des besoins et inhérents à la nature informelle de ces derniers, tels que :

- L'ambiguïté : informations données au niveau des besoins ayant différentes interprétations possibles;
- L'imprécision : besoins flous ou insuffisamment définis, structurés ou détaillés suite à des reformulations par exemple ambiguës;
- L'omission : besoins omis par oubli objectif ou subjectif, menant ici aussi à des incohérences ou des imprécisions;
- La redite : besoins exprimés plus d'une fois ou reformulés en d'autres termes induisant encore des incohérences ou de l'ambiguïté;
- L'incohérence : besoin en contradiction avec un ou plusieurs besoins. Nous pouvons aussi citer le bruit, la non-mesurabilité, la non-intelligibilité ou non traçabilité [Lamsweerde et al., 2009].

Dans cette thèse, on ne fait pas de différence entre besoin et exigence d'où on va utiliser IB au lieu de IBE.

Pour la représentation des besoins [Sabri et al., 2013] dans les étapes de collecte, de formalisation et de traitement, quatre catégories de modèles sont proposées dans la littérature, à savoir :

- Modèles basés sur des schémas relationnels : le modèle E/A [Luján-Mora et Trujillo, 2003], Schéma en étoile [Feki, Ben Abdallah et al. 2006]
- Modèles de requêtes : les besoins des décideurs dans ces modèles sont formulés via des requêtes [Ghozzi, Ravat et al. 2005][Ghozzi, 2010][Ravat, 2007]
- Modèles de tableaux : la collecte des besoins des décideurs dans les modèles de tableaux se fait via des tableaux n-dimensionnels contenant les concepts de faits, de dimensions, de mesures, de paramètres, de hiérarchies et d'attributs (Bargui, Fekiet al. 2009) [Abdelhédi et Zurfluh, 2013]
- Modèles de buts : utilisent un modèle de buts pour représenter les besoins évalués [Mazon, Trujillo et al., 2005] [Giorgini, Rizzi et al., 2005] [Gam El Gholli, 2008].

Nous distinguons également des travaux qui combinent deux ou plusieurs types des modèles afin de collecter et formaliser les besoins. [Prakash et Gosain, 2003] : Requête-Buts

Au cours des dernières années [Boukhari et al., 2014], plusieurs travaux visent à proposer un modèle adapté aux besoins décisionnels s'adressant à des décideurs non informaticiens. Néanmoins, nous constatons le manque de modèles conceptuels permettant de faire abstraction des contraintes techniques de stockage de données et de mieux appréhender les besoins de ces décideurs [Golfarelli et al., 2002].

La représentation de l'imperfection (imprécision ou incertitude) et son exploitation dans des systèmes d'information forment une thématique majeure de l'intelligence artificielle mais aussi un gage de fiabilité des dits systèmes. De fait, la gestion de l'imperfection des données fait partie de la gestion de la qualité. [Zoghلامي et al., 2013].

Si le décideur est non-informaticien et/ou non-expert dans les modèles de conception, ses besoins exprimés vont présenter des imprécisions qui vont entacher la conception. D'où l'importance de l'étude de cette imprécision pour améliorer la qualité de l'expression des besoins décisionnels.

Ce problème a été déjà évoqué dans les travaux de [Ghozzi, 2010] mentionnant que la gestion des données incertaines dans le langage de requête est nécessaire. Aussi [Nebot et al., 2009] ont noté que l'utilisation des ontologies est plus flexible dans le sens qu'elles donnent des définitions des données générées incomplètes, inexactes et implicites et enfin [Golfarelli et al., 2009] avait mentionné que les besoins incertains ou peu claires peuvent être corrigées par l'utilisation des ontologies pour améliorer la conception de DW. Mais ces travaux et d'autres n'ont pas proposé de solutions ni étudié le problème d'une façon profonde. Ainsi, nous avons concentré les travaux de cette thèse sur l'évaluation de **l'imprécision dans l'expression des besoins décisionnels** que l'on propose d'entamer en faisant appel aux outils d'intelligence artificielle et plus précisément au système expert d'abord puis à l'ontologie floue.

La première solution a impliqué d'abord un langage SQL, un traducteur (Syntax-Directed Translation Schemes) et une base de connaissance (**Système expert**). La base des règles est stockée dans une table appelée MKT (Meta Knowledge Table). Contrairement à la majorité des travaux antérieurs dans ce domaine et qui se concentrent sur un projet logiciel (framework) ou une phase de projet comme la phase de conception du projet. Les travaux précédents proposaient des modèles de solutions utilisant soit un langage flou, soit un système expert. Cette première proposition considérée comme une solution naïve nous a permis de remplacer, par la suite, le système expert flou par **une ontologie floue** et de proposer une solution plus rigoureuse. La littérature propose plusieurs approches [Boukhari et al., 2014] et méthodes pour l'expression des besoins des utilisateurs.

**Approches dirigées par les buts** : L'approche orientée but doit sa naissance à Yue [Yue et al., 1987]. Elle s'intéresse à l'utilisation des buts pour le recueil, l'analyse, la validation et la gestion des besoins. "Un but est un objectif que le système considéré doit atteindre". Différentes modèles suivent l'approches orientées buts : GQM [Basili et al., 1992], i\* Framework [Siu-Kwong et al., 1996] et [Siu-Kwong et al., 1997], KAOS [Dardenne et al., 1993], TROPOS [Bresciani et al., 2004].

**Approches à base de scénarii** : Dans les approches à base de scénarii, les besoins sont décrites à l'aide de scénarii visant à la compréhension du système par les participants.

Les événements de scénarii sont des interactions entre l'utilisateur et le système. Un scénario peut être défini comme l'ordre des actions ou des événements pour un cas spécifique d'une certaine tâche générique qu'un système doit accomplir [Rolland et al., 1998]. Les scénarios sont exprimés dans différents langages : informels (langage naturel), semi-formels (diagramme des cas d'utilisation (UML), tableaux, etc.) ou formels (langages basés sur des grammaires régulières, diagrammes d'états, etc.).

**Approches basées sur les automates et réseaux de Pétri** : le modèle de traitements de méthodologie Merise [Rochfeld et al., 1983] qui représente un bon exemple de méthode d'expression de besoins avec un réseau de Petri. Aussi, dans la littérature [Sabri & al., 2012], et [Larbi et al., 2017] quatre groupes de modèles sont disponibles pour représenter les besoins d'un système d'aide à la décision (SAS), à savoir:

- (i) modèles **Entité-relation** [Carmé et al., 2010], [Abai et al., 2013] et [Boukhari et al., 2014]
- (ii) **modèles UML** [Golfarelli et al., 2009], [Prat et al., 2006] et [Golfarelli et al., 2011];
- (iii) modèles à base de **requêtes** [Maté et al., 2012], [Golfarelli et al., 2011] et [Pardillo et al., 2011];
- (iv) modèles **de buts** [Song et al., 2007] et [Hachaichi et al., 2013].

L'analyse des besoins est une étape très importante du processus décisionnel. Les imperfections des données des systèmes décisionnels sont dues à de nombreuses raisons telles que la difficulté d'avoir des modèles réalistes ou des limitations d'instruments et de processus d'acquisition. En outre, une représentation inadéquate des informations collectées peut également être une source de dégradation de la qualité des données; la représentation devient trop éloignée de l'objet du monde réel que nous voulons modéliser. À partir d'une mauvaise interprétation des besoins, toutes les conceptions de l'entrepôt de données (DW) et l'analyse OLAP qui en découle seront fausses. Aussi, le succès et l'échec d'un projet d'entrepôts de données (DW) sont principalement liés à la phase conceptuelle selon la plupart des chercheurs dans ce domaine. Lors de l'analyse des besoins du système décisionnel, de nombreux problèmes récurrents apparaissent et les besoins en matière de modélisation des difficultés sont détectés. Aussi, nous rencontrons le problème associé à l'expression des besoins par les décideurs non-informaticiens et non- experts sur les modèles de conception. L'ambiguïté de l'expression des besoins décisionnels conduit à une mauvaise interprétation des besoins résultant d'une insuffisance de conception d'entrepôts de données et d'analyse OLAP incorrecte. Par conséquent, de nombreuses études ont mis l'accent sur l'inclusion des données vagues dans les systèmes d'information en général [Sequeira & al., 2014], [Musa & al., 2015], [Pivert & al., 2012] [Fasel & al., 2014] mais peu d'études ont examiné ce cas dans les entrepôts de données [Musa & al., 2015] [Fasel & al., 2014]. On va décrire l'une des lacunes des approches actuelles de conception d'entrepôts de données qui est l'étude de l'imprécision et surtout dans l'analyse des besoins et comment les systèmes experts et les ontologies peuvent nous aider à le surmonter.

Nous focalisons notre étude sur l'imprécision dans l'expression des besoins pour cela on présente un état de l'art sur ce sujet montrant que rares sont les travaux qui tiennent

compte de l'imprécision dans l'étude de cette phase cruciale du processus décisionnel. On ajoute un état d'art des approches basées sur l'ontologie vue que ce genre d'approche vise améliorer la qualité de la conception du DW en introduisant l'ontologie et quelques travaux de cette catégorie évoquent le problème de l'imprécision dans les besoins décisionnels. Selon notre connaissance aucune étude rigoureuse de l'imprécision dans ce domaine n'a été faite.

## **1.4 Objectifs**

Afin d'évaluer l'imprécision des besoins décisionnels et pour améliorer la qualité de la conception de l'entrepôt de données et la qualité de l'analyse OLAP, cette thèse vise à proposer des solutions (naïve et rigoureuse) pour évaluer l'imprécision des besoins. Donc, la qualité de la données extraites (erreurs, imprécisions, etc.) et l'évaluation précise des besoins des décideurs (éliminer l'ambiguïté, l'incertitude et le flou) avant de choisir la base de production est devenu, au cours des dernières années, un axe important de la recherche dans le domaine de la prise de décision en général et la conception des systèmes décisionnel (SD) en particulier.

Notre objectif est de proposer des solutions d'évaluations de l'imprécision de l'expression du besoin décisionnel à base de requête, ces méthodes proposées s'appuient principalement sur les techniques d'intelligence artificielle

## **1.5 Méthodologie suivie**

Dans le but de l'évaluation de l'imprécision des besoins décisionnels et en vue d'améliorer la qualité de conception de l'entrepôt de données. Nous avons ainsi commencé par une revue de littérature sur l'imprécision qui nous a montré l'absence d'étude d'imprécision des besoins décisionnels d'une part, la présence d'étude d'imprécision des données des entrepôts et ce sont généralement des solutions structurelles d'autre part. Nous avons constaté aussi que la majorité des solutions traitant l'imprécision, dans d'autres cas, été des solutions ontologiques.

Ainsi, nous avons commencé par la proposition de solution structurelle consistant à ajouter des métabases (dictionnaire de données additive) à la structure de l'entrepôt en vue de permettre l'évaluation de l'imprécision des besoins. Nous avons ensuite entamé l'étude en visant une solution naïve se basant sur un système expert. Vu que la technologie de système expert est de moins en moins utilisée et remplacée par d'autres nouvelles technologies telles que les ontologies et malgré les résultats encourageants de la solution naïve, nous avons décidé de s'orienter vers les solutions ontologiques.

Par la suite, nous avons proposé une solution ontologique et précisément un modèle ontologique GLMR qui nous permet à partir des hypothèses de notre problème d'avoir une classification floue ou une affectation floue du nuage de points existants. Ce qui représente en fait une évaluation du besoin flou. Pour mettre en évidence cette proposition. Les trois solutions étaient appliquées dans l'évaluation de l'imprécision des risques (soit d'accidents dans le transport intelligent, soit de gestion de projet intelligent). Les solutions proposées visent seulement le cas où les besoins décisionnels sont exprimés

sous forme de requêtes. Pour les autres cas, il faut soit adapter ses solutions, soit penser à proposer d'autres solutions.

## 1.6 Contributions

Nous avons commencé par l'étude bibliographique [Larbi & al., 2017] sur l'imprécision dans les besoins décisionnels. Nous avons constaté dans cette étude qu'aucun des travaux ne tient compte de l'imprécision en tant qu'élément d'information de critère de comparaison des travaux dans le domaine de l'analyse de la prise de décision. Nous proposons donc une enquête sur l'imprécision des besoins décisionnels. Nous nous concentrons sur les approches basées sur les besoins et sur les ontologies. Par conséquent, dans la littérature traitant de l'imprécision des systèmes de décision, les études se sont concentrées sur l'imprécision des données et non sur l'imprécision des besoins de décision. L'étude que nous avons menée a montré que l'imprécision n'était pas largement couverte dans les approches d'analyse des besoins, et ce en étudiant les œuvres rares qui ont cité l'imprécision dans leurs travaux et qui n'ont pas proposé de solutions par la suite.

1. L'étude bibliographique sur les solutions structurelles traitant l'imprécision nous a permis de s'orienter vers la solution de Fasel quoique sa solution traite l'imprécision dans les données du Datawarehouse et non pas dans les besoins décisionnels. Ainsi, notre première contribution [Larbi & al., 2018] représente une solution structurelle inspirée de la solution de Fasel. Cette solution proposée consiste à ajouter un méta schéma (méta structure) à l'entrepôt de données permettant d'évaluer les besoins décisionnels flous et préservant la performance du support.

2. Notre deuxième contribution [Larbi & al., 2013] consiste en un modèle (un système expert flou) permettant d'évaluer les requêtes flexibles et impliquant un langage SQL, un traducteur (Syntax Directed Translation Schemes) et une base de connaissance (système expert flou) :

Cette contribution vise à implémenter une interface basée sur la dérivation de requêtes floues en requêtes booléennes à l'aide d'un schéma de traduction guidée par la syntaxe (STDS) qui nous permet de mieux interpréter une requête floue et en utilisant un système expert flou. Afin de mettre en œuvre cette proposition, nous l'appliquerons au cas de l'évaluation de risque dans un projet intelligent.

3. Notre troisième contribution [Larbi & al., 2018 ] est la continuité de la solution naïve à base de système expert. Elle consiste à proposer une solution permettant l'évaluation de l'imprécision dans la prise de décision à base d'ontologie floue:

L'imprécision dans la prise de décision peut être due à l'ambiguïté de l'expression des besoins décisionnels. Afin d'évaluer l'expression dans les besoins décisionnels et afin d'améliorer la qualité de conception des entrepôts de données, cette contribution présente une solution rigoureuse basée sur l'ontologie floue.

4. La quatrième contribution (présentée en Annexe) consiste à proposer une solution semi-automatique afin de mettre en évidence l'utilité d'introduire la sémantique dans

la base de données XML et surtout pour l'enrichissement des réponses aux besoins. Nous l'avons appliquée sur les bases de données pharmaceutiques, et les résultats préliminaires sont très encourageants. Nous avons montré à travers cette solution l'utilité d'utilisation de l'ontologie pour améliorer la réponse d'une requête en utilisant le lien de similarité entre les médicaments. Elle a représenté un facteur de motivation pour entamer la solution ontologique proposée pour l'évaluation de l'imprécision des besoins décisionnels.

## 1.7 Organisation du manuscrit

Ce mémoire de thèse est organisé de six chapitres. Le premier chapitre présente une introduction générale sur le sujet de thèse. Le deuxième chapitre concerne la conception d'entrepôts de données & Modélisation des besoins décisionnels. Le troisième chapitre concerne un background sur la modélisation des besoins décisionnels en général et sur l'évaluation des besoins décisionnels flous en particulier. Le quatrième chapitre concernant l'état d'art sur l'évaluation des besoins décisionnels flous à base de connaissance ou en utilisant une solution structurelle. Le cinquième chapitre présente l'évaluation des besoins décisionnels flexibles avec les approches proposées, les applications et enfin une synthèse. Le chapitre six présente une conclusion générale.

Après la présentation de la problématique, le contexte de l'étude et les objectifs des travaux ainsi que la présentation de différentes contributions qui font l'objet du **premier chapitre**, le **chapitre 2** présente l'axe de recherche, il s'agit de la conception de data warehouse en général, et spécifiquement la modélisation des besoins décisionnels et plus précisément, cette thèse étudie l'imprécision dans les besoins décisionnels et son impact sur le design de l'entrepôt de données d'une part et sur les opérations OLAP et la visualisation des Cubes OLAP d'autre part.

Le **chapitre 3** présente l'évaluation de besoins décisionnels flexibles utilisant une solution structurelle, ou un système expert (base de connaissance) puis utilisant une ontologie. On présente ensuite en **chapitre 4** une revue de littérature sur ce thème, décrivant les travaux existants sur ce sujet.

Le **chapitre 5** présente les différentes contributions : l'évaluation des BDFs utilisant le changement de structure, utilisant une base de connaissance (Système expert, une ontologie).

Le **chapitre 6** présente la conclusion générale de ce mémoire, un récapitulatif et une évaluation de nos travaux suivie d'une série de perspectives. Il montre le positionnement du sujet de thèse qui se situe au carrefour regroupant l'informatique décisionnelle, l'ingénierie des besoins, la logique floue et l'ingénierie des connaissances visant à évaluer le besoin imprécis dans le but d'assurer la qualité de conception du data warehouse par la suite en particulier et la qualité du processus décisionnel en général.





Savoir pour prévoir, afin de pouvoir.  
Auguste Comte

# CHAPITRE II

## Conception d'entrepôts de données & Modélisation des besoins décisionnels

<b>2.1 Conception d'entrepôts de données .....</b>	<b>09</b>
<b>2.1.1 Concepts de base .....</b>	<b>09</b>
<b>2.1.2 Architecture d'un SID .....</b>	<b>12</b>
<b>2.1.3 Méthodologie de développement de DW .....</b>	<b>15</b>
<b>2.1.4 Modélisation multidimensionnelle .....</b>	<b>15</b>
<b>2.2 Modélisation des besoins décisionnels .....</b>	<b>16</b>
<b>2.2.1 Ingénierie des besoins décisionnels .....</b>	<b>16</b>
<b>2.2.2 Ingénierie des besoins &amp; Ingénierie des SIDs .....</b>	<b>18</b>
<b>2.2.3 Classification des besoins .....</b>	<b>20</b>
<b>2.2.4 Langages de modélisation des besoins .....</b>	<b>21</b>
<b>2.2.5 Cycle de vie de l'ingénierie des besoins .....</b>	<b>22</b>
<b>2.2.6 Conclusion .....</b>	<b>23</b>
<b>2.3 Conclusion du chapitre .....</b>	<b>23</b>





# CHAPITRE II

## Conception d'entrepôts de données & Modélisation des besoins décisionnels

Savoir pour prévoir, afin de pouvoir.  
Auguste Comte



En raison de la mondialisation, les entreprises doivent être en mesure de livrer une vive concurrence sur les marchés mondiaux, qui est de plus en plus complexe et changeante. Les entreprises sont confrontées à une concurrence de plus en plus forte, des clients de plus en plus exigeants, Pour faire face aux nouveaux enjeux économiques, l'entreprise doit s'appuyer sur une information pertinente en exploitant efficacement le volume d'informations provenant soit de leurs systèmes opérationnels, soit de leur environnement extérieur, pour soutenir la prise de décision. Les entrepôts de données sont apparus pour intégrer et stocker l'information utile aux décideurs. La modélisation des besoins devient de plus en plus primordial pour garantir la qualité du design du data warehouse. Dans ce chapitre, on présente dans la section 1, la conception d'entrepôts de données en mettant en évidence l'informatique décisionnelle et le processus décisionnel. La section 2 est consacrée à la modélisation des besoins décisionnels, l'ingénierie des besoins (I.B), cycle e vie de l'I.B et surtout le langage de modélisation des besoins.

### 2.1 Conception d'entrepôts de données

Tout d'abord, on va définir ce que c'est un système, un système d'information et un système décisionnel avant de définir l'entreposage de données.

**2.1.1 Concepts de base:** Avant d'entamer cette étude, on va définir les concepts de base se rapportant au domaine du décisionnel, Il s'agit de système, système d'information, système d'information opérationnel et système d'information décisionnel.

a. **Système:** « Ensemble d'éléments en interaction dynamique, organisé en fonction d'un but » [Rivière et al., 2014].

b. **Système d'information:** est l'ensemble des ressources (matériels, logiciels, données, procédures, humains,...) structurés pour acquérir, traiter, mémoriser, transmettre et rendre disponible l'information (sous forme de données, textes, sons, images, ...) dans et entre les organisations [Rivière et al. , 2014].

Le SI [Reix et al., 1995] représente l'ensemble des ressources (humaines, matérielles, logicielles) organisées pour:

- Collecter l'information : Enregistrer une information (support papier, informatique...) avant son traitement
- Mémoriser l'information (stockage): Conserver, archiver (utilisation ultérieure ou obligation légale)
- Traiter l'information : effectuer des opérations (calcul, tri, classement, résumé,...)
- Diffuser : transmettre à la bonne personne (éditer, imprimer, afficher, une info après traitement).

Les éléments du système sont eux-mêmes des systèmes (ou sous-systèmes) et qui sont : Système de décision, Système d'information, Système opérant et Environnement. Comme tout système, l'entreprise est un système. L'entreprise peut se décomposer en trois sous systèmes : système de décision, système d'information et système opérant

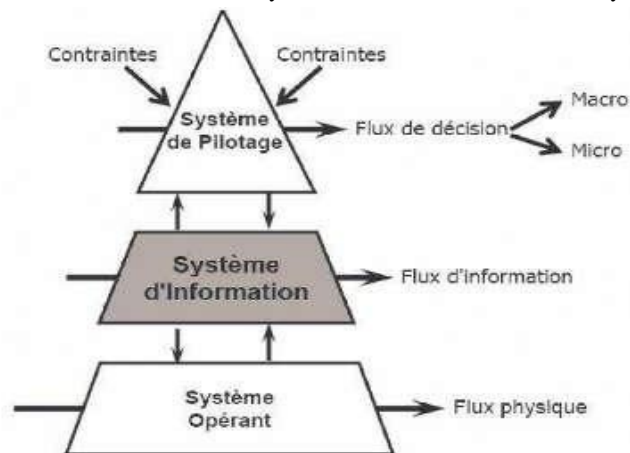


Figure 2.1 Les trois grands types de systèmes d'information [Reix et al., 1995]

• On retrouve comme types de systèmes ; Fonctionnels, Décisionnels et pour la communication. Nous avons aussi deux grandes catégories de système d'information [Ferragu et al., 2013]:

c. **Système d'information opérationnel (SIO)** a pour objectif premier de servir de support à la réalisation des activités d'un ensemble de processus métier. Par exemple, un SIO dédié au processus de vente.

d. **Système d'information décisionnel (SID)**, par opposition à un SIO dont l'objectif est l'exécution d'un processus métier, un SID a pour but l'évaluation de la performance des processus. Il a pour vocation de faciliter la prise de décision en fournissant des réponses à des questions telles que : quelle fut l'évolution du chiffre d'affaires et de la marge brute pour chaque catégorie de produits entre le premier semestre de cette année et celui de l'année précédente ?

Le tableau 2.1 illustre les principales différences entre les SIO et les SID.

Critère	SIO	SID
Objectif	Exécution de processus métier	Évaluation de la performance des processus métier
Mode d'interaction entre les utilisateurs et la base de données	Insertion, mise à jour, suppression et sélection de données	Sélection de données
Périmètre d'interaction entre les utilisateurs et la base de données	Transactions unitaires	Sélection de données en masse
Type d'utilisation	Prédéfinie, prévisible	Non prédéfinie, imprévisible
Complexité des requêtes des utilisateurs	Faible	Elevée
Optimisé pour	La performance des transactions unitaires	La performance des requêtes de sélection des données en masse
Fréquence de mise à jour de la base de données	Mises à jour en temps réel, au fur et à mesure de l'exécution des processus métier	Mises à jour périodiques en mode « batch »
Historique des données utilisées	Données courantes	Données courantes mais aussi et surtout historiques
Degré de normalisation des données	Hautement normalisé (3e forme normale)	Dénormalisé

Tableau 2.1 Différences entre les SIO et les SID [Golfarelli, 2010]

Les origines des SID remontent au début de l'informatique et des systèmes d'information qui ont, tous deux, connu une grande et complexe évolution. Cette évolution se poursuit à ce jour. Parmi les différentes définitions données du décisionnel, on trouve [Sfaxi et al., 2014]: La Donnée est le résultat direct d'une mesure, l'Information est une donnée à laquelle un sens et une interprétation ont été donnés, la Connaissance est le résultat d'une réflexion sur les informations analysées et la Sagesse est l'état d'esprit général de discernement final sur le contenu et de jugement de bon sens

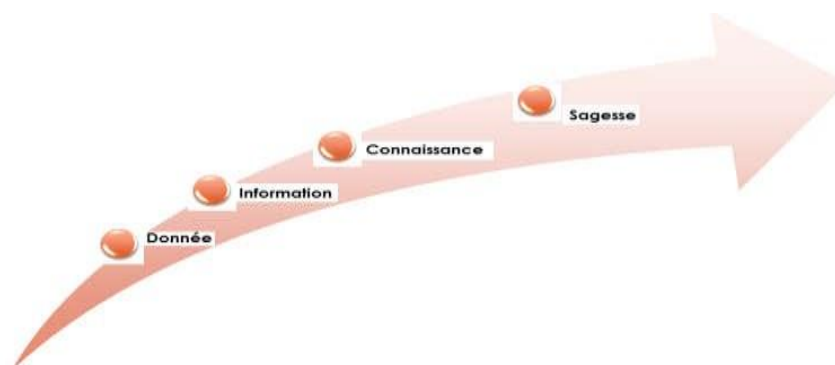


Figure 2.2 Métriques d'aide à la décision

### e. Informatique décisionnelle

L'informatique décisionnelle s'intéresse plus à améliorer les performances décisionnelles de l'entreprise en répondant aux demandes et besoins d'analyse des décideurs. Un décideur, à titre d'exemple, peut se poser dans sa gestion de clientèle les questions suivantes : Qui sont mes clients? Pourquoi sont-ils mes clients ? Comment les conserver ou les faire revenir? Ces clients sont-ils intéressants pour moi ?

L'informatique décisionnelle est donc l'ensemble des méthodes, moyens et outils informatiques utilisés pour piloter une entreprise et aider à la prise de décision. Les systèmes décisionnels permettent de générer de la connaissance à partir des données produites par les systèmes opérationnels (bases de données relationnelles). [Net01]

### 2.1.2 Architecture d'un SID

Les outils décisionnels sont basés sur l'exploitation d'un système d'information décisionnel alimenté grâce à l'extraction de données diverses à partir des données de production et d'informations concernant l'entreprise. Un outil appelé ETL (Extract, Transform and Load) est ainsi chargé d'extraire les données dans différentes sources, de les nettoyer et de les charger dans un entrepôt de données. Enfin des outils d'analyse décisionnelle permettent de modéliser des représentations à base de requêtes afin de constituer des tableaux de bord, on parle ainsi de reporting [Net01]

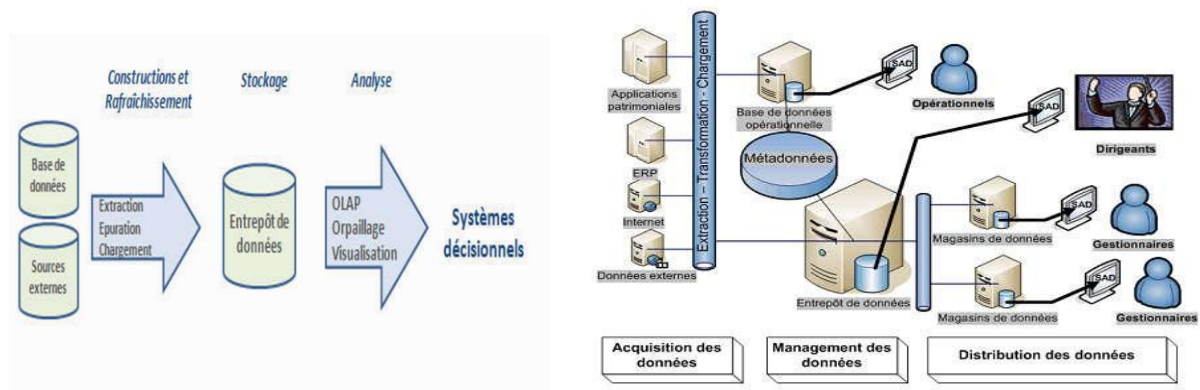


Figure 2.3: Architecture d'un système décisionnel [Lebraty et al., 2006]

#### a. Les différentes composantes du décisionnel :

En relation étroite avec les nouvelles technologies de l'information et des télécommunications, le système décisionnel se manifeste à différents niveaux selon leurs utilités et leurs missions principales, comme illustré dans la figure suivante :

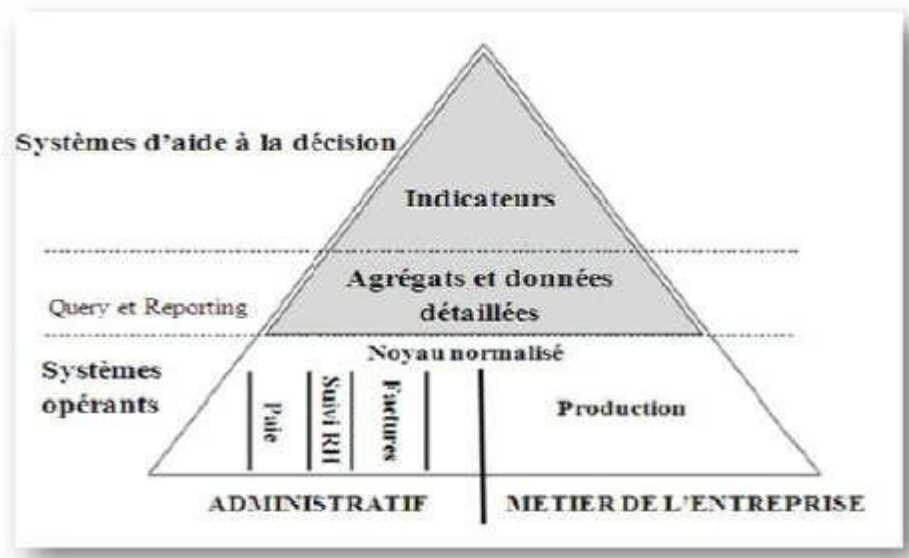


Figure 2.4 Les différentes composantes du décisionnel

#### a. Historique des systèmes d'information décisionnels

Les primitives de l'informatique décisionnelle se sont développées dans les années 70. Historiquement, les grandes firmes ont été les premières à comprendre la valeur ajoutée des outils d'aide à la décision. En effet, disposant de quantités considérables d'informations dans leurs bases de données opérationnelles, elles ont, en premier lieu, commencé par les interroger directement, par des requêtes. Cette solution de remplacement a vite montré ses limites aussi bien en temps qu'en ressources humaines et matérielles. De nouveaux outils apparaissent alors. Ces outils ont connu l'évolution suivante :

- L'infocentre : l'infocentre est une copie à l'identique des bases de données dans un nouvel environnement (années 70 et 80).
- L'EIS : Executive Information System proposant les premiers tableaux de bord dans les années 1990.
- Les entrepôts de données qui sont considérés comme étant le lieu de stockage des gros volumes de données devant être analysés.
- Les bases de données multidimensionnelles (OLAP), une base où chaque indicateur est analysé en fonction de plusieurs critères ou dimensions.
- et enfin le Business Intelligence qui regroupe les fonctions d'analyse des données et de reporting.

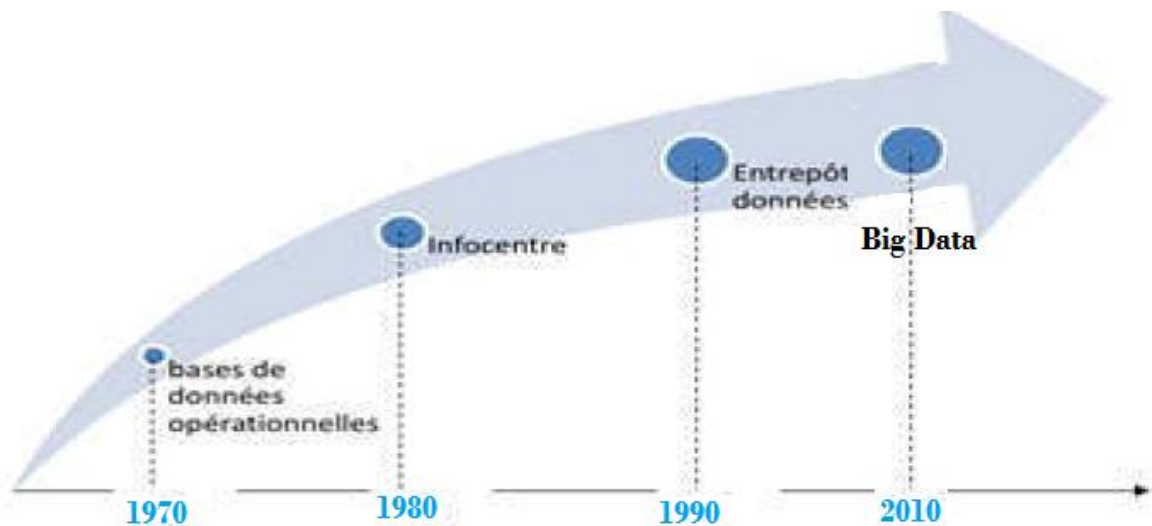


Figure 2.5 Evolution des bases de données décisionnelles

### b. Data Warehouse

Bill Inmon définit le Data Warehouse, dans son livre considéré comme étant la référence dans le domaine “Building the Data Warehouse” comme suit: « Le Data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et évolutives dans le temps, organisées pour le support d’un processus d’aide à la décision.» [Inmon et al., 2001]. Le Data warehouse est aussi un ensemble de méthodes, techniques et outils rassemblant des données issues de sources multiples au sein d’un modèle cohérent à des fins d’analyse et d’aide à la décision. L’architecture et la logistique du système d’information utilisant un data warehouse transforme le flux de données opérationnelles en informations décisionnelles. Un système né au début des années 1990.

**Orienté sujet :** le Data Warehouse est organisé autour des sujets majeurs de l’entreprise, contrairement à l’approche transactionnelle utilisée dans les systèmes opérationnels, qui sont conçus autour d’applications et de fonctions telles que : cartes bancaires, solvabilité client..., les DWs sont organisés autour de sujets majeurs de l’entreprise tels que : clientèle, ventes, produits.... Cette organisation affecte forcément la conception et l’implémentation des données contenues dans le Data Warehouse. Le contenu en données et en relations entre elles diffère aussi. Dans un système opérationnel, les données sont essentiellement destinées à satisfaire un processus fonctionnel et obéit à des règles de gestion, alors que celles d’un Data Warehouse sont destinées à un processus analytique.

**Intégrée :** le Data Warehouse va intégrer des données en provenance de différentes sources. Cela nécessite la gestion de toute incohérence.

**Évolutives dans le temps :** Dans un système décisionnel, il est important de conserver les différentes valeurs d’une donnée, cela permet les comparaisons et le suivi de l’évolution des valeurs dans le temps, alors que dans un système opérationnel la valeur d’une donnée est simplement mise à jour. Dans un DW chaque valeur est associée à un moment

**Non volatiles** : c'est ce qui est, en quelque sorte la conséquence de l'historisation décrite précédemment. Une donnée dans un environnement opérationnel peut être mise à jour ou supprimée, de telles opérations n'existent pas dans un environnement Data Warehouse.

**d. Etapes de conception d'un Entrepôt de Données (ED)** : On retrouve comme phases de processus décisionnel : L'analyse des besoins, la modélisation conceptuelle, la modélisation logique, le processus ETL (Extraction Transformation Loading) et la modélisation physique

### **2.1.3 Méthodologie de développement de DW:**

Trois catégories de méthode pour le développement d'un Système d'Information Décisionnel (SID) : ascendantes, descendantes et mixtes.

#### **a. L'approche descendante**

Elle propose de définir le schéma de l'entrepôt en fonction des besoins d'analyses et suppose que les données disponibles permettront la mise en œuvre d'un tel schéma (schéma de l'entrepôt), ou tout du moins que la confrontation avec les données réelles se fera dans un second temps.

#### **b. L'approche ascendante**

Elle ne prend pas en compte les besoins des utilisateurs. Elle consiste à construire le schéma de l'entrepôt de données à partir des sources et suppose que le schéma construit pourra répondre à tous les besoins d'analyses. Par exemple, des auteurs qui proposent une méthodologie semi-automatique pour construire un schéma du data warehouse à partir des schémas entité-relation qui représentent les bases de données sources.

#### **c. L'approche mixte**

L'approche mixte quant à elle combine les deux méthodes précédentes. Cependant elle présente l'inconvénient de ne pas avoir de méthodes d'analyses des besoins utilisateurs. L'avantage de cette démarche est la combinaison de la méthode descendante et ascendante ; elle intègre à la fois les données des sources de productions et les besoins des décideurs dans la définition du modèle décisionnel. [Sfaxi et al., 2014]

### **2.1.4 Modélisation Multidimensionnelle**

L'entrepôt de données est typiquement modélisé par des modèles multidimensionnels (encore appelés cubes de données). En fonction de la manière dont le cube est stocké, deux approches existent pour construire des systèmes multidimensionnels: Relationnel OLAP (ROLAP), et Multidimensionnel OLAP (MOLAP).

Ainsi, on va présenter dans la section suivante la modélisation des besoins décisionnels pour mettre en évidence la phase, longtemps ignorée, du processus décisionnel qui est l'analyse des besoins.

## 2.2 Modélisation des besoins décisionnels

Pendant la dernière décennie, les chercheurs ont donné plus d'importance à une phase aussi importante et influente que les autres phases sur le processus décisionnel. Il s'agit de l'analyse des besoins. Dans cette section, nous commençons par présenter l'ingénierie des besoins et particulièrement l'aspect décisionnel, ensuite les classes des besoins et les langages de représentation de besoins.

### 2.2.1 Ingénierie des Besoins Décisionnels (IBD)

L'Ingénieur des Besoins Décisionnels représente le processus chargé de définir les besoins correspondant à la mission et aux objectifs de l'organisation. Ces besoins, exprimés par des décideurs sont recensés, négociés, validés, spécifiés et répertoriés dans des documents de spécifications des besoins.

#### a. Concept du besoin [Boukhari et al., 2014]

Dans la littérature anglo-saxonne, les auteurs ne différencient pas réellement les concepts de besoin et d'exigence. En effet, selon Chemuturi : «*Une exigence est un besoin, une attente, une contrainte ou une interface de toutes les parties prenantes, qui doit être satisfaite par le système logiciel proposé*». Le besoin est alors plutôt vu comme une attente basique voire vitale, le minimum absolu sans lequel le système ne sera pas utilisable.

**Besoin** : « *stakeholder requirement* » ou « *user requirement* » considérée alors comme la représentante des parties prenantes commanditaires. Il exprime et décrit une situation, un vœu, un but ou une contrainte vis-à-vis du système à réaliser. Les besoins sont le plus souvent définis de façon informelle, dans le langage de la maîtrise d'ouvrage, afin de simplifier la communication avec ces parties prenantes. Ils reflètent le monde du problème. Le **Besoin** selon [Abran et al., 2004] est une description d'une propriété ou des propriétés du système qui doivent être remplies. Souvent, les besoins peuvent être classés dans les deux catégories suivantes. [Boukhari et al., 2014]

– **Besoins fonctionnels** : les besoins fonctionnels spécifient une fonction que le système ou un composant du système doit être capable d'exécuter. Les auteurs de cette définition soulignent que le système peut être vu comme un ensemble de composants.

**Exemple** : Dans un établissement, un système doit permettre à la scolarité de l'université de vérifier les moyennes des étudiants qui ont obtenu le master informatique.

– **Besoins non fonctionnels** : Un besoin non fonctionnel est défini par un attribut ou une contrainte du système comme la flexibilité, la performance, la sécurité, etc. qui ne porte pas sur la fonction de ce système.

**Exemple** : Le système doit permettre aux étudiants de ne voir uniquement que leurs propres moyens

**b. Concept Exigence:** « System requirement », une exigence est définie comme « *un énoncé qui prescrit une fonction, une aptitude ou une caractéristique à la quelle doit satisfaire un produit ou un système dans un contexte donné* » [Scukanec et al., 2007]. Les besoins traduisent donc les besoins, elles sont établies de fait dans le langage de la maîtrise d'œuvre. Elles représentent les propriétés attendues du système qui vont ainsi contraindre la conception de la solution. Ces besoins sont organisés dans un référentiel, i.e. l'ensemble des besoins qui reflète les consensus et décisions prises par les parties prenantes et est à la base d'un document contractuel définissant les rôles et responsabilités des commanditaires et prestataires. En ce qui concerne les projets de développement d'applications logicielles, la distinction entre besoin et exigence est d'autant plus importante qu'elle permet de mettre l'accent sur le problème à solutionner en considérant le point de vue de chaque partie prenante pour éviter que les seules descriptions du système à réaliser soient exprimées en termes de solution [Hull et al., 2010].

Il existe deux catégories de besoins (ISO15288 2008) à savoir : les besoins fonctionnels et les besoins non-fonctionnels.

**Les besoins fonctionnels** font référence aux fonctionnalités et aux services que le système doit fournir. Lorsqu'il s'agit par exemple d'une application de gestion, elles traduisent les opérations à automatiser et sont fortement liées au fonctionnement de l'entreprise.

**Les besoins non-fonctionnels** contraignent et précisent comment le système doit satisfaire les besoins fonctionnels en termes de performance, sécurité, fiabilité, ergonomie, niveau de maintenabilité, portabilité, réutilisabilité, etc. Elles peuvent concerner également certains aspects liés à la gestion du projet de développement comme le coût et les délais de réalisation. Leur identification est importante puisqu'elles servent de critères de sélection permettant à l'équipe de développement de faire des choix en termes de conception et d'implémentation.

La première référence à la nécessité d'actions d'ingénierie des besoins et des besoins était faite au milieu des années 70 : « the requirements for a system do not arise naturally; instead, they need to be engineered and have continuing review and revision » [Bell et al., 1976] et depuis l'an 2000, un grand nombre de méthodes et de langages ont été proposés, suivants différentes orientations (buts, scénarios, etc.). Cependant, quelle que soit la méthode proposée, l'objectif des travaux en IB est d'améliorer la qualité des besoins en atténuant les défauts, omissions ou erreurs souvent rencontrés dans la phase d'expression des besoins et inhérents à la nature informelle de ces derniers, tels que [Loucopoulos et al., 1989]:

- **L'ambiguïté** : informations données au niveau des besoins ayant différentes interprétations possibles;
- **L'imprécision** : besoins flous ou insuffisamment définis, structurés ou détaillés suite à des reformulations par exemple ambiguës;

- **L'omission** : besoins omis par oubli objectif ou subjectif, menant ici aussi à des incohérences ou des imprécisions;
- **La redite** : besoins exprimés plus d'une fois ou reformulés en d'autres termes induisant encore des incohérences ou de l'ambiguïté;
- **L'incohérence** : besoin en contradiction avec un ou plusieurs besoins.

Nous pouvons aussi citer **le bruit, la non-mesurabilité, la non-intelligibilité ou la non-traçabilité** [Lamsweerde et al., 2009].

### 2.2.2 Ingénierie des Besoins et Ingénierie des SIDs

Dans [Rolland, 2003], l'Ingénierie des Besoins (IB) est définie comme un processus qui permet l'exploration des objectifs des acteurs afin de dériver les besoins. Nous distinguons trois (ou quatre) types d'approches d'IB :

- (i) approche à base de scénarios
- (ii) approche à base de buts
- (iii) et approche couplant à la fois les scénarios et les buts
- (iv) Approche basée sur les automates [Jilani, 2010]

Un scénario décrit «un ensemble fini d'interactions entre des agents dont le comportement est dirigé par des objectifs distincts» [Ben Achour et al., 1999]. Chaque interaction implique deux agents et permet d'atteindre un but donné. Un but peut être défini comme «quelque chose que quelqu'un espère réaliser dans le futur» [Ben Achour et al., 1999]. Le couplage des deux concepts consiste à rattacher un but à un et à un seul scénario.

À l'instar de l'IB des Systèmes d'Information (SI), l'IBD représente le processus par lequel les analystes-concepteurs décisionnels et les décideurs construisent le SID à travers la découverte des besoins. La réussite d'un projet décisionnel dépend fortement de la démarche d'IBD adoptée. Cette dépendance est liée aux activités définies ainsi qu'aux modèles et documents proposés pour la réalisation de ces activités.

Malgré la diversité des travaux proposés relatifs à ce sujet, l'IBD nécessite encore des investigations et des améliorations au niveau des modèles et au niveau de la démarche. Les démarches d'IBD retenues pour cet état de l'art concernent deux aspects de l'IBD :

#### Aspect global des approches d'IBD

- Aspect portant sur les modèles de représentation des besoins utilisés par ces démarches. L'ingénierie des besoins et des exigences (IB) [Amokrane et al., 2016] est une discipline commune à toutes les ingénieries métiers (génie logiciel, mécanique, électronique, etc.) comme à l'Ingénierie Système qui la reconnaît comme la phase décisive la plus en amont dans tout projet d'ingénierie d'un système i.e. de produits et/ou de services satisfaisant les besoins de différentes parties prenantes.

Le processus, regroupant toutes les activités d'IB, est défini par [Loucopoulos et al., 1989] comme étant “the systematic process of developing requirements through an iterative process of analyzing a problem, documenting the resulting observations, and checking the

accuracy of the understanding gained”. Il suppose donc l’application systématique et répétitive de techniques d’analyse, de documentation et de vérification et de validation des besoins. Dans le domaine du Génie Logiciel [Amokrane et al., 2016]: «l’ingénierie des besoins est concernée par l’identification de buts assignés au système envisagé et l’opérationnalisation de ces buts en contraintes imposées au système et aux agents qui en assureront le fonctionnement. L’ingénierie des besoins peut être vue comme le processus qui permet de transformer une idée floue en spécification précise des besoins servant de support à la spécification du système et de ses interfaces avec l’environnement ».

L’IB guide et accompagne le passage d’une description initiale, souvent peu claire voire contradictoire que nous appelons **besoin**, à une spécification plus formelle et plus contractuelle de ce besoin, appelée **exigence**. Ceci se fait au moyen d’activités, entre autres, d’expression, de documentation et d’évaluation continue des besoins et de construction d’un référentiel de besoins.

**L’Ingénierie des Besoins (IB):** dans le cadre des systèmes d’information, elle est définie par Van Lamsweerde [80] définit trois objectifs de l’ingénierie des besoins:

- (i) l’identification des buts que le système envisagé doit accomplir,
- (ii) l’opérationnalisation de ces buts sous forme de fonctions et de contraintes et
- (iii) l’assignation des fonctionnalités aux agents.

**Processus d’ingénierie des Besoins** [Mylopoulos et al., 1992]

Généralement, un processus d’IB comporte plusieurs étapes consistant à découvrir, analyser, valider et faire évoluer l’ensemble des besoins relatifs aux fonctionnalités du système. Selon [Annoni, 2006], un processus d’IB inclut les étapes suivantes : élicitation, modélisation, analyse, validation et gestion des besoins (figure ). Ces étapes sont détaillées dans ce qui suit :

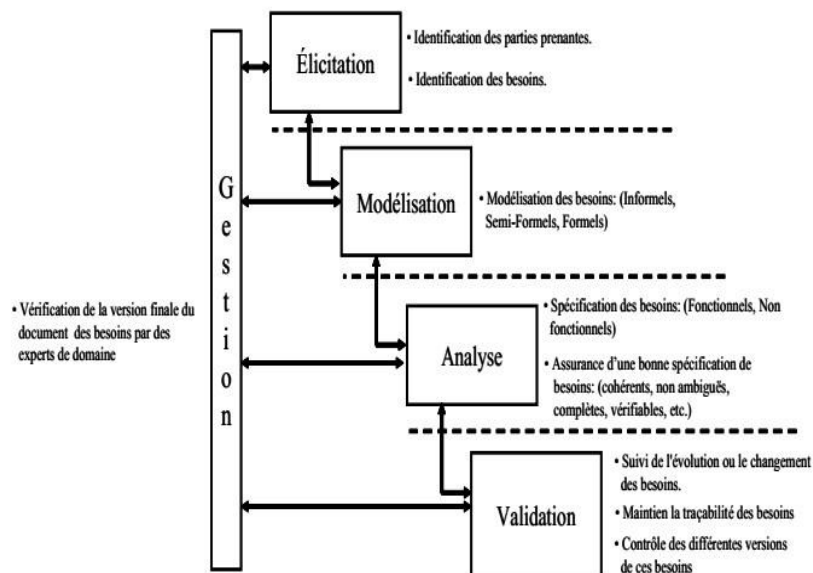


Figure 2.6 Processus d’ingénierie des besoins.

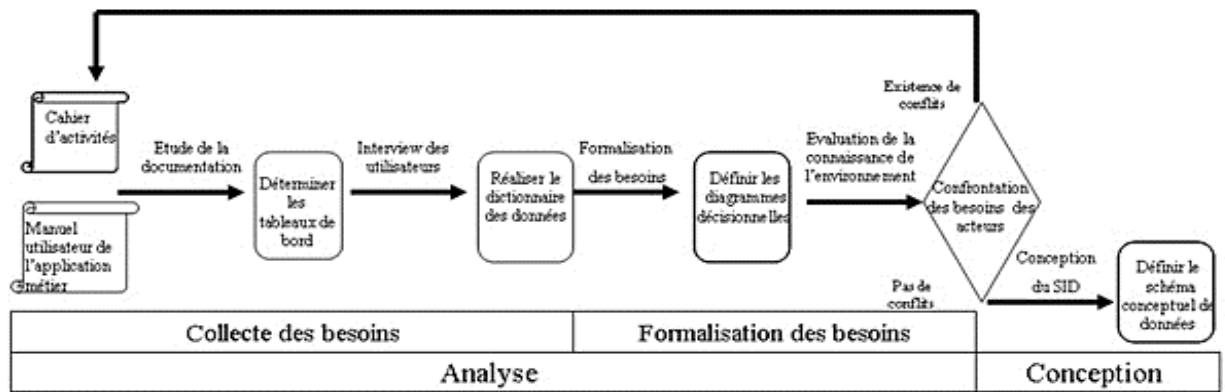


Figure 2.7 Processus de collecte et de formalisation des besoins utilisateurs [Thomsen et al., 2002]

### 2.2.3 Classification des besoins [Annoni, 2006]:

Nous classons les besoins exprimés par les différents intervenants dans un projet SID en quatre types :

- ✓ **analytique** : besoins liés aux analyses de données, à la qualité des données et aux processus de consolidation, d’historisation, d’archivage et de rafraîchissement des données,
- ✓ **fonctionnel** : besoins liés aux trois principales tâches, soient, **l’alimentation** (règles de gestion,...), le stockage (volume et format des données) et la **présentation** (restitution et diffusion),
- ✓ **non fonctionnel** : besoins liés à la sécurité des données, aux performances d’interrogation et de restitution de l’information,
- ✓ **stratégique** : besoins liés aux différentes politiques de l’entreprise.

Nous classons les acteurs du SID en trois groupes :

- le groupe **Utilisateurs** exprime les besoins **analytiques**, **fonctionnels** (règles de gestion, règles de calcul) ainsi que **non fonctionnels** (ergonomie d’outil),
- le groupe **Equipement** exprime les besoins fonctionnels liés au **stockage** et à la disponibilité ainsi que des besoins **non fonctionnels** liés à la sécurité des données et à la maintenance,
- le groupe **Pilotage** exprime les besoins **analytiques**, **fonctionnels**, **non fonctionnels** mais aussi des **besoins liés à la stratégie** de l’entreprise.

### 2.2.4 Langages de modélisation des besoins [Jilani, 2010].

Plusieurs langages de modélisation des besoins ont été proposés dans la littérature et plusieurs d’entre eux sont mis en pratique dans l’industrie. Voici une classification en trois catégories des langages de modélisation des besoins.

#### Langages informels (langage naturel)

Les langages informels se basent sur le langage naturel avec ou sans règles de structuration. Bien qu'elles ne requièrent aucune formation particulière et qu'elles garantissent la facilité de compréhension (due à la manière familière de communiquer) entre les parties prenantes du système, elles introduisent des risques d'ambiguïté, des malentendus et des incohérences car ni leur syntaxe, ni leur sémantique ne sont explicitement définies. Parmi ces langages, nous pouvons citer : le questionnaire, l'interview et le cahier des charges.

#### **Langages semi-formels**

Les langages semi-formels sont basés sur des notations graphiques avec une syntaxe précise. Ces langages facilitent la communication entre les parties prenantes. Leur principal inconvénient est le manque de sémantique précise, laissant des risques d'ambiguïté et d'incohérence des besoins. Les modèles orientés buts ou cas d'utilisation UML sont des exemples de modèles semi-formels.

#### **Langages formels**

Les langages formels sont basés sur des notations mathématiques qui fournissent un cadre précis et non ambigu pour la modélisation des besoins à l'exemple de la méthode B. L'avantage de ces langages est de permettre la vérification des besoins. Leur inconvénient est qu'elles sont difficiles à manipuler par l'ensemble des parties prenantes qui ne maîtrisent généralement pas ces formalismes.

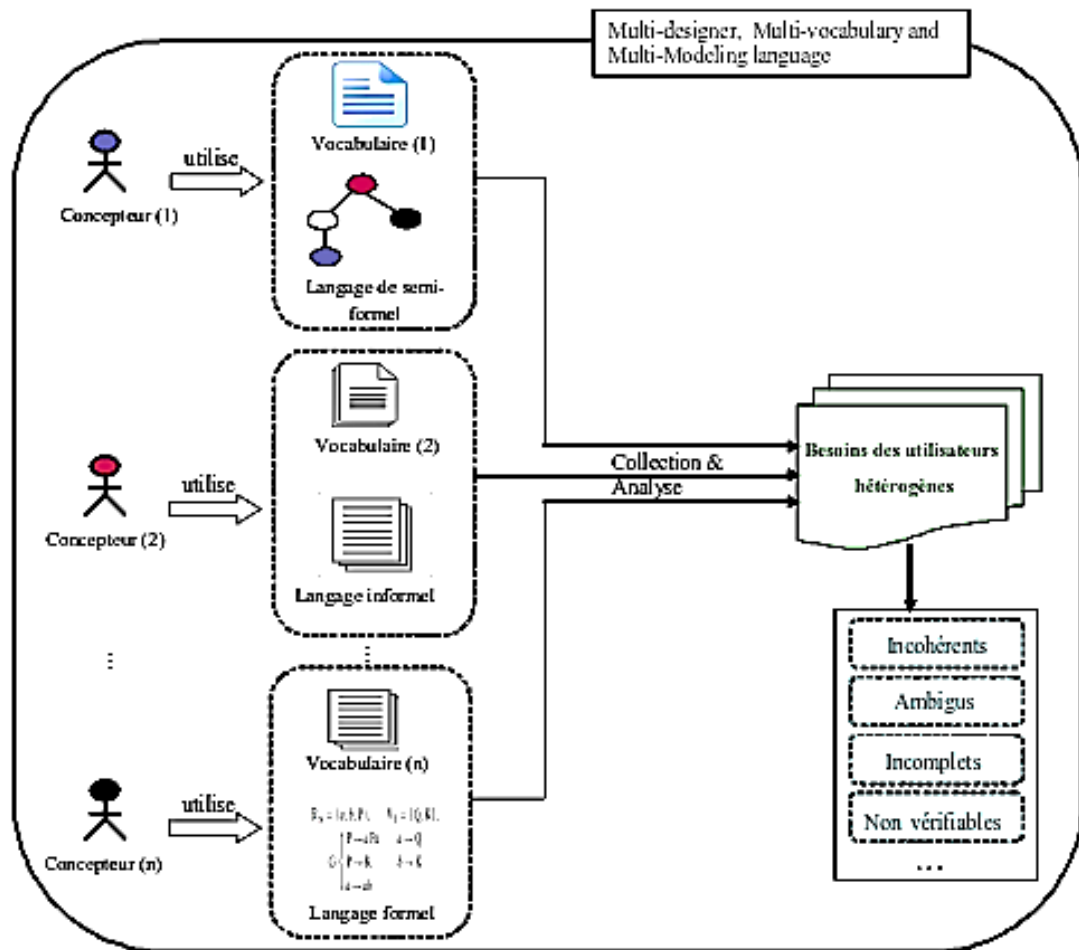


Figure 2.8 Hétérogénéité des vocabulaires et des langages de modélisation des besoins [Boukhari et al., 2014].

## 2.2.5 Cycle de vie de l'ingénierie des besoins

Différentes études de l'ingénierie des besoins [Bellatrèche et al., 2007] proposées pour prendre en charge l'intégration, la validation et la validation des problèmes de persistance des besoins. Nous analysons les travaux les plus importants selon leur contribution au cycle de vie de l'ingénierie des besoins, où nous avons identifié quatre étapes principales: Source Sélection, Intégration des sources, mécanisme de raisonnement et la persistance des besoins. Nous avons identifié pour chaque étape de la démarche les éléments suivants afin de mettre en évidence notre contribution:

**Etape 1: Sélection des sources:** nous avons défini quatre critères pour cette étape: désamorçage du schéma Global (Gs), autonomie des schémas locaux (Ls), hétérogénéité du formalisme (Fh) et Planification des tâches (St). L'hétérogénéité du formalisme est généralement gérée en associant différents langages de modélisation formels, semi-formels et informels comme entrées.

**Étape 2: Intégration des Sources (SIn):** nous avons identifié le fichier niveaux d'intégration suivants pour cette étape: Intégration linguistique (Li), Intégration conceptuelle (Ci), Intégration formaliste (Fi), Planification de l'intégration des tâches (Si).

Pour le niveau linguistique, l'intégration des vocabulaires est rarement considérée. L'intégration du concept est réalisée en unifiant les schémas locaux définis à partir du schéma global (généralement une ontologie conceptuelle)

**Étape 3: Mécanisme de raisonnement (MR):** le but de ce processus est d'identifier de nouvelles relations cachées entre les besoins et de vérifier leur cohérence.

**Étape 4: Persistance des besoins (RP):** peu d'études s'intéressent au stockage des besoins. La particularité de nos travaux [Jilani, 2010] concerne la persistance des besoins des utilisateurs dans les bases de données ontologiques, ce qui permet une réutilisation efficace, les mécanismes de gestion et d'interrogation.

### 2.2.6 Conclusion

Ce chapitre montre le lien étroit entre la conception de l'entrepôt de données et la modélisation du besoin décisionnel. En fait, Ignorer la phase de modélisation du besoin peut engendrer des erreurs au niveau de la conception du DW et peut par la suite entraîner un processus décisionnel erroné. D'où l'importance qu'on doit attribuer à cette phase en vue d'améliorer la qualité du design du DW.

## 2.3 Conclusion du chapitre

Dans ce chapitre, Nous allons présenter le processus décisionnel regroupe la modélisation conceptuelle, logique et physique de l'entrepôt de données. Il regroupe aussi la phase d'analyse des besoins qui n'a pas eu suffisamment d'études de la part des chercheurs dans ce domaine. Dans le chapitre suivant, nous focalisons l'étude sur l'évaluation du besoin. Nous avons choisi qu'elle sera en langage formel (SQL par exemple) pour traiter l'imprécision du besoin. Dans le chapitre suivant, nous avons présenté le background sur la thématique de cette thèse, se rapportant aux approches qu'on propose comme une solution à ce genre de problème.





Le besoin est le maître des arts, il crée le génie.  
Proverbe latin

# CHAPITRE III

## Evaluation de Besoins Décisionnels Flous

<b>3.1 Evaluation des BDFs utilisant une solution structurelle.....</b>	<b>25</b>
3.1.1 Modèle multidimensionnel .....	25
3.1.2 Modélisation d'un entrepôt de données .....	25
3.1.2 Métadonnée et méta modèle d'entrepôts de données .....	27
3.1.3 Conclusion .....	27
<b>3.2 Evaluation des BDFs utilisant un système de connaissance .....</b>	<b>28</b>
3.2.1 Evaluation des BDFs utilisant un système expert .....	28
3.2.1.1 Logique classique & logique floue .....	28
3.2.1.2 Extension de la logique floue (CWP) .....	33
3.2.1.3 Evaluation de requêtes flexibles à base de système expert.....	34
3.2.1.4 STDS : Schéma de translation dirigée par la syntaxe .....	37
3.2.1.5 Conclusion .....	37
3.2.2 Evaluation des BDFs utilisant une ontologie floue.....	38
3.2.2.1 Définition d'Ontologie .....	38
3.2.2.2 Composantes de l'ontologie .....	39
3.2.2.3 Cycle de vie d'une ontologie .....	41
3.2.2.4 Les langages de description des ontologies .....	42
3.2.2.5 Ontologies floues.....	47
3.2.2.6 Conclusion (Evaluation de BDFs à base d'ontologie) .....	52
<b>3.3 Conclusion du chapitre .....</b>	<b>52</b>



# CHAPITRE III

---

## Evaluation de Besoins Décisionnels Flous

---

Aucune origine n'est belle. La beauté véritable est au terme des choses.

Sous le signe de minerve (édition 1954) - Charles Maurras

---

Afin d'avoir un système d'information très fiable, deux solutions sont possibles : soit on restreint le système d'information à une portion du monde réel, c'est-à-dire qu'on utilise seulement la partie sur laquelle on possède des informations fiables, soit on définit un système d'information qui accepte des informations imparfaites. C'est cette dernière solution qui est employée dans la plupart des systèmes d'information.

Afin qu'il soit générique, un processus doit être ouvert pour permettre l'intégration des connaissances. Pour que les connaissances soient exploitables par le processus, elles doivent être formalisées et stockées d'une manière informatique, tout en prenant en compte leurs imperfections [Olteanu, 2008]

L'Ingénierie des Connaissances (IC) est une branche de l'IA issue de l'étude des Systèmes Experts (SE). Si ces derniers n'avaient pour objet que la résolution automatique de problèmes, les Systèmes à Base de Connaissances (SBC) qui leur ont succédé sont censés permettre le stockage et la consultation de connaissances, le raisonnement automatique sur les connaissances stockées (sans préjugé sur le type de raisonnement à mener), la modification des connaissances stockées, et, avec le développement des réseaux, le partage de connaissances entre systèmes informatiques.

On découple alors la représentation des connaissances de celle des mécanismes inférentiels [Fürst, 2002]. Mais d'abord, avant d'étudier l'évaluation d'imprécision à base d'ontologie on envisage de l'étudier utilisant, en premier lieu, une solution structurelle puis un système expert.

Les lacunes dans la qualité des données, qui sont toutes sources de différents types d'incertitude, peuvent être le résultat de plusieurs facteurs [Thakur & al., 2011]: Exactitude et erreur: écart par rapport aux valeurs réelles ; Imprécision: imprécision des concepts utilisés pour décrire l'information (p. ex. près et loin sont des propriétés métriques vagues d'un objet spatial décrivant la distance par rapport à l'observateur) ; Incohérence: conflits découlant de l'information. L'expression d'imprécision n'est pas limitée au langage naturel mais peut également être présente dans des langages structurés. Pour cibler l'imprécision dans les requêtes OLAP, nous considérons la structure générique suivante d'une telle requête:

```
SELECT [Attribute1, Attribute2 ..]
FROM [Table1, Table2 ..]
WHERE [Predicate: -operator-value attribute] [attribute - operator - value]
connector GROUP BY [OLAP operator: drilldown, ..] (attribute)
```

L'imprécision dans une requête OLAP peut être localisée dans des clauses peut inclure des parties inexacts: valeur d'attribut du prédicat, connecteur, opérateur et agrégation. Par exemple, considérez la catégorie Bus en fonction de la relation " Non sécuritaire ". On peut dire que le « Bus dangereux » si le nombre d'accidents de ce bus dans une période donnée dépasse  $\alpha$  et le nombre de décès ne dépasse pas  $\beta$ . Le problème de l'imprécision se pose s'il est réduit de 1 ou 2 est-ce que l'appartenance à cette catégorie reste ou non et avec quel degré d'appartenance? Même si nous ajoutons 1 ou 2 à  $\alpha$ , quel est le degré d'adhésion du Bus dans cette classe "Unsafe". Dans ce travail, nous nous intéressons précisément à l'inexactitude et l'imprécision dans la valeur prédicat de la requête exprimant la décision requiert. Pour le reste du chapitre, La section 1 présente les principaux concepts se rapportant à la solution structurelle, à savoir : les concepts du modèle multidimensionnel, la modélisation d'entrepôts de données et surtout les schémas de la modélisation logique et les concepts de méta donnée et méta modèle du DW. La section 2 présente les concepts de base de la solution à base de connaissance. Elle se répartie en 2 sous-section dont la première présente les concepts pour la solution à base de système expert telle que : la logique floue, variables linguistiques, fonctions d'appartenance, modificateurs linguistiques, langage d'interrogation flou et schéma de translation dirigé par la syntaxe. La seconde sous-section est consacrée aux concepts de la solution ontologique à savoir, l'ontologie, composants de l'ontologie, cycle de vie d'une ontologie, langage de description d'une ontologie, l'ontologie floue et ses concepts et le modèle en oignon.

---

## 3.1 Evaluation des BDFs utilisant une solution structurelle

---

Deux concepts fondamentaux caractérisent le modèle multidimensionnel: dimensions et faits. Le Fait représente le sujet d'analyse (Table de Fait), Les dimensions représentent les axes d'analyse (Tables de dimension)

### 3.1.1 Modèle multidimensionnel

- a. Concept de fait :** Une table de faits est la table centrale d'un modèle dimensionnel, où les mesures de performances sont stockées. Une ligne d'une table de faits correspond à une mesure.
- b. Concept de dimension :** Les tables de dimension sont les tables qui accompagnent une table de faits, elles contiennent les descriptions textuelles de l'activité. Une table de dimension est constituée de nombreuses colonnes qui décrivent une ligne.
- c. Concept d'hierarchies :** Il est important d'avoir des hierarchies bien définies dans un SID. L'importance provient du fait que la prise de décision commence par des vues générales puis les informations se détaillent de plus en plus. En plus, si des outils OLAP sont utilisés pour l'analyse des données, il est ainsi possible de réaliser des agrégations automatiques des données en s'appuyant sur les hierarchies définies.
- d. Concept de Granularité:** La granularité est le niveau de détail des données dans un entrepôt de données. La granularité détermine le volume des données ainsi que le type des requêtes que l'utilisateur peut poser.

### 3.1.2 Modélisation d'un entrepôt de données

Trois schémas principaux sont utilisés pour modéliser les systèmes ROLAP : (1) le schéma en étoile, (2) le schéma en flocon de neige, (3) Schéma en constellation [Kimball, 2002].

**a. Schéma en étoile :** Ce modèle représente visuellement une étoile, on parle de modèle en étoile (Star Schéma), où tous les faits sont définis dans une simple table relationnelle. La table des faits est généralement normalisée, par contre les tables de dimension sont dé-normalisées. Les requêtes, définies sur ce schéma, sont appelées requêtes de jointure en étoile. (star-join Requêtes),

```
SELECT <Liste de projection> <Liste d'agrégation> FROM <Nom de la table des faits> <Liste de noms de tables de dimension> WHERE <Liste de prédicats de sélection & jointure> GROUP BY <Liste des attributs de tables de dimension>
```

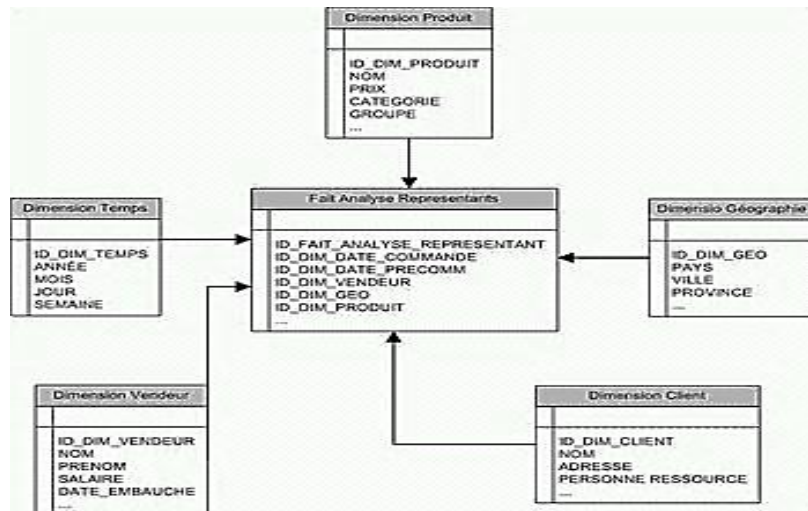


Figure 3.1 Exemple d'un schéma en étoile

Les caractéristiques des requêtes de jointure en étoile sont :

- Il y a des jointures multiples entre la table des faits et les tables de dimension.
- Il n'y a pas de jointure entre les tables de dimensions.
- Chaque table de dimension impliquée dans une opération de jointure a plusieurs prédicat de sélection sur ses attributs descriptifs.

**b. Schéma en flocon de neige:** Le flocon est simplement une étoile dont les branches sont elles-mêmes décomposées en sous-hiérarchies. Modéliser en flocon, c'est donc conserver le cœur du modèle en étoile, à savoir les tables de faits et affiner la modélisation des tables de dimensions pour les éclater en sous-tables. La figure 3.2 illustre ce modèle.

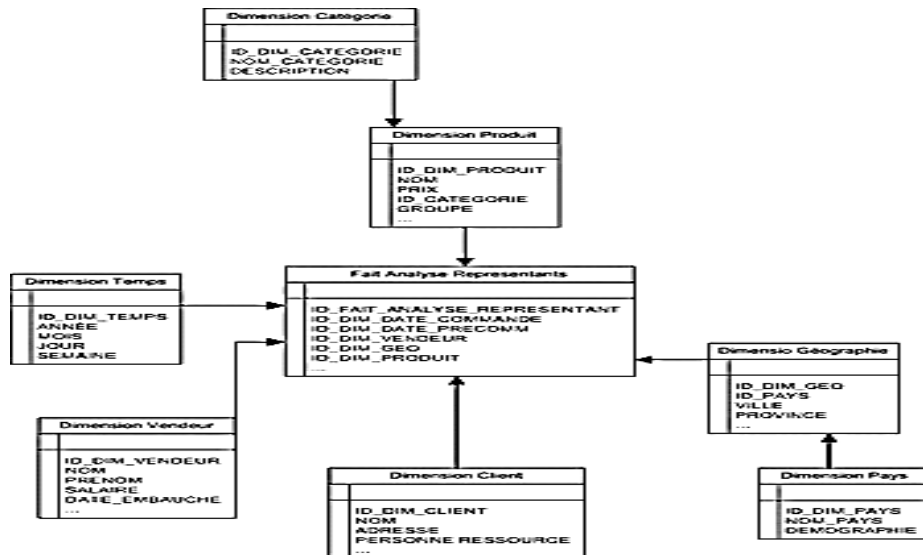


Figure 3.2 Exemple d'un schéma en flocon de neige

Ce modèle présente l'inconvénient d'être touffu et parfois difficile à comprendre, et la base correspondante est lente et moins performante que celle correspondante au modèle en étoile. Ceci est dû essentiellement aux multiples jointures intermédiaires entre tables de dimension hiérarchisées.

**b. Schéma en constellation:** Il s'agit de fusionner plusieurs modèles en étoile qui utilisent des dimensions communes.

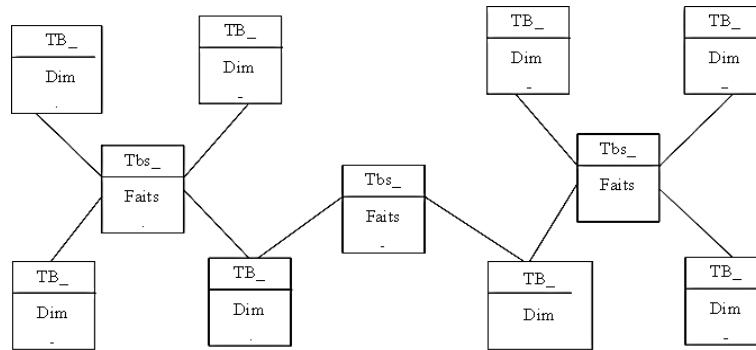


Figure 3.3 Exemple d'un schéma en constellation

### 3.1.3 Métadonnée et méta modèle d'entrepôts de données

Les métadonnées de l'entrepôt de données sont des éléments d'information stockés dans un ou plusieurs référentiels de métadonnées à usage spécifique qui incluent (a) des informations sur le contenu de l'entrepôt de données, leur localisation et leur structure, (b) des informations sur les processus qui se déroulent à l'arrière-plan de l'entrepôt de données, concernant le rafraîchissement de l'entrepôt avec des données propres, mises à jour, sémantiquement et structurellement rapprochées, (c) des informations sur la sémantique implicite des données (par rapport à un modèle d'entreprise commun), ainsi que sur tout autre type de données permettant à l'utilisateur final d'exploiter les informations de l'entrepôt, d) des informations sur l'infrastructure et les caractéristiques physiques des composants et les sources de l'entrepôt de données, et (e) des informations comprenant des statistiques de sécurité, d'authentification et d'utilisation qui aident l'administrateur à ajuster le fonctionnement de l'entrepôt de données, le cas échéant.

### 3.1.4 Conclusion

Cette section présente un background de la solution structurelle, commençant par le changement de structure où le passage généralement du modèle en étoile vers le modèle en flocon de neige. Aussi, l'utilisation de métadonnées, pour l'évaluation d'imprécision et l'optimisation de performance de l'entrepôt en minimisant les opérations de jointure causées par le changement de structure.

---

## **3.2 Evaluation des BDFs utilisant un système de connaissance**

### **3.2.1 Evaluation des besoins décisionnels flexibles Utilisant un système expert flou**

---

La logique vient du grec logos, qui signifie raison ou discours. La logique n'est pas empirique et ne permet pas une connaissance du monde. Elle est plutôt réflexive et tente d'expliquer et de formaliser les mécanismes et le fonctionnement de l'esprit. La logique est donc une théorie du raisonnement et permet de déterminer si une inférence est valide ou non [Belna, 2005]. Parmi les applications de la logique floue les systèmes experts flous.

#### **3.2.1.1 Logique classique & logique floue**

Dans le cadre de la logique classique [Djaaboub, 2009], une proposition est soit vraie, soit fausse (1 ou 0). Par exemple, la logique classique peut facilement partitionner la température d'une pièce en deux sous-ensembles, «moins de 15 degrés» et «15 degrés ou plus». Les figures 3.4 et 3.5 montrent le résultat de cette partition. Toutes les températures de moins de 15 degrés sont alors considérées comme appartenant à l'ensemble «moins de 15 degrés». On leur affecte une valeur de 1. Toutes les températures atteignant 15 degrés ou plus ne sont pas considérées comme appartenant à l'ensemble «moins de 15 degrés». On leur attribue une valeur de 0. Cependant, le raisonnement humain s'appuie fréquemment sur des connaissances ou des données inexacts, incertaines ou imprécises. Une personne placée dans une pièce dont la température est soit de 14.95 degrés soit de 15.05 degrés, ne fera certainement pas de distinction entre ces deux valeurs. Cette personne sera pourtant capable de dire si la pièce est «froide» ou «chaude», sans pour cela utiliser de température limite ni de mesure précise. La logique floue permet de définir des sous-ensembles, comme «froid» ou «chaud», en introduisant la possibilité pour une valeur d'appartenir plus ou moins à chacun de ces sous-ensembles. La logique floue permet de faire intervenir les notions d'imprécision et d'incertitude dans un système. L'incertitude et l'imprécision peuvent également être prises en compte dans le cadre de la logique floue quand on utilise une connaissance issue d'un expert humain.

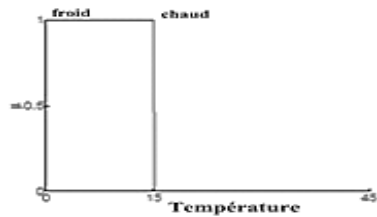


Figure 3.4 Deux sous-ensembles selon la logique classique

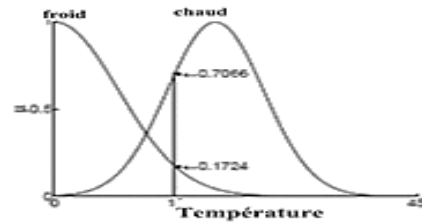


Figure 3.5 Deux sous-ensembles selon la logique floue

**Logique floue :** La logique floue consiste à attribuer intuitivement un degré auquel un prédicat flou peut convenir à un objet. On peut ainsi définir, de la même façon que pour la théorie des ensembles ordinaires, des relations entre des ensembles flous. La différence réside principalement en ce que ces ensembles ont des données vagues, qui sont représentées par des intervalles.

**Caractéristiques :** Cette théorie permet de formaliser plusieurs aspects du langage naturel, ce qui n'était pas le cas de la logique classique, notamment des notions vagues de quantité comme "très" ou "peu", la logique floue soulève cependant des problèmes épistémologiques, comme par exemple l'idée d'une appartenance à un ensemble à un certain degré. Les éléments de base de la logique floue sont les variables linguistiques, les fonctions d'appartenance et les déductions aux inférences

**Variables linguistiques :** Une variable linguistique est un triplet  $(V, X, T_V)$ , dans lequel  $V$  est une variable définie sur un ensemble de référence  $X$ . L'ensemble  $T_V = \{A_1, A_2, \dots\}$ , fini ou infini, contient des sous-ensembles flous normalisés de  $X$ , utilisables pour caractériser  $V$ . Selon l'utilisation qui doit être faite, le nombre d'éléments de  $T_V$  est plus ou moins grand.

Exemple : Afin de décrire la température d'une pièce par une variable linguistique, on peut utiliser l'ensemble des termes suivants :  $T(V) = \{\text{froid, chaud, très chaud}\}$ . En considérant que l'univers du discours est l'intervalle  $[0, 45]$ , on peut définir les termes linguistiques comme suit : «froid» est «une température environ inférieure à 10 degrés», «tiède» est «une température d'environ 17 degrés» et «chaud» est «une température environ supérieure à 24 degrés». Ces termes peuvent être caractérisés par les fonctions d'appartenance représentées sur la figure 3.6 . N'importe quelle fonction d'appartenance, par exemple la fonction «froid», définit un nombre flou.

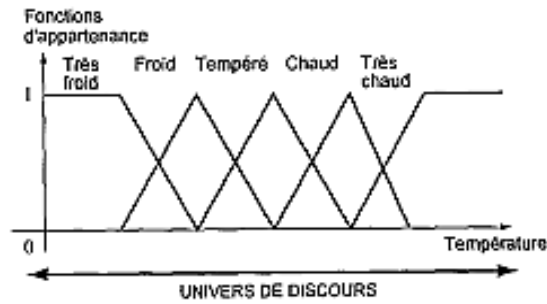


Figure 3.6 Variables linguistiques

Les ensembles flous sont décrits par leur fonction d'appartenance  $\mu_A(x) \in [0, 1]$

**Fonctions d'appartenance :**

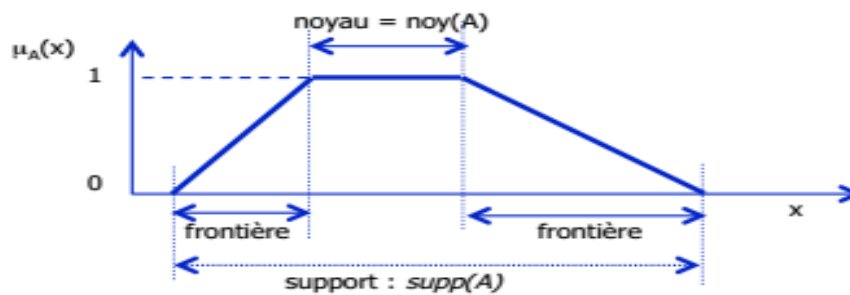
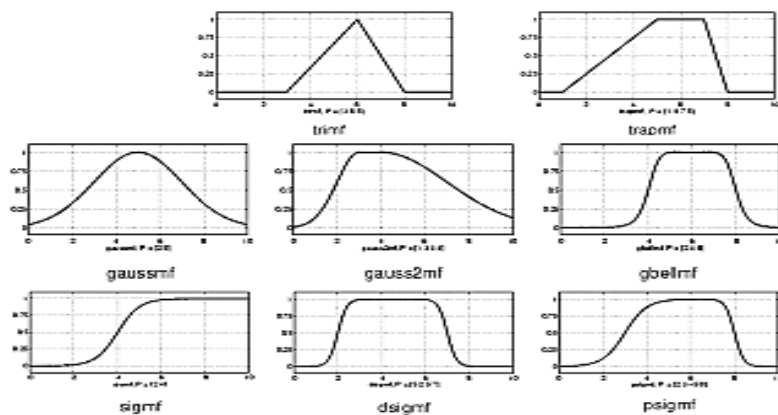


Figure 3.7 Fonction d'appartenance

Au lieu d'appartenir à l'ensemble « vrai » ou à l'ensemble « faux » de la logique binaire traditionnelle, la logique floue admet des degrés d'appartenance à un ensemble donné. Le degré d'appartenance à un ensemble flou est matérialisé par un nombre compris entre 0 et 1. Une valeur précise de la fonction d'appartenance liée à une valeur de la variable est notée  $\mu$  et appelée « facteur d'appartenance ».

Exemple : Les fonctions d'appartenance peuvent théoriquement prendre n'importe quelle forme. Les plus utilisées sont :



Triangulaire,  
rapézoïdale,  
  
Gaussienne,  
  
Sigmoides...

Figure 3.8 Formes de fonctions d'appartenance

### c. Modificateurs linguistiques

Un modificateur linguistique peut changer la forme d'un ensemble flou, ou en créer un nouveau. Il y en a de nombreux types, et il existe de nombreuses méthodes de calcul. Certains modificateurs comme très, extrêmement, un\_peu, plutôt, légèrement, intensifient ou diluent la fonction d'appartenance. D'autres, comme absolument et généralement, changent le degré de flou global. Un autre groupe de modificateurs, comme « au-dessus de », « en dessous de » et non (l'ensemble non complémentaire), restreignent le domaine d'un ensemble flou existant et d'autres encore, Comme « environ », « près de », et « alentour de », diffusent un ensemble flou existant, ou créent un ensemble flou à partir d'une valeur scalaire. Selon [Cox, 1992], en général, il n'y a pas de «principe premier» définissant précisément la manière d'opérer d'un modificateur donné.

L'action d'un modificateur est conçue de façon heuristique; il doit changer la forme d'un ensemble flou d'une manière qui soit en étroite concordance avec la sémantique.

Les modificateurs linguistiques jouent le même rôle en raisonnement approximatif que les adjectifs et les adverbes dans le langage naturel. Ils modifient la signification de l'ensemble flou auquel ils s'appliquent [Smithson, 1986].

#### Modificateur très

Un des modificateurs les plus anciens, et le plus fréquemment utilisé, est très, employé dans les requêtes comme un adjectif d'ensemble flou. Par exemple,

```
SELECT Nom FROM Société WHERE chiffre_d'affaires_annuel IS très Élevé
```

Le modificateur très intensifie la fonction d'appartenance d'un ensemble flou existant, en en prenant son carré. L'équation suivante donne la formule de transformation :

$$\mu_{\text{très.W}}(x) = (\mu_{\text{W}}(x))^2$$

Comme les valeurs d'appartenance sont dans l'intervalle [0,1], leur carré les diminue (sauf pour zéro et un). Ceci provoque une légère courbure de la fonction, vers le bas, pour les valeurs intermédiaires (les valeurs appartenant à l'ensemble]0,1[). Le modificateur très a aussi la propriété appréciable de conserver le support de l'ensemble flou. La figure 3.9 montre l'effet de très sur l'ensemble flou de base Élevé

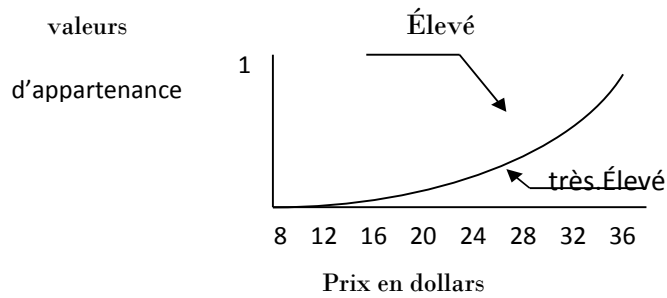


Figure 3.9 Valeurs d'appartenance

### Opérations sur les ensembles flous

Soit  $A$  et  $B$  deux ensembles flous définis dans l'univers  $U$ . Les opérations classiques de la théorie des ensembles peuvent à présent être étendues aux ensembles flous grâce aux opérateurs  $op_1$  et  $op_2$  précédemment définis, qui permettent d'obtenir leurs fonctions d'appartenance respectives

Opération	Fonction d'appartenance (générale)	Fonction d'appartenance (opérateurs de Zadeh)
Intersection	$\forall x \in U, \mu_{A \cap B}(x) = op_1(\mu_A(x), \mu_B(x))$	$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$
Union	$\forall x \in U, \mu_{A \cup B}(x) = op_2(\mu_A(x), \mu_B(x))$	$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$
Complément	$\forall x \in U, \mu_{\bar{A}}(x) = 1 - \mu_A(x)$	$\mu_{\bar{A}}(x) = 1 - \mu_A(x)$
Différence	$\forall x \in U, \mu_{A-B}(x) = op_1(\mu_A(x), \mu_{\bar{B}}(x))$ ( $car A - B = A \cap \bar{B}$ )	$\mu_{A-B}(x) = \min(\mu_A(x), \mu_{\bar{B}}(x))$

Tableau 3.1 Opérations sur les ensembles flous

### Déductions aux inférences :

Prenons l'exemple de la conduite automobile à l'approche d'un carrefour contrôlé par des feux tricolores : Si vous êtes un conducteur normal, le processus de conduite met en œuvre les quelques règles élémentaires suivantes :

si le feu est rouge;	si ma vitesse est élevée,	et si le feu est proche	alors je freine fort.
si le feu est rouge,	si ma vitesse est faible,	et si le feu est loin,	alors je maintiens ma vitesse.
si le feu est orange,	si ma vitesse est moyenne,	et si le feu est loin,	alors je freine doucement.
si le feu est vert,	si ma vitesse est faible,	et si le feu est proche	alors j'accélère.

Tableau 3.2 quelques règles du processus de conduite

Prenons la première des 4 règles. Elle est typiquement "floue". Si nous essayons de la transposer dans un monde plus mathématique "moins flou", cela donnerait (en forçant "légèrement" le trait): Si le feu est rouge, si ma vitesse dépasse 85,6 km/h et si le feu est à moins de 62,3 mètres, alors j'appuie sur la pédale de frein avec une force de 33,2 Newtons ! Et bien la logique floue formalise le monde exactement comme le fait votre cerveau. Elle apprécie les variables d'entrées de façon approximative (faible, élevée, loin, proche), fait de mêmes pour les variables de sorties (freinage léger ou fort) et édicte un ensemble de règles permettant de déterminer les sorties en fonction des entrées.

**Applications (Système expert) :** Parmi les applications de la logique floue on retrouve les systèmes experts flous. Ces systèmes peuvent être composés de plusieurs couches de règles avec des opérateurs ET et Ou. Les variables de sortie résultats de l'exécution d'un nombre de règles peuvent être appliquées aux entrées d'autres règles. Les applications de ces systèmes sont diverses : les contrôles linéaires et non linéaires, l'aide à la décision, la reconnaissance de forme, le système financière, la recherche opérationnelle, etc.

### **Langages d'interrogation floue**

La base formelle par l'extension de l'algèbre relationnelle étant établie, il est nécessaire, pour la mettre en pratique, de définir un langage de requête. Pour les bases de données actuelles, SQL est le plus souvent utilisé, c'est pourquoi les principaux travaux pour développer de tels langages ont étendu SQL pour le support des requêtes floues.

Une classification détaillée des langages de requêtes flexibles (basés ou non sur SQL) a été établie par [Rosado et al., 2006]. Elle divise principalement ces langages en deux catégories selon qu'ils s'adressent à des bases de données floues ou à des bases de données classiques.

Parmi ces langages, on retiendra les deux plus connus [Galindo, 2008] :

**FSQL :** Une équipe de l'université de Málaga dirigée par J. Galindo a mis au point ce langage qui est une extension de SQL permettant d'écrire des conditions flexibles dans les requêtes adressées à une base de données relationnelle floue.

**SQLf :** Le projet Pilgrim à l'université de Rennes 1 a parmi ses objectifs de développer le langage SQLf [Bosc, 1995].

### **Imprécision dans les requêtes**

Pour introduire cette imprécision dans les requêtes, plusieurs approches sont proposées. Nous en retiendrons trois. La première divise chaque requête en deux parties : une sélection booléenne ordinaire et une composante permettant le classement des résultats (cas notamment des requêtes à préférences comme celles du langage Preference SQL. Dans la deuxième, les conditions imprécises sont traduites en conditions booléennes se rapportant à des intervalles au lieu de valeurs uniques. Quant à la troisième approche, l'interprétation des conditions imprécises par des ensembles flous. Pour le classement des

résultats, elle utilise, comme précédemment, une notion de distance des n-uplets par rapport à la condition imprécise. Cependant, les ensembles flous constituent un outil plus précis et naturellement plus adéquat. [Bosc, 1992]

### **3.2.1.2 Extension de la logique floue (CWP)**

Le père de la logique floue, Lotfi Zadeh, a proposé une nouvelle direction dans cette ligne de recherche, à savoir étendre la logique floue au calcul avec mots et perceptions (CWP). Le domaine a été introduit dans l'article de fond de Zadeh «De l'informatique avec les nombres à l'informatique avec les mots - de la manipulation des mesures à la manipulation des perceptions» et développé dans des articles ultérieurs. [Mendel & al., 2010] CWP fournit un cadre pour développer des systèmes computationnels ayant la capacité de calculer avec la signification des expressions NL, c'est-à-dire avec la capacité de calculer avec des descriptions imprécises du monde de la même manière que les humains le font. Dans CWP, un granule est un amas d'éléments qui sont rassemblés par indistinction, similarité, proximité ou fonctionnalité [Zadeh, 1979]. La limite d'un granule est floue. Le flou des granules nous permet de modéliser la façon dont les concepts humains sont formés, organisés et manipulés dans un environnement d'imprécision, d'incertitude et de vérité partielle [Zadeh, 1997]. Un granule sous-tend le concept d'une variable linguistique [Zadeh, 2008]. Une variable linguistique est une variable dont les valeurs sont des mots ou des phrases en LN [Zadeh, 1975]. L'idée de quantificateurs flous linguistiques a été introduite par [Zadeh, 1983]. Dans un article récent, écrit par Zadeh avec d'autres chercheurs Mendel, Lawry et Zadeh, [Mendel & al., 2010], ils distinguent deux niveaux de développement de CWP:

Au niveau 1, les porteurs d'informations sont des nombres et des mots, par exemple, "  $X = 5$  ", "  $X$  est petit " et " Si  $X$  est petit alors  $Y$  est grand ". Les objets de calcul peuvent être considérés comme des instructions d'affectation, qui affectent des valeurs à des variables. Le niveau 1 n'exige aucune connaissance de la sémantique des langues naturelles. Au niveau 2, les supports d'information sont des nombres, des mots et des propositions ou des collections de propositions tirées d'une langue naturelle comme «La plupart des Suédois sont grands» et «Il est très improbable qu'il y ait une baisse significative du prix de pétrole dans un proche avenir ". Le niveau 2 exige une certaine compréhension de la façon dont la signification d'une proposition peut être composée à partir des significations de ses constituants.

Par conséquent, le défi actuel consiste à étendre le paradigme existant avec de nouvelles capacités pour modéliser la signification des expressions linguistiques complexes. Le concept de résumé linguistique flou peut être considéré comme une contribution à cet objectif. Le résumé linguistique flou a la forme générale ( $w$ , "  $Q$  objets dans la base de

données satisfaisant D sont S ") où Q est appelé le quantificateur, D est une propriété floue caractérisant l'ensemble flou de référence d'objets, S est une autre propriété floue des objets appelés le résumé, et  $w \in [0, 1]$  est le degré de validité de la clause linguistique pour représenter la signification dans le contexte spécifique. L'une des limites de ces approches est qu'elles représentent des perceptions directement liées aux données et qu'elles ne permettent pas d'interpréter les perceptions de second ordre.

### 3.2.1.3 Evaluation de requêtes flexibles à base de système expert

**a. Système expert :** Un système expert est un logiciel qui reproduit le comportement d'un expert humain accomplissant une tâche intellectuelle dans un domaine précis. On peut souligner les points suivants:

- les systèmes experts sont généralement conçus pour résoudre des problèmes de classification ou de décision (diagnostic médical, prescription thérapeutique, régulation d'échanges boursiers, ...),
- les systèmes experts sont des outils de l'intelligence artificielle, c'est-à-dire qu'on ne les utilise que lorsqu'aucune méthode algorithmique exacte n'est disponible ou praticable,
- un système expert n'est concevable que pour les domaines dans lesquels il existe des experts humains. Un expert est quelqu'un qui connaît un domaine et qui est plus ou moins capable de transmettre ce qu'il sait : ce n'est par exemple pas le cas d'un enfant par rapport à sa langue maternelle.

Un système expert est composé de deux parties indépendantes :

\* une base de connaissances elle-même composée d'une base de règles qui modélise la connaissance du domaine considéré et d'une base de faits qui contient les informations concernant le cas que l'on est en train de traiter. Pour le cas du modèle d'estimation de risque, il y a deux experts, qui déterminent le risque associé à un projet, compte tenu des valeurs prévues pour la durée, le budget et l'effectif.

La première pile [Bosc, 1995] établie par Bill Smith, un chef de projet confirmé, contient quatre règles : si durée projet est longue alors risque est acceptable.

si durée\_projet est assez\_longue alors risque est important. si budget\_projet est élevé alors risque est important.

si effectif\_budget est confortable alors risque est élevé.

La seconde pile, proposée par Mary Williams, directrice de recherche de la division Marketing, contient trois règles :

si durée\_projet est longue alors risque est important si budget\_projet est élevé alors risque est acceptable si effectif\_budget est confortable alors risque important

\* un moteur d'inférences capable de raisonner à partir des informations contenues dans la base de connaissance, de faire des déductions, etc.

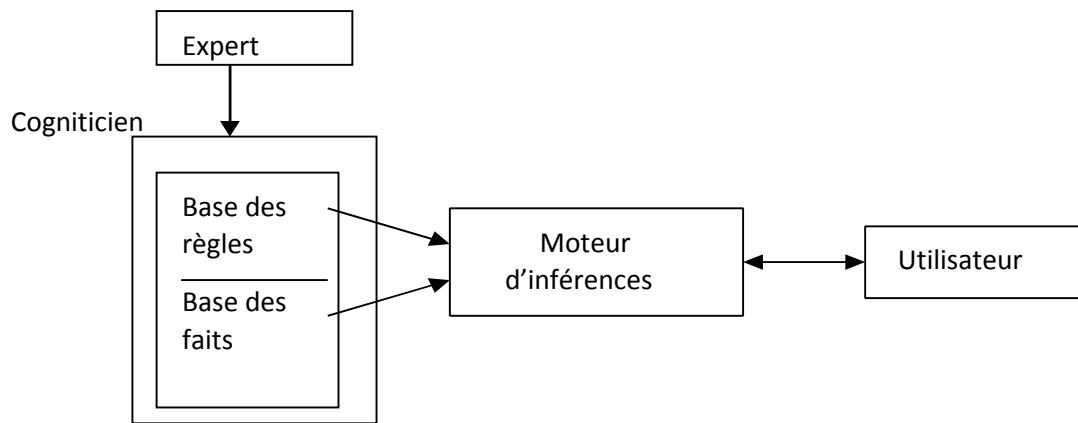


Figure 3.10 structure d'un système expert

Le rôle du cogniticien est de soutirer leurs connaissances aux experts du domaine et de traduire ces connaissances dans un formalisme se prêtant à un traitement automatique, c'est-à-dire en règles. Ces deux tâches sont aussi délicates l'une que l'autre. En effet, un expert est la plupart du temps inconscient de la majeure partie de son savoir ; et s'il arrive à en exprimer une partie, c'est souvent sous une forme qui ne se laisse pas facilement formaliser. De plus l'extraction de connaissances peut avoir un effet perturbant sur l'expert : il est bien connu que si l'on demande à un maître aux échecs de réfléchir à sa manière de jouer, on observera une baisse de ses performances dans les jours ou les semaines qui suivent cet effort d'introspection.

Le système expert est souvent complété par des interfaces plus ou moins riches permettant un dialogue avec les utilisateurs, l'idéal étant une interface en langage naturel.

#### b. Requête SQL Classique & requête floue

**Principes de requête SQL Classique :** Le langage de requêtes structurées, ou SQL (Structured Query Language) est un langage de 4<sup>ème</sup> génération. Il comprend un jeu complet d'instructions pour interroger et mettre à jour une base de données relationnelle.

#### **Evaluation de requêtes flexibles : Logique et sémantique des requêtes**

À l'évidence, ce genre de langage de haut niveau est un puissant outil d'élaboration de requêtes pour les bases de données relationnelles. Néanmoins, bien que SQL puisse facilement exprimer le mécanisme d'une requête, il n'est pas capable d'en appréhender la sémantique [Cox & al, 1998]. Il permet de regrouper et segmenter des ensembles de données de bien des manières, mais les lignes de partage sont toujours clairement définies. Ces frontières, à l'instar des brusques bifurcations dans les systèmes experts, peuvent nous contraindre à prendre des décisions arbitraires concernant les appartenances de concepts

qui sont intrinsèquement imprécis. Cette technique d'interrogation souffre les mêmes limitations et rencontre les mêmes problèmes que les systèmes à base de connaissances conventionnels utilisant des règles non floues.

En SQL, la logique de la requête se situe dans la partie WHERE. Le résultat de cette expression, comme indiqué précédemment, doit être égal à VRAI pour qu'une ligne puisse être sélectionnée. Pourtant la logique d'une expression SQL n'est pas aussi binaire que nombre de personnes le pensent. Comme un champ d'une colonne d'une base de données peut avoir la valeur VIDE (c'est-à-dire, en fait, qu'aucune valeur n'a été attribuée), SQL fonctionne selon une logique trivalente : vrai, faux et inconnu. Une valeur manquante (plus précisément, égale à VIDE) donnera "inconnu" comme résultat logique, du fait de la prépondérance accordée à VIDE devant VRAI et FAUX.

L'objectif essentiel de l'utilisation de la logique floue est ici d'ajouter une structure "intentionnelle" à la requête elle-même. On peut utiliser des prédicats flous pour élargir les spécifications dans l'instruction WHERE, de façon à inclure la notion de frontières approximatives. Cela permettra d'exprimer ce que nous souhaitons obtenir, plutôt que ce que nous allons nécessairement obtenir, du fait de la nature arbitrairement limitative de l'instruction WHERE classique. L'intentionnalité est exprimée en améliorant la sémantique de la requête, de manière à mettre en évidence sa signification réelle.

L'interrogation flexible de BD peut se faire dans plusieurs domaines où la souplesse et la flexibilité sont les caractéristiques essentielles des données interrogées et mises à jour.

### 3.2.1.4 STDS : Schéma de translation dirigée par la syntaxe

Un STDS est une grammaire qui engendre un langage, et qui possède pour chaque règle de production un élément de translation attaché. Chaque fois qu'on appelle une production pour faire une dérivation, on appelle aussi la translation engendrée par l'élément attaché. L'exemple suivant fait une simple translation:

Soit  $(x, xR)$  une translation où  $xR = y$  est l'image miroir de  $x$  /  $x \in \{0,1\}^*$

- Soit la grammaire  $G = (\Sigma, N, P, S)$

$\Sigma = \{0,1\}^*$  ;  $S = \{S\}$  ;  $N = \{S\}$

Soit les règles de production :

$P : S \rightarrow 0S \dots\dots\dots S \rightarrow S0$

$S \rightarrow 1S \dots\dots\dots S \rightarrow S1$

$S \rightarrow \lambda \dots\dots\dots S \rightarrow \lambda$

- Soit la chaîne d'entrée  $x = 001$  et le schéma de translation:

$(S, S) \rightarrow (0S, S0) \rightarrow (00S, S00) \rightarrow (001S, S100) \rightarrow (001, 100) \rightarrow (x, y)$

$Y = 100$  est le translate de  $x = 100$  et  $y = xR$

Plus formellement, un STDS est un 5-uplets  $T = (N, \Sigma, \Delta, R, S)$

- $N$  : ensemble fini de symboles non terminaux

- $\Sigma$  : ensemble fini d'alphabet d'entrée
- $\Delta$  : ensemble fini d'alphabet de sortie
- $S$  : Axiome de la grammaire (non terminal)
- $R$  : ensemble fini de production de type  $A \rightarrow \alpha, \beta$  avec  $\alpha \in (NY\Sigma)^*$  et  $\beta \in (NY\Delta)^*$ .

Les non terminaux de  $\beta$  sont une permutation de  $\alpha$ .

### 3.2.1.5 Conclusion

Dans cette section, on a présenté tout d'abord la logique classique, puis la logique floue avant d'entamer l'une des applications de cette dernière qui est le système expert. Il sera l'objet de notre deuxième proposition. On a aussi vu l'extension de la logique floue proposé par Pr. Zadeh, le père de la logique floue et présenté dans [Mendel & al., 2010] montarnt une nouvelle direction de l'intelligence artificielle. Cette extension est appelée CWP, qui est à l'origine de l'ontologie GLMP qu'on a considéré comme point de départ de notre troisième proposition ontologique: le modèle GLMR. On a enfin présenté l'évaluation de requête flexible qui est l'axe principal de cette these. On s'est limité ici sur l'évaluation à base de système expert. Cette section présente donc les principaux concepts qui ont un rapport avec notre deuxième proposition. La section suivante présente les concepts se rapportant à la solution ontologique à savoir l'ontologie floue et la classification des ontologies de domaine (le modèle en oignon).

---

## 3.2 Evaluation des BDFs utilisant un système de connaissance

### 3.2.2 Evaluation des besoins décisionnels flexibles utilisant une ontologie floue

Rien de parfait n'existerait si l'ÊTRE INFINI n'existait pas, en tant que totalité harmonieuse inchangeable, de tout ce qui est.

---

Les ontologies, à l'origine d'une branche de la philosophie qui s'intéresse à la nature et l'organisation de la réalité, correspondent à ce qu'Aristote appelait la Philosophie première, c'est-à-dire la partie de la métaphysique qui s'intéresse à l'être en tant qu'être, par opposition aux philosophies secondes qui s'intéressent à l'étude des manifestations de l'être. En informatique, la littérature fournit un tas de définitions du mot ontologie. Ces définitions, dans leur diversité, offrent des points de vues à la fois différents et complémentaires. Dans cette section, nous décrivons quelques-unes avant de présenter les composantes et langages de description des ontologies, nous permettant de mettre en oeuvre notre solution ontologique. On présente ensuite les ontologies floues et l'importance de ce concept dans l'évaluation du besoin flou.

#### 3.2.2.1 Définition d'Ontologie

##### a. Ontologie informatique

Les ontologies informatiques sont des outils qui permettent de représenter précisément un corpus de connaissances sous une forme utilisable par une machine. Elles représentent un ensemble structuré de concepts. Les concepts sont organisés dans un graphe dont les relations peuvent être des relations sémantiques et/ou des relations de composition et d'héritage (au sens objet) [Guergour, 2007].

Il existe de nombreuses définitions pour la notion de l'ontologie, parmi eux [Bensafi, 2013]:

**b. Définition de Neches et al(1991):** «une ontologie définit les termes et les relations de bases qui composent le vocabulaire d'un domaine, bien que les règles de combinaison des termes et les relations pour définir l'extension du vocabulaire». Cette définition indique en quelque sorte qu'est-ce qu'on fait pour construire une ontologie, elle identifie les termes de base et les relations entre termes, et les règles pour combiner les termes, selon cette définition, une ontologie inclut les termes qui sont définies d'une manière explicite, plus des connaissances qui peuvent être portées par les termes.

Quelques années après, vient la définition qui nous semble être la plus célèbre et la plus citée est celle de Gruber:

**c. Définition de Gruber(1993):** «une ontologie est une spécification explicite d'une conceptualisation». Le terme 'conceptualisation' réfère à un modèle abstrait d'un certain phénomène de la réalité et qui permet d'identifier les concepts pertinents de ce

phénomène. Le terme ‘explicite’ signifie que le type des concepts utilise ainsi que les contraintes sur leur emploi, sont réellement définies d'une manière claire et précise.

**d. Définition de Guarino et al(1995):** «Une ontologie est une théorie logique qui permet une spécification explicite et partielle d'une conceptualisation » . Ou le terme conceptualisation désigne l'idée de la réalité qu'un individu ou un groupe d'individus peuvent avoir.

**e. Définition de Borst (1997) :** a modifié légèrement la définition de Gruber en citant qu'une ontologie est définie comme étant « une spécification formelle d'une conceptualisation partagée ». Cette définition précise d'une part, le fait que l'ontologie doit être ‘formelle’, c'est à dire exprimée sous forme d'une logique pouvant être exploitable par une machine. D'autre part, elle doit être partagée dans la mesure où elle doit capturer des connaissances partagées entre différents individus.

**3.2.2.2 Composantes de l'ontologie :** Les ontologies rassemblent les connaissances propres à un domaine particulier. Ces connaissances sont formalisées en mettant en jeu les composants suivants: concepts, relations, axiomes et instances.

**Concepts :** Un concept peut représenter un objet, une idée, ou bien une notion abstraite [Uschold et al., 1995]. Ils sont appelés aussi classes de l'ontologie dans certain travaux. Un concept peut être divisé en trois parties : **un terme (ou plusieurs), une notion et un ensemble d'objets.**

**Le terme (ou bien label) d'un concept** est l'expression linguistique utilisée couramment pour y faire référence. La notion désigne ce qui est appelé, au sens de la représentation des connaissances,

**L'intention** du concept. Elle contient sa sémantique qui est définie à l'aide de propriétés (relations et attributs), de règles et de contraintes.

**L'ensemble d'objets** définis par le concept forme ce qui est appelé **l'extension** du concept.

Il s'agit des objets auxquels le concept fait référence, autrement dit, de ses instances.

**Relations :** Les relations sémantiques unissent les concepts du segment analysé de la réalité entre eux et traduisent les associations existantes. Formellement, elles sont définies comme étant tout sous ensemble d'un produit de  $n$  ensembles, c'est-à-dire:  $R \subset C_1 * C_2 * \dots * C_n$ . Les ontologies généralement contiennent que des relations binaires. Le premier argument d'une relation binaire est dit domaine, alors que le deuxième argument est dit Co-domaine. Cela permet de désigner la façon dont la relation doit être lue. Ces relations sont caractérisées par un terme (voire plusieurs), une signature qui précise le nombre d'instances de concepts que la relation lie et leurs types. Elles englobent les associations suivantes : sous-classe de (généralisations spécialisation) ; partie de (agrégation ou composition) ; associée-à ; instance de ; est un; etc.

On distingue alors, les relations taxonomiques (dite aussi de subsomption) et les relations associatives.

### **Relation de subsomption**

La relation de subsomption « est-un » (is-a) a un statut particulier car elle structure la **hiérarchie ontologique**. A ce titre, elle est implicite. Un concept  $C_1$  (concept père) subsume un concept  $C_2$  (concept fils) si toute propriété sémantique de  $C_1$  est également une propriété sémantique de  $C_2$  et si  $C_2$  est plus spécifique que  $C_1$ . Ainsi, l'extension d'un concept est forcément plus réduite que celle de son concept père. Son intention est par contre plus riche. Selon [Baneyx, 2007], la relation de subsomption est définie dans la littérature de plusieurs manières. :

**Définition intensionnelle** : un concept  $C_1$  subsume un concept  $C_2$  si tout individu décrit par  $C_2$  l'est aussi par  $C_1$ , autrement dit si l'ensemble des propriétés d'un individu dont la description est définie par  $C_2$  contient l'ensemble des propriétés spécifiées par  $C_1$ .

**Définition extensionnelle** : un concept  $C_1$  subsume un concept  $C_2$  si l'ensemble des individus dénotés par  $C_1$  contient l'ensemble des individus dénotés par  $C_2$ .

**Définition logique** : un concept  $C_1$  subsume un concept  $C_2$ , si être un individu décrit par  $C_2$  implique être un individu décrit par  $C_1$ . La relation de subsomption n'est pas la seule relation qui permette de structurer la hiérarchie ontologique, la relation de méronymie, « partie-tout » (part of) est souvent utilisée.

L'**héritage multiple** est une propriété qui peut être définie sur la relation de subsomption : un concept d'une ontologie peut avoir plusieurs pères par la relation de subsomption.

L'héritage multiple implique que le concept hérite des propriétés de tous ses pères.

**Axiomes** : Les axiomes permettent de modéliser des assertions acceptées comme vraies, à propos des abstractions du domaine traduites par l'ontologie. Leur inclusion dans une ontologie peut avoir plusieurs objectifs: interviennent dans la définition des significations des composants d'ontologie, les contraintes sur les valeurs des attributs, les arguments de relations et dans l'inférence de nouvelles informations, etc. D'après [Gómez-Pérez et al., 2004], Les axiomes formels sont utilisés pour vérifier la consistance de l'ontologie.

**Instances (ou individus)** : Elles constituent la définition extensionnelle de l'ontologie. Ils représentent les éléments singuliers véhiculant les connaissances à propos du domaine.

### **Les types d'ontologies [Khalfi, 2009]**

**Ontologie de haut niveau** : Une ontologie de haut niveau décrit des concepts très généraux comme l'espace, le temps, la matière, les objets, les événements, les actions, etc. Ces concepts ne dépendent pas d'un problème ou d'un domaine particulier, et doivent être, du moins en théorie, consensuels à de grandes communautés d'utilisateurs.

Exemples d'ontologies de haut niveau : Dolce ou Sumo.

**Ontologie de domaine :** Une ontologie de domaine décrit le vocabulaire ayant trait à un domaine générique (Ex. : l'enseignement, la médecine...), notamment en spécialisant les concepts d'une ontologie de haut niveau.

**Ontologie de tâche :** Une ontologie de tâche décrit le vocabulaire concernant une tâche générique (ex. : enseigner, diagnostiquer...), notamment en spécialisant les concepts d'une ontologie de haut niveau [Guarino, 1998].

Certains auteurs emploient le nom « ontologie du domaine de la tâche » pour faire référence à ce type d'ontologie.

**Ontologie d'application :** L'ontologie d'application contient des concepts dépendants d'un domaine et d'une tâche particuliers, qui sont généralement subsumés par des concepts de ces deux ontologies. Ces concepts correspondent souvent aux rôles joués par les entités du domaine lors de l'exécution d'une certaine activité [Goh et al., 1999]. Il s'agit donc ici de mettre en relation les concepts d'un domaine et les concepts liés à une tâche particulière, de manière à en décrire l'exécution (ex : apprendre les statistiques, effectuer des recherches dans le domaine de l'astronomie, etc.)

**3.2.2.3 Cycle de vie d'une ontologie :** Puisque les ontologies sont destinées à être utilisées comme des composants logiciels dans des systèmes répondant à des objectifs opérationnels différents, leur développement doit s'appuyer sur les mêmes principes que ceux appliqués en génie logiciel. Ainsi, les ontologies doivent être considérées comme des objets techniques évolutifs et possédants un cycle de vie qui nécessite d'être précisé. Dans ce contexte, les activités liées aux ontologies sont d'une part, des activités de gestion de projet (planification, contrôle, assurance qualité) et d'autre part, des activités de développement (spécification, conceptualisation, formalisation); s'y ajoutent des activités transversales de support telles que l'évaluation, la documentation, la gestion de la configuration [Blazquez, 1998]. Un cycle de vie inspiré du Génie Logiciel est proposé dans [Baneyx, 2007] basé sur [Dieng, 2001] et [Gandon, 2006].

Il comprend une étape initiale de détection et de spécification des besoins qui permet notamment de cerner précisément le domaine de connaissances, une étape de conception qui se subdivise en trois phases, une étape de déploiement et de diffusion, une étape d'utilisation, une étape incontournable d'évaluation et enfin une sixième étape consacrée à l'évolution et à la maintenance du modèle. Après chaque utilisation significative, l'ontologie et les besoins doivent être réévalués et l'ontologie peut être étendue et si nécessaire en partie reconstruite. La validation du modèle de connaissances est au centre du processus et se fait de manière itérative [Baneyx, 2007].

Fernandez et ses collègues [Fernández-López, 1997] insiste sur le fait que les activités de documentation et d'évaluation sont nécessaires à chaque étape du processus de construction. L'évaluation précoce permettant de limiter la propagation d'erreurs. Le

processus de construction peut être intégré au cycle de vie d'une ontologie comme l'indique la figure suivante

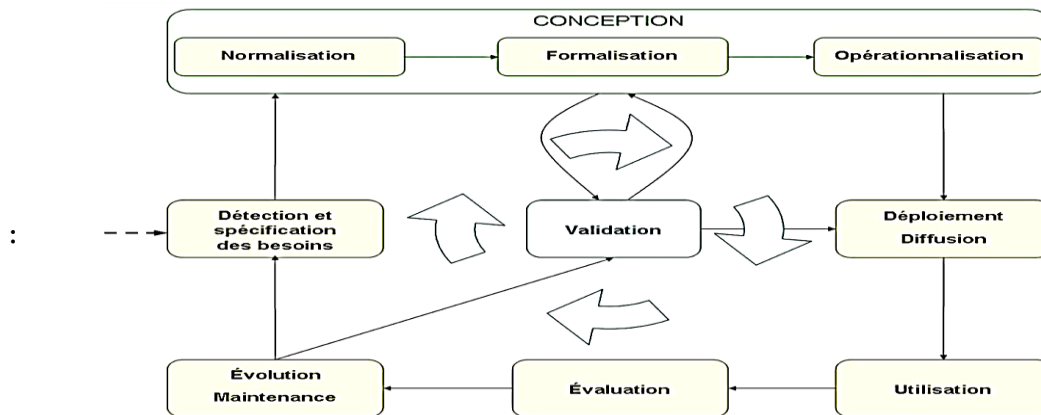


Figure 3.11 Cycle de vie d'une ontologie

### 3.2.2.4 Les langages de description des ontologies

Pour rendre les ontologies compréhensibles par les machines, il faut avoir un langage qui décrit ces ontologies. Dans ce contexte, plusieurs langages ont été développés, nous en citons RDFS et OWL.

#### a. RDFS (Resource Description Framework Schéma)

RDF Schéma ou RDFS est un langage extensible de représentation des connaissances. Il appartient à la famille des langages du Web sémantique publiés par W3C (World Wide Web Consortium). RDFS fournit des éléments de base pour la définition des ontologies ou des vocabulaires destinés à structurer des ressources RDF. La première version de RDFS a été proposée en Mars 1999, et la recommandation finale a été publiée par W3C en février 2004. Les principaux composants de RDFS sont intégrés dans un langage d'ontologies plus expressif, OWL. Parmi les principales caractéristiques de RDFS, nous citons : Classes et sous-classes :

- `rdfs:Class` : permet de déclarer une ressource RDF comme une classe pour d'autres ressources.

Un exemple simple de `rdfs:Class` est `Personne`. Une instance de la classe `Personne` est une ressource liée à cette classe en utilisant un prédicat `rdf:type`. La définition de `rdfs:Class` est récursive. `rdfs:Class` est la classe de toutes les classes [Mazziéri, 2004].

- `rdfs:subClassOf` : permet de définir des hiérarchies de classes. Par exemple, « Tout Agent est une personne » : `rdfs:classe Personne rdfs:subClassOf Agent`.

- Propriétés : RDFS précise la notion de propriété définie par RDF en permettant de donner un type ou une classe au sujet. Pour cela, RDFS ajoute les notions de « domain », correspondant au domaine de définition d'une fonction, et « range », son ensemble d'arrivée :

- `rdfs:domain` : définit la classe des sujets liée à une propriété.

- `rdfs:range` : définit la classe ou le type de données des valeurs de la propriété.

Avec RDF Schema, nous pouvons définir les concepts de l'ontologie et les relations entre ces concepts. Cependant, il est impossible de raisonner sur ces définitions car la sémantique reste limitée. Ce manque de sémantique est expliqué par l'absence des moyens qui permettent d'exprimer des contraintes sur les données.

De plus, nous remarquons que RDF Schéma ne permet pas, au niveau des propriétés, de distinguer entre une simple propriété (comme Âge et Adresse) et une propriété objet qui définit une relation entre deux concepts. Nous pouvons conclure alors que RDF Schéma n'est pas un langage complet de description d'ontologies. C'est ce manque de sémantique qu'OWL vise de combler par l'apport d'un vocabulaire plus riche.

### **b. OWL (Ontology Web Language)**

Ontologie Web Langage (OWL) est, comme RDFS, un langage profitant de l'universalité syntaxique de XML (eXtensible Mark up Langage). Fondé sur la syntaxe de RDF/XML, OWL offre un moyen de décrire des ontologies Web : Il est justement un langage d'ontologies. Si RDF et RDFS apportent à l'utilisateur la capacité de décrire des classes (avec des constructeurs) et des propriétés, OWL intègre, en sus, des outils de comparaison des propriétés et des classes : identité, équivalence, contraire, cardinalité, symétrie, transitivité, disjonction, etc. Ainsi, OWL offre aux machines une plus grande capacité d'interprétation du contenu Web grâce à un vocabulaire plus large et à une vraie sémantique formelle. OWL se compose de trois sous langages d'expressivité croissante : OWL Lite, OWL DL et OWL Full.

La syntaxe du langage OWL est fondée sur RDF/XML, ce qui nous explique la présence de tous les éléments du langage RDF Schema. D'une façon générale, les améliorations apportées par OWL représentent surtout des éléments qui permettent de mieux exprimer la valeur sémantique de l'ontologie : les éléments qui expriment des restrictions. Parmi les améliorations d'OWL, nous trouvons :

- La notion de restriction (owl : restriction) qui permet de définir une contrainte,
- La notion de cardinalités (owl : MinCardinality et owl : MaxCardinality) qui permet de définir les cardinalités minimale et maximale, et
- La notion de propriétés d'objet (owl : ObjectProperty) qui permet de définir les relations entre les concepts.

Comme en RDF Schema, nous pouvons définir avec OWL tous les concepts de l'ontologie et les relations entre ces concepts. De plus, nous pouvons éviter la majorité des insuffisances de RDF Schema. Nous pouvons conclure qu'OWL représente une meilleure solution pour la description des ontologies comportant plusieurs contraintes sémantiques. Il existe actuellement des moyens adéquats pour décrire les ontologies aux différents agents et machines : RDF Schéma pour les ontologies simples, ou l'un des sous langages

d'OWL pour les ontologies de plus en plus complexes et qui portent plusieurs contraintes sémantiques. Ces moyens, malgré leur jeunesse relative, sont déjà source d'un réel enthousiasme pour atteindre l'objectif ultime : l'opérationnalisation d'ontologies [Ghorbel & al., 2008].

### **3.2.2.5 Ontologies floues**

Les ontologies modélisent donc la réalité à l'aide de ses différents composants. Cependant, le monde réel comporte plusieurs autres composants que l'on ne peut pas présenter par les ontologies précises. Afin de remédier à ce problème, on peut penser à l'application des notions de la logique floue aux ontologies, ce qui donne des ontologies floues. Ceci permettrait aux ontologies, dont le rôle, rappelons-le, consiste à clarifier les domaines d'étude, d'éliminer les ambiguïtés qui peuvent exister dans l'univers de discours et de rapprocher leurs modèles de la perception humaine des choses. En effet, les ontologies floues doivent permettre:

- de modéliser la représentation humaine des connaissances,
- de modéliser l'incertitude et l'imprécision qui existent dans les domaines d'étude,
- d'améliorer les performances des systèmes.

#### **Définition d'une ontologie floue**

Les ontologies floues sont souvent définies en se basant sur la logique de description (LD) [Carbonell & al. , 2007], mais pas sur leurs composants et sur la différence de ces derniers par rapport à ceux des ontologies précises. Cependant, l'utilisation de la logique floue dans les ontologies floues suscite de plus en plus d'intérêt de la part de la communauté scientifique. Ses apports sont montrés par plusieurs travaux pour différents domaines [Ghorbel & al. , 2008]. Les auteurs de [Zhai & al. , 2008] essayent de définir les ontologies floues dans un modèle ontologique flou en se basant sur la logique floue, sans distinction entre composants précis et composants flous. Les composants d'ontologies floues (i.e., concepts flous et relations floues) sont souvent définis en se basant sur la logique floue. Contrairement à l'auteur de [Straccia, 2006] qui se limite à définir un composant flou à partir de ses instances, nous visons à définir un composant flou en se basant sur l'intégration des incertitudes et des imprécisions à la définition d'un composant précis.

Nous voyons que la définition d'une ontologie floue se base donc sur l'intégration de la logique floue à la définition des ontologies précises pour représenter les incertitudes et les imprécisions. Nous pouvons dire qu'une ontologie floue est une ontologie [Gruber, 2008] qui possède, en plus des composants de base (i.e., concepts, relations sémantiques, relations conceptuelles, axiomes et instances), de nouveaux composants conformes à la logique floue. Ces composants assurent la représentation des parties floues d'un univers de discours.

### **a. Intérêt du domaine des ontologies floues**

L'utilisation de la logique floue et des ontologies floues suscite de plus en plus d'intérêt de la part de la communauté scientifique. Leurs apports sont montrés par plusieurs travaux pour différents domaines. Dans [Sanchez & al., 2006], les auteurs montrent l'apport de la logique floue, en général, et des ontologies floues, en particulier, pour la conception d'un système intelligent de recherche d'informations dans le domaine du Web sémantique. Pour mieux répondre à une interrogation floue du Web, ce système transforme les requêtes des utilisateurs en requêtes floues avant de les traiter. Dans [Gaëlle & al., 2007], les auteurs mettent en relief l'apport des ontologies floues dans l'annotation des tableaux. Ils ont utilisé les idées développées par Gagliardi [Gagliardi & al., 2005] concernant l'annotation de tableaux guidée par une ontologie, sur la base d'égalités de mots entre les termes de l'ontologie et ceux du Web. Ils proposent deux méthodes de traitement selon que les données sont numériques ou symboliques. Une annotation floue est utilisée pour les données symboliques. Dans [El Sayed & al., 2007], les ontologies floues sont utilisées pour traiter le problème de recherche d'informations sur des documents hétérogènes contenant du texte et des images. Les auteurs ont montré comment une ontologie floue pourrait être utilisée pour permettre une recherche conceptuelle par le contenu.

**b. Composants d'une ontologie floue** Une ontologie floue comprend alors des concepts exacts et des concepts flous, des relations exactes et des relations floues, des axiomes et des instances.

Les concepts et les relations ont les mêmes définitions et jouent les mêmes rôles que dans une ontologie classique. Nous ne trouvons des changements que sur les concepts flous, les relations floues, les axiomes et les instances.

#### **i) Les concepts flous**

Dans une ontologie floue, un concept flou est considéré comme un ensemble flou. Ses instances possèdent ainsi des degrés d'appartenance prenant leur valeur dans l'intervalle  $[0,1]$ . Ces degrés sont calculés conformément aux formules de calcul relatives à la fonction d'appartenance définie pour le concept [Straccia, 2006]. Autrement dit, un concept flou est un concept qui possède une propriété qui peut être représentée par une variable linguistique (telle que présentée par [Zadeh, 1965]).

**Exemple :** une ontologie floue avec des concepts flous

Le schéma ontologique de la figure suivante comporte trois concepts exacts (Personne, Homme et Femme) et quatre concepts flous (Homme-Jeune, Homme-Agé, Femme-Jeune et Femme-Agée).

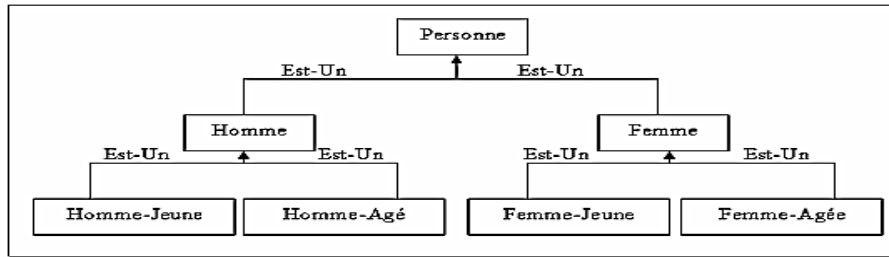


Figure 3.12 Exemple d'une ontologie floue avec des concepts flous

### Les concepts exacts

**Personne** : il représente l'ensemble des personnes. Ses propriétés sont : Nom, Prénom, Adresse et Âge. **Homme** : il représente l'ensemble des hommes. Il est une spécialisation du concept «Personne», **Femme** : il représente l'ensemble des femmes. Il est aussi une spécialisation du concept « Personne ».

### ii) Relations floues

Une relation floue est aussi définie comme un ensemble flou. Ses instances possèdent ainsi des degrés d'appartenance prenant leur valeur dans l'intervalle [0,1]. Ces degrés sont calculés conformément aux formules de calcul relatives à la fonction d'appartenance définie pour la relation.

**Exemple** : une ontologie floue avec des relations floues

Le schéma ontologique de la figure 3.13 représente deux concepts exacts (Aéroport et Hôtel) et deux relations floues qui relient entre ces concepts (Proche-De et Loin-De). Dans cet exemple, l'utilisateur de l'ontologie doit préciser les degrés d'appartenance des relations « Proche-De » et « Loin-De » lors de la déclaration des instances.

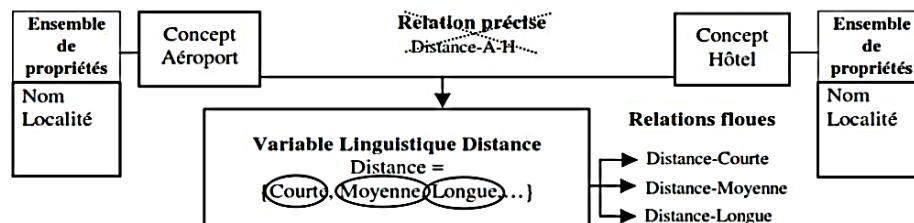


Figure 3.13 Exemple d'instances de relations floues

**Les axiomes** : A part leurs utilisations classiques, les axiomes servent à exprimer les formules de calcul des fonctions d'appartenance.

**Exemple** : Pour le schéma ontologique de la figure 3.13 nous devons déclarer des axiomes pour définir les formules de calcul des degrés d'appartenance des différents concepts flous. Les axiomes pour déclarer les formules de calcul des degrés d'appartenance du concept flou « Homme-jeune »

$$\text{Degré\_Appartenance} = (\text{Age} - 12) / (18 - 12) \text{ Si Age} \in [12 ; 18]$$

$$\text{Degré\_Appartenance} = 1 \text{ Si Age} \in ]18 ; 30]$$

$$\text{Degré\_Appartenance} = (-\text{Age} + 40) / (40 - 30) \text{ Si Age} \in ] 30 ; 40]$$

$$\text{Degré\_Appartenance} = 0 \text{ Ailleurs}$$

**Les instances :** Dans les ontologies floues, nous trouvons deux types d'instances : les instances des concepts et des relations exactes et les instances des concepts et des relations floues.

➤ Les instances des concepts et des relations exacts sont des réalisations « exactes » des éléments d'une ontologie floue. Elles ne sont pas caractérisées par des degrés d'appartenance à des ensembles flous : ce sont des instances ordinaires.

➤ La construction d'une ontologie floue

➤ Une ontologie floue possède des composants complexes et délicats. Sa construction n'est pas une affaire aisée, compte tenu de la complexité des notions utilisées. C'est pour cela qu'une méthode claire de conception d'une ontologie floue s'impose. Dans la littérature nous ne trouvons pas une telle méthode. Mais, nous pouvons la définir en s'inspirant des méthodologies de conception d'ontologies classiques.

➤ Nous voyons que la conception d'une ontologie floue doit passer les mêmes étapes (évaluation des besoins, conceptualisation, ontologisation et opérationnalisation) de la conception d'une ontologie classique, mais avec des travaux supplémentaires.

**L'évaluation des besoins :** A part le fait que l'équipe de travail doit fixer les objectifs attendus, déterminer ses choix et prendre les décisions nécessaires, elle doit répondre à la question suivante: Est- ce que le domaine d'étude considéré nécessite vraiment la construction d'une ontologie floue?

**La conceptualisation :** Lors de la conception d'une ontologie floue, l'ontologiste doit accorder une attention particulière aux incertitudes et imprécisions qui peuvent exister dans le domaine d'étude. Ces incertitudes et imprécisions nécessitent plus de travaux d'analyse pour qu'on puisse les recenser et les comprendre. A la fin de cette étape, on obtient une description informelle (même en langage naturel) des composants exacts (concepts et relations exacts) et flous (concepts et relations floues) de l'ontologie.

**L'ontologisation :** L'ontologiste doit maintenant définir les solutions adéquates pour représenter les incertitudes et les imprécisions trouvées. Il doit décrire formellement, à côté des concepts et des relations exacts, les concepts et les relations flous, en tant qu'ensembles flous, ainsi que leurs fonctions d'appartenance.

**L'opérationnalisation :** Les outils utilisés à l'implémentation doivent être capables de supporter tous les composants flous de l'ontologie, et surtout les formules de calcul des fonctions d'appartenance pour qu'il puisse automatiser les tâches lourdes de calcul des degrés d'appartenance.

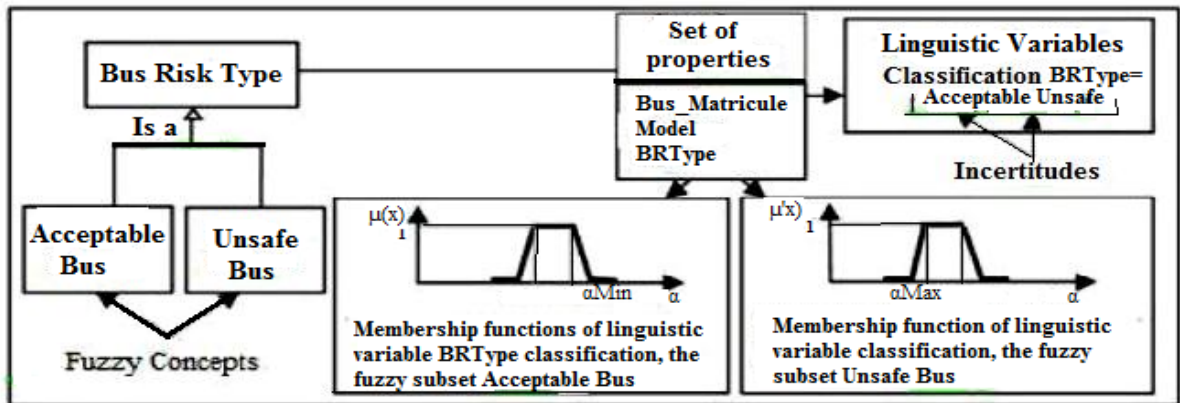


Figure 3.14 Composants de l'ontologie floue

### Classification des ontologies de domaine : le modèle en oignon

Toutes les ontologies de domaine ne sont pas identiques. Nous distinguons les trois catégories d'ontologies suivantes : [Jean, 2007]

- ✓ les Ontologies Conceptuelles Canoniques (OCC) contenant les ontologies dont les définitions ne contiennent aucune redondance. Les OCC adoptent une approche de structuration de l'information en termes de classes et de propriétés et leur associent des identifiants réutilisables dans différents langages ;
- ✓ les Ontologies Conceptuelles Non Canoniques (OCNC) contenant les ontologies contenant non seulement des concepts primitifs (canoniques), mais aussi des concepts définis (non canoniques), c'est-à-dire qui peuvent être construits à partir de concepts primitifs et/ou définis à l'exemple de OWL;
- ✓ les Ontologies Linguistiques (OL) contenant les ontologies qui définissent l'ensemble des termes qui apparaissent dans la description langagière d'un domaine. Outre des définitions textuelles, un certain nombre de relations linguistiques (synonyme, hyperonyme, hyponyme, etc.) sont utilisées pour capturer de façon approximative et semi formelle les relations entre les mots. Un exemple d'OL est WordNet.

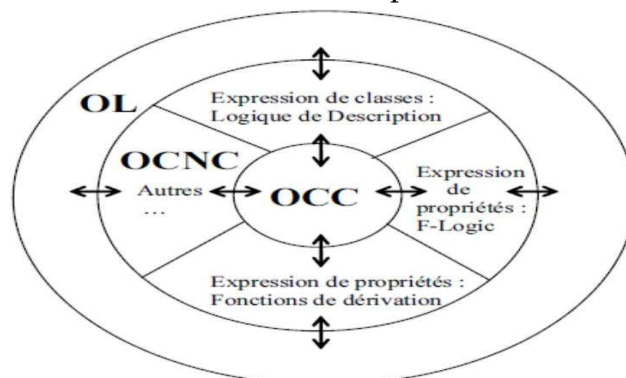


Figure 3.15 Le Modèle en oignon [Jean, 2007]

### 3.2.2.6 Conclusion (Evaluation de besoins décisionnels flexibles à base d'ontologie)

Cette section a mis en évidence l'importance des ontologies surtout pour la recherche sémantique de requêtes et pour l'évaluation du besoin flou. En fait, les ontologies ont contribué à l'apparition des moteurs de recherche sémantiques [Ben Mustapha, 2009]. Parmi ceux-ci, les moteurs de recherche contextuels se basent sur un nombre déterminé d'ontologie de domaine. La recherche dans ces moteurs est limitée à un domaine bien déterminé. Il est bien évident qu'il s'avère difficile de concevoir et de construire des ontologies relatives à tous les domaines pour être exploités par les moteurs de recherche sémantiques. Les ontologies modulaires peuvent apporter une réponse à cette limitation.

### **3.3 Conclusion du chapitre**

Ce chapitre a présenté l'ensemble des concepts et des techniques qu'on pense utiliser dans les approches proposées. Dans le chapitre suivant, on va faire une revue de littérature des trois types d'approches qu'on vise étudier pour proposer des solutions à ce problème. On va chercher les travaux existants dans ce thème de recherche.



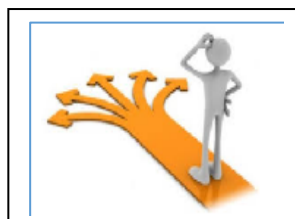


Le besoin est le maître des arts, il crée le génie.  
Proverbe latin

# CHAPITRE IV

## Evaluation de BDF- Etat d'art

<b>4.1 Evaluation des BDFs : Imprécision dans les besoins décisionnels .....</b>	<b>54</b>
<b>4.1.1 Positionnement par rapport aux approches existantes .....</b>	<b>54</b>
4.1.2 Approches ascendantes .....	54
4.1.3 Approches descendantes .....	55
4.1.4 Approches mixtes .....	55
4.1.5 Récapitulatif .....	56
4.1.6 Conclusion .....	58
<b>4.2 Revue de littérature : Evaluation des besoins décisionnels flexibles &amp; Solution structurelle</b>	<b>59</b>
4.2.1 Analyse des besoins décisionnels et traitement de l'imprécision .....	59
4.2.2 Imprécision des données dans l'entrepôt de données .....	59
4.2.3 Revue de littérature .....	59
4.2.4 Synthèse .....	63
4.2.5 Solution structurelle (modèle FASEL) .....	63
4.2.6 Conclusion .....	67
<b>4.3 Revue de la littérature : Evaluation des besoins décisionnels flexibles &amp; Système expert</b>	<b>68</b>
4.3.1 Outils décisionnels utilisant les approches flous .....	69
4.3.2 Systèmes experts, gestion des connaissances, prise de décision et ingénierie des besoins	69
4.3.3 Revue de littérature .....	70
4.3.4 Requêtage d'une base de données flexible .....	71
4.3.5 Etude Comparative .....	74
4.3.6 Conclusion .....	75
<b>4.4 Revue de la littérature : Evaluation des besoins décisionnels flexibles &amp; Ontologie</b>	<b>76</b>
4.4.1 Conception de DW basée sur les ontologies .....	78
4.4.2 Récapitulatif .....	79
4.4.3 Conclusion .....	79



---

## 4.1 Evaluation des besoins décisionnels flexibles: Revue de littérature

---

Avant d'entamer l'étude de revue de littérature pour le sujet de l'évaluation des besoins décisionnels flexibles à base de connaissance, on propose de faire une recherche bibliographique de l'évaluation et les solutions structurelles en vue de proposer à l'issu de cette étude une solution structurelle qui représente notre première proposition de solution à ce type de problème.

### 4.1.1 Positionnement par rapport aux approches existantes [Larbi & al., 2017]

Avant d'étudier différents travaux dans le domaine de l'analyse des besoins, nous présenterons d'abord les différentes revues de littératures existantes sur ce sujet. Parmi les plus récents revues de littérature, les travaux de [Sabri et al., 2013], qui ont développé une revue de littérature de l'analyse des besoins en utilisant les critères suggérés:

- (a) Complétude de l'approche analytique,
- (b) Moyens de collecte des besoins,
- (c) Modèles de formulation des besoins,
- (d) Orientation et appui aux utilisateurs sur les besoins de leur expression,
- (e) Distinction des acteurs du SID,
- (f) Classification des besoins collectés,
- (g) Traitement des besoins,
- (h) Capitalisation et réutilisation des connaissances,
- i) Documentation de l'analyse des processus d'analyse des besoins décisionnels.

Ils visaient à définir une méthode rapide d'analyse et de conception des besoins métier d'un système décisionnel dans le cadre de la réutilisation. Aussi, [Elamin, 2014] effectue une étude exhaustive des travaux proposés afin de comparer leurs méthodes en fonction d'un ensemble de critères significatifs.

Nous concluons à travers toutes ces études comparatives de différentes revues de littérature que ces travaux n'ont pas considéré l'étude de l'imprécision dans l'expression des besoins, par contre quelques travaux ont mis en évidence l'importance la sémantique des données et des liens entre ces données et les besoins pour réduire les lacunes dans la conception du DW. Ainsi, on a procédé à l'étude de quelques travaux pour les trois types d'analyse des besoins se basant sur le traitement du besoin et l'imprécision des besoins comme critère de comparaison principal.

### 4.1.2 Approches ascendantes (approches axées sur les données)

L'approche ascendante utilise uniquement le schéma des sources pour générer des schémas candidats multidimensionnels sans tenir compte, au départ, des besoins des décideurs.

Après l'étude, on peut constater que les approches ascendantes ignorent, en général, les besoins d'analyse (et par la suite l'imprécision des besoins). Elles visent principalement à concevoir des diagrammes multidimensionnels à partir de schémas sources et supposent que les schémas résultants peuvent répondre aux besoins des décideurs. Dans [Mazon & al., 2008] ont proposé d'utiliser le Framework  $i^*$  pour modéliser les objectifs et les besoins de leur approche d'architecture pilotée par le modèle (MDA) pour le développement d'entrepôts de données. L'étude de [Thenmozhi & al., 2012] propose un cadre plus complet qui implique l'expression des besoins métier dans le format du langage naturel; la conciliation des besoins et de l'ontologie globale (représentant la source) pour dériver des concepts intéressants et l'utilisation d'un algorithme de raisonnement efficace pour extraire les éléments multidimensionnels de l'ontologie.

Parmi les points forts de ce type d'approche est que c'est le plus rapide pour définir un modèle DW et le plus simple et même stable. Mais les modèles produits ne pourraient pas nécessairement répondre aux besoins des utilisateurs, ce qui est un inconvénient de l'approche.

**4.1.3 Approches descendantes (approches axées sur les besoins):** L'approche descendante ne tient compte que des besoins des décideurs. Elle s'appuie sur la spécification de ces besoins pour définir les sujets et lignes d'analyse. Après le processus de développement du diagramme multidimensionnel, la correspondance entre le schéma des résultats et la source de données est établie. Contrairement à l'approche précédente, l'approche descendante est guidée par les besoins des décideurs [Trujillo & al., 2004] et [Prat & al., 2006]. Aussi, dans ce cas on a remarqué que l'imprécision des besoins était absente de ces travaux.

Parmi les points forts de ce type d'approche, c'est la participation de l'utilisateur final qui est essentielle dans les projets décisionnels pour assurer le succès de l'utilisation du DW, mais elle présente quelques faiblesses.

**4.1.4 Approches mixtes (approches axées sur les données et les besoins):**

Cette approche combine les deux processus précédents. Aussi appelée approche hybride, elle consiste à développer un schéma multidimensionnel à partir des besoins des décideurs (habituellement formalisé par des requêtes analytiques) en appariant le schéma à la source de données. La confrontation entre les deux schémas est assurée par le concepteur. Concevoir un diagramme multidimensionnel selon l'approche mixte tient compte à la fois du schéma des sources et des besoins des décideurs. Voici une étude comparative des diverses approches hybrides récentes étudiées afin d'explorer l'incorporation de l'imprécision dans ces travaux.

Les travaux les plus récents sur l'analyse des besoins dans le processus de conception de

DW optent pour l'approche hybride compte tenu de ses avantages. Concevoir un schéma multidimensionnel selon l'approche mixte tient compte à la fois du schéma source et des besoins des décideurs. En raison de sa nature progressive, cette approche nous est apparue la plus appropriée pour notre processus qui s'adresse aux non informaticiens [Abdelhedi & al., 2014]. Là aussi, l'étude d'imprécision des besoins est absente.

#### **4.1.5 Récapitulatif**

Voici un résumé des travaux récents sur l'analyse des besoins décisionnels pour montrer que la majorité des études n'ont pas tenu compte du traitement de l'imprécision des besoins:

Travail	Automatique	Types Approches			language expression besoins	Input	Niveau Conceptionl			Utilisation ressource sémantique	Traitement Imprécision
		B U	T D	M			Conceptuel	Logique	Physique		
Jensen, 2004	Auto	✓					X				Oui
Romero, 2007	Semi	✓			Formal	Domain Onto.	X			Oui	Non
Song et al., 2008	Semi	✓			/		X	X		Oui	Non
Golfarelli & 2009	Semi	✓			XML	NC	X	X	X		Oui
Pinet & al.2009	Semi	✓			/		X	X	X		Non
Carmè et al 2010	Auto	✓			/		X	X			Non
Mazón &al. 2011		✓				NC	X	X	X	Non	Non
R. Pérez M.,2012		✓					X				Non
Tangora 2013	Semi	✓			Formal	NC	X				Non
Hachaichi,2013	Auto	✓			NC		X			Non	Non
A/aziz, 2014	Semi	✓			/		X			Non	Non
						Requirements Collection Model					
Kimball & al., 2008			✓					X			Non
Phipps, 2002	Semi		✓			NC	X			Non	Non
Vrdoljak, 2003			✓				X				Non
Winter & 2003			✓		Naturel	Non	X			Non	Non
Trujillo et al., 2004			✓		UML/XML		X				Non
Prakash &al.,2005			✓		Formel	Requête SQL	X				Non
Mazon & al.2008			✓		Naturel	Non	X			Non	Non
Prat, 2006			✓		Object/Constraint		X			Non	Non
Romero & al 2009	Auto		✓		Requête SQL		X			Oui	Non
Nebot 2009	Semi		✓		Formel	NC				Oui	Non
Jovanovic &al., 2012	Semi		✓		Requête		X				Non
Cavero & al., 2003				✓		-	X				Non
Giorgini &al.2005	Semi			✓	Naturel	Goals Model	X			Non	Non
Ghozzi &al.,2005				✓	Naturel	Requête	X				Oui Perspectiv
Soussi & al.,2005				✓	Naturel	Table	X				Non
Annoni, 2007				✓	Table	Decisional Diagram	X				Non
El Golli, 2008				✓	Naturel	Goals	X				Non
Soler et al. 2008				✓	Formel		X				Non
Zepeda & al., 2008	Auto			✓	Formel		X				Non
Romero &al. 2010				✓	Requête SQL	Requête	X				Non
Mazón &al.2009	Auto			✓			X			Non	Non
Romero & al, 2011	Semi			✓			X	X		Oui	Non
S. Rizzi P. 2010				✓			X				NC
Atigui & al., 2012				✓			X				Non
Di Tria 2012	Semi			✓	Formel	BMM Model	X			Non	Non
Khouri 2012	Semi			✓	Formel	Ontologie Domaine	X			Oui	Non
Carvero & al. 2013				✓	Formel	-	X				Non
Tenmozhi 2013	Auto			✓	Naturel		X			Oui	Non
Di Tria 2014				✓	Formel	-	X			Non	Non

(NC : Non cité)

Tableau 4.1 Synthèse des approches [Larbi & al., 2017]

Après l'étude comparative des différentes approches, nous pouvons noter:

Certaines approches traitent de l'imprécision [Song & al., 2007] et [Carmè & al., 2009] et d'autres non [Khouri & al., 2012] et [Trujillo & al., 2004]. Bien que [Prakash & al., 2005] et [Romero & al., 2009] avaient utilisé les requêtes pour la collecte des besoins, mais ils n'ont pas pris en considération les requêtes floues.

Pour les formulations de besoins, les auteurs ont utilisé plusieurs modèles: le modèle I\* [Giorgini & al., 2005], [Mazon & al., 2008], [Soler, 2008]; modèle BMM [Di Tria, 2012], modèle sur mesure [Golfarelli et al., 2009], Tables [Annoni, 2006], [Romero & al., 2009]; Modèle GDI [Prakash & al., 2005]. Cette étude concerne plusieurs types de conception d'entrepôts: mixte [Winter,

2003]; [Giorgini & al., 2005]; [Di Tria, 2012] et [Cravero & al., 2014], ascendant [Golfarelli et al., 2009], descendant [Romero & al., 2009], [Mazon & al., 2008], [Prakash & al., 2005].

Certaines solutions utilisent l'ontologie pour résoudre ce problème [Romero & al., 2009] d'autres non [Golfarelli & al., 2009], [Jensen & al., 2004]. La majorité des travaux étudiés utilisent comme expression de besoins un langage naturel [Pardillo & al., 2011], [Thenmozhi & al., 2013] et [Carmè & al., 2009], d'autres utilisent un langage formel [Tebourski & al., 2013] , [Jovanovic & al., 2014] et [Giorgini & al., 2005] à moins que Romero & al. qui ont utilisé un langage de requête et [Annoni & al., 2007] qui ont utilisé des tables.

**4.1.6 Conclusion :** En récapitulation, on a des études qui proposent le traitement des besoins [Jovanovic & al., 2014] et d'autres qui ne le font pas [Romero & al., 2009]. Nous remarquons aussi que l'imprécision a été mentionnée mais n'a pas été soigneusement étudiée par [Ghozzi & al., 2008], [Nebot & al., 2009] , [Golfarelli et al., 2009] et [Jensen & al., 2004]. Comme des opportunités pour le travail de thèse de [Ghozzi & al., 2008], la gestion des données incertaines dans le modèle, le langage de requête et la méthodologie qui prennent en compte les inexactitudes dans le monde réel. [Nebot & al., 2009] a noté que les ontologies utilisées ressemblent aux diagrammes de bases de données, mais qu'elles sont plus flexibles en ce sens qu'elles fournissent des définitions incomplètes, inexactes et implicites pour les données générées. [Golfarelli et al., 2009] avait mentionné que des ontologies d'utilisation pouvaient remédier à des besoins incertains ou peu clairs afin d'améliorer la conception DW. A notre connaissance, il n' y a pas eu d'étude de conception d'entrepôts de données basée sur une ontologie floue utilisant des besoins à base de requêtes et prenant en compte l'imprécision lors de la phase d'analyse des besoins. La solution que nous proposerons après, comme solution au problème d'imprécision dans l'expression des besoins, est basée sur l'ontologie. Mais d'abord, on commence par traiter les solutions structurelles qui représentent une solution fréquente pour les bases de données floues et les DWs floues c.a.d l'imprécision des data ce qui nous a aider à penser à proposer une solution structurelle pour ce genre de problème. On va donc faire, dans ce qui suit, une étude sur l'imprécision dans les données d'entrepôts pour pour pouvoir voir la possibilité de similarité entre les données et les besoins dans l'évaluation de l'imprécision.

## 4.2 Revue de littérature : Evaluation des besoins décisionnels flexibles & Solution structurelle

Cette section montre que l'évaluation des besoins flexibles n'a pas été étudiée d'où l'absence de solution structurelle. Par contre, il y a des travaux sur l'évaluation des données floues dans les entrepôts de données floues utilisant des solutions structurelles. Afin de traiter l'imprécision dans l'expression des besoins pour la conception/mise à jour de l'entrepôt de données, nous présenterons d'abord une étude bibliographique sur l'évaluation des données floues utilisant des solutions structurelles.

### 4.2.1 Analyse des besoins décisionnels et traitement de l'imprécision

Après une étude comparative antérieure [Larbi A. & al., 2017] des différentes approches, il a été constaté que la majorité des travaux existants sur les besoins décisionnels ne tenaient pas compte de l'imprécision dans l'expression des besoins. Pour cela on a entamé une étude de l'évaluation de l'imprécision présente dans les données des entrepôts.

**4.2.2 Imprécision des données dans l'entrepôt de données** La plupart des études sur le flou dans l'ED, se concentrent sur l'imprécision des données de l'ED et proposent des solutions structurelles: la plupart des problèmes découverts pour intégrer des concepts flous dans les systèmes d'entrepôt de données sont résolus en les intégrant dans la structure d'une table Meta.

**4.2.3 Revue de littérature :** Dans [Sapir & al., 2008], la solution est d'ajouter des tables aux tables de dimensions floues.

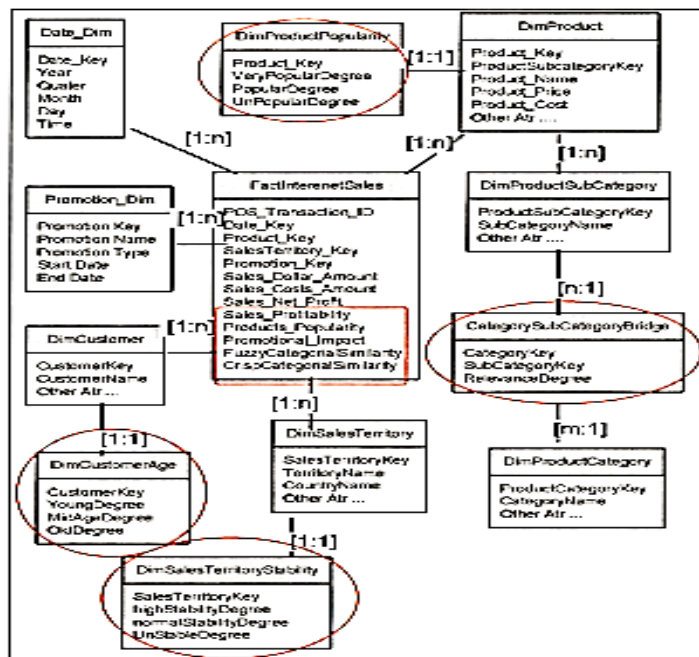


Figure 4.1 Nouveau schéma d'entrepôt de données flou incluant une hiérarchie floue

La signification technique derrière la hiérarchie floue est que les niveaux de l'enfant et des parents auront une relation multiple à multiple. Ce genre de relation n'est pas commun dans le DW classique, mais Kimball a mentionné la faisabilité d'une telle relation dans un DW. Ce type de relation peut se produire, par exemple, lorsque plusieurs clients partagent le même compte bancaire et ont en même temps d'autres comptes bancaires privés. La solution à ce problème est d'ajouter une table de "pont" entre le parent et le niveau d'enfants. La table passerelle contient des combinaisons de la clé parent et de la clé enfant et, en outre, le degré d'appartenance est ajouté entre les deux éléments de niveau. Le concepteur précisera le degré d'appartenance entre chaque catégorie et sous-catégorie. La spécification d'une hiérarchie floue se fait à l'aide d'un type de variable floue qui contient une matrice de similarité. La matrice de similarité définit le degré d'appartenance dans la table de transition, qui relie le niveau parent et le niveau enfant.

Par conséquent, une étiquette floue représentant la popularité du produit (p. ex. unPopular, Popular, VeryPopular) est utile. Cet attribut peut être estimé en comparant la somme des ventes du produit au taux de vente d'autres produits.

Dans [Boyadzhieva & al., 2014], les auteurs présentent un entrepôt de données mis en œuvre sur la base d'un système de gestion de base de données qui prend en charge les opérations sur des données floues intuitionnistes. Pour analyser les concepts intuitionnistes flous (IF), on applique une méthode de comparaison des données IF et implémente de nouveaux opérateurs dans le modèle de données relationnelles IF.

L'exemple d'application de la méthode est fourni à travers quelques requêtes vers un schéma en étoile de l'IFDW.

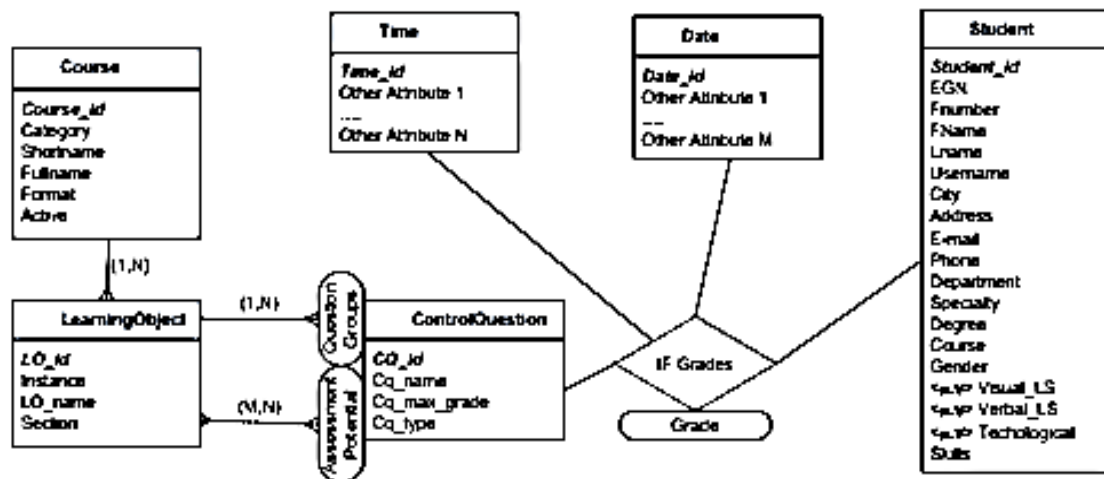


Figure 4.2 Schéma conceptuel IFDW

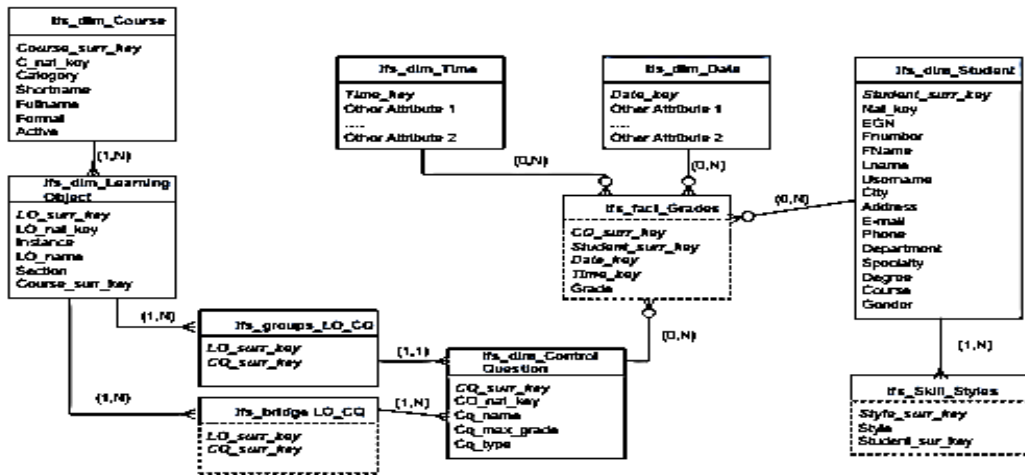


Figure 4.3 Schéma logique IFDW

Dans [Delgado & al., 2015], Les entrepôts Web (WW) sont des DW dont les sources de données proviennent du Web. Les auteurs définissent une WW flexible, qui peut être configurée en fonction de différents domaines, à travers la sélection des sources web et la définition du traitement des données caractéristiques. Un système de gestion des processus métier (BPM) permet de modéliser et d'exécuter les processus métier (BP) en supportant l'automatisation des processus. Pour soutenir le processus de construction flexible WW, on propose un niveau de deux BP: un processus de configuration pour prendre en charge la sélection des sources Web et la définition des schémas et des mappages, et un processus d'alimentation qui prend la configuration définie et charge les données dans le WW. En résumé, les auteurs proposent d'établir les Mappings entre DW et le système intégré.

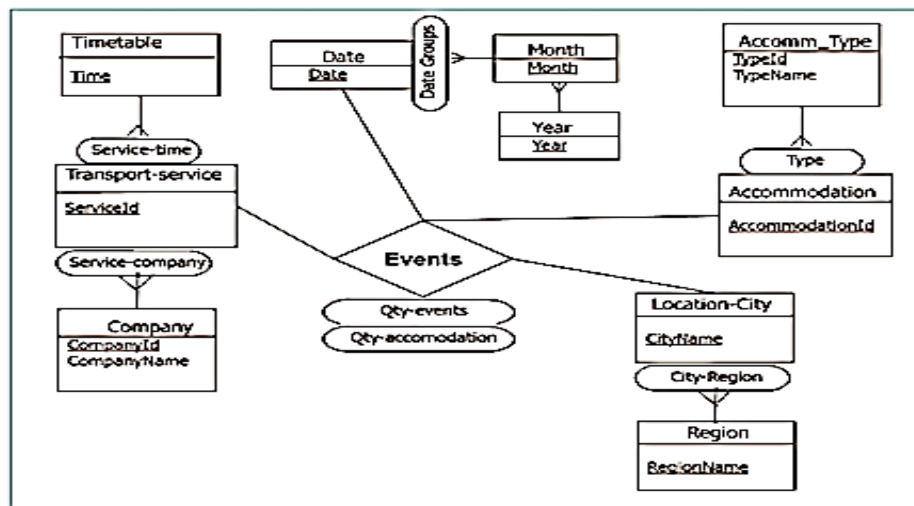


Figure 4.4 Modèle conceptuel multidimensionnel pour Fait Events

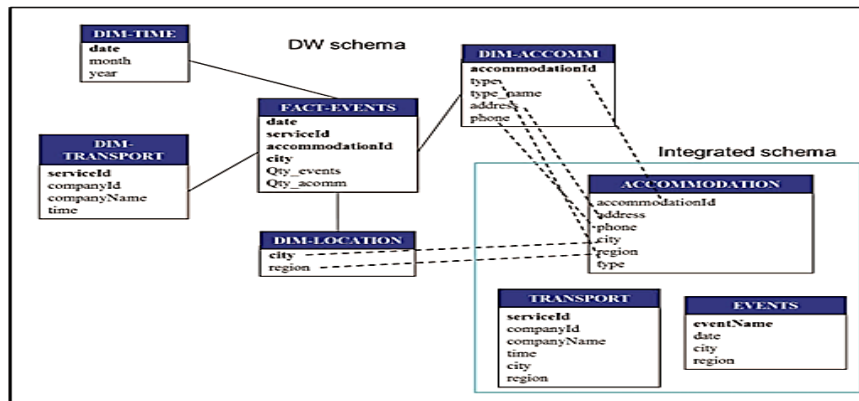


Figure 4.5 Correspondance entre le DW et schéma intégré

Dans [Fasel & al., 2014], les auteurs ont proposé un modèle d'entrepôt de données flou. En utilisant ce modèle, les données peuvent être classifiées à la fois floues et claires. Pour cette raison, le Fuzzy DW prend en charge les analyses qualitatives et quantitatives, sans affecter le schéma de base du data warehouse. De plus, la requête peut être basée sur le langage naturel grâce à l'utilisation directe de classifications terminologiques floues. À la lumière de l'étude, ils ont conclu que des mesures floues et claires peuvent être utilisées pour l'analyse et la prise de décision [Fasel, 2012]. Ce modèle a défini un DW flou en ajoutant deux tables (FMT et FCT) au DW classique afin de définir l'attribut flou (appelé TA). Pour chaque table (FT ou DT) contenant un attribut flou, il faut ajouter les deux tables FCT et FMT (voir figure 4.6).



Figure 4.6 Les tables ajoutées FCT & FMT

- On a: T<sub>Ai</sub> : Target Attribut
- FMT : Fuzzy Membership Table
- FCT : Fuzzy Classification Table
- CMA : Class Membership Attribute
- MDA : Membership Degree Attribute



Figure 4.7 Propriétés des tables FCT & FMT

#### 4.2.4 Synthèse

Catégorisation	Approches	Problèmes discutés	Solution avec concept proposé
Les concepts flous dans la TF	<ul style="list-style-type: none"> <li>Sapir, et. al. [Sapir &amp; al., 2008], [Boyadzhieva &amp; al., 2014] et al., Delgado &amp; al., 2015]</li> </ul>	<ul style="list-style-type: none"> <li>Inflexible pour les variantes de concept flous dans [Sapir &amp; al., 2008]</li> <li>Utilise un schéma en flocon de neige partiel [Boyadzhieva &amp; al., 2014]</li> </ul>	La flexibilité donnée par la structure de méta tables proposée améliore ce concept pour une application sur différentes portées d'application. Grâce à l'utilisation de la structure des tables méta, un seul cube OLAP peut contenir les données nettes et floues et permet d'analyser les deux données en même temps.
Les concepts flous dans les TDs	<ul style="list-style-type: none"> <li>Sapir, et. al. [Sapir &amp; al., 2008]</li> </ul>	<ul style="list-style-type: none"> <li>Lors de l'application de concepts flous dans les hiérarchies de dimensions, les degrés d'appartenance doivent être normalisés [Sapir &amp; al., 2008]</li> </ul>	L'utilisation de la structure de table de méta proposée permet de classer les attributs de dimension sans intégrer les concepts flous dans la structure hiérarchique des dimensions. Par conséquent, l'agrégation des faits sur la dimension peut toujours être exécutée sur les attributs de catégorie nets. La cohérence de l'agrégation est garantie et les concepts flous n'ont pas besoin d'être normalisés. Les concepts flous peuvent être intégrés à tous les niveaux d'une hiérarchie de dimensions
Pas de remplacement ou modification de valeurs d'origine	<ul style="list-style-type: none"> <li>[Delgado &amp; al., 2015]</li> </ul>	<ul style="list-style-type: none"> <li>Les concepts flous sont normalisés dans [Delgado &amp; al., 2015]</li> <li>La complexité de la navigation est augmentée par rapport aux dimensions dans [Delgado &amp; al., 2015]</li> </ul>	Le modèle de table de méta propose une intégration de concept flou dans un modèle d'entrepôt de données existant. Par conséquent, un seul cube OLAP est utilisé pour présenter à la fois des valeurs nettes et floues. La complexité de la synchronisation de plusieurs cubes et de la navigation peut être omise.
	Fasel & al., 2012], [Fasel & al., 2014]	Les concepts flous sont appliqués à tous les niveaux	L'approche proposée simplifie l'intégration et l'agrégation des concepts flous et fournit un entrepôt de données floues plus flexible

Tableau 4.2 Etude Comparative

Parmi les recherches les plus récentes sur ce sujet, nous trouvons les travaux de [Fasel & al., 2014], ceci nous permettra d'adapter son approche pour traiter l'imprécision des besoins dans le cas Fuzzy DW.

Ainsi, on a opté pour choisir le modèle Fasel dans le sens de l'adapter au cas du besoin flou pour traiter l'imprécision dans le besoin à base de requête vu que dans [Fasel & al., 2014] les concepts flous sont appliqués à tous les niveaux et cette solution simplifie l'intégration et l'agrégation des concepts flous.

#### 4.2.5 Solution structurelle (modèle FASEL)

[Fasel, 2012] a traité le problème d'imprécision sur des entrepôts flous et a proposé d'ajouter deux tables à l'attribut susceptible d'être flou appelé (Target Attribut):

Table des classifications floue (FCT ou TCF) ; Table d'appartenance floue (FMT ou TAF)

Il n'a pas envisagé le cas où la valeur de prédicat est floue et l'entrepôt de données n'est pas « fuzzy » et il n'a pas pris en considération le profil du décideur (requête à préférence),

•  $FDW = \{Dim; Fact ; FCT_{TA} ; FMT_{TA} \}$  Les deux tables sont définies comme suit (voir la figure 4): FCT est une table composée de termes linguistiques et de leurs identifiants

uniques appelée table de classification floue.

- $FCT_{TA} = \{Identifier; CMA_{TA}\}$  ; Lorsque CMA est un attribut de membre du groupe pour un attribut cible TA, représenté par CMA est un attribut qui a un ensemble de termes linguistiques  $T_1; \dots; T_k$ , TA auquel l'attribut cible peut appartenir (domaine de l'attribut  $Dom_{TA}$ )

- FMT est une table qui stocke les valeurs représentant le degré auquel une valeur est liée à un terme linguistique s'appelle Fuzzy Membership Table  $FMT_{TA} = \{Identifier; Identifier\ of\ TA; Identifier\ of\ CMA_{TA}; MDA_{TA}\}$ , Où MDA est le degré d'appartenance d'un attribut cible TA, est un attribut qui a un ensemble de degrés d'adhésion de l'attribut cible TA. La valeur d'un degré d'adhésion est calculée par une fonction d'appartenance et représentée par la fonction d'adhésion.

$\mu(TA) = MD [0;1]$  où MD (taux d'appartenance de TA) est la mesure dans laquelle les valeurs d'un attribut cible TA sont liées aux termes linguistiques CMA. [Fasel, 2012] a construit le modèle FDW en deux étapes: Définir d'abord les éléments de classification puis construire le modèle FDW (Figure 4.8)

- Afin de créer un entrepôt de données flou, une méthode est présentée qui peut guider la transformation d'un entrepôt de données croustillant en entrepôt de données flou. L'entrée dans la méthode est un entrepôt de données classique et la sortie est un entrepôt de données flou. Il s'agit d'une méthode en deux étapes: dans la première étape, les éléments de la classification sont désactivés et dans la deuxième étape, l'entrepôt de données flou est construit. La figure montre les tâches et l'ordre dans lequel elles sont exécutées.

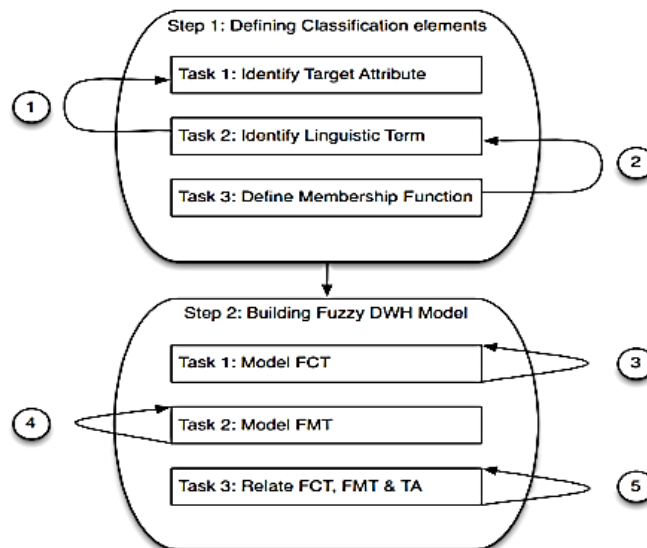


Figure 4.8 Modèle FDW [Fasel & al., 2014]

Exemple. Un entrepôt de données contient la dimension Client. Il s'agit d'une dimension et hiérarchie unique contenant une table client. Chaque client a les attributs nom, adresse

et date de naissance. A partir de l'attribut anniversaire, l'âge du client peut être calculé à l'aide de la fonction aujourd'hui-anniversaire.

**Définition des éléments de classification:** Cette étape a pour but de désagréger les éléments de classification qui sont utilisés dans la deuxième étape pour construire le modèle d'entrepôt de données flou. Il implique trois tâches, identifie l'attribut cible et identifie les termes linguistiques et définit les fonctions d'adhésion. Les détails sont les suivants:

**Première Tâche:** La première tâche est d'identifier ce qui doit être classifié, c'est-à-dire identifier l'attribut cible qui contient les valeurs qui sont destinées à être classifiées de façon floue. Ceci est fait en tenant compte de la contribution de l'utilisateur final. Dans le cas le plus simple, un attribut cible est identifié. Par exemple, considérez l'âge du client comme un attribut cible.

**Deuxième Tâche:** La seconde tâche est d'identifier l'ensemble des termes linguistiques utilisés pour classer les instances d'un attribut cible. Répétez cette tâche pour tous les attributs cibles. Il est représenté par le loop itératif 1 dans la figure 4.8 . Il y a deux possibilités:

**Cas 1 - Termes linguistiques distincts:** Il s'agit du cas le plus simple dans lequel les termes linguistiques sont distincts, c'est-à-dire qu'il existe un ensemble unique de termes linguistiques. Formellement, instance TA (1): Classes floues (1) Pour l'exemple de l'âge du client, considérez l'ensemble de termes linguistiques suivants pour les jeunes, les personnes d'âge moyen et les personnes âgées.

**Cas 2 - Différents termes linguistiques pour un attribut cible:** Il s'agit d'un cas où il existe plus d'un ensemble de termes linguistiques pour classer les instances des attributs cibles. Dans ce cas, les instances de l'attribut cible appartiennent à plus d'un terme linguistique tel qu'identifié par les utilisateurs métier. Formellement, instance TA (1): Classes floues (M)

Pour l'exemple d'âge du client, considérons que les deux termes linguistiques suivants sont identifiés. Ces ensembles sont le jeune âge, l'âge moyen, la vieillesse et l'adolescence, l'adulte et l'aîné. Les termes linguistiques peuvent déjà exister dans un modèle d'entrepôt de données classique sous la forme d'instances d'un attribut de catégorie de dimension. Dans ce cas, ces termes peuvent être utilisés pour classer les instances de l'attribut cible.

**Troisième tâche:** La troisième tâche consiste à définir une fonction d'appartenance (représentée par) pour chaque terme linguistique. Il se fait de telle manière que les valeurs peuvent être déterminées sur une échelle de 0 à 1. Répétez la tâche pour chaque terme linguistique identifié. Il est représenté par la boucle itérative 2 en figure 4.8. Il pourrait arriver que pour les utilisateurs différents? Un attribut cible appartient au même ensemble de termes linguistiques avec des diplômes de membres différents. L'affaire est la suivante:

$$\mu_{old}(customer\ age) = \begin{cases} customer\ age \geq 65, & MD_{customer\ age} = 1 \\ customer\ age \leq 40, & MD_{customer\ age} = 0 \\ else, & MD_{customer\ age} = \frac{customer\ age - 40}{65 - 40} \end{cases}$$

$$\mu_{middle}(customer\ age) = \begin{cases} customer\ age \geq 65, & MD_{customer\ age} = 0 \\ customer\ age \leq 20, & MD_{customer\ age} = 0 \\ 30 \leq customer\ age \leq 40, & MD_{customer\ age} = 1 \\ 20 \leq customer\ age \leq 30, & MD_{customer\ age} = \frac{customer\ age - 20}{30 - 20} \\ else, & MD_{customer\ age} = \frac{65 - customer\ age}{65 - 40} \end{cases}$$

$$\mu_{young}(customer\ age) = \begin{cases} customer\ age \geq 30, & MD_{customer\ age} = 0 \\ customer\ age \leq 20, & MD_{customer\ age} = 1 \\ else, & MD_{customer\ age} = \frac{30 - customer\ age}{30 - 20} \end{cases}$$

Figure 4.9 Membership Function of Customer\_age [Fasel, 2012]

Le concept flou ne peut être agrégé que sur les niveaux dimensionnels de la région dimensionnelle

La surface de l'attribut dimensionnel est agrégée à l'aide d'une fonction moyenne et, par conséquent, une agrégation du concept flou à l'aide de la moyenne arithmétique est une méthode valable pour agréger le concept flou. Dans ce cas, aucune table supplémentaire ne doit être créée.

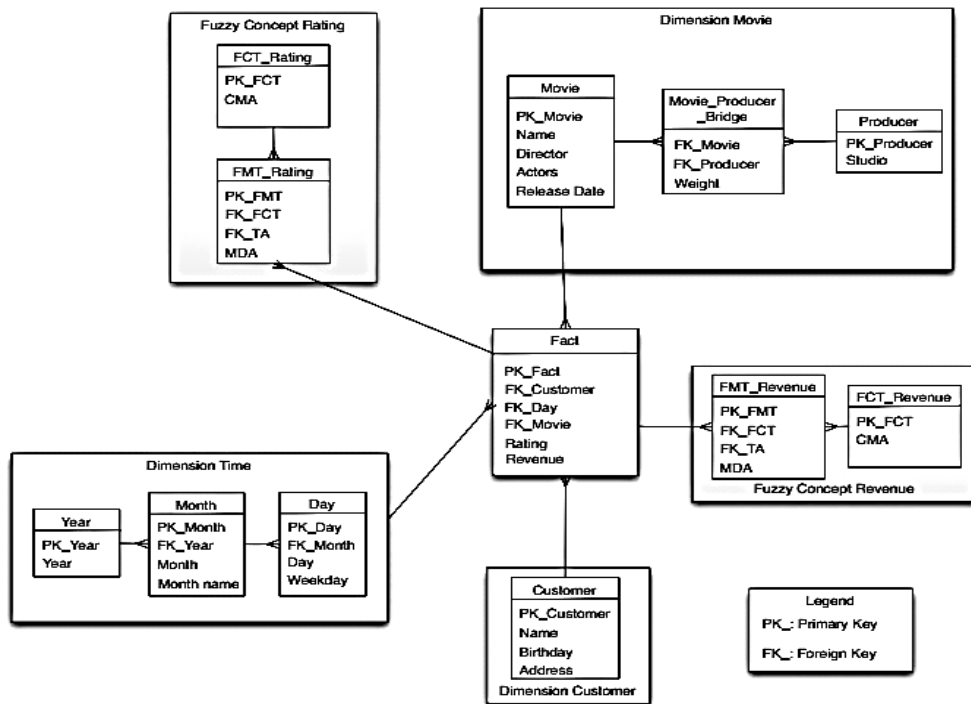


Figure 4.10 Les tables FCT et FMT additionnelles dans une solution [Fasel, 2012]

Afin de spécifier un concept flou dans le Méta schéma, les caractéristiques suivantes doivent être spécifiées: Attribut cible du concept flou et persistance de l'attribut cible (volatile ou persistant). Dans le cas d'un attribut cible volatile, on spécifie l'attribut cible de base et fonction d'appartenance pour chaque attribut membre de classe.

Pour l'agrégation du concept flou on spécifie la fonction d'agrégation ; le concept de propagation flou pour le cas de propagation du concept flou et le concept volatile flou pour les cubes résultants des opérations.

#### 4.2.6 Conclusion

Vu que la solution Fasel vise à évaluer les données floues du DW mais aussi que cette solution n'est pas opérationnelles si le nombre des attributs cibles TA est important où l'augmentation d'opération de jointures des tables additives et les TDs rend le système moins ou non performant. On pense inspirer cette proposition de la solution structurelle [Fasel, 2012], et l'adapter pour évaluer les besoins flexibles lorsque l'entrepôt de données n'est pas nécessairement flou et aussi remédier au manque de performance en rendant les tables additives FMT et FCT des méta tables indépendantes de la structure du DW.

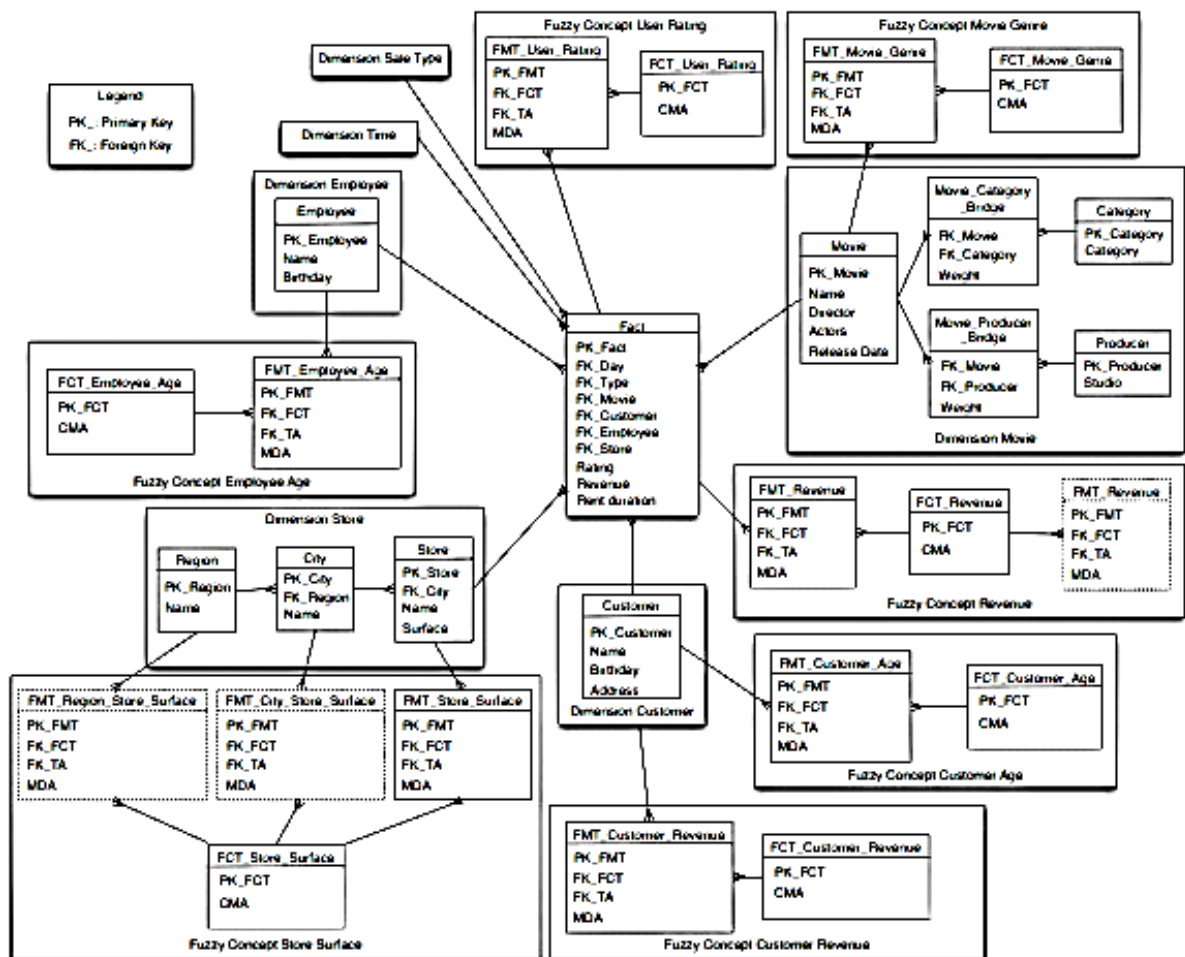


Figure 4.10a Les tables FCT et FMT additionnelles avec le nombre important de TA [Fasel, 2012]

---

### 4.3 Revue de la littérature : Evaluation des besoins décisionnels flexibles & Système expert

---

D'après l'étude de Simon (1997) [Supriadi & al., 2018], les décisions sont regroupées en deux types de base: les décisions programmées et non programmées. Après l'étude de Simon dans les années 1950, Gorry et Morton (1971) ont continué à classer les décisions en trois niveaux: (i) les décisions structurées, dont les ingrédients, ou les variables, qui constituent une décision sont connus et peuvent être mesurés quantitativement; (ii) décision non structurée, à savoir que les ingrédients, ou variables, qui constituent une décision ne peuvent pas être mesurés quantitativement; et (iii) Décision semi-structurée, qui se situe entre des décisions structurées et non structurées. La plupart des décisions d'affaires sont généralement sous la forme semi structurée (Simon, 1997).

Prendre des décisions dans le monde réel se déroulerait dans un environnement avec des objectifs et des contraintes inconnus ou imprécis qui ont besoin d'outils pour aider le processus (Kaya et Kahraman, 2010).

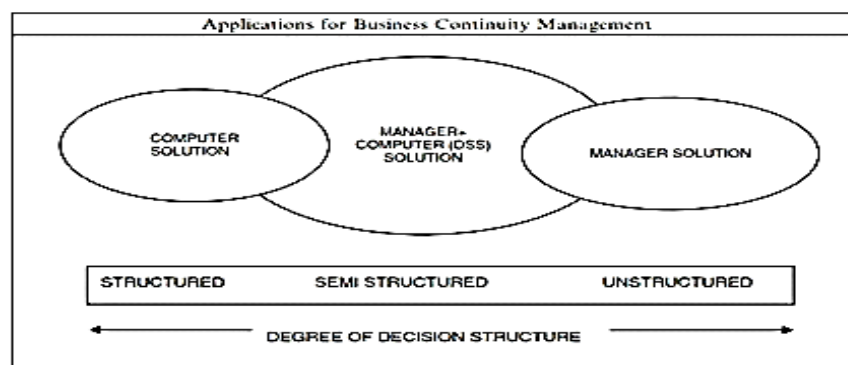


Figure 4.12 Les caractéristiques DSS

De plus, Buyukozkan et Feyzioglu (2004) ont observé que différentes méthodes de décision ont été développées pour surmonter les problèmes liés à l'incertitude, qui sont résumés comme suit:

- Modes probabilistes: ceux-ci incluent la simulation de Monte Carlo et les arbres de décision. L'analyse de Monte Carlo utilise le processus de simulation pour obtenir une gamme de solutions à un problème. L'arbre de décision est un diagramme qui fournit une approche structurée à la prise de décision qui incorpore l'incertitude des résultats.
- Modèles de notation et liste de contrôle: Diverses alternatives sont notées et notées sur une variété de questions qualitatives.
- Approches comportementales: Ce sont des outils conçus pour amener les gestionnaires à un consensus sur l'alternative à choisir, et inclure des méthodes telles que la méthode Delphi qui est une méthode de prévision qualitative utilisant un panel d'experts. Ceux-ci

sont particulièrement utiles pour les études au début, où seules des informations qualitatives sont disponibles.

□ **Processus de hiérarchie analytique (AHP):** Ce sont des outils de décision basés sur des comparaisons appariées de critères.

□ **Logique floue:** Elle traite des problèmes où une source d'imprécision est impliquée (Zadeh, 1965). En général, le concept de probabilité est lié à la fréquence d'occurrence des événements, capturée par des expériences répétées dont les résultats sont enregistrés, tandis que les ensembles flous fournissent le cadre approprié pour évaluer la possibilité d'événements plutôt que leur probabilité.

□ **Analyse de sensibilité:** Elle examine comment la solution optimale et la valeur objective optimale sont affectées par les changements des paramètres d'incertitude (valeurs et probabilités) considérés comme importants.

□ **Analyse de scénarios:** Cette technique a été largement préférée et utilisée par de nombreux décideurs. Ici, une combinaison de valeurs possibles des paramètres d'incertitude est supposée pertinente pour différents points de vue (par exemple pessimiste, neutre et optimiste), et le scénario résultant est résolu. En résolvant le problème à plusieurs reprises pour différents scénarios et en étudiant les solutions obtenues, le décideur observe les sensibilités et décide heuristiquement d'une solution appropriée.

#### **4.3.1 Outils décisionnels utilisant les approches flous**

Selon [Bashiri & al., 2011], en raison d'informations incomplètes ou indisponibles dans la situation réelle, les données (attributs) sont souvent moins déterministes et la majorité de ces attributs peuvent être évalués par la perception humaine et le jugement humain. Par conséquent, ces attributs sont généralement flous imprécis.

Au cours des dernières décennies, des méthodes de prise de décision ont été développées et, parallèlement à cette croissance, une variété d'algorithmes hybrides a également été développée. Un grand nombre de ces méthodes hybrides ont été créées dans un environnement flou.

Généralement, les problèmes de prise de décision vagues dans le monde réel sont exprimés par la pensée humaine et la perception subjective en mots, plutôt que par la probabilité et les statistiques. C'est la raison pour laquelle les nombres flous et les ensembles flous ont été utilisés pour rendre ce processus plus réaliste. De plus, des modèles d'analyse de décision multicritères ont été utilisés dans un environnement flou.

#### **4.3.2 Systèmes experts, gestion des connaissances, prise de décision et & ingénierie des besoins**

La recherche dans le domaine de gestion d'ingénierie de connaissance (EKM) a identifié les défis avec la source systématique de connaissances d'ingénierie pour l'optimisation du

processus de conception. Dans ce contexte [Quintana-Amate & al., 2015], l'ingénierie basée sur les connaissances (KBE) est reconnue comme un domaine clé dans le domaine de l'EKM et désignée par la communauté des chercheurs comme une solution potentielle pour réaliser la capture et la réutilisation efficaces des connaissances d'experts.

Les auteurs dans [Bosc & al., 1998] ont précisé que la tâche principale du système expert est d'analyser et de décider sur un certain type de problème et de donner des mesures de traitement correspondantes. Ces problèmes peuvent être des maladies, ou des défaillances mécaniques, et peuvent être n'importe quelle autre défaillance, même de la surface, telle que l'analyse des données de forage traitées. Les systèmes experts [Agarwal, 2014] sont fondamentalement développés pour aider à résoudre des problèmes complexes en raisonnant sur des connaissances déjà connues comme un expert humain. Il ne suit pas la procédure suivie dans la programmation conventionnelle par un développeur.

Il y a certains facteurs qui peuvent mener à la défaillance des systèmes experts et parmi eux l'ingénierie des besoins pour les systèmes experts est l'un. Lors du développement de systèmes experts, les développeurs accordent moins d'attention au processus d'ingénierie des besoins. Au lieu de cela, l'ingénierie des besoins est cruciale pour rassembler toutes les besoins nécessaires à un système expert. Si les besoins ne répondent pas à tous les souhaits et besoins du client, alors dans ce cas le système expert est considéré comme un échec même s'il fonctionne parfaitement. C'est pourquoi, pour le développement réussi des systèmes d'experts il faut établir le nécessaire, qui met l'accent sur le processus d'ingénierie des besoins des systèmes d'experts.

Ce que l'on peut remarquer à travers ce préambule, c'est que dans la prise de décision, l'ingénierie basée sur les connaissances est désignée comme une solution potentielle pour un grand nombre de problème. Aussi, on retrouve que le besoin est mutuelle entre l'ingénierie des besoins et les systèmes experts en général et les bases de connaissances en particulier.

### **4.3.3 Revue de littérature**

#### **a. Système expert & conception de Base de données [Dogac & al.,1989]**

Dans son papier, [DOGAC,1989] a proposé un système expert généralisé pour la conception des bases de données (GESDD) comprenant :

- i) Un système expert de génération de méthodologies pour la conception des bases de données, appelé ESGM;
- ii) Un système expert pour la conception de bases de données, appelé ESDD.

ESGM fournit un outil permettant à l'expert en conception de base de données de spécifier différentes méthodologies de conception ou de modifier celles qui existent déjà. Le concepteur de la base de données utilise ESDD en conjonction avec l'une de ces méthodologies pour concevoir une base de données à partir de la phase de spécification des besoins et produire un schéma logique dans l'un des modèles de données bien connus,

soit le modèle de données hiérarchique, le modèle de données réseau ou le modèle de données relationnelles. Le concepteur de la base de données interagit avec l'ESDD dans la phase de spécification des besoins pour présenter les informations sur une entreprise particulière et pendant la phase de conception pour tester le schéma conceptuel produit par l'ESDD. Le système est évolutif en ce sens qu'il est possible de modifier une méthodologie existante ou d'en ajouter une nouvelle.

Parmi les cas traités, La prise de décision en gestion universitaire [Noodin & al , 2016] dans le milieu universitaire reflète un leadership efficace pour déterminer la qualité des diplômés.

#### c. **Besoin & système expert**

Dans [Joseph & al., 1998], On a mis au point un système d'aide à la décision qui fournit une approche structurée pour faciliter la validation et la vérification des ensembles de besoins et améliorer la qualité de la conception qui en résulte en réduisant les risques. De plus, une mise en œuvre automatisée de cette approche a été réalisée à l'aide du logiciel Advanced Integrated Requirements Engineering System (AIRES).

La corroboration pour l'application de ce système d'aide à la décision a été obtenue par des tests automatisés et manuels effectués sur la spécification du système d'un projet de développement et d'intégration logicielle à grande échelle. L'approche d'aide à la décision, appelée méthode d'identification des risques fondée sur les tests ("BRIM), a été utilisée pour détecter quatre grands types de besoins: **l'ambiguïté, les conflits, la complexité et les facteurs techniques.**

Dans [Kahramana & al., 2015], la prise de décision floue est utilisée lorsqu'il existe des données vagues et incomplètes pour la solution multicritères

#### 4.3.4 Requêtage d'une base de données flexible

a. Tahani a été l'un des premiers à préconiser l'utilisation d'ensembles flous pour interroger des bases de données conventionnelles. Son idée consistait à permettre l'expression de conditions imprécises à l'intérieur de requêtes considérées comme des ensembles flous. Tahani a suggéré l'extension du bloc de base SEQUEL [Tahani & al., 1977] (à l'époque, c'était le langage de base de données relationnelle le plus connu pour supporter la comparaison imprécise entre une valeur d'attribut et une constante ou entre deux valeurs d'attribut (jointures)). Ces prédicats élémentaires peuvent être combinés en utilisant les connecteurs AND et OR fonctionnant comme une intersection et une union d'ensembles flous (min / max).

b. Dans [Bosc & al., 1988] et [Bosc & al., 1994], l'objectif est d'introduire du flou dans le bloc de base de SQL. Cela peut être réalisé à deux niveaux principaux: dans les prédicats et dans la façon dont ils sont combinés (connecteurs). L'expression technique de cette extension théorique se retrouve dans les généralisations du SQL au SQL flou, tel SQLf (f pour fuzzy) introduit par [P. Bosc, 1995]. Une requête en SQLf typique est :

```
SELECT THRESHOLD 0.6 Name FROM Inhabitants  
WHERE Salaire = très élevé AND DistancePlace = faible
```

c. Dans [Jarzabek & al., 1998], un langage de requête de programme, appelé PQL, qui facilite la conception de SPA flexibles (Static Program Analyzers). PQL est une notation conceptuelle, indépendante du langage source, qui permet de spécifier les requêtes de programme et les vues de programme.

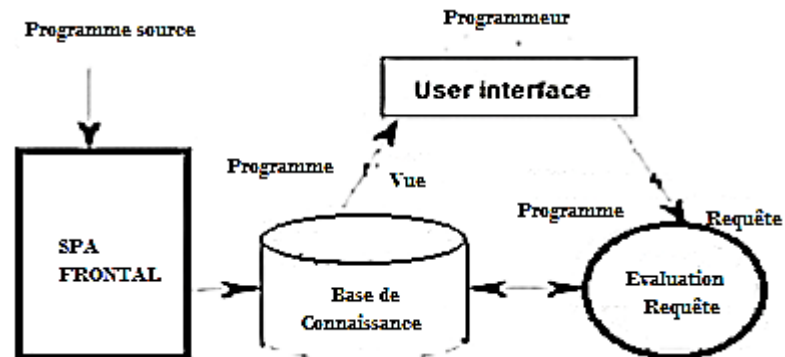


Figure 4.13 Architecture d'analyseur programme statistique (SPA)

Dans [Bartel & al., 2006], le logiciel iSQLF est basé sur le langage sqlf développé par l'équipe de recherche Badins. Ce langage est en fait un superset de SQL. En effet, le langage étend le bloc SQL, permettant ainsi aux conditions de la clause WHERE d'être floues et ces requêtes seront faites sur un SGBD conventionnel flou.

d. Dans [Aude & al., 2007], un prototype de logiciel iSQLf a été construit avec une interface Web. Le but du projet était de reconstruire complètement l'application en utilisant des technologies plus récentes comme Java et les servlets.

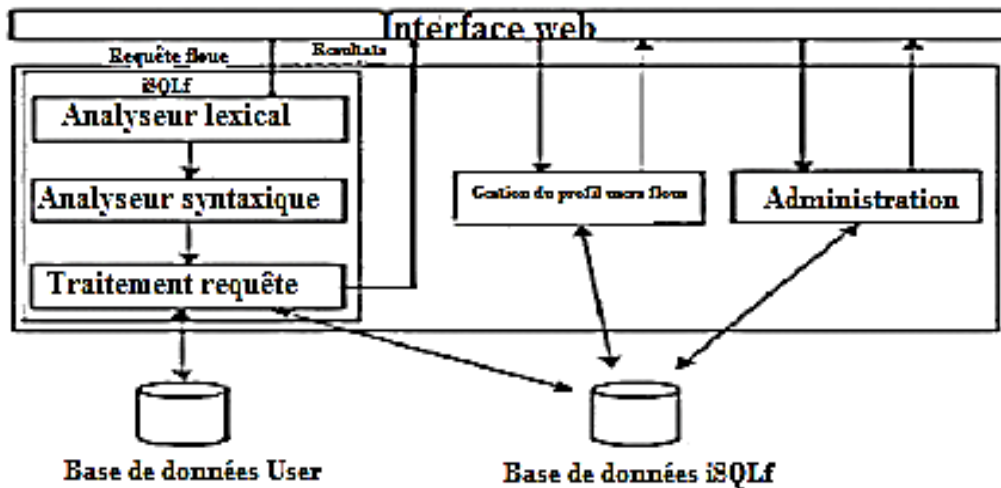


Figure 4.14 Schéma général de l'Application (2007)

e. Les auteurs dans [Gonzalez & al., 2009] ont proposé une mise à niveau en accord avec la norme SQL2003, proposant ces nouvelles fonctionnalités floues: stockage de table floue, type de données floues multiset et déclaration de fusion floue.

### Employé bien payé

pId	name	age	profession	salary	depId	
1001	Abraham	20	Engineer	1100	a001	0,10
1002	Billbad	30	Technician	1150	b002	0,15
1005	Esteban	60	Technician	2000	a001	1,00
1007	Gedeon	35	engineer	2250	c003	1,00
1008	Ester	45	Technician	1750	d004	0,75
1009	Isaias	55	laborer	1250	a001	0,25
1010	Josue	18	engineer	1500	b002	0,50
1015	Oscas	23	laborer	1250	a001	0,25

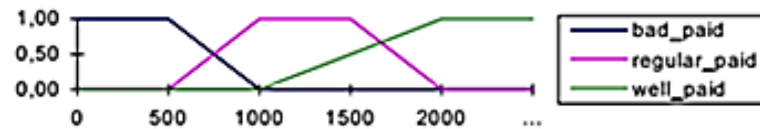


Tableau 4.3 Représentation des prédicats flous

f. Dans [Tamani & al., 2011], la personnalisation de l'interrogation flexible des informations, dédiés aux réseaux de transport multimodal est abordé dans le présent document. Tamami & al. ont proposé une approche qui consiste à développer une couche sémantique floue pour étendre les fonctionnalités des systèmes d'information au raisonnement. Cette couche est constituée de fonctions de raisonnement qui combinent des données d'application de domaine avec des axiomes d'ontologie.

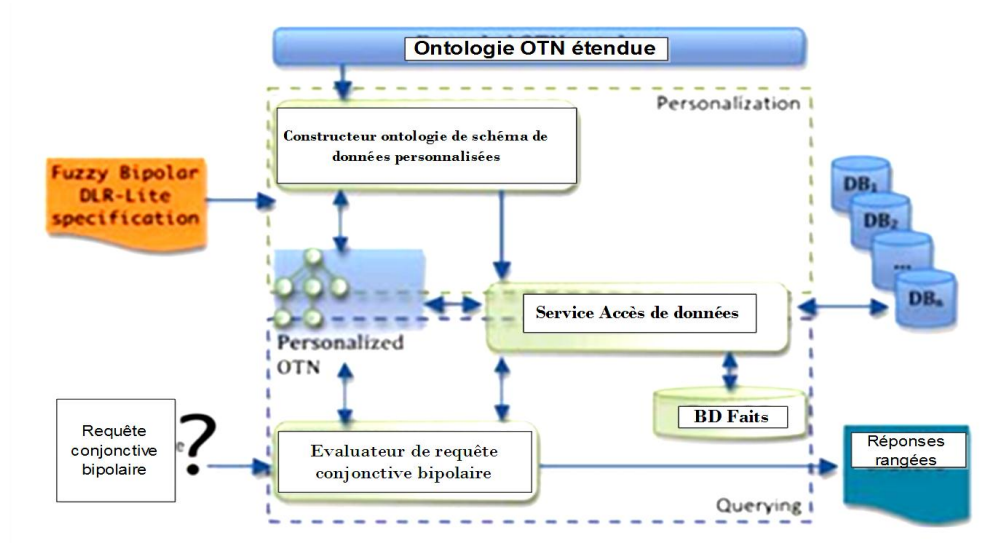


Figure 4.15 Architecture de base du prototype développé

g. [Smits & al., 2012] ont présenté la mise en œuvre d'une requête floue, intégré à une base de données relationnelle de gestionnaires. Ce système repose sur le langage SQLf pour exécuter des requêtes impliquant des prédicats flous connectés par conjonctions, disjonctions, quantificateurs flous graduellement par d'autres opérateurs. [Bosc & al., 1998]

**Requête SQL floue** : Plusieurs études [Gonzalez & al., 2009] comme OMRON, FQUERY, ISKREOT, FSQL, SQLf, SOFTSQL,... et d'autres ont appliqué la théorie des ensembles flous à la base de données pour résoudre ce problème. SQLf [Bosc & al., 1995] est le plus complet car il a des instructions SQL plus longues et c'est la seule façon de mettre à jour SQL: 1999 [Goncalves & al., 2001].

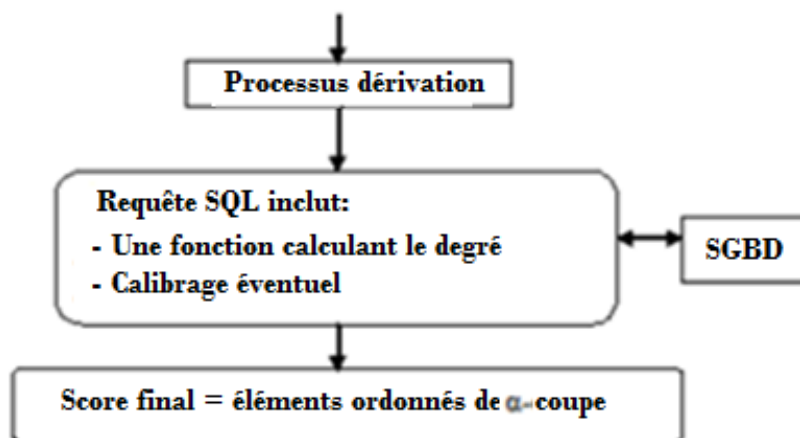


Figure 4.16 Architecture générale de l'évaluation de requêtes SQL Floues (bypass)

**4.3.5 Etude Comparative** : Voici une étude comparative des travaux importants déjà cités.

Approches	Problèmes discutés	Solution avec les concepts proposés
<b>Tahani</b>	Permettre l'expression de conditions imprécises à l'intérieur de requêtes	Extension du bloc de base SEQUEL
<b>Bosc</b>	Analyser et de décider sur des problèmes présentant des défaillances	Introduire du flou dans le bloc de base de SQL (Prédicat – Connecteur)
<b>Bartel</b>	Conditions de la clause WHERE sont floues	le logiciel iSQLF est basé sur le langage sqlf Conditions de la clause WHERE peuvent être floues
<b>Aude</b>	Evaluer la requête floue via interface web	un prototype de logiciel iSQLf a été construit avec une interface Web.
<b>Gonzalez</b>	Stockage flou et fusion flou	nouvelles fonctionnalités floues: stockage de table floue, type de données floues multiset et déclaration de fusion floue.
<b>Tamani</b>	Personnalisation de l'interrogation flexible	développer une couche sémantique floue pour étendre les fonctionnalités des systèmes d'information au raisonnement

Tableau 4.4 Comparaison des solutions proposées

Comme on vient de voir, la majorité des travaux présentés se basent sur les travaux de [Bosc, 1998]. On n'est pas intéressé à l'interface web [Bartel, 2006 ; Aude, 2007], comme on ne veut pas traiter les requêtes à préférence et personnalisée [Tamani, 2011] , on veut pas traiter les problèmes de stockage flou. Pour tout cela, on s'intéresse aux travaux de Bosc. On s'intéresse aussi à l'étude, plus tard, du connecteur flou dont Bosc le présente dans ses travaux. On a ainsi décider d'apporter une extension à la solution de Bosc et ce en introduisant un traducteur utilisant une grammaire et de l'appliquer sur l'exemple détaillé dans [Cox & al., 1994].

#### **4.3.6 Conclusion**

Cette section a présenté une revue de littérature pour le traitemnt de requête floue. On a remarqué que la majorité des travaux ont proposé comme solution des logiciels (langage) tels que SQLf, iSQLf et SEQUEL qui représente des extensions du bloc de base SQL

Vu que les besoins décisionnels à base de requêtes peuvent présentés le flou au niveau autre que la valeur des prédicats (Traitée par la majorité des travaux étudiés), à savoir le connecteur ou l'opérateur. Vu que [Bosc, 1998] a déjà envisagé dans son approche l'étude du connecteur flou et pour les autres raisons déjà citées, nous optons à améliorer cette approche par l'introduction du traducteur.

---

## 4.4 Revue de la littérature : Evaluation des besoins décisionnels flexibles & Ontologie

---

Cette section montre que malgré la présence de travaux ayant traité les besoins dans la conception de DW basée sur les ontologies mais ces travaux n'ont pas envisagé l'étude de l'imprécision des besoins ou l'étude des besoins imprécis. Pour cela, on a décidé d'étaler l'étude à d'autres problèmes qui proposent des solutions ontologiques afin de faire un rapprochement en vue de proposer une solution ontologique à notre problème.

### 4.4.1 Conception de DW basée sur les ontologies : Etat d'art

Dans [Giorgini & al., 2005] et [Mazon & al., 2008] ces travaux traitent les besoins des utilisateurs en réconciliation avec les sources de données. Ils suivent un procédé dans lequel un schéma multidimensionnel est obtenu à partir des besoins des utilisateurs, puis, ils appliquent justesse entre les sources de données multidimensionnelles et modèles obtenus. [Mazon & al., 2008] proposent une approche hybride en conformité avec le paradigme MDA. Dans leur papier, [Romero & al., 2010] ont présenté une nouvelle approche pour appuyer la conception de DW basée sur AMDO: une méthode automatique pour identifier les concepts susceptibles de jouer des rôles multidimensionnels, à partir d'une ontologie représentant leur domaine d'activité. AMDO effectue une analyse approfondie des sources de données. Depuis des approches axées sur l'offre sont connus pour générer trop de résultats, ils ont proposé d'exploiter la sortie de AMDO au moyen de la recherche de motifs.

[Khoury & al., 2012] ont défini une ontologie basée sur les besoins des utilisateurs de construction de Base de données cohérente avec la structure de DW (application par objectifs). Ils ont proposé d'étendre l'utilisation des ontologies, sur la même manière que pour les sources de données, afin de clarifier (éliminer toute ambiguïté) et d'unifier les besoins des utilisateurs. Ce travail occulte la phase d'analyse de données et aussi il ne répondait pas aux besoins exprimés par les utilisateurs. Toutefois, cette proposition avait permettre au concepteur d'exprimer ses besoins en langue naturelle et à former des buts sans reformulation SQL, la réutilisation de ce modèle n'est pas systématique. Les auteurs dans [Sabri et al., 2013] ont proposé une méthode pour l'analyse des besoins décisionnels, comme un processus et des modèles qui prennent en charge les besoins de découverte, de spécification, de négociation et de validation des besoins. Cette approche visait à guider et orienter les concepteurs analystes au long du processus décisionnel. L'étude en [Thenmozhi & al., 2012] transforme chaque source de données en format de l'ontologie, et ce en utilisant des outils RDBtoOnto, JXML2OWL et OntoLT par les auteurs. Les auteurs proposent une approche hybride de l'ontologie basée pour dériver un schéma de DW multidimensionnel. Dans [Thenmozhi & al., 2013], les besoins des utilisateurs et les sources de données sont représentés à l'aide d'ontologie, cette solution présente un outil axé sur l'ontologie qui permet de dériver automatiquement le modèle conceptuel et le modèle

logique de l'entrepôt de données à partir de la source de données et des besoins de l'entreprise. Les auteurs utilisent l'ontologie et ses capacités de raisonnement à différentes étapes de l'outil proposé pour faciliter la tâche de conception. [Maté & al., 2012] ont présenté une proposition pour l'application de modules, spécialement conçus pour DW. Ils ont défini leur proposition, et ont fourni des lignes directrices comment l'appliquer correctement. Ils ont également montré un exemple d'application et ils ont réalisé une expérience, avec leur proposition, y compris les participants allant de nouveaux utilisateurs aux experts de la modélisation i\*. Cette solution applique l'analyse des besoins basées sur les objectifs, mais ne concerne pas ceux basés sur les modèles ou sur les requêtes.

Des travaux récents sont intéressés à DW basée ontologie en phase analytique ou opérationnel [Soussi & al., 2005], mais peu d'études sont intéressés à ce type de structure dans la phase de conception. De nombreuses études sur la conception de l'entrepôt de données basé sur l'ontologie [Soussi & al., 2005] , [Ghozzi & al., 2005] , [Ravat , 2007] et [Di Tria & al., 2012] reposent sur les objectifs, mais il y a moins d'études basées sur les requêtes ou utilise des modèles de cas [Romero & al., 2010] et [Jovanovic & al., 2012]. Dans [Golfarelli & al., 2011], les besoins pour les analyses de données sont souvent floues et incertaines, principalement parce que les processus de décision sont flexibles structurés et mal partagés entre les grandes organisations, mais aussi à cause des difficultés entre les utilisateurs et les analystes. Selon les auteurs travaillant les quatre qualités du processus de développement de logiciels, parmi lesquels la robustesse qui signifie la flexibilité du processus, à savoir sa capacité à réagir rapidement et en douceur aux changements imprévus dans l'environnement. Les ontologies utilisées permettront d'améliorer plusieurs aspects de la conception DW. Dans [Pardillo & al., 2011], l'auteur présente les situations dans lesquelles les ontologies peuvent aider les concepteurs dans le développement. Ils sont intéressés par certaines de ces situations:

(i) Incomplétude dans les modèles multidimensionnels: connaissance ontologique peut enrichir un modèle multidimensionnel dans les aspects qui ne sont pas prises en compte lors de l'analyse des besoins ou l'alignement des données source. Il existe certains types de données qui peuvent être obtenues directement à partir d'autres sources de données "publiques".

(ii) Raisonnement sur requêtes OLAP : L'idée principale est d'enrichir l'expressivité OLAP interrogeant avec la capacité d'inférence fournie par la fondation ontologique de l'entrepôt de données.

#### 4.4.2 Récapitulatif : Voici une comparaison des approches à base d'ontologie :

Types d'approches	Auteurs	Groupe de modèles	Langage d'expression des besoins	Trait. des besoins
A	Romero, 2007	Ontologie de domaine	Formel	Oui
A	Thenmozhi&al. 2012	Ontologie	Naturel	Non
M	Romero, 2011	Format XML+Onto.	Formel	Non
M	Thenmozhi & al. 2013	Ontologie	Formel	Non
M	Di Tria, 2014	Ontologie de Modèle Graph-based	Formel	Oui
M	Khoury, 2012	Ontologie de domaine	Naturel	Oui
A	Pardillo, Mazon, 2011	Ontologie	Formel	Oui

Tableau 4.5 Approches basées sur les ontologies

En récapitulation de l'étude comparative, on n'a pas trouvé d'étude sur l'imprécision des besoins malgré qu'il y a des travaux qui ont traité le besoin en utilisant l'ontologie [Romero,2007], [Pardillo, 2011] et [Khoury,2012] mais ils n'ont pas traité le besoin imprécis.

Parmi les travaux déjà traités, et en perspective du travail de thèse de Ghazzi, 2010], Elle a évoqué le problème en disant que la gestion de données incertaines dans le modèle, dans le langage d'interrogation et dans la méthodologie qui prend en compte les inexactitudes dans le monde réel. Aussi [Nebot & al., 2009] ont projeté de proposer une solution ontologique traitant les données implicites et ont noté que les ontologies utilisées ressemblent aux modèles de base de données, mais qu'elles sont plus flexibles en ce sens qu'elles donnent des définitions de données générées incomplètes, inexactes et implicites. Comme perspectives, ils envisagent de spécifier le schéma de l'analyse multidimensionnelle en termes d'axiomes de l'ontologie mais, on n'a pas trouvé de suite.

Parmi les plus récentes travaux sur les données floues et utilisant la nouvelle approche de Zadeh [Mendel, 2010], ceux de [Martinez & al, 2016] présentent une contribution pour résoudre le problème de l'organisation de la représentation de l'expérience dans les systèmes computationnels qui sont capables de générer des descriptions linguistiques pertinentes de données pour des utilisateurs et des contextes spécifiques. Nous prétendons que, typiquement, les connaissances d'experts modélisées dans ces systèmes sont limitées à l'une des dimensions de la signification du langage naturel. Pour cela, ils ont construit une ontologie qui aidera les concepteurs à modéliser leur expérience dans le domaine d'application.

On a ainsi décidé à adapter les travaux de [Martinez & al., 2016] pour les raisons suivantes:

\* Quoique leur modèle ne concerne pas le système décisionnel, ni le besoin décisionnel. Le modèle [Martinez, 2016] permet de traiter le besoin en langage naturel.

- \* Ce modèle permet de traiter l'imprécision en exploitant les différents concepts du modèle ontologique GLMP(Granular Linguistic Model of Perception)
- \* Ce modèle permet la representation en plus de la couche linguistique (qui permet d'évaluer l'imprécision de la valeur du prédicat d'un requête-besoin), et peut envisager les deux autres couches du modèle en oignon (OCC et OCNC)

#### **4.4.3 Conclusion**

La qualité de la conception de l'entrepôt de données est étroitement liée à l'analyse des besoins (fonctionnels et non-fonctionnels) et les approches utilisées pour réaliser cette conception. L'expression inexacte, l'ambiguïté, les besoins manquants ou inadéquates sont la cause des défauts de conception. Cette étude met en évidence un point très important qui est l'étude de l'imprécision et que les chercheurs n'ont pas donné l'importance qu'elle mérite. Nous visons à travers cette étude de déclencher une discussion sur ce point focal pour améliorer la conception de DW et afin de satisfaire les différents acteurs dans le processus décisionnel.

#### **4.5 Conclusion du chapitre**

L'absence d'approches ontologiques pour l'évaluation de l'imprécision du besoin à base de requête nous n'a pas empêchés de chercher des solutions ontologiques d'autres problèmes qui peuvent nous servir de point de départ pour proposer notre approche. Le modèle GLMP permettant d'évaluer la perception était, selon notre étude, le cas le plus approprié pour l'adapter et l'étendre pour permettre l'évaluation du besoin imprécis.

Dans le chapitre suivant, on va présenter les trois approches proposées et les applications réalisées pour mettre en œuvre ces propositions.



Rien n'est sans risque, et qui ne tente rien n'a rien.

Henri-Frédéric Amiel ; 1855.

# CHAPITRE V

## Evaluation de BDF

### CONTRIBUTIONS & APPLICATIONS

5.1 Contributions et approches proposées	80
5.2 Approche 1 : Evaluation des BDFs utilisant le changement de structure	82
5.2.1 Présentation de l'approche 1 (Solution Structurelle)	82
5.2.2 Exemple illustratif : Evaluation de risque dans le transport routier	86
5.2.3 Expérimentation	88
5.2.4 Conclusion	91
5.3 Approche 2 : Evaluation des BDFs utilisant un système expert	92
5.3.1 Présentation de l'approche 2 (Solution à base de système expert)	92
5.3.2 Exemple illustratif : Evaluation de risque de gestion intelligente de projet	96
5.3.3 Expérimentation	98
5.3.4 Conclusion	100
5.4 Approche 3 : Evaluation des BDFs utilisant une ontologie floue	101
5.4.1 Présentation de l'approche 3 (Solution Ontologique)	101
5.4.2 Exemple illustratif : Evaluation de risque dans le transport routier	107
5.4.3 Expérimentation	108
5.4.4 Conclusion	115
5.5 Conclusion du chapitre : Récapitulatif des trois solutions (Structure, Système expert, Ontologie)	116





# CHAPITRE V

---

## Évaluation des besoins décisionnels flexibles (BDF)



DECISIONAL REQUIREMENT

Partant d'une étude primaire pour évaluer l'imprécision dans l'expression des besoins décisionnels qui est basée sur un langage de requête floue [Larbi & al., 2018] et qui suppose de le nombre d'attributs cibles (TA) concernés par les prédicats flous est limité ainsi on propose une solution structurelle qui consiste à ajouter des méta tables au dictionnaire de données. Cette solution devient non faisable une fois le nombre de TA devient grand, nécessitant un espace mémoire important. D'où la nécessité de chercher une autre solution au problème se basant sur le raisonnement et non pas sur la structure de données. Ainsi, on a proposé une solution naive se basant sur le système expert flou. Cette solution nous a permis d'évaluer l'imprécision et de proposer une solution se basant sur les connaissances et les règles fournies par l'expert. Le problème se pose à nouveau si on veut passer de l'évaluation de termes d'un domaine qui est assuré par le système expert (Linguistique Informatique) vers le vocabulaire canonique des concepts d'un domaine [Jean, 2007], on passe ainsi à la solution ontologique [Larbi & al., 2018] et on aura la possibilité de passer de la couche linguistique aux autres couches du modèle en oignon (Ontologie Conceptuelle Canonique et Ontologie Conceptuelle Non Canonique).

### 5.1 CONTRIBUTIONS et APPROCHES PROPOSEES

On peut simplifier notre démarche comme suit ;

- i) Le décideur aura un ou plusieurs besoins décisionnels qui ne sont pas précis et demande à l'analyste de satisfaire ces besoins
- ii) L'analyste a déjà élaboré ces besoins sous forme de requêtes en langage formel (SQL par exemple). Ces requêtes sont sauvegardées en Workload (Charge de travail) et parmi leurs caractéristiques est qu'elles ne changent que rarement (La mise à jour des besoins décisionnels n'est pas fréquente)
- iii) Il procède ainsi au chargement de la charge de travail. Puis il procède au chargement aussi des méta tables contenant les matrices mentionnant le lien entre requête floue et l'attribut cible (TA : Target attribut), ou lien entre requête classique et l'attribut.
- iv) Si le nombre de TA est restreint, il peut procéder à l'évaluation en utilisant la solution structurelle qui ne va pas entacher la performance du support à travers le

nombre important de jointure entre les tables additives et les tables de dimension correspondantes.

v) Sinon, en consultant les matrices chargées, il saura si le flou est présent seulement dans la valeur de prédicat et que l'évaluation est seulement pour les termes linguistiques, dans ce cas, il fait appel à la solution à base de système expert.

vi) Dans le cas contraire, il fait appel à la solution ontologique.

Donc ces trois solutions sont plus tôt complémentaires et permette de satisfaire le décideur dans le cas ou on aura un ou plusieurs besoins (requêtes) flous.

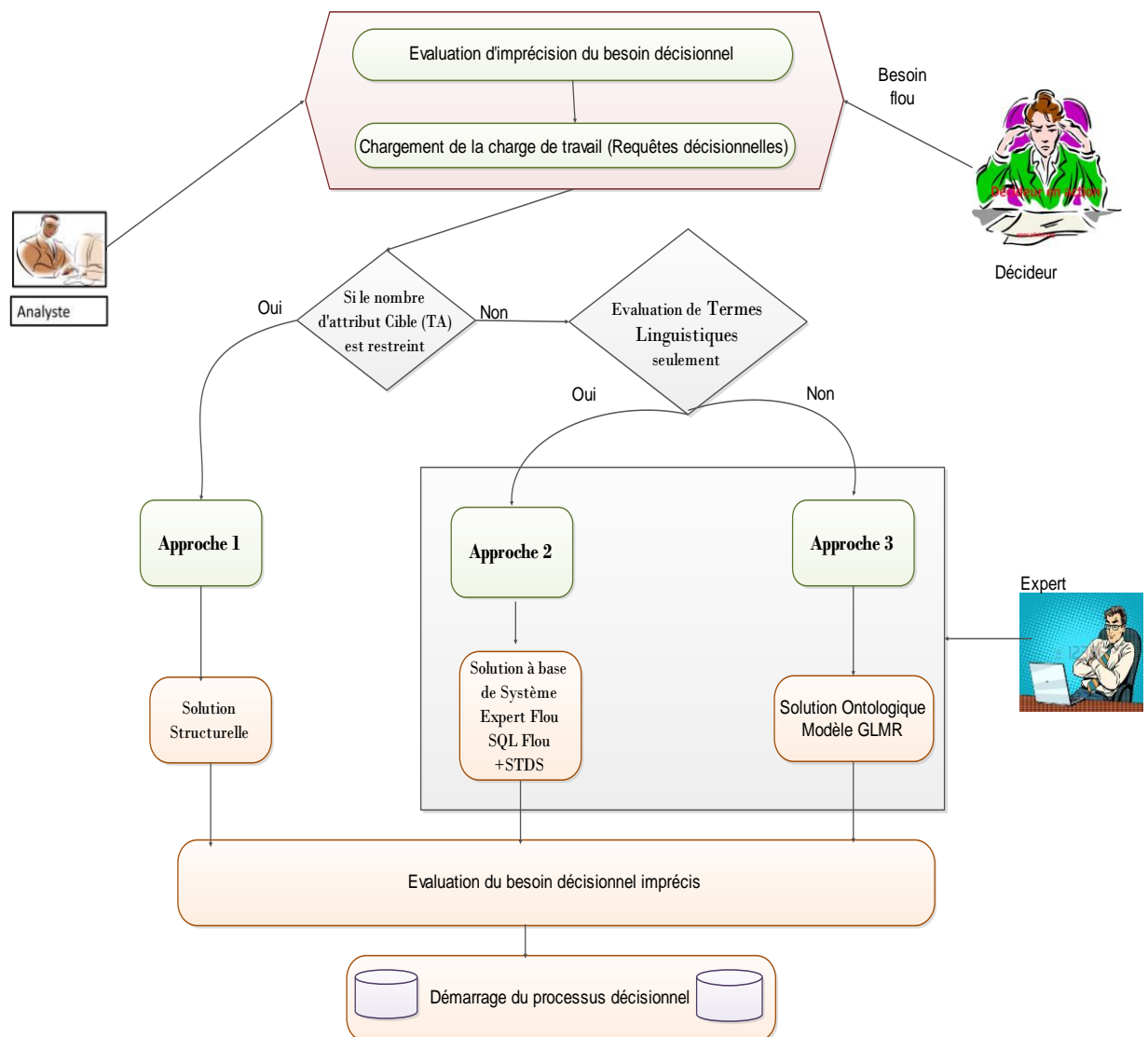


Figure 5.1 Démarche générale de notre approche.

---

## 5.2. Approche 1 : Evaluation des BDFs utilisant le changement de structure

---

La littérature traitant de l'imprécision des systèmes décisionnels s'est intéressé particulièrement l'imprécision des données, mais pas à l'imprécision des besoins fondés sur les requêtes. Afin d'évaluer l'imprécision des besoins basés sur les requêtes et dans le but d'améliorer la qualité de conception de l'entrepôt de données. Notre démarche s'est intéressée à l'extension du modèle FASEL, qui traite uniquement l'évaluation des données floues, mais pas l'aspect du besoin flou.

Ce travail présente une solution primaire qui représente une proposition structurelle dans le cas de la mise à jour de l'entrepôt de données.

Actuellement, il existe deux travaux [Chiu & al., 2014] et [Chen & al., 2009] qui ont traité uniquement l'imprécision des requêtes dans les bases de données (i.e., l'aspect OLTP) en négligeant les requêtes décisionnels dans les entrepôts de données (i.e., l'aspect OLAP).

### 5.2.1 Présentation de l'approche (Solution structurelle)

FASEL Model	Notre proposition
DW est floue	DW n'est pas flou
Attribut cible flou	L'expression du besoin contient l'attribut cible et est flou.
Ajouté des tables FCT et FMT pour chaque attribut flou	Ajouté des tables FCT et FMT pour chaque attribut cible (besoin flou)
Pas d'ajout d'autres tables	Ajouté les métras tables MT1, MT2 et 2 Tables – matrices
Les fonctions d'appartenance dépendent de l'attribut cible	Les fonctions d'appartenance dépendent des paramètres de l'expert $MT1 = \{FCT\_BRtype, FMT\_BRtype, ..\}$ ; $MT2 = \{\alpha, \beta, \alpha_{Min}, \alpha_{Max}, \beta_{Min}, \beta_{Max}\}$ ; MT3, MT4 (Voir Figure 5.10)

Tableau 5.1 Comparaison entre le modèle FASEL et notre proposition  
[Larbi & al., 2018]

Cette solution proposée nous permet d'améliorer la conception de l'entrepôt de données et/ou d'effectuer des requêtes OLAP et de réécrire des requêtes flexibles, en particulier lorsque le DW n'est pas flou et que l'imprécision est présente dans l'expression du besoin du décideur. La solution proposée présente pour le DW non flou.

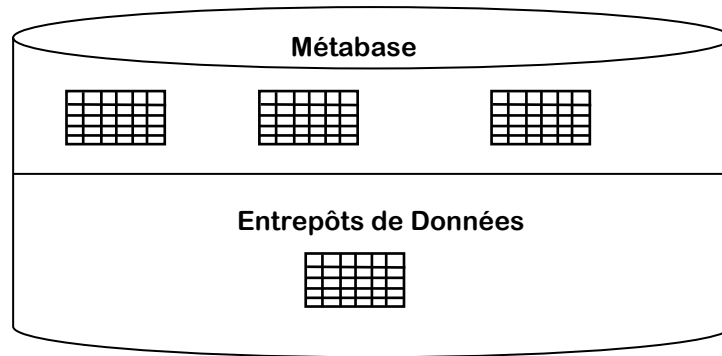


Figure 5.2 Métabase & Entrepôts de données

Différemment du modèle Fasel où pour chaque table incluant un attribut flou nous ajouterons deux tables FCT et FMT, pour notre cas nous proposons d'ajouter les deux méta tables FCT et FMT pour la table contenant un attribut « soit-disant fuzzy » (i.e. la valeur du prédicat associée à l'attribut cible est floue).

Exemple : Le type de risque du bus est Inacceptable ; BRType est inacceptable

Pour connaître cet attribut, nous proposons d'ajouter une Meta table (ou matrice) qui liste les attributs participant à l'évaluation de la requête du besoin flou.

Nous proposons également d'ajouter une méta table pour définir les paramètres définis par le décideur permettant cette évaluation.

Exemple :  $MT_2 = \{ \alpha, \beta, \alpha_{Max}, \alpha_{Min}, \text{Seuil} \}$

De plus, les fonctions d'appartenance dépendent du paramètre et non de l'attribut cible comme pour le modèle FASEL. Enfin, les fonctions d'appartenance peuvent dépendre de deux ou plusieurs paramètres, contrairement à celles définies par le modèle Fasel où ces fonctions ne dépendent que de l'attribut cible.

Exemple :  $TARF = \{(A_1, Req_3), (A_2, Req_1, \dots)\}$

a. Lorsque l'attribut concernant les besoins flexibles est flou, la solution sera l'adaptation du modèle de Fasel [Fasel & al., 2012]. Dans chaque structure de table floue (contenant un ou plusieurs attributs flous), il y aura un changement structurel (par exemple, le modèle en étoile deviendra un modèle en flocon de neige):

- Ajout d'au moins deux tables (FMT et FCT) pour chaque table de dimensions (TD) qui contient au moins un attribut flou. Les nouvelles tables comprennent les définitions de concepts flous.
- Ajouter deux autres tables de métadonnées floues à la table de métadonnées existante, l'une définissant toutes les tables ajoutées et l'autre les paramètres de données flexibles définis par le décideur.

b. Bien que l'approche FASEL soit la solution lorsque les TAs des besoins flous ne sont pas, on ne peut pas considérer la solution Fasel comme solution.

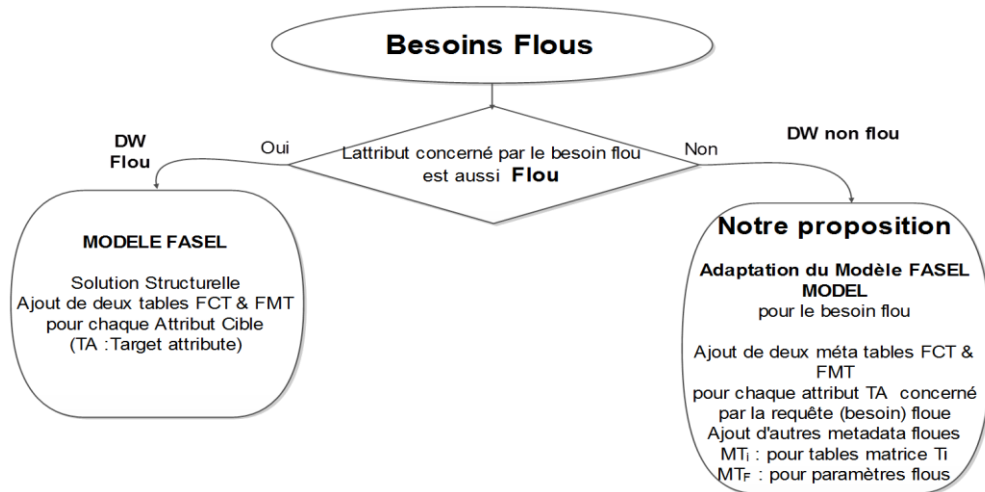


Figure 5. 3 Solution proposée

Notre proposition consiste donc à ajouter les structures (tables) au dictionnaire de données et non pas à la structure globale de l'entrepôt. Ainsi, on a à ajouter:

$$MT1 = \{FCT\_BRtype, FMT\_BRtype, ..\}$$

**Table Attributs–Requêtes floues** (valeur de prédicat flou) : C'est une matrice du genre :

$$TARF = \{(A1, Req3), (A2, Req1), (A4, Req1)\}$$

Pour dire que la requête 1 de la charge de travail comprend deux attributs cibles TA1 =A2 et TA2=A4 qui sont associés à des valeurs de prédicats floues.

Table pour les paramètres du flou (Calibre, ..) fixés par l'expert

$$\text{Exemple } MTF2 = TPF = \{\text{calibre}=2, \alpha=2, \beta=4, \alpha_{\text{Min}}=1, \dots\}$$

**Table Requêtes-attributs** : C'est une matrice mentionnant les attributs participants dans une requête donnée.

Exemple :  $TRA = \{(Req1, A1), (Req1, A2), (Req1, A4)\}$  pour dire que les attributs A1, A2 et A4 participent dans la requête Req1.

Deux Tables FCT et FMT pour chaque attribut cible TA qui ont le même rôle que pour le modèle FASEL.

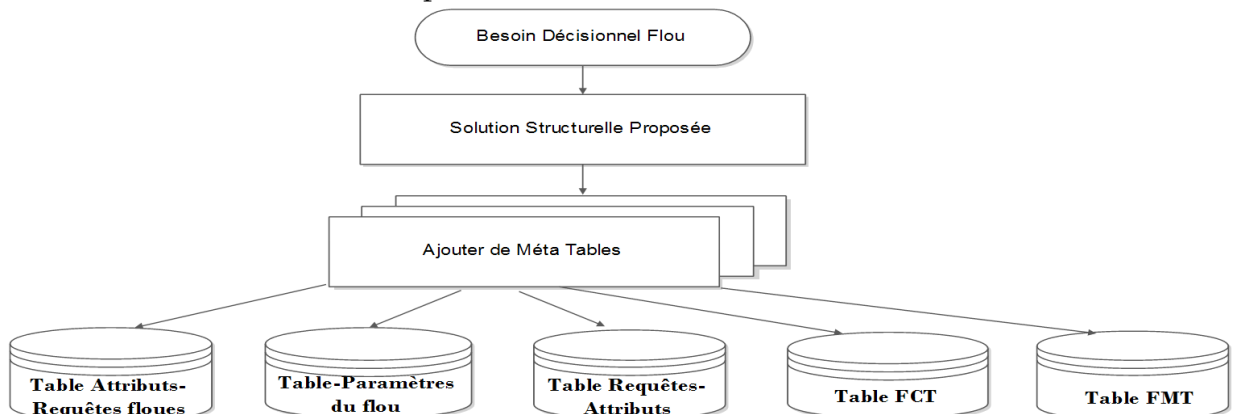


Figure 5.4 Tables additives

Ce travail met en évidence les besoins décisionnels où l'expression de ces besoins par les décideurs, qui ne sont pas nécessairement des informaticiens, génère des besoins ambigus, vagues ou imprécis et qui peuvent affecter n'importe quelle phase du processus de prise de décision. Pour remédier à ce problème et compte tenu de l'importance de cette phase et de son impact sur l'ensemble du processus, nous procédons à déterminer l'imprécision et le flou dans ce cas en envisageant une solution basée uniquement sur l'ontologie floue. Pour illustrer cette solution, on l'a appliquée à l'évaluation des risques dans le transport en utilisant un ensemble d'indicateurs pour classer les types de risques. On projette prochainement, de présenter la base de métadonnées en modèle ontologique au lieu de la simuler, dans notre solution primaire, le dictionnaire de données en fichiers plats ou en d'autres formats. Le schéma de la solution sera proche du modèle BDBO [Jean, 2007]

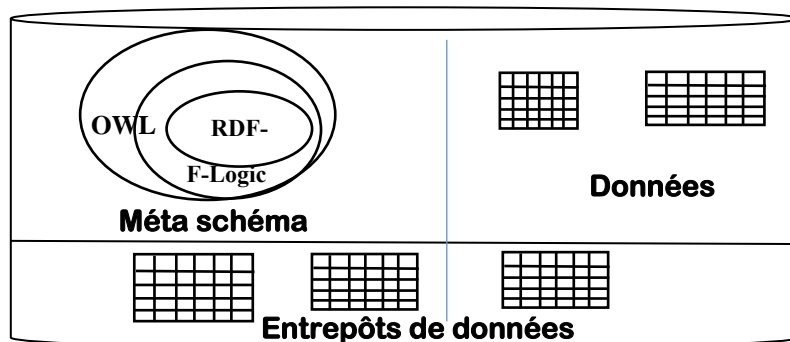


Figure 5.5 Entrepôts de données à base d'Ontologie (EDBO)

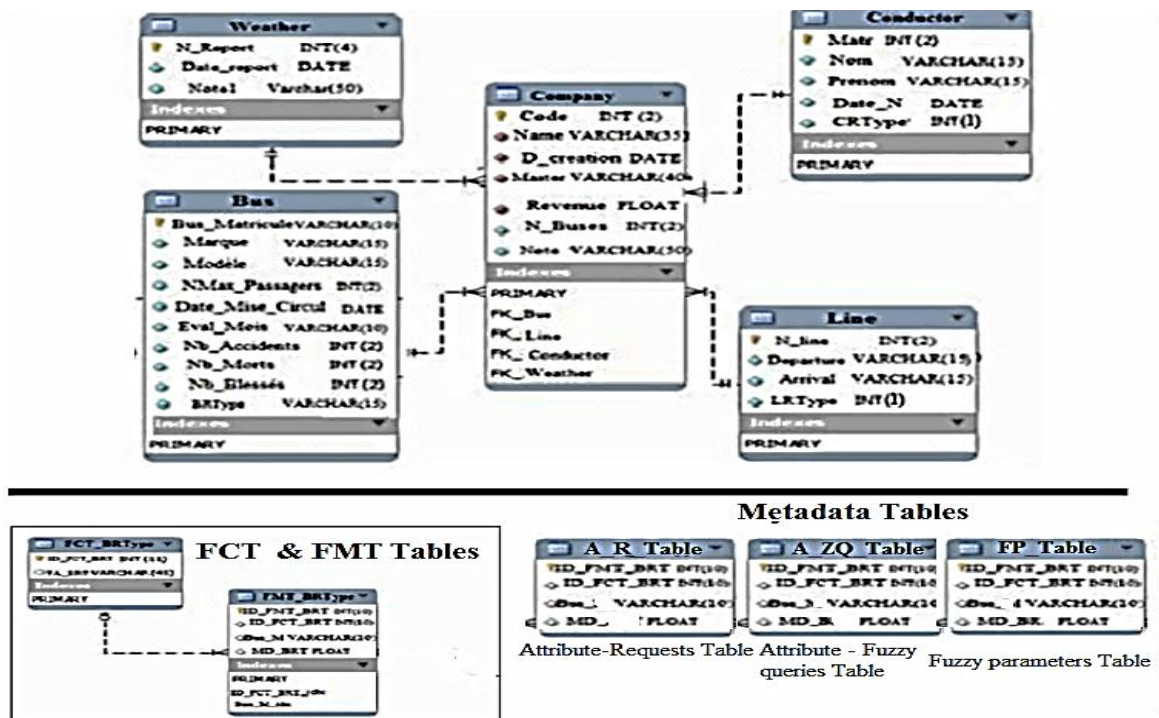


Figure 5.6 Nouvelle structure issue de notre solution

La proposition consiste donc à ajouter les structures (tableaux) au dictionnaire de données et non pas à la structure globale de l'entrepôt.

La solution FASEL permet de résoudre le problème du besoin flou par redéfinition de l'attribut cible (Target Attribute) mais le problème qui se pose c'est que la solution est gourmande en espace mémoire. En plus que l'entrepôt occupe un espace considérable, les tables à ajouter FCT et FMT pour chaque TA concerné par la requête floue ou le prédicat flou vont augmenter davantage cet espace.

Dans notre étude de cas, nous aurons pour l'analyse OLAP, par exemple, des bus inacceptables qui seront retirés par le décideur (mis au rebut) ou des conducteurs inacceptables qui seront licenciés. Les occurrences qui en résultent changent chaque fois que nous changeons les paramètres de l'évaluation des requêtes imprécision de spécification des besoins BI.

### 5.2.2 Exemple Illustratif : Evaluation du risque dns le transport routier intelligent

Notre étude de cas traite les risques d'accident au vu des dommages catastrophiques causés par les accidents de la route en Algérie, , notamment le grand nombre de morts chaque année (13 morts par jour en moyenne) et en essayant de classer les principaux facteurs de ce fléau, nous trouvons des difficultés à classer les acteurs du transport (Bus inacceptable, ligne négligeable, conducteur indésirable,...) car cette classification est basée sur des données historiques (data warehouse) de certains paramètres préalablement définis par l'expert routier. Ce genre de problème trouve dans notre proposition la solution pour lever le flou et affecter les observations aux classes existantes (Affectation floue).

Année	Nombre d'Accidents	Moyenne	Morts	Bléssés	Morts/Acc.	Bléssés/Acc	Morts/Jour
2012	42477	116	4447	69141	0,10	1,63	12,18
2011	41467	114	4598	66361	0,11	1,60	12,60
2010	32873	90	3660	52435	0,11	1,60	10,03

Tableau 5.2 Statistiques des accidents en Algérie

Ce problème n'est pas spécifique à l'Algérie par exemple en France, il y a le même fléau bien que la sécurité routière s'est améliorée ces dernières années.

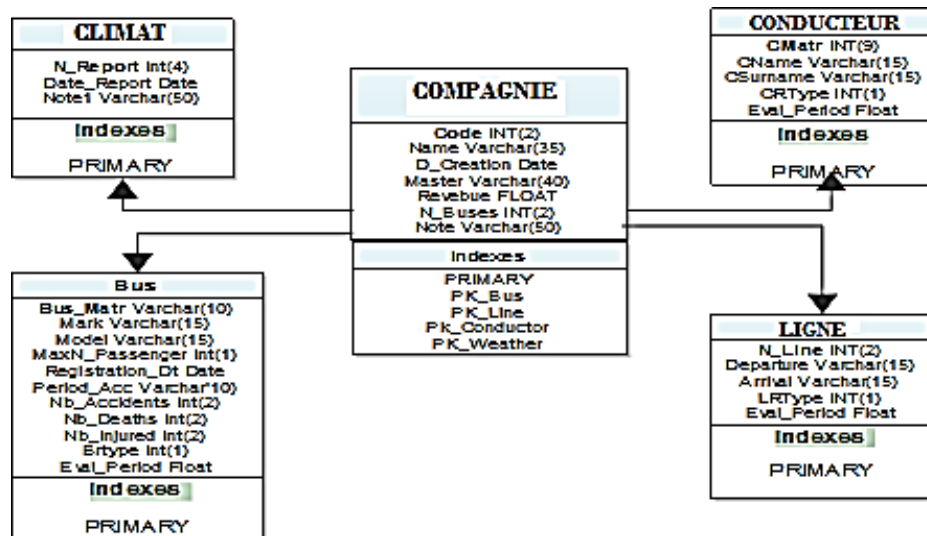


Figure 5.7 Classes de DW existante [Larbi & al., 2018]

Dans cette section, nous nous concentrons sur la description des mises à jour possibles d'un modèle de données d'entrepôt lorsqu'une décision est évaluée différemment. Pour illustrer ce changement dans l'évaluation des besoins de prise de décision et les impacts qu'ils ont sur l'évolution des DWs en partie sur le schéma et l'autre sur les données, nous proposons un DW existante sur les risques du transport routier. Le diagramme multidimensionnel de la figure 5.15 est utilisé pour analyser par exemple BRTYPE. Cet attribut indique le type de risque du bus selon un rapport d'un formulaire papier de déclaration d'incident / accident au contrôleur du transport de voyageurs d'une entreprise de transport privé ou public. Selon l'exigence du décideur, ce formulaire de notification change: Par exemple, si notre classification d'un transport routier de premier acteur (Bus) dépend uniquement du nombre d'accidents et de décès, nous examinerons les paramètres  $\alpha$  et  $\beta$  c'est-à-dire les attributs Nb\_accidents (A) et Nb\_Deaths (D; tué); si l'évaluation du bus (Eval\_period) dépend également du nombre de blessures, nous ajouterons l'attribut Nb\_Injured (I), le même accident dépend également d'autres facteurs tels que les valeurs climatiques (température (T), pluie (R), vent, neige, etc. ) ou l'état de la route (RS) qui provoquera une mise à jour DW en ajoutant des attributs (Temp, quantité de pluie, vitesse du vent, visibilité,...).

Exemple: Le décideur veut connaître le "conducteur indésirable" du dernier semestre.

Le besoin du décideur consiste à connaître, dans une période de temps déterminée, les bus et les chauffeurs de l'entreprise "à risque" (Inacceptable, Unsafe, etc.), pour notre exemple la période est le semestre. Les besoins décisionnels imprécises sont les suivants: bus dangereux, conducteur non désiré, ligne inacceptable, etc. Cette classification permet de prendre la décision adaptée à chaque acteur (Bus, Driver, Line,...).

### 5.2.3 Expérimentation: Evaluation de risque dans le transport intelligent (Solution Structurelle)

Nous avons créé notre DW simple avec Oracle puis nous l'avons rempli avec quelques exemples donnés par la direction de transport puis nous avons défini tous les concepts flous avant de charger le DW à notre application java pour traiter l'imprécision dans les besoins de prise de décision. Le traitement des besoins flexibles consiste à créer les tables additionnelles et à fournir au système les différents paramètres données par l'expert. Ainsi, pour chaque acteur (Bus, Conducteur,...) on cherche à déterminer le type de risque (remplir les attributs BRType/CRType et Eval\_period). La réponse à la requête du besoin est alors directe sans réécriture.

Nous appliquerons l'exemple présenté précédemment à notre modèle de solution. Pour des raisons de simplification, la présentation se limite à deux classes importantes: le bus et le conducteur et d'abord l'expert en transport doit définir deux paramètres  $\alpha$  et  $\beta$  (variant d'un expert à l'autre), puis définit les classes des acteurs (bus, conducteur,...). Par exemple, les classes de bus sont établies par l'expert (Négligible, Acceptable, Unsafe ou Unwanted, Unacceptable).

Pour qu'un bus fasse l'objet d'une inspection minutieuse, il doit y avoir, dans une période donnée (par exemple trimestre ou semestre), l'indicateur dépasse une certaine valeur  $\alpha$  et un nombre de décès pour la même période supérieur à  $\beta$ . Dans ce cas, le BRType est inacceptable. Si le Bus dépasse un certain seuil trimestriel  $\alpha$ , le cas est qualifié de "grave" si les causes ne sont pas liées aux conducteurs, ni à la route. Par conséquent, dans ce cas, le Bus est appelé "bus à risque ou dangereux" et doit être envoyé au service d'entretien. Si les accidents du bus n'ont causé que des dommages matériels, nous avons un Bus inacceptable ". Si nous avons moins deux accidents et sans dommages matériels ni blessés, le Bus est considéré comme "acceptable". Sinon, le risque est négligeable. De la même manière, nous procédons au classement du conducteur et de la ligne et avec d'autres paramètres.

Chaque évaluation d'imprécision aura une incidence directe sur la conception du DW. Pour l'analyse, tout changement d'évaluation aura une incidence directe sur l'analyse OLAP. Ainsi, une éventuelle mise à jour peut ajouter une dimension (Climat, indiquer le sol, etc.). Cela équivaut à augmenter le niveau de détail du tableau des faits (les données seront plus détaillées), car les mesures contenues dans le tableau des faits seront décrites par une dimension supplémentaire et présenteront donc plus de descripteurs. Nous pouvons avoir l'ajout de plus de détails, par la présence d'un doubleur, c'est-à-dire l'ajout d'un attribut à la table Conducteur.

$BRType = x \rightarrow BRType = f(A, D)$  or  $f(A, D, I)$  or  $f(A, D, I, T)$  etc. ..

Au lieu de cela, une autre mise à jour possible consiste à supprimer une dimension, ce qui réduit le niveau de détail de la table de faits (les données seront moins détaillées), car les mesures contenues dans la table de faits seront décrites par une dimension inférieure. Par exemple, il est possible de supprimer la dimension Climat, par exemple. Ainsi,  $BRType = f(A, D, I)$  seulement, c'est-à-dire que BRType serait

identifié en termes de A, D et I seulement. Il peut également y avoir suppression d'une mesure telle que l'attribut de la table de dimension Conductor du doubleur. On peut aussi avoir l'ajout de niveaux de granularité enrichissant une hiérarchie existante c'est le cas par exemple si le DW est flou au niveau de la table de dimension Bus et surtout de l'attribut BRType flou qui génère une hiérarchie à évaluer. Encore une fois, cette hiérarchie peut subir des changements si nécessaire pour le décideur en ajoutant un niveau de granularité plus facile à évaluer. Ce dernier cas n'est pas considéré dans ce document. La détermination des paramètres spécifiant un type d'acteurs est déterminée par le décideur dans l'ontologie. Les mêmes règles de raisonnement pour assigner un type d'acteur ou un autre. (Bus inacceptable, Conducteur indésirable...) Ces besoins décisionnels flexibles ne peuvent pas avoir à répondre à partir du schéma DW déjà existant puisque le classement d'un acteur nécessite une analyse attentive de plusieurs paramètres (utilisant plusieurs attributs tels que le nombre d'accidents, le nombre de décès, le nombre de blessés) puis selon les valeurs obtenues, l'affectation du résultat à une classe n'est pas toujours évidente car elle dépend de l'évaluation du décideur. Nous proposons donc une solution ontologique spécialement pour le cas où les données du DW ne sont pas floues. Si le DW est inexistant, l'évaluation de l'imprécision des besoins aide le concepteur à choisir la base de production et tout changement dans l'évaluation peut influencer sur ce choix, soit au niveau de la base de production elle-même (tables de sources supplémentaires), soit dans les cas d'une même base de production. Si le DW existe, l'évaluation de l'imprécision des besoins, l'évaluation des requêtes floues d'aide à la décision pour déterminer la base de production requise et les tableaux ou informations supplémentaires à ajouter à la base existante afin de satisfaire le décideur.

On suppose qu'un attribut participant à l'évaluation du flou de besoin "Bus non désiré" est l'attribut "BRType" et n'a ajouté que deux tables FCT et FMT. Heureusement; cependant le modèle Fasel nécessite beaucoup d'espace (tableaux supplémentaires et a de plus en plus Fuzzy Meta Table), c'est un très fréquent Lorsque nous avons seulement deux ou trois ou même des attributs cibles de four-four concernant l'exigence flexible requête qui a corrigé le défaut de cette solution.

Dans notre étude de cas, nous aurons, par exemple, des Bus inacceptables à retirer par le décideur (détruits) ou des conducteurs inacceptables qui seront renvoyés. Les occurrences résultantes changent chaque fois que nous changeons les paramètres de l'évaluation des requêtes imprécision de la spécification des besoins BI. Exemple. Un concept flou est ajouté à l'attribut dimensionnel BRType en tant qu'attribut cible. Le concept flou BRType contient quatre termes linguistiques: "inacceptable", "dangereux", "acceptable", "négligeable".

Il est caractérisé comme un concept flou adaptatif et les fonctions des membres sont les suivantes: Si l'attribut Target BRType = Num (1) n'est pas flou, nous avons par exemple: BRType=1 pour représenter le cas de "Inacceptable\_Bus", BRType=2 pour

représenter le cas de "Unwanted\_Bus", BRTYPE=3 pour représenter le cas de "Acceptable\_Bus".

$$\begin{aligned} \mu(\text{BRTYPE}=1) &= \text{If } A > \alpha \text{ and } D > \beta : \text{Eval\_period}=1 \\ & \text{If } A > \alpha \text{ and } M \leq \beta \text{ Eval\_period}=(A-\alpha/\alpha\text{Max}-\alpha); \\ & \text{If } A \leq \alpha : \text{Eval\_period} = 0 \end{aligned}$$

Figure 5.8 Règles de risque appliquées pour le cas de BRTYPE=1

Cela veut dire que si le nombre d'accidents  $A \leq \alpha$ , on aura un degré d'appartenance à cette classe (Bus inacceptable) est nul c.a.d que pour ce cas le bus est acceptable. Si par exemple  $A > \alpha$  and  $M \leq \beta$ , dans ce cas on calcule le degré d'appartenance qui est  $(A-\alpha/\alpha\text{Max}-\alpha)$

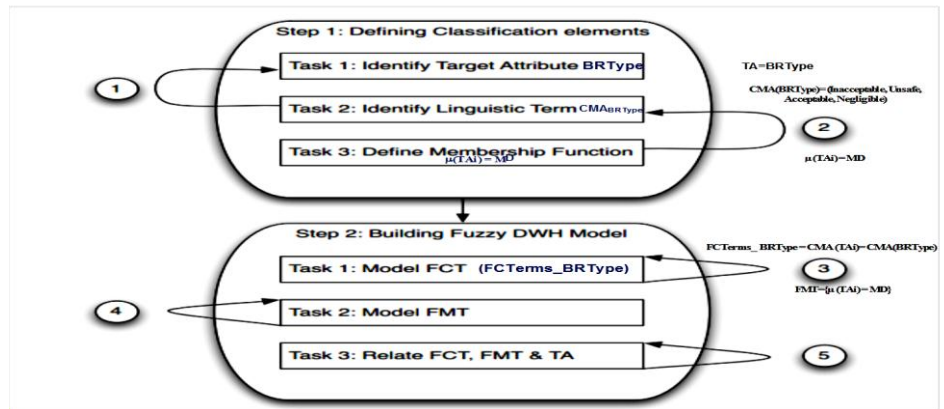


Figure 5.9 Tables FCT et FMT ajoutées pour l'attribut cible TA=BRTYPE

On définit les deux tables FCT et FMT qui seront dans notre cas :  
 $TA = \text{BRTYPE}$  ;  $CMA(\text{BRTYPE}) = \{\text{Inacceptable, Unsafe, Acceptable, Negligible}\}$   
 $MD = \mu(\text{BRTYPE}) = \text{Eval\_period}$   
 $FCT = CMA(\text{BRTYPE})$  et  $FMT = \{\mu(\text{BRTYPE})\}$

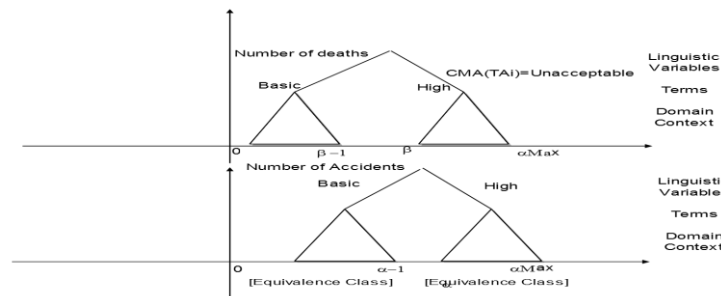


Figure 5.10 Classe d'équivalence

Les tables Meta MT1 et MT2 ajoutées contiennent:  $MT1 = \{FCT\_BRtype, FMT\_BRtype, ..\}$ ;  $MT2 = \{\alpha, \beta, \alpha_{Min}, \alpha_{Max}, \beta_{Min}, \beta_{Max}\}$  Il est donc possible d'effectuer des opérations d'analyse OLAP classiques sans avoir recours à des évaluations de requêtes et à des réécritures puisque le DW n'est pas flou. Ainsi, à chaque changement de paramètres d'expert, l'analyse DW et OLAP sera influencée et sera mise à jour. Notre entrepôt comprend 10 sociétés de transport avec 180 Bus et pour une période de 06 mois, il y a eu 10000 feuilles d'itinéraire/semestre, agrégées par semaine soit 1430

incidents et plus de 160 accidents, 50 décès et 600 blessés dont 20 bus concernés. Après l'application de notre solution et pour une période de 06 mois, nous avons détecté les résultats suivants:

Select Count(BRTy Select Count(BRType) As Bus_Number (Seuil =0.2)		Select Count(BRType) As Bus_Number (Seuil =0.7)	
	Bus Number	Bus Number	Decision
Inacceptable	03	01	Doit être retiré du service
Peu sûr	04	02	Doit être envoyé à la maintenance
Acceptable	10	09	/
Negligeable	03	06	/
TOTAL	20	20	
Select Count(CRType)As Conductor_Number(Threshold=0.2)		Select Count(CRType) As Conductor_Nb (Threshold = 0.7)	
	Conductor_Number	Conductor_Numb	Decision
Inacceptable	02	01	Révocation
Peu sûr	04	03	It must warm
Acceptable	28	24	/
Negligeable	06	12	/
TOTAL	40	40	

	Bus	Conducteur	Ligne
Inacceptable	3	2	1
Unsafe	4	4	0
Acceptable	10	28	6
Negligible	3	6	3

Tableau 5.3 Fonction Count(BRType)

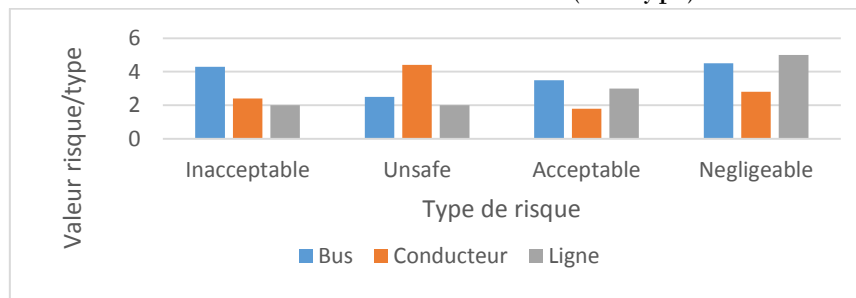


Figure 5.11 Evaluation de risque de transport routier (Cas d'accident)

**5.2.4 Conclusion :** L'idée principale est d'enrichir l'expression de recherche du besoin avec les informations contenues dans les tables de métadonnées de l'entrepôt de données. De plus, nous avons précisé la forme du rapport final du besoin en utilisant une requête exacte et nous n'avons pas tenu compte de l'incertitude ou autre facteur d'imperfection mais seulement de l'imprécision linguistique. Quelle que soit cette expérimentation (étude) qui n'était pas exhaustive (nous avons l'intention d'utiliser un vrai DW) et que nous n'avons pas pris en compte le facteur de ligne (Trajet) et le facteur climatique, car nous ne répartissons pas l'étude sur une longue période, les résultats sont très encourageants. L'avantage de cette solution est qu'elle est simple à implémenter mais le problème rencontré est que cette approche consomme beaucoup d'espace mémoire en plus du grand espace déjà occupé par le DW. Pour cette raison, on a décidé de passer de la solution structurelle à la solution à

base de connaissance avec moins d'espace et pour bénéficier du raisonnement et de la connaissance de l'expert.

---

### 5.3. Approche 2 : Evaluation des BDFs utilisant un système expert flou

---

Dans ce qui suit nous présentons notre approche basée le système expert flou, On présente le schéma de traduction dirigée par la syntaxe, l'algorithme de traduction SQL flou puis nous présentons notre solution appliquée à l'évaluation de risque dans la gestion intelligente du projet et finalement une conclusion.

#### 5.3.1 Présentation de l'approche proposée

L'approche proposée est composée à la fois d'un système expert flou (une base de connaissance et des règles d'inférence) et un schéma de translation dirigée par la syntaxe. La base de connaissance et les règles d'inférence pour le cas d'évaluation de risque dans la gestion intelligente de budget est tirée du livre [Bosc, 1995] où il a présenté explicitement ce cas à l'issue d'une étude réelle.

#### STDS : Schéma de translation dirigée par la syntaxe

Nous allons définir un STDS que nous avons proposé comme composant de notre solution Fuzzy SQL. Un STDS est une grammaire qui génère un langage, et a une règle de production pour chaque élément de traduction attaché. Chaque fois qu'une production fait appel à un bypass, également connu sous le nom de translation générée par l'élément attaché. Plus formellement, un SDTS est un 5-tuple  $T = (N, \Sigma, \Delta, R, S)$

- $N$  : Ensembles finis de symboles non-terminaux:
- $\Sigma$  : Ensemble fini d'alphabet d'entrée
- $\Delta$  : Ensemble fini d'alphabet de sortie
- $S$  : axiome de la grammaire (non terminale)
- $R$  : L'ensemble fini de production est de type  $A \rightarrow \alpha, \beta$  avec  $\alpha \in (N \cup \Sigma)^*$  and  $\beta \in (N \cup \Delta)^*$ . Non-terminal  $\beta$  est une permutation de  $\alpha$ .

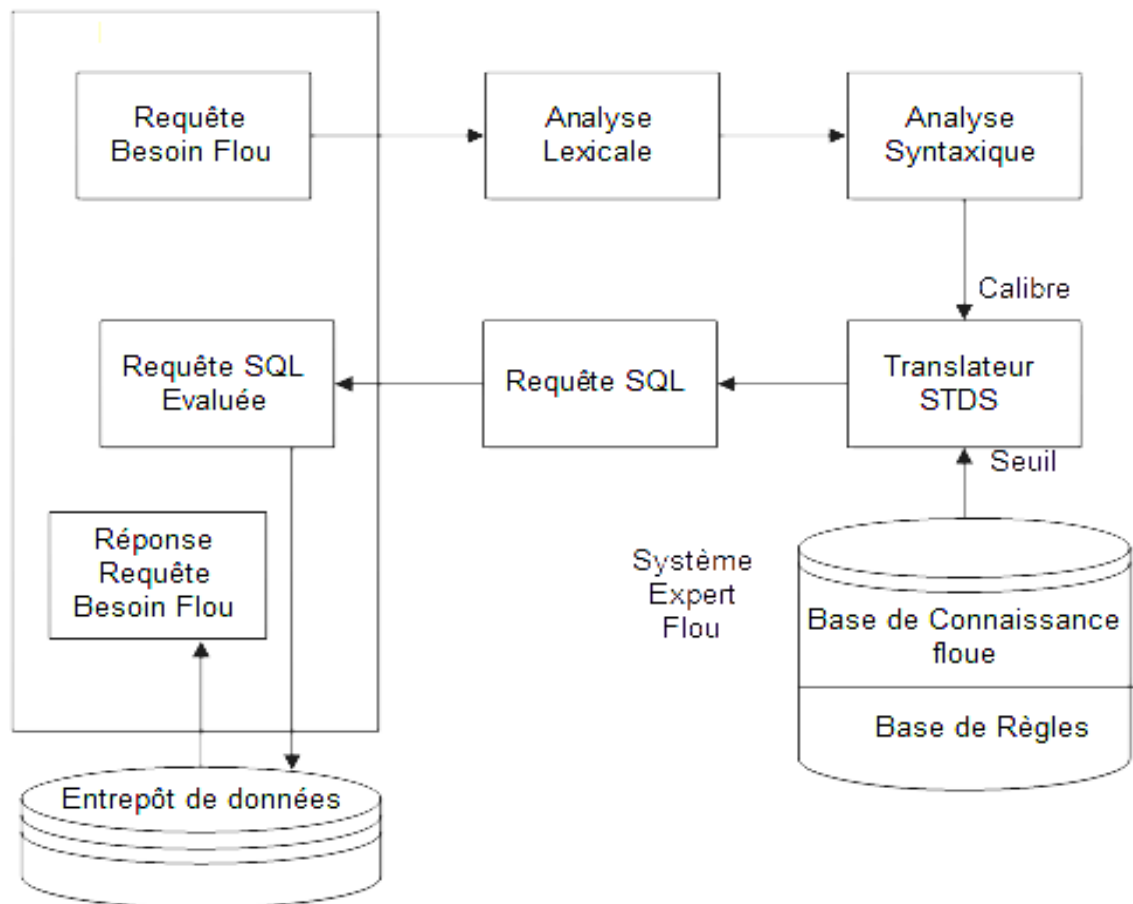


Figure 5.12 Architecture de notre interface – Niveau d’analyse de requête SQL Floue.

[Larbi & al., 2013]

Le traducteur nous permet à travers les productions grammaticales utilisées de ne rien faire pour les clauses communes entre SQL et SQL Flou et d'effectuer les changements nécessaires pour les autres cas. Ces changements concernent les utilisations du seuil et du calibre pour transformer (ou) les valeurs qualitatives en valeurs quantitatives.

#### **Grammaire utilisée dans notre approche**

Le programme utilisateur peut utiliser Fuzzy SQL, le langage de manipulation des données (DML) pour interroger et mettre à jour l’entrepôt de de données (DW). Ainsi, la première étape lorsqu'un utilisateur formule sa requête est de savoir si cette requête est lexicalement correcte et que les mots appartiennent au lexique du traducteur SQL flou. S'il y a des erreurs, un message d'erreur s'affiche. Après cette étape, nous procédons à un analyseur pour vérifier que l'utilisateur s'est conformé aux règles d'écriture d'une requête dans Fuzzy SQL. Ces règles ne sont que légèrement différentes des règles du standard SQL. S'il y a des erreurs, nous analysons la requête entrée par l'utilisateur (SELECT, INSERT, DELETE ou UPDATE) clause par clause pour déterminer où se trouve l'erreur. Il affiche ensuite un message d'erreur.

Pour traduire la requête SQL floue saisie par l'utilisateur, nous avons utilisé la grammaire Backus Naur Form (BNF). Voici une partie de la grammaire utilisée :

```

<début> ::= <requête_select> | <requête_insert> | <requête_delete> | <requête_update>
<requête_select> ::= SELECT (<calibre>)? (<all_distinct>)? <selection>
<clause_from> (<clause_where>)?
<selection> ::= '*' | <liste_colonne>(',<liste_colonne >)*
<liste_colonne> ::= <agrégat> '*' | <agrégat><colonne> | <colonne>
<colonne> ::= (<nom_table>'.') ?<nom_colonne>
<clause_from> ::= FROM <nom_table> (',<nom_table>)*
<clause_where> ::= WHERE <conditions_recherche>
<conditions_recherche> ::= (<condition_booléenne> | <condition_floue>)
(<and_or>(<condition_booléenne> | <condition_floue>))*
<condition_booléenne> ::= <colonne> IS [NOT] NULL |
<colonne><opérateur><exp1> | <colonne> [NOT] LIKE
<exp2> | <colonne> [NOT] BETWEEN <exp1> AND
<exp1> | <colonne> [NOT] IN ('<exp1>(',<exp1>)*')
<condition_floue> ::= <colonne> IS <terme_flou>
<calibre> ::= \d+
<all_distinct> ::= ALL | DISTINCT
<agrégat> ::= AVG | MAX | MIN | SUM | COUNT
<nom_table> ::= \w+
<nom_colonne> ::= \w+
<opérateur> ::= < | <= | > | >= | != | <>
<exp1> ::= [\w\s\.\']*+
<exp2> ::= [\w'%']*+
<terme_flou> ::= (\w+.)?(\w+.)?\w+
<and_or> ::= AND | OR
<requête_insert> ::= INSERT INTO <nom_table>
('(<nom_colonne>(',<>nom_colonne)*')? (VALUES
('(<exp1>(',<exp1>)*') | <sous_select>)
<sous_select> ::= SELECT (<all_distinct>)? <selection> <clause_from>
(<clause_where>)?
<requête_delete> ::= DELETE FROM <nom_table> (<clause_where>)?
<requête_update> ::= UPDATE <nom_table> SET <nom_colonne>='<exp1>
(<clause_where>)?

```

Figure 5.13 Une partie de la grammaire utilisée [Larbi & al., 2013]

**Algorithme de translation SQL Flou:** L'algorithme utilisé pour traduire une requête SQL Flou en une requête SQL classique est le suivant:

```

Algorithme principal
Début Lire la requête
      Analyser la requête
Si la requête est lexicalement correcte alors
  Si c'est une requête SELECT alors
    Appeler le module SELECTVALIDE
  Sinon
    Si c'est une requête INSERT alors
      Appeler le module INSERTVALIDE
    Finsi Sinon
    Si c'est une requête DELETE alors
      Appeler le module DELETEVALIDE
    Finsi Sinon
      Appeler le module UPDATEVALIDE
    Finsi
  Sinon
    Si la requête est lexicalement fausse alors
      Afficher un message d'erreurs
    Finsi
  Sinon Si la requête est vide alors
    Afficher un message d'erreurs
    Finsi
  Finsi
Fin

```

Algorithme 5.1 : Algorithme Principal [Larbi & al., 2013]

Cet algorithme appelle les algorithmes SelectValide lorsqu'on utilise la clause Select. De même pour les clauses insert, delete et update, le programme principal appelle respectivement les algorithmes insert, delete et update.

<pre> <b>Algorithme SELECTVALIDE</b> Début Si la requête est lexicalement correcte alors   Si il y a une(des) condition(s) floue(s) alors     Traduire chaque condition floue en condition booléenne   Si la condition floue ne peut être traduite alors     Afficher un message d'erreurs   Finsi   Exécuter la requête booléenne et extraire les résultats   Si il y a une(des) erreur(erreurs)     Afficher un message d'erreurs   Finsi   Calculer la fonction d'appartenance de chaque champ   d'une colonne qui entre dans une condition floue   Si le nombre de conditions floues &gt;= 2 alors     Pour chaque ligne faire       Calculer le min entre les champs des colonnes qui entrent       dans les conditions floues     Finpour   Trier les résultats   S'il y a un calibre alors     Si le calibre &lt; le nombre de lignes trouvées alors       Sélectionner le nombre de résultats voulus     Finsi   Finsi </pre>	<pre>       Afficher les résultats     Sinon       Si il y a un calibre alors         Ne pas prendre en compte le calibre       Finsi       Exécuter la requête booléenne et extraire les résultats       Afficher les résultats     Finsi     Sinon       Analyser la clause SELECT       Si la clause SELECT est syntaxiquement fausse alors Afficher       un message d'erreurs     Finsi       Analyser la clause FROM       Si la clause FROM est syntaxiquement fausse alors         Afficher un message d'erreurs       Finsi       Analyser la clause WHERE       Si la clause WHERE est syntaxiquement fausse alors         Afficher un message d'erreurs       Finsi Finsi     Fin </pre>
---	---

Algorithme 5.2 : Algorithme SELECTVALIDE [Larbi & al., 2013]

<p>Algorithme INSERTVALIDE</p> <p>Début</p> <p>Si la requête est lexicalement correcte alors</p> <p>Si il y a une(des) condition(s) floue(s) alors</p> <p>Traduire chaque condition floue en condition booléenne</p> <p>Si la condition floue ne peut être traduite alors</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Exécuter la requête booléenne</p> <p>Si il y a une(des) erreur(erreurs)</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Sinon</p> <p>Exécuter la requête booléenne</p> <p>Si il y a une(des) erreur(erreurs)</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Finsi</p> <p>Sinon</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Fin</p>	<p>Algorithme DELETEVALIDE</p> <p>Début</p> <p>Si la requête est lexicalement correcte alors</p> <p>Si il y a une(des) condition(s) floue(s) alors</p> <p>Traduire chaque condition floue en condition booléenne</p> <p>Si la condition floue ne peut être traduite alors</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Exécuter la requête booléenne</p> <p>Si il y a une(des) erreur(erreurs)</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Sinon</p> <p>Exécuter la requête booléenne</p> <p>Si il y a une(des) erreur(erreurs)</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Finsi</p> <p>Sinon</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Fin</p>	<p>Algorithme UPDATEVALIDE</p> <p>Début</p> <p>Si la requête est lexicalement correcte alors</p> <p>Si il y a une(des) condition(s) floue(s) alors</p> <p>Traduire chaque condition floue en condition booléenne</p> <p>Si la condition floue ne peut être traduite alors</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Exécuter la requête booléenne</p> <p>Si il y a une(des) erreur(erreurs)</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Sinon</p> <p>Exécuter la requête booléenne</p> <p>Si il y a une(des) erreur(erreurs)</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Finsi</p> <p>Sinon</p> <p>Afficher un message d'erreurs</p> <p>Finsi</p> <p>Fin</p>
---	---	---

Algorithme 5.3  
INSERTVALIDE  
[Larbi & al., 2013]

Algorithme 5.4  
DELETEVALIDE  
[Larbi & al., 2013]

Algorithme 5.5  
UPDATEVALIDE  
[Larbi & al., 2013]

### 5.3.2 Exemple illustratif : Evaluation de risque de projet intelligent (Solution à base de Système Expert)

**Description de la solution :** Le système expert proposé contient [Bosc, 1995]:

a) **La base des règles :** Ces règles sont sous la forme :

R1 : si < budget\_projet est élevé > ou < durée\_projet est longue > alors risque est acceptable ;

R2 : si < durée\_projet est assez\_longue > ou < budget\_projet est élevé > ou < durée\_projet est longue > ou < effectif\_budget est confortable > alors risque est important

R3 : si < durée\_projet est assez\_longue > ou < budget\_projet est élevé > ou < durée\_projet est longue > et < effectif\_budget est confortable > alors risque est élevé.

Où les conclusions sont de la forme d'une liste projets de la base de données entreprise. Les faits : la base des faits est la mémoire de travail du système expert. Pour notre cas, la base des faits est la base de données de l'entreprise, elle est de type symboles où notre système utilise des faits symboliques ou réels, sans utiliser des variables.

b) Le **moteur d'inférences** : un moteur d'inférences est un mécanisme qui permet d'inférer des connaissances nouvelles à partir de la base de connaissances du système. Pour notre cas, on utilise la technique du **chaînage arrière**. Le mécanisme de chaînage

arrière consiste à partir du fait que l'on souhaite établir, à rechercher toutes les règles qui concluent sur ce fait, à établir la liste des faits qu'il suffit de prouver pour qu'elles puissent se déclencher puis à appliquer récursivement le même mécanisme aux faits contenus dans ces listes.

### **Exemple**

**Salaire** : possède les étiquettes linguistiques définies sur les distributions de possibilité trapézoïdales suivantes :

**BAS**(50, 80, 120,180), **MOYEN**(150, 300, 400,550), **ÉLEVÉ**(400,600,800,1000).

Ces valeurs sont en Dinars Algériens.

**Nombre\_heures** : possède les étiquettes linguistiques : **PETIT**(0, 2, 7, 8), **MOYEN**(5, 10, 15, 20), **ÉLEVÉ**(15, 20, 25, 30).

Le logiciel fournit une interface simple et interactive fuzzy SQL, qui permet d'éditer les requêtes SQL et fuzzy SQL. En effet, il étend les concepts de base du standard SQL en ajoutant de nouveaux concepts flous définis dans le fuzzy SQL. Fuzzy SQL comme n'importe quel demandeur doit analyser la requête avant de l'exécuter. Pour l'analyse lexicale, syntaxique et sémantique, si une erreur se produit, elle sera traitée et affichée. Ainsi on peut représenter ces critères selon le modèle relationnel comme suit :

**DÉPARTEMENT** (*num\_d*, *nom*, *matricule\_directeur*, *date\_début*, *local*)

**POSTE** (*code\_poste*, *profession*, *études*, *expérience*)

**EMPLOYÉ** (*matricule*, *num\_d* , *code\_poste*, *nom*, *prénom*, *sexe*, *adresse*, *date\_naissance*,

*salaire*, *études*, *rendement*, *Expérience*, *ssoc\_superviseur*)

**PROJET** (*num\_p*, *num\_d*, *nom*, *budget*, *durée\_de\_réalisation*, *effectif\_projet* , *effectif\_budget*,

*nombre\_de\_phases*)

**DÉPENDANT** (*matricule*, *prénom*, *sexe*, *date\_naissance*, *lien\_parenté*)

**TRAVAILLE\_DANS** (*matricule*, *num\_p*, *nombre\_heure*)

Remarque : Les attributs écrits en italique et qui sont soulignés, représentent des clés étrangères.

Exemple : pour chercher les projets de risque acceptable la requête sera sous la forme:

```
Select num_p From projet
Where risque is acceptable;
```

### 5.3.3 Expérimentation

Nous avons développé une solution dans le cadre de l'évaluation de l'imprécision ou l'estimation du risque pour la gestion intelligente du projet. Cette solution présente deux composants:

- \* Un langage de requête flou pour que le décideur ait des réponses à ses requêtes, même si elles sont flexibles.

- \* Un système expert pour lui permettre de juger du sort du projet en estimant le risque de ce dernier, en utilisant la base de connaissances du système. Nous avons été en mesure de valider notre modèle en appliquant l'évaluation des risques pour la gestion intelligente du projet en utilisant les valeurs empiriques établies par les experts du Groupe "Metus System". Les résultats obtenus étaient très encourageants.

Le logiciel SQL Flou fournit une interface SQL interactive floue simple. Il permet d'éditer des requêtes SQL (floues, classiques ou combinées) pour nous permettre d'évaluer les risques des projets intelligents.

#### Le Framework SQL flou

SQL flou est un programme indépendant qui sert comme une interface entre l'utilisateur et l'entrepôt de données. L'utilisateur introduit une requête SQL flou et le programme SQL flou communique avec le DW dans le but d'obtenir le résultat final. Le SGBD choisi dans l'implémentation du SQL flou est ORACLE, à cause de sa variabilité, sa grande extension, son niveau de sécurité très élevé.

SQL flou a été réalisé avec le langage de programmation C# de la plateforme .Net en utilisant l'IDE Visual Studio 2010. Il supporte les systèmes d'exploitation Windows 7.0, et pour l'utiliser il faut avoir Oracle version 8i ou ultérieure et le .Net Framework version 2.0 ou ultérieure. Voici un aperçu sur une des fonctionnalités de notre interface :

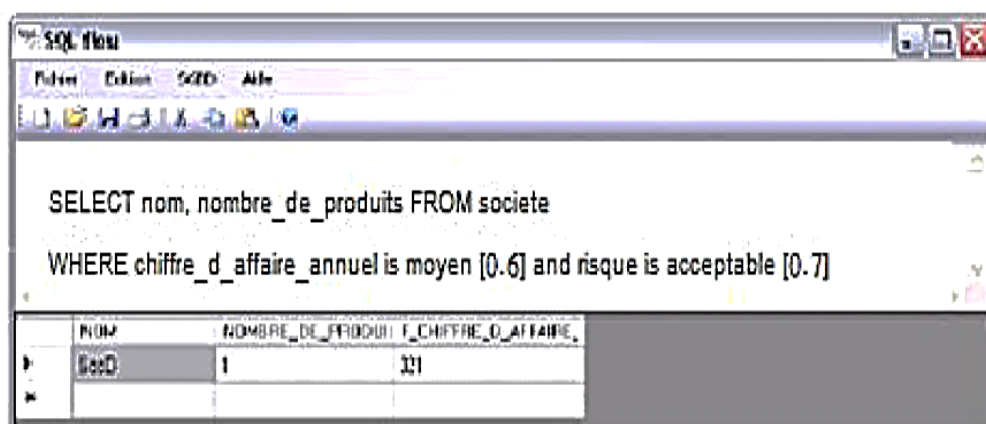


Figure 5.14 : Evaluation de requête floue utilisant notre framework avec un seuil=0.7 [Larbi & al, 2013]

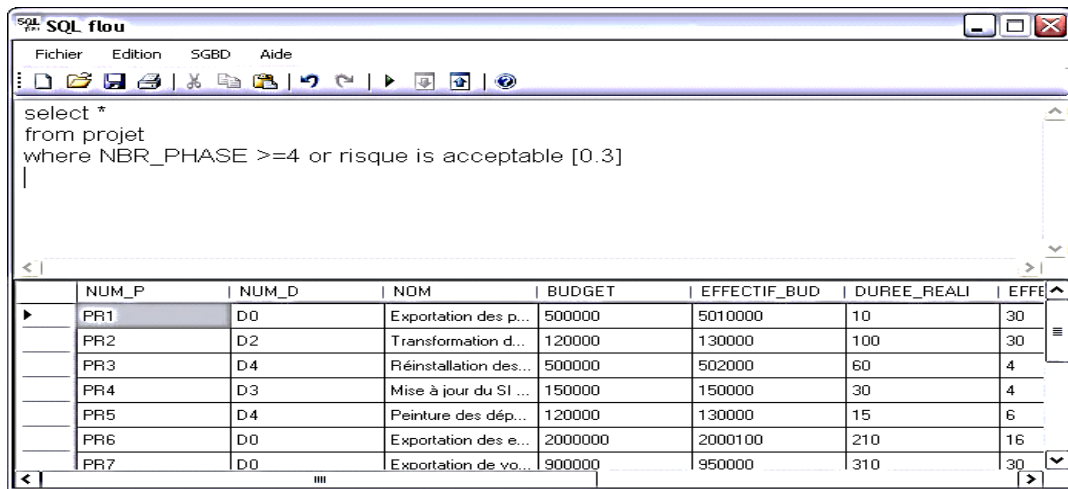


Figure 5.15 : Evaluation de requête floue utilisant notre framework avec un seuil =0.3 [Larbi & al, 2013]

Ainsi, en faisant varier le seuil et en interrogeant à chaque fois l’ED, nous aurons les résultats suivants :

Seuil	Nombre d'entreprise en entrée	nombre d'entreprise résultat
0.6	200	1
0.8	200	0
0.3	200	5
0.2	200	10

Tableau 5.4 Influence du seuil sur la réponse de requête floue

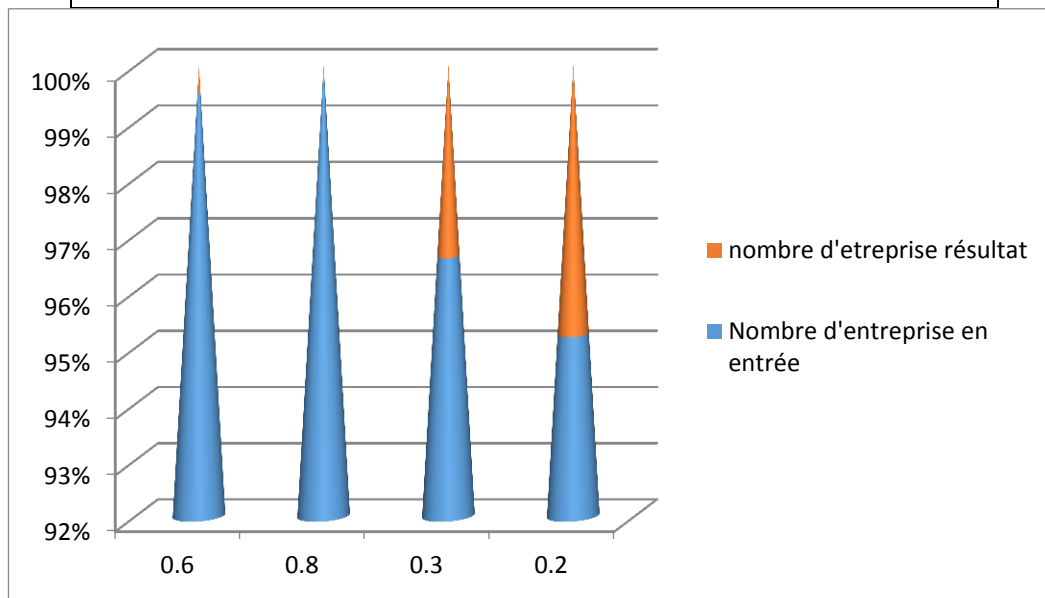


Figure 5.16 Influence du seuil sur la réponse à une requête floue

Les résultats primaires de cette solution naive sont encourageants et nous en avons permis d'évaluer l'imprécision de la valeur du prédicat de la requête représentant le besoin du décideur. On a testé cette solution seulement pour une seule valeur de prédicat floue pour le cas de requête floue et pour le cas de requête mixte (un prédicat exacte et un autre flou) en variant à chaque fois le seuil.

#### **5.3.4 Conclusion**

Cette contribution montre qu'il est réaliste de définir un langage de requête supportant des termes imprécis en tant qu'extension d'un langage de requête relationnelle régulier utilisant un STDS. Dans le contexte de SQL, cela est essentiel pour les utilisateurs qui sont familiers avec SQL trouveront un langage enrichi. Nous voulions que Fuzzy SQL soit expressif, un langage source simple. Cette solution ne touche pas le cas des requêtes complexes floues où il y a une structure imbriquée ou autre. Nous avons testé tous les cas d'erreurs (erreurs lexicales, syntaxiques et quelques cas d'erreurs simples sémantiques). Le résultat a été encourageant. Nous développons également une interface graphique qui permet aux utilisateurs de construire des requêtes sans connaissance ni des langages de requêtes floues ni de la logique floue. Ce travail est une étude préliminaire visant à définir un système supportant des données floues et dans la seule nouveauté est l'introduction d'un traducteur (STDS) permettant d'introduire des requêtes d'inclusion graduelle. Ce travail peut être amélioré en utilisant un compilateur et visant les parties non encore traitées par les travaux antérieurs.

---

## 5.4 Approche 3 : Evaluation des BDFs utilisant une ontologie floue

---

Cette section présente une solution rigoureuse basée sur l'ontologie floue. Basée sur la dernière théorie de Zadeh [Zadeh, 2008], les auteurs dans [Martinez-Cruz & al., 2016] et [Martinez-Cruz & al., 2014] proposent une solution consistant à utiliser des ontologies pour fournir « une compréhension de la façon dont la signification d'une proposition peut être composée de la signification de ses constituants ». Une des limites de cette solution est la présence du flou dans la partie d'adjectif de la phrase seulement. Dans un certain sens, notre proposition peut être considérée comme une continuation de ce travail. Nous limitons notre étude, dans cette section, au cas de l'opérateur NEAR-. A notre connaissance, ce cas n'a pas encore été traité dans le contexte de l'entrepôt de données.

### 5.4.1 Présentation de l'approche

Nous avons limité l'étude dans cette section uniquement à l'opérateur flou des requêtes des besoins. Mais avant de présenter notre solution, nous précisons les causes qui nous ont conduits à proposer une nouvelle solution: Le modèle GLMP n'évalue pas le besoin flou même s'il est présenté en langage naturel lorsque la flexibilité n'est pas présente dans (la valeur de prédicat de la requête i.e dans l'adjectif de la phrase).

- Le modèle GLMP ne considère que l'entrée des données des bases de données, mais le cas de la prise de décision et de l'utilisation de l'entrepôt de données n'est pas pris en compte.

- En utilisant d'autres modèles ontologiques de la littérature, nous ne pouvons pas résoudre le cas où le besoin est plus complexe et nous n'avons pas la possibilité d'exprimer le besoin en langage naturel, qui est le cas le plus courant et le plus simple pour les décideurs. L'algorithme de Memon pour évaluer l'opérateur flou NEAR ne peut pas nous suffire, en effet, nous devons traiter NEAR Negative seulement et adapter la solution NEAR qui dans notre cas a besoin d'autres paramètres non mentionnés par la solution [Memon & al., 2007]. Nous avons l'intention de voir d'autres opérateurs, à notre connaissance qui ne sont pas encore étudiés. Ainsi, la nouvelle solution proposée est basée sur l'ontologie floue appelée GLMR "ontologie du modèle linguistique granulaire de l'ontologie des besoins - GLMR" et présente deux cas (Voir Figure 5.17):

1<sup>st</sup> Cas: "Entrepôt de données flou": on a deux situations:

- \* Lorsque l'attribut concernant les besoins flexibles est flou., ce cas n'est pas présenté dans ce mémoire de thèse.

- \* Pour le cas où les attributs flous ne concernent pas les besoins flexibles, nous proposons comme solution notre approche ontologique.

2<sup>ème</sup> cas: "Entrepôt de données non flou": Dans notre cas, l'entrepôt de données n'est pas nécessairement flou, aussi la table de dimension n'est pas forcément floue, (comme

la solution proposée dans [Fasel, 2012]. La solution proposée est notre approche ontologique : Ontologie GLMR

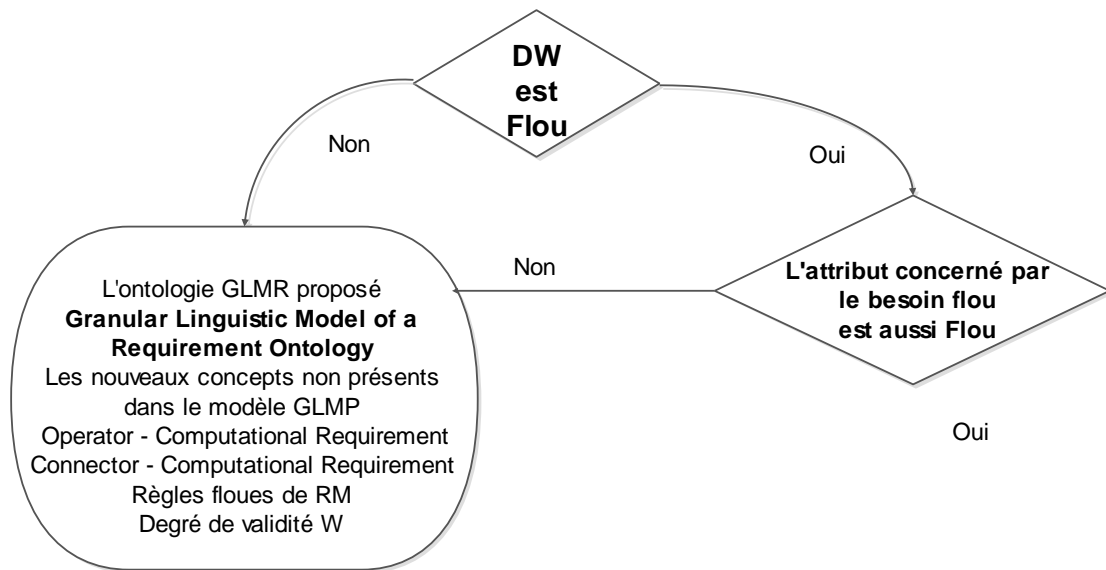


Figure 5.17 Approche ontologique proposée GLMR

Le besoin de décision représenté dans le formulaire de requête présente une partie flexible, dans notre cas, il s'agit de l'opérateur flou NEAR- contenu dans le prédicat de requête. Avant de gérer cette exigence, l'utilisateur de l'entrepôt de l'entreprise de transport chargera l'ensemble de données pertinent (pour la période concernée) et pour enrichir la requête et évaluer la partie floue, il consulte l'ontologie GMLR en fonction des paramètres fournis par le transport expert. Une fois l'information traitée par le raisonneur d'ontologie, déjà conçue et implémentée, nous aurons comme résultat, RM (Requirement mapping) et O-CPR (Opérateur-Calcul Perception de l'exigence), qui nous permet d'affecter chaque valeur acquise à un ou plusieurs des classes prédéfinies avec un degré d'appartenance pour chaque classe prédéfinie, afin que nous puissions répondre à la demande et satisfaire le besoin de décision. Notre solution proposée présente trois acteurs: le décideur, l'utilisateur, l'expert en plus du concepteur d'ontologies qui ne peut être que l'utilisateur du framework. Comme support, nous utilisons l'ontologie GLMR, l'entrepôt de données à partir duquel nous chargeons notre dataset, un pc et un moyen de transmission entre l'utilisateur et le décideur ou entre l'expert et l'utilisateur. Comme le montre la figure 5.20, au début, nous avons en entrée l'ensemble de données (les données correspondantes dans le semestre) prises à partir de l'entrepôt de données de la société de transport pour lequel, nous voulons classer le bus de la compagnie en fonction du type de risque de chaque bus. Une fois, l'ensemble de données à la demande de la décision doit être chargé. Nous collectons également certains paramètres relatifs aux risques du transport routier de l'expert routier ( $pv_1$ ,  $pv_2$ ).

L'ontologie GLMR est utilisée pour évaluer l'imprécision présente, dans ce cas, au niveau de l'opérateur, en utilisant l'algorithme NEAR. Une fois le traitement effectué, nous aurons les CPR\_BRTypes correspondant aux données acquises de l'expert en transport. Ces sorties du type f (Di, li, MDi), qui permettent à l'utilisateur de classer toutes les données acquises et qu'il transmet au décideur pour la prise de décision. Premièrement, nous étudions le modèle GLMP pour définir notre modèle de proposition appelé modèle GLMR et cela pour plusieurs raisons: La raison la plus importante est que contrairement à toutes les solutions d'ontologie floue existantes qui traitent uniquement les ontologies linguistiques, le modèle GLMP nous permet d'étendre l'étude aux ontologies canoniques et non-canoniques du modèle de l'oignon [Jean, 2007]. D'autre part, l'expression du besoin est généralement en langage naturel sur lequel l'ontologie GLMP est basée. Enfin, le modèle GLMR nous permet de représenter les requêtes OLAP, alors que GLMP n'a pas l'intention de représenter. Ici, quelques différences entre les deux modèles:

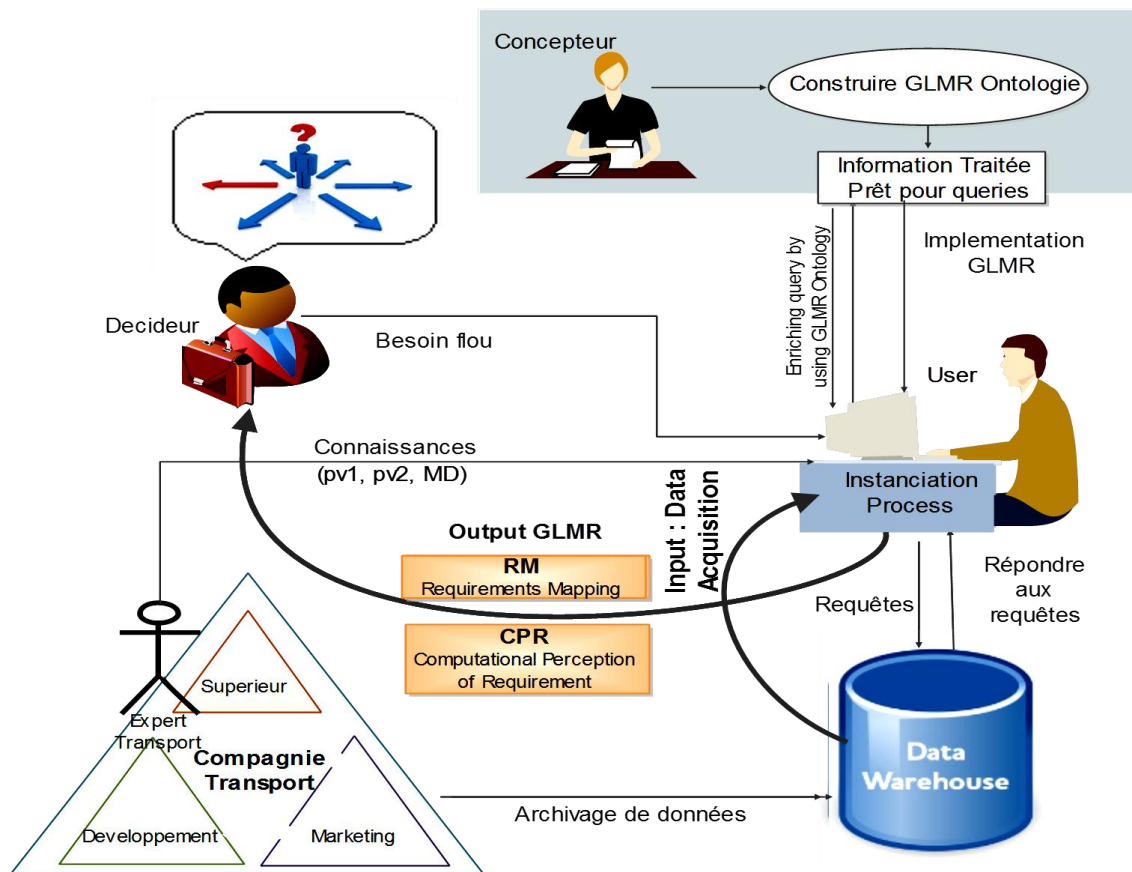


Figure 5.18 Architecture de notre système

**Pour expliquer le système proposé :** Existe-il une différence entre notre modèle et le modèle GLMP ou seulement c'est une extension de celui-ci ? Pour répondre, on va dresser un tableau comparatif mentrant les points communs et les différences entre les deux modèles :

Modèle GLMP	Modèle GLMR
1. La table de dimension est floue	1. La Table de dimension n'est pas nécessairement floue
2. Traiter la représentation des connaissances de l'expert	2. Traiter les besoins du décideur
3. La représentation computationnelle utilise seulement un langage naturel	3. Les besoins à base de requêtes ne sont pas nécessairement écrites en langage naturel
4. Imprecision est présente dans le texte représentant la perception de l'expert.	4. L'imprécision est présente dans le texte représentant la besoin du décideur.
5. Le modèle GLMP ne supporte pas les clauses de requêtes OLAP	5. Le modèle GLMR prend en charge les clauses de requête OLAP
6. Représente le modèle de perception	6. Représente le modèle du besoin
7. GLMP permet d'interpréter les perceptions de second ordre	7. GLMR permet d'interpréter les besoins du second ordre
8. INPUT: Valeurs fournies par un capteur de température ou par une BD.	8. INPUT: Valeurs fournies par l'entrepôt de données.
9. OUTPUT : Générer des phrases décrivant cette valeur	9. OUTPUT : Correspond à la valeur en entrée d'une classe avec un degré d'appartenance

Tableau 5.5 Comparaison entre GLMP & GLMR

On peut dire que notre modèle GLMR n'est autre que le modèle GLMP seulement dans le cas où la valeur de prédicat de requête du besoin est la seule partie de la requête qui présente le flou. Ce qui n'est pas toujours le cas.

Notre modèle supporte les besoins décisionnels flexibles, en ajoutant de nouveaux concepts: (Tableau 5.6)

#### A. Le concept O-CPR: Computational Perception of Requirement - Operator case:

Il est appelé O-CPR (Operator - CPR) qui définit les opérateurs entre les termes linguistiques. On représente ce concept comme suit :

O-CPR (E, B, O, L), où O: opérateur flou;  $E = (e_1, e_2, \dots, e_n)$  est un vecteur de prédicats ou d'expressions linguistiques (mots ou phrases en Langage naturel (LN) ou en Pseudo ou en langage SQL) qui représente l'ensemble du domaine linguistique du CPR. Chaque  $e_i$  décrit la valeur de CPR dans chaque position avec un degré de granularité spécifique. Ces phrases peuvent être soit faciles, comme  $a =$  "La température est proche de 18" ou plus complexe, par exemple,  $a =$  "Parfois, la température ambiante n'est pas proche de 25";  $W = (w_1, w_2; \dots, W_n)$  est un vecteur de degrés de validité  $w_i [0; 1]$  attribué à chaque  $e_i$  dans le contexte spécifique. Le concept de validité dépend de l'application, par exemple, il est fonction de la véracité et de la pertinence de chaque phrase dans son contexte d'utilisation.

Pour des raisons de simplification, nous présentons notre solution GLMR en se contentant du cas où le flou se produit au niveau de l'opérateur. L'attribut TA dans la requête contenant l'opérateur flou se présente sous deux formes: Nombre ou Date. Dans cette étude, nous traiterons uniquement le cas du nombre En fait, le schéma précédent, nous avons utilisé le composants suivants:

**Operator-Computational Perception of Requirement (O-CPR):**

L'opérateur Perception de calcul utilise l'algorithme de l'opérateur flou soit l'algorithme NEAR, soit NEAR Négatif.

O-CPR est un quadruplet (E, B, O, L) où:

A (a1, ..., an) sont les attributs de la cible actuelle lorsque les opérateurs sont flexibles. B (B = OP2) la valeur fixe du prédicat, O (o1, o2, ..., on) est l'opérateur cible (OP), dans ce cas c'est ou bien "NEAR" ou "Not NEAR" et avec un degré de validité MD.

Exemple: Select \* from Bus where D near 9 AND I Near 14

**• Operator - Requirement Mapping (O-RM) :**

Nous avons utilisé les RMs pour créer et agréger les CPRs. Il existe plusieurs types de RMs.

O-RM est un triplet (U, Y, T) où:

U: est une entrée de RM (Z = A, B, MD (MD-k, k = 0.2))

Y: release de O-CPR ; TO: Traitement de l'Algorithme NEAR

**• ALgorithme NEAR- :**

La procédure de calcul ajoute des marges à la valeur de prédicat PV, m1 et m2 où: m1 = a - b à gauche; (9) m2 = a + b (10) à droite avec a=pv; (10) b=pv-1 (11)

L'algorithme calcule les valeurs des deux côtés avec la formule: m\_left = (md \* diff1) + m1; (12) avec md est le degré d'appartenance (Initialement égal à 0.8)

diff1 = PV - m1; (13) diff2 = m2 - PV (14) et m1, m2: marges de PV

L'algorithme compare les valeurs de TA (ie: attributs) par les éléments définis de m\_left à pv. Si la plage est vide, l'algorithme ajoute 0,2 à md et recalcule les marges.

En différence de la solution donnée par [Memon & al., 2007], qui donne comme solution L'intervalle [m\_left, m\_right], la solution dans notre cas est [m \_ left, PV], c'est plutôt la solution de NEAR NEGATIVE.

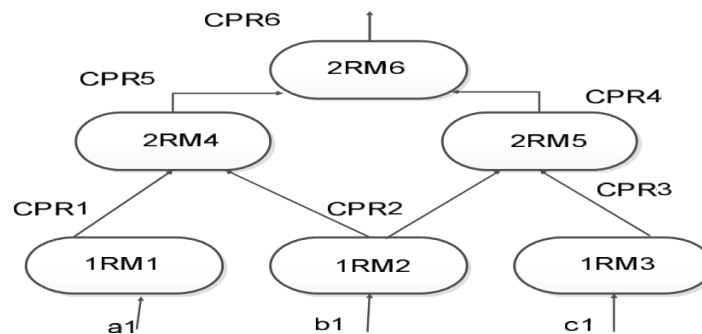


Figure 5.19 Schéma global de notre Ontologie GLMR

□ Opérateur flou – Cas du Nombre:

Pour le cas du Mapping des besoins (RM) et en particulier pour l'évaluation de l'opérateur flou, il est proposé d'avoir en entrée un attribut cible TA (attribut existant dans un prédicat incluant un opérateur flou), une valeur fixe B en général, un

opérateur flou (NEAR dans notre cas) et une valeur initiale de degré d'appartenance MD (dans notre cas MD = 0,8). Ainsi, après l'application de l'algorithme de l'opérateur flou NEAR, le résultat dans la sortie sera l'intervalle m1, m2 ou m1, PV qui est la plage de validité de l'opérateur avec un degré de validité.

L'algorithme NEAR (figure 8) pour le type de données NOMBRE est adapté à notre solution comme suit:

\* La marge au pv (valeur de prédicat), c'est-à-dire m1 est ajoutée dynamiquement sur le côté gauche, compte tenu de la valeur qu'il contient. Ceci est effectué à l'aide des formules (9):  $m1 = a - b$ ; pour le côté gauche de l'échelle,  $a = pv$ ; (10)  $b = pv - 1$  (11)

\* L'opérateur NEAR- est censé obtenir la valeur NEAR pour pv, ainsi le degré d'appartenance cible (md) est initialement fixé à 0.8. Les valeurs jusqu'à md sont calculées sur un côté par rapport à la marge ajoutée à pv, à l'aide des formules:  $m\_left = (md * diff) + m1$  ; où, md = membership degree (initialement 0.8)  $diff = pv - m1$  ; m1 : marge de pv

\* L'algorithme compare les valeurs opl (colonne) ligne par ligne, jusqu'à la fin de la table, aux éléments de l'ensemble définis par NEAR, c'est-à-dire valeurs de m\_left à pv, en ajoutant des tuples correspondants à l'ensemble résultat.

\* Il est possible que le jeu de résultats soit vide car aucune valeur dans la plage n'existe dans opl. L'algorithme vérifie l'ensemble de résultats vide, et dans ce cas, diminue la cible md de 0.2 et passe à l'étape 3. C'est la raison pour laquelle de grandes marges sont ajoutées au pv.

\* Dans le cas où il n'y a pas de valeurs dans opl qui sont entre m1 et pv (et donc le résultat est vide), l'algorithme récupère les deux valeurs les plus proches (tuples) à pv, une de chaque côté, comme résultat. L'algorithme renvoie un résultat vide uniquement si la table référencée à la requête est vide. Le graphique de la figure 5.23 illustre le fonctionnement de l'opérateur NEAR- pour le type de données NUMBER.

Nous présentons d’abord une comparaison entre notre modèle et celui du modèle GLMP. Ainsi, le schéma global de la solution proposée est le suivant:

Les concepts dans le Modèle GLMP	Les Concepts dans le modèle GLMR (Cas Operateur)
<p>CP (Computational Perception) CP (A,W)</p> <p>A = (a1, a2,.....,an) est un vecteur d'expressions linguistiques (mots ou phrases en NL) Predicate value is a fixed value As “hot” in the sentence Temperature is “hot”</p> <p>W = (w1, w2; ....., wn) is a vector of validity degrees wi [0; 1] assigned to each ai in the specific context.</p> <p>CP</p> <p>N’ existe pas</p>	<p>CPR (Computational Perception of Requirement) CPR(E,F, W, L, <math>\alpha</math>)</p> <p>E= (e1, e2,.....,en) est un vecteur de prédicats de requêtes de besoins ou d'expressions linguistiques (mots ou phrases en NL ou en SQL ou en Pseudo) F = (f1,f2,...,bn) est un vecteur de valeurs de predicat (Valeurs fixes pour le cas d’opérateur flou, valeur floue pour le case de valeur de prédicat flou); W=(w1, w2; ....., wn) = MD=(MD1, MD2; ....., MDn) est un vecteur de validité des degrés wi [0;1] assignés à chaque ei pour differents cas. V-CPR (Predicate Value)</p> <p>O-CPR (Operateur) (Operateur flou tel que “NEAR” ou “Not NEAR”) L (m_left) est un vecteur des valeurs obtenues m_left après application de l’algorithme d’évaluation de l’opérateur flou C-CPR (Connector) (Fuzzy Connector such as “and if ‘possible”) <math>\alpha</math> (Mu1,Mu2) sera traité ulterieurement CPR Item : ((pvi,wi),..., (pvj,wj))</p>

Tableau 5.6 Comparaison entre les concepts de GLMP & GLMR

Le second prédicat est automatiquement exécuté car il n'inclut aucune valeur flexible.

#### 5.4.2 Exemple illustratif : Evaluation de risque dans le transport routier

Nous pouvons simplifier ces règles après consultation avec l'expert en transport routier en éliminant le paramètre nombre d'accidents avec la prise en compte dans l'étude uniquement des cas où le nombre d'accidents est au maximum 2 et en ignorant certains paramètres tels que le paramètre climatique ou le paramètre état de la route. Il est donc limité à un modèle de solution à deux paramètres (le nombre de morts D et le nombre de blessés I).

Les modèles routiers de sécurité mettent l'accent sur le processus de génération de l'accident et les victimes qui surviennent et se situent à l'intérieur du système de transport routier. Pour concevoir et estimer des modèles de surveillance de la sécurité routière, le modélisateur dispose d'une grande liberté, le résultat dépendra des objectifs [Seo & al., 2009].

Notre exemple d'illustration traite le risque du secteur des transports en particulier l'accident qui est le premier danger vécu dans le transport et le plus dangereux.

Les différents acteurs de l'accident sont: Le conducteur, le bus et la ligne. Il y a d'autres acteurs comme le climat dont nous ne discuterons pas dans ce document. Ces critères sont inclus dans les normes à titre indicatif seulement pour limiter les critères de décision. Les indicateurs d'accidents sont à la base des programmes de sécurité routière. Les données sur les accidents sont collectées depuis de nombreuses années comme point de départ pour améliorer la sécurité routière.

Les significations des situations dangereuses sont les suivantes: inacceptable: doit être éliminé; non désiré: le risque n'est acceptable que lorsque la réduction de ce dernier est impossible; dans ce cas, l'accord de l'opérateur ou des régulateurs de réseau est impératif; acceptable: Le risque est acceptable avec une surveillance appropriée et l'accord de l'exploitant ou du régulateur; Négligeable: Acceptable sans condition.

Appliquons donc cette matrice dans notre cas du transport routier. Ces règles ont été validées par des experts en ingénierie du trafic. Dans ce contexte, nous voyons que ce travail représente l'étude de l'imprécision dans les besoins décisionnels qui se présente dans la valeur du prédicat du besoin à base de requête. Cette étude propose une solution structurelle qui s'appliquera à l'exemple portant sur le risque du transport routier.

La règle proposée est: R1 (Risk Type is Unacceptable): (Deaths NEAR- pvD AND Injured NEAR- pvI) OR (Deaths >=pvD AND Injured >= pvI)

### 5.4.3 Expérimentation

Pour réaliser une application de manipulation d'ontologies floues, nous avons choisi d'utiliser le langage JAVA avec l'environnement de développement intégré Netbeans permet également de supporter différents autres langages, aussi d'apporter des bibliothèques (Plug-ins)

Langages de programmation et l'environnement de développement intégré (EDI)

**Langages de programmation (Java):** Le langage Java est un langage de programmation informatique orienté objet créé par James Gosling et Patrick Naughton, employés de Sun Microsystems, avec le soutien de Bill Joy (cofondateur de Sun Microsystems en 1982), présenté officiellement le 23 mai 1995 au SunWorld. La société Sun a été ensuite rachetée en 2009 par la société Oracle qui détient et maintient désormais Java. La particularité et l'objectif central de Java est que les logiciels écrits dans ce langage doivent être très facilement portables sur plusieurs systèmes d'exploitation tels que UNIX, Windows, Mac OS ou GNU/Linux, avec peu ou pas de modifications. Pour cela, divers plateformes et frameworks associés visent à guider, sinon garantir, cette portabilité des applications développées en Java.

**Environnement de développement intégré (Netbeans) :** NetBeans est un environnement de développement intégré (EDI), placé en open source par Sun en juin 2000 sous licence CDDL et GPLv2 (Common Development and Distribution License). En plus de Java, NetBeans permet également de supporter différents autres langages, comme Python, C, C++, JavaScript, XML, Ruby, PHP et HTML. Il comprend toutes les

caractéristiques d'un IDE moderne (éditeur en couleur, projets multilingage, refactoring, éditeur graphique d'interfaces et de pages Web).

Conçu en Java, NetBeans est disponible sous Windows, Linux, Solaris (sur x86 et SPARC), Mac OS X ou sous une version indépendante des systèmes d'exploitation (requérant une machine virtuelle Java).

Un environnement Java Development Kit JDK est requis pour les développements en Java. NetBeans constitue par ailleurs une plate-forme qui permet le développement d'applications spécifiques (bibliothèque Swing (Java)). L'IDE NetBeans s'appuie sur cette plate-forme. L'IDE NetBeans s'enrichit à l'aide de plugins permettant la lecture et la manipulation des ontologies décrites en RDFS ou en OWL, pour d'autres raisons, désire aller rapidement à l'implémentation. Une familiarité avec XML et Java est requise. Jena soutient sérialisation des graphes RDF à une base de données relationnelle RDF / XML, ..

**Jena :** Jena est un API java open source développé par le laboratoire de HewlettPackard

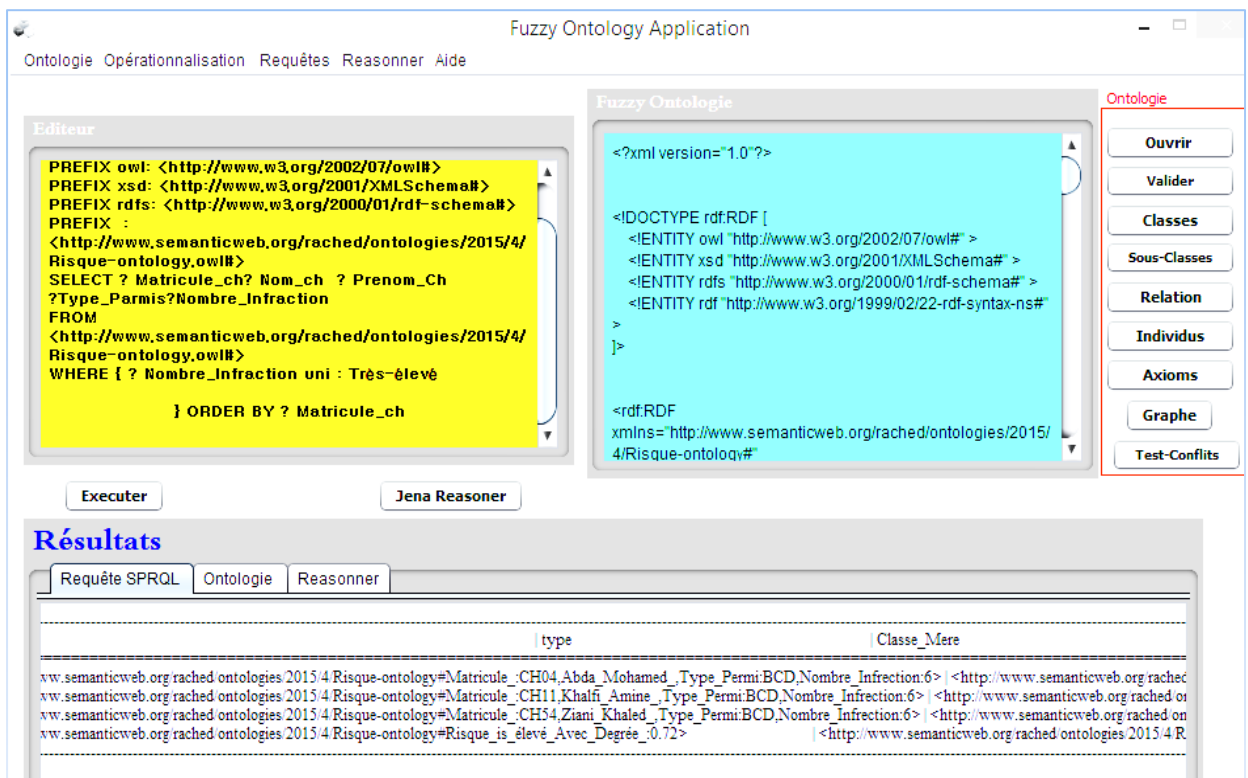


Figure 5.20 Interface Risque-Ontologie

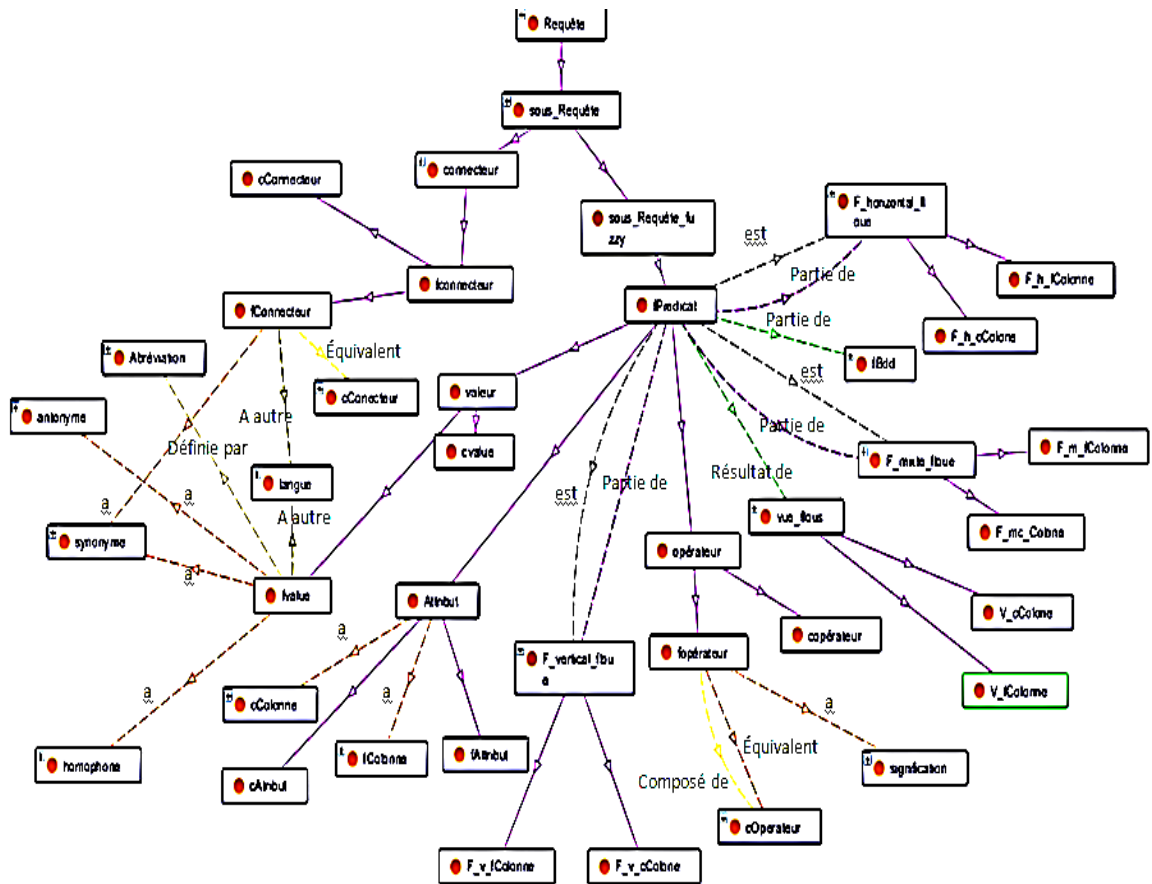


Figure 5.21 Schéma de spécification de la requête floue

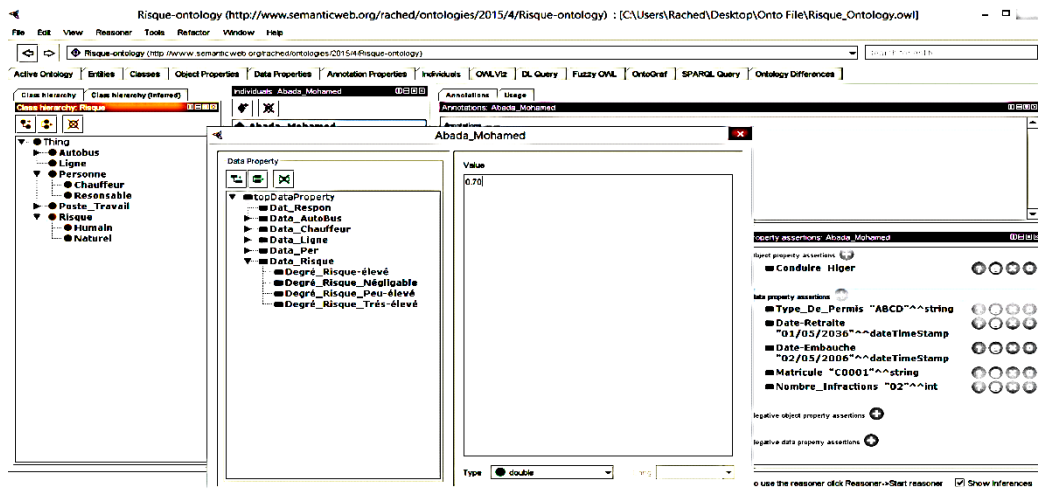


Figure 5.22 Application Fuzzy Ontology

### Expérimentation & résultats obtenus

**Besoin décisionnel:** Le décideur veut avoir la liste des Bus inacceptables dans le semestre. `Select * from Bus where BRType="Inacceptable"`

Pour notre exemple d'illustration, nous avons Brtype est le type de risque du bus qui peut prendre des valeurs {Inacceptable\_Bus, Unwanted\_Bus, Bus négligeable, Acceptable\_Bus} Règle 1: Par période de temps:

On peut voir que la requête a une partie flexible dans la valeur de prédicat mais par l'application de la règle 1, le: flou sera présent dans l'opérateur comme:

Select \* from Bus where D~9 AND I~14

CPR (Computational Perception of Requirement): O-CPR(E, F, W, L)

$$E = \{e_1, e_2, \dots, e_n\} = \{(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)\}$$

For our case:  $E = \{e_1, e_2\} = \{(a_1, b_1), (a_2, b_2)\}$

$$E = \{e_1, e_2\} = \{(a_1, b_1), (a_2, b_2)\}$$

$$e_1 = d1 \text{ NEAR- } pvD \text{ AND } I \text{ NEAR- } pvI$$

$$= \underbrace{d1 \text{ NEAR- } 9}_{a_1} \text{ AND } \underbrace{I \text{ NEAR- } 14}_{b_1}$$

$$e_2 = d2 \text{ NEAR- } pvD \text{ AND } I2 \text{ NEAR- } pvI$$

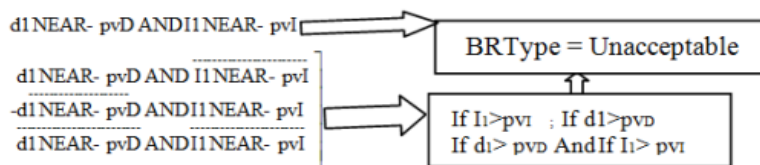
$$= \underbrace{d2 \text{ NEAR- } pvD}_{a_2} \text{ AND } \underbrace{I2 \text{ NEAR- } pvI}_{b_2}$$

$$F = \{pvD, pvI, ?pvI\} = \{pvD, pvI\} = \{10, 15\}$$

$$L = \{m\_left1, m\_left2, ? m\_left\} = \{m\_leftD, m\_leftI\}$$

$$MD = W = \{MD1, MD2, ?, MDr\} = \{MD1, MD2\}$$

Il y a quatre solutions possibles pour le cas de **BRType = Inacceptable**



Nous notons que C peut prendre le cas échéant la valeur de pvD ou pvI et clarifier cette solution, nous supposons cet exemple de règles suivantes:

R1 : BRType = Iacceptable\_bus , le nombre de morts <= 9 AND le nombre de blessés <= 14).

R2 : BRType = Unwnated (Unsafe) bus ; le nombre de morts <= 5 AND le nombre de blessés <= 5).

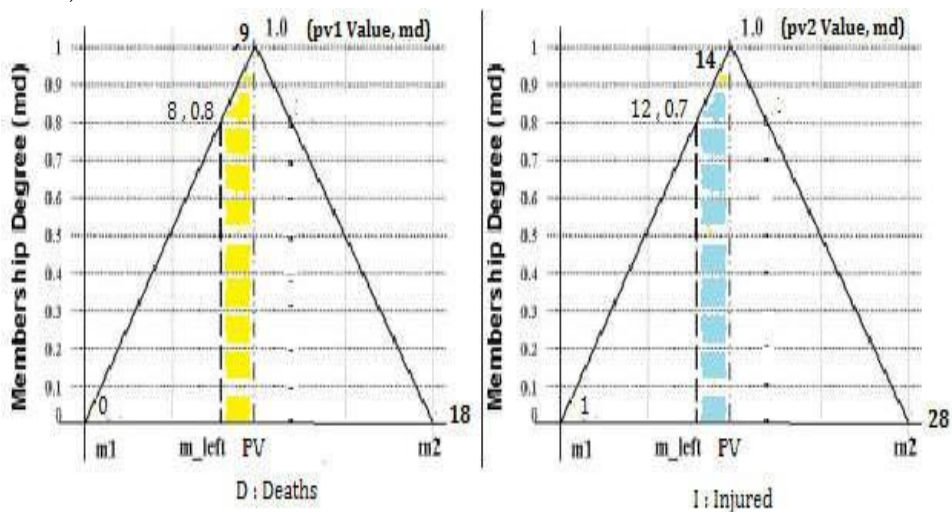


Figure 5.23 Marge cible initial de l'opérateur NEAR Negatif

Le comportement du système, lorsqu'une requête flexible est exécutée dans le système, est analysé dans ce cas. Une partie de la DW pour la période semestrielle est développée et utilisée uniquement à des fins expérimentales.

Select \* from Bus where BRType="Inacceptable"

"Le besoin du décideur c'est de savoir, dans une période de temps spécifique, les bus de l'entreprise" à risque "(Inacceptable, Unsafe, etc.) ce qui implique l'utilisation des règles données par l'expert en transport:

Select \* from Bus where nb\_deaths is NEAR- 9 AND nb\_injured is NEAR- 14

Avec l'application de l'algorithme de l'opérateur flou NEAR-, nous aurons:

$m\_leftD \leq nb\_death \leq pvD$  and  $m\_leftI \leq nb\_injured \leq pvI$ .

Ainsi, la requête sera écrite:

Select \* from Bus where nb\_deaths is between (m\_leftD AND pvD) AND nb\_injured is between (m\_leftI AND pvI)

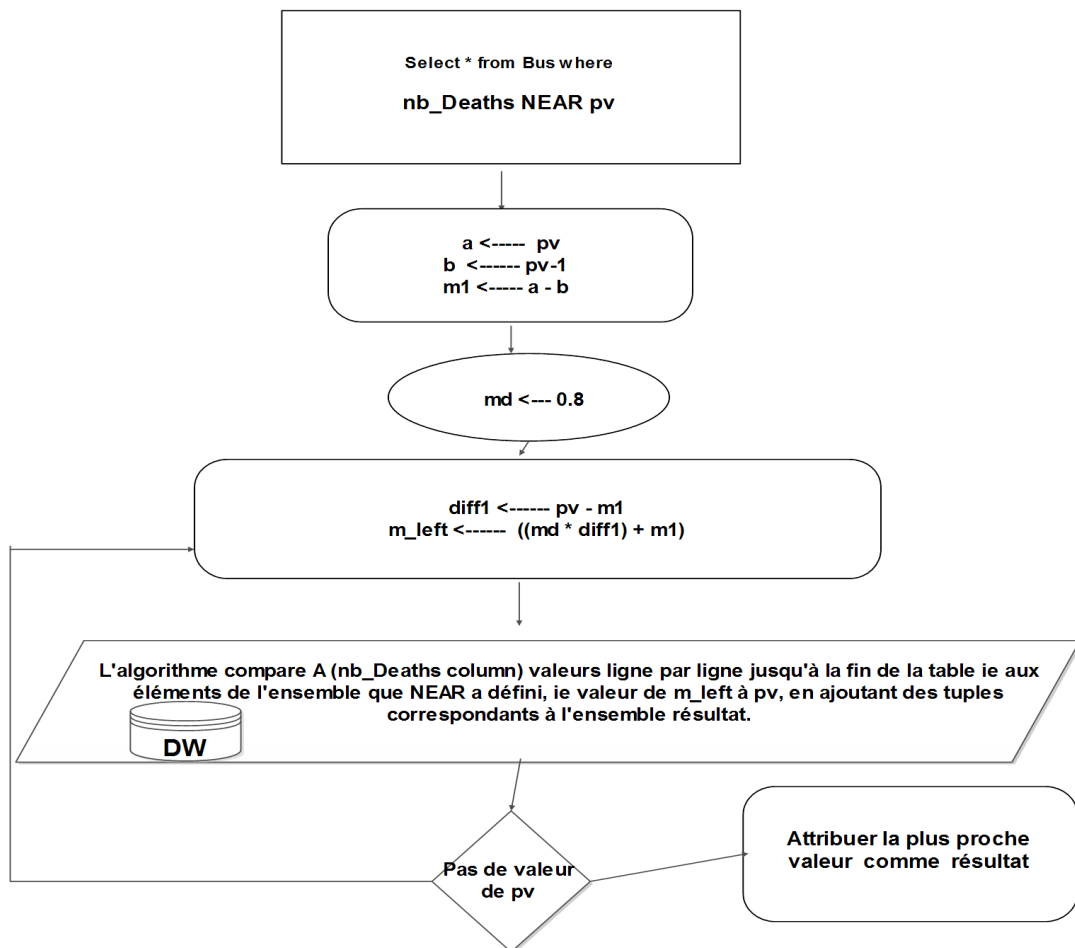


Figure 5.24 Algorithme Operateur Flou NEAR-

Nous pouvons donc calculer le degré d'appartenance aux différentes classes.

Nous avons créé notre DW simple dans l'éditeur d'ontologie Protégé puis le peuplons avec quelques exemples donnés par la direction de transport puis nous avons défini tous les concepts flous à l'ontologie GLMR avant de charger la partie DW concernée, à notre application Java pour adresser l'imprécision dans les besoin décisionnelles et surtout dans la partie opérateur.

Dans cet exemple, un bus avec au moins 2 accidents causant au moins huit morts et au moins 12 blessés est considéré comme un bus à risque (bus inacceptable).

Par conséquent, R1 (type de risque bus est inacceptable): le nombre de décès NEAR-10 et le nombre de blessures NEAR-15 sont interprétés par notre solution ontologique comme suit, R1:  $D = [8, 10]$  and  $I = [12, 15]$ . The values for  $D > 10$  and  $I > 15$  belong automatically to the desired class.

**Exemple 1:** Nous supposons avoir un nombre de décès de 9 ou 5 et respectivement un nombre blessé est de 14 ou 10 et nous étudierons l'appartenance du couple à la classe Bus Inacceptable pour  $d1 = \{9,5\}$  et  $I1 = \{14, 10\}$ .

Bus Class	Inacceptable					
Requête	Select * from Bus where BRType="Inacceptable" i.e. Select * from Bus where nb_deaths is close to 9 AND nb_injured is close to 14					
Hypothèses	D=9 ; I=14 ; a=pv ; b=pv-1 ; m1=a - b where pv=pvD =9 or pv=pvI= 14 Diff1 = a - m1 (diffD, diffI) ; diff1 = pv - m1					
Réécriture de requête entre ontologie GLMR	Select * from Bus where nb_deaths is between (m_leftD AND pvD) AND nb_injured est entre (m_leftI AND pvI)					
Application GLMR	Resultats : CP <sub>Deaths</sub> ; CP <sub>Injured</sub> ; RMB <sub>Rtype</sub> <input type="checkbox"/> MD <sub>D</sub> ; MD <sub>I</sub> ; MD (D AND I)=MIN(MD <sub>D</sub> , MD <sub>I</sub> )					
NEAR- Algorithm Application	Calculate : m_left= around ((md * diff) + m1 )					
	Diff <sub>D</sub>	md <sub>D</sub>	m_left <sub>D</sub>	Diff <sub>I</sub>	md <sub>I</sub>	m_left <sub>I</sub>
Initial Value	8	0.8		13	0.7	
1st iteration	8	0.8	7.1 ~ 7	13	0.7	10.1 ~ 10
2nd iteration	8	0.6	5.8 ~ 6	13	0.5	7.5 ~ 8
3rd iteration	8	0.4	4.2 ~ 4	13	0.3	4.9 ~ 5
4th iteration	8	0.2	2.6 ~ 3	13	0.1	2.3 ~ 2
Result	Dans le cas où il n'y a pas de valeurs dans A (Nb_deaths, Nb_injured) qui sont entre ml et pv (et donc le jeu de résultats est vide), l'algorithme récupère la valeur la plus proche (tuple) de pv comme résultat.: m_leftD=7(Valeur trouvé); m_leftI=12(Valeur Nearest)					
Degree membership	0.8	0.7	mdD = (7-1)/(10-1) = 6/9 = 0.67 ; mdI = (12-1)/(15-1)= 11/14 = 0.79 ; MD = MIN (mdD , mdI) = 0.67			
Hypotheses	D=4 ; I=10					
NEAR- Algo.	Diff <sub>D</sub>	md <sub>D</sub>	m_left <sub>D</sub>	Diff <sub>I</sub>	md <sub>I</sub>	m_left <sub>I</sub>
Initial Value	4	0.8		9	0.7	
1st iteration	4	0.8	4.2 ~ 4	9	0.7	7.3 ~ 7
2nd iteration	4	0.6	3.4 ~ 3	9	0.5	5.5 ~ 6
3rd iteration	4	0.4	2.6 ~ 3	9	0.3	3.7 ~ 4
4th iteration	4	0.2	1.8 ~ 2	9	0.1	1.9 ~ 2
Result	m_leftD =4 (Found value) ; m_leftI 8 (Nearest value)					

Degree membership	0.8	0.7	$mdD = (4-1)/(9-1) = 3/8 = 0.373$ $mdI = (8-1)/(14-1) = 7/13 = 0.54$ $MD = \text{MIN}(mdD, mdI) = 0.37$
$f1 = \{11, 13\}$ belongs to “Unacceptable bus” class with MD = 0.75 $f2 = \{4, 10\}$ belongs to the “Unacceptable bus” class with MD = 0.37 Likewise using this solution in the case D=2, I=4 $f3 = \{2, 4\}$ belong to “Unacceptable Bus” class with md=0, Thus: CPRI1: f1, MD =0.75; CPRI2 : f2, MD =0.37; CPRI3 : f3,MD =0			

Tableau 5.7 CPRI Calcul des bus de l'entreprise pour la classe de bus Inacceptable

Example 2: Pour  $d1 = \{2,4\}$ ,  $I1 = \{4,4\}$ , Bus Class = {Unwanted Bus}, on a :

Classe Bus	Type : Unwanted					
Query	Select * from Bus where BRType="Unwanted" i.e. Select * from Bus where nb_deaths is close to 5 AND nb_injured is close to 5					
Hypotheses	D=2 ; I=4 (Input) a=pv ; b=pv-1 ; m1=a - b where c=pvD =5 or c=pv=pvI= 10 DiffI = a - m1 (diffD, diffI) ; diffI = pv - m1					
NEAR- Algorithm Application	Calculate : m_left=around ( ( md * diff) + m1)					
	DiffD	mdD	m_leftD	DiffI	mdI	m_leftI
Initial Value	4	0.8		09	0.7	
1st iteration	4	0.8	4.2 ~ 4	09	0.7	3.8 ~ 4
2nd iteration	4	0.6	3.4 ~ 3	09	0.5	3.0
Result	m_leftD =3 ; m_leftI 7 (Found values)					
Degree membership	0.8	0.7	$mdD = 0$ ; $mdI = (4-1)/(5-1) = 3/4 = 0.75$ $MD = \text{MIN}(mdD, mdI) = 0$			
Hypotheses	D=4 ; I=4					
NEAR- Algo.	DiffD	mdD	m_leftD	DiffI	mdI	m_leftI
Initial Value	4	0.8		9	0.7	
1st iteration	4	0.8	4.2 ~ 4	9	0.7	3.8 ~ 4
2nd iteration	4	0.6	3.4 ~ 3	9	0.5	3.0 ~ 3
3rd iteration	4	0.4	2.6 ~ 3	9	0.3	2.2 ~ 2
4th iteration	4	0.2	1.8 ~ 2	9	0.1	1.4 ~ 1
Result	m_leftD =4 (Found value) ; m_leftI 4 (Nearest value)					
Degree membership	0.8	0.7	$mdD = (4-1)/(5-1) = 3/4 = 0.75$ $mdI = (41)/(5-1) = 3/4 = 0.75$ ; $MD = \text{MIN}(mdD, mdI) = 0.75$			
For the “Unwanted Bus” class, we have $f1 = \{11,13\}$ belongs to “Unacceptable bus” class with MD=0						
Ainsi : CPRI1 : f1, MD =0 ; CPRI2 : f2, MD =0.75 ; $f2 = \{4, 10\}$ belongs to the “Unwanted bus” class with MD = 0.75 $F3 = \{2, 4\}$ belongs to the “Unwanted Bus” class with MD= 0 ; $F4 = \{4, 4\}$ belong to this class with $md = \text{min}(0.75, 0.75) = 0.75$						
CPRI3 : f3, MD =0 ; CPRI4 : f4, MD =0						

Tableau 5.8 Calcul du CPRI pour les bus de la comganie pour la classe « unwanted bus »

Cependant, il y a encore beaucoup de travail à faire, pour généraliser cette solution, pour permettre l'évaluation des parties flexibles à la valeur du prédicat de la requête, les opérateurs incluant les opérateurs OLAP et les connecteurs. L'idée principale est d'enrichir l'expression d'interrogation avec la capacité d'inférence fournie par l'ontologie floue ajoutée à l'entrepôt de données. Nous avons construit notre benchmark (entrepôt de données) représentant les accidents de la route en Algérie, en observant les tarifs réels dans un parc de 180 bus et pour une période de 06 mois, il y avait 10000 feuilles de route / semestre, agrégées par semaine soit 1430 occurrences et plus de 160 accidents, 50 morts et 600 blessés dont 20 concernés. Après application de

notre solution pendant une période de 06 mois et pour seulement 20 Bus, nous avons détecté les résultats suivants (Tableau 8):

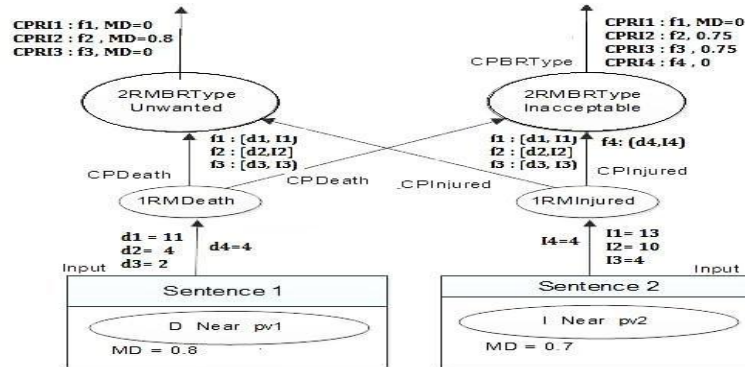


Figure 5.25 Solution GLMR pour exemples de Bus Risk Type

Bien que l'ensemble de données ne soit pas grand, les premiers résultats sont très encourageants

Tableau 5.9 Résultat expérimental

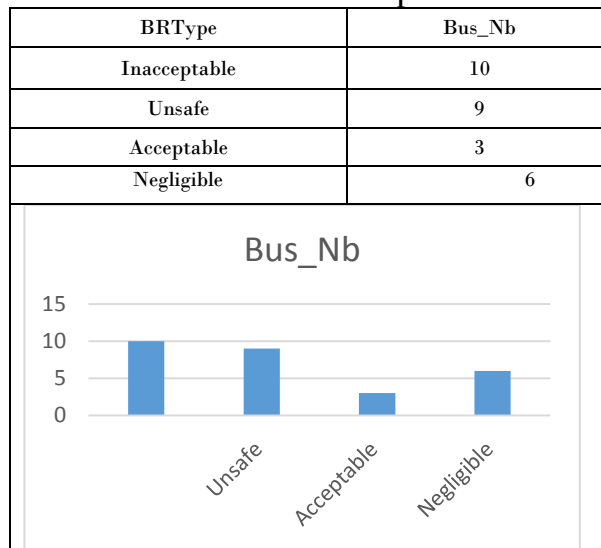


Figure 5.26 Résultat expérimental

Select Count(BRTType) As Bus_Number (Threshold =0.7)		
BRTType	Bus_Nb	Decision
Inacceptable	10	A mettre hors service
Dangereux	09	A envoyer à la maintenance
Acceptable	03	/
Negligeable	06	/
TOTAL	20	

Tableau 5.10 Issue des résultats de l'expérimentation

#### 5.4.4 Conclusion

Ce travail met en évidence l'une des études sur les besoins décisionnels dans laquelle l'expression de ces besoins par les décideurs, qui ne sont pas nécessairement des experts en informatique, génère des besoins ambigus ou imprécises pouvant affecter l'ensemble du processus décisionnel. Nous évaluons l'ambiguïté et l'imprécision dans

ce cas en considérant une solution floue basée sur l'ontologie. Pour illustrer cette proposition, elle a été appliquée à la gestion des risques du transport en utilisant un ensemble d'indicateurs pour classer les types de risques. Nous présentons notre modèle GLMR proposé pour l'opérateur NEAR-flou. Dans les travaux futurs, nous allons améliorer cette solution pour utiliser les algorithmes d'opérateur Skyline [Agarwal & al., 2017].

## **5.5 Conclusion du chapitre : Solutions (Structure, Système expert, Ontologie)**

En évaluant l'imprécision des requêtes décisionnelles en utilisant les trois propositions, on peut déceler les résultats suivants :

- \* La solution structurelle nous a permis d'évaluer l'imprécision quoique cette solution est coûteuse en espace mémoire de plus en plus qu'il y a d'attributs cibles TA concernés par les besoins décisionnels flexibles. Ainsi pour chaque attribut cible, on a à ajouter deux métables FMT et FCT et aussi deux autres tables supplémentaires pour définir les paramètres de l'expert et la table de correspondance attribut-requête floue.

- \* La solution à base de système expert nous a permis de lever l'ambiguïté du besoin imprécis en se basant sur les paramètres de l'expert et après avoir réécrit la requête flexible du besoin décisionnel. Malgré que cette solution nous ait été bénéfique mais vu que cette technologie a été améliorée par l'apparition des ontologies, on a vu nécessaire au lieu d'améliorer la solution primaire proposée de passer à une solution ontologique.

- \* On a d'abord de tester l'utilité de la solution ontologique pour enrichir la solution de la requête et ce à travers la proposition d'une ontologie médicale (voir Annexe) permettant d'améliorer la réponse de requête XML (Complément : Contribution 4)

- \* La solution ontologique GLMR proposé, inspiré de la solution de Martinez-Cruz sur les phrases mentionnant les perceptions en mot de l'utilisateur, nous a permis, en premier lieu, d'adapter la solution GLMP pour évaluer l'imprécision au niveau de la valeur de prédicat de requête du besoin décisionnel. Puis, on a proposé une solution rigoureuse permettant d'évaluer l'imprécision se présentant au niveau de l'opérateur de la requête décisionnelle, on a choisi pour mettre en évidence cette proposition l'opérateur NEAR Négatif et on l'a appliquée sur le cas d'évaluation de risque de transport routier. Les résultats primaires été très encourageants. Les tests de cette solution ontologique sur d'autres parties de la requête décisionnelle sont en cours.

Pour les trois solutions, l'évaluation de l'imprécision a donné des valeurs acceptables par l'expert de transport de le 2<sup>ième</sup> cas d'étude et logique selon les marges données par [Bosc, 1995] pour le 1<sup>er</sup> d'étude.

Bien que l'approche proposée est encore dans son début vu qu'il reste encore à définir les concepts de connecteur flou ; d'opérateur OLAP flou, etc.; mais les résultats préliminaires sont très encourageants.





# CHAPITRE VI

---

---

## CONCLUSION GENERALE

---

---

6.1 Conclusion générale.....117

6.2 Perspectives.....120

---

---



## **6.1 CONCLUSION GENERALE**

La conception d'un entrepôt de données est très différente de celle d'une base de données. En plus, les besoins des utilisateurs de l'entrepôt ne sont aussi clairs que ceux des utilisateurs des systèmes OLTP. La réécriture des requêtes a attiré l'attention de nombreux chercheurs car elle est en relation avec plusieurs problèmes de gestion de données : optimisation de requêtes, intégration des données, conception des entrepôts de données. Pour répondre aux problèmes d'évaluation et d'optimisation des requêtes en vue de spécifier les besoins des décideurs (requêtes classiques) surtout en cas où ces besoins sont vagues ou comprennent des parties imprécises ou indéterminées (requêtes flexibles). On peut faire appel à la logique floue ou aux systèmes experts ou même les ontologies pour pouvoir lever l'indétermination.

Les imprécisions des données sources (Base de production) des systèmes décisionnels peuvent conduire à des analyses OLAP erronées. Donc, la qualité des données extraites (erreurs, imprécision, etc.) et l'évaluation précise des besoins du décideur (Lever l'ambiguïté, l'incertitude et l'imprécision) avant de choisir la base de production est devenue dans ces dernières années un axe important de recherche en décisionnel en général et dans la conception ascendante du SAD en particulier. Les imperfections des données en systèmes décisionnels sont dues à de nombreuses causes comme la difficulté d'avoir des modèles proches de la réalité ou des limitations des instruments et des processus d'acquisition. De plus, une représentation inadaptée de l'information recueillie peut être aussi une source de dégradation de la qualité des données, la représentation devenant trop éloignée de l'objet du monde réel que l'on souhaite modéliser. L'imprécision peut être due à l'ambiguïté d'expression des besoins du décideur. En partant d'une mauvaise interprétation des besoins, toutes les conceptions du DW et les analyses OLAP qui s'en suivent seront erronées. Par conséquent, de nombreux travaux se sont intéressés à la prise en compte des données vagues dans les Systèmes d'Information mais peu de travaux qui ont étudié ce cas dans les entrepôts de données. Ainsi les travaux présentés dans ce mémoire vise à proposer des solutions à base d'ontologie, de logique floue pour évaluer l'imprécision des besoins décisionnels en vue d'améliorer la conception de data warehouse et aussi pour affiner l'analyse OLAP entamée.

Quoique la solution ontologique est rigoureuse mais la proposition n'est qu'à son début, vu qu'on peut exploiter les mécanismes de raisonnement offerts par les ontologies pour identifier les besoins (à base de requête) présentant l'imprécision non pas dans la valeur du prédicat mais notamment dans l'opérateur ou le connecteur ou même l'opérateur OLAP.

Pour mettre en évidence, le problème de l'imprécision des besoins décisionnels et pour illustrer les solutions proposées pour l'évaluation des besoins décisionnels flexibles, on a choisi de traiter l'évaluation de risque. Parmi les cas d'évaluation de risques :

- On a choisi d'abord la **gestion de projet intelligent et estimation du risque** qui a été largement explicitée dans le livre « Logique floue » du [Bosc,1995]. Cette étude associait l'évaluation de risque d'un projet, budgétisation du capital, et aide à la décision stratégique, dans un même système utilisant sa propre gestion de bases de données relationnelles. Cette conception fit prendre conscience que le modèle relationnel fournit non seulement une méthode cohérente de stockage des composants du système, mais aussi une manière souple et très évolutive de présenter l'information sous de nouveaux aspects, sans avoir à prédéfinir toutes les relations possibles.
- On a choisi **l'évaluation de risque dans le transport intelligent**, qu'on l'a mis en application dans deux solutions proposées (Solution structurelle et solution ontologique):

La première proposition est une solution structurelle qui nous a permis de répondre à la requête telle que :

Le décideur veut savoir quel conducteur appartient à la catégorie « unwanted » pour le dernier semestre, c.a.d :

```
Select Count (BRType) As Bus_Number (Threshold =0.2) From Company C, Bus B  
where C.Pk_bus=B.Bus_Matr AND BRType="Inacceptable"
```

Cette solution permet à travers un changement de structure d'évaluer l'imprécision qui se présente au niveau de la valeur de prédicat. Le changement de structure permet, à travers des tables supplémentaires, d'évaluer et de déterminer cette imprécision.

- La deuxième solution est une solution à base de système expert appliquée à l'estimation de risque, où un système expert indépendant qui contient un jugement d'experts cohérent, au sujet du risque à lancer un projet, en fonction d'un certain nombre de caractéristique, On peut ainsi répondre à une requête, par exemple :

- `Select num_p From projet Where risque is acceptable;`

Cette requête a été utilisée par Bosc pour la recherche dans les bases de données relationnelles et on l'a repris pour l'utiliser comme étant une requête décisionnelle. Pour notre première proposition, la solution consiste à associer un système expert (une base de connaissance) et un traducteur (STDS) pour pouvoir interagir avec le décideur et la base de données de l'entreprise.

- La troisième proposition est une solution ontologique qui permet en utilisant une ontologie GLMR de déterminer l'imprécision au niveau de la requête du besoin décisionnel.

En comparant les trois solutions, on aura le tableau suivant:

Approches	Points Forts	Points faibles
Approche structurelle	<ul style="list-style-type: none"> <li>* Simple à réaliser</li> <li>* Traite un nombre restreint de TA</li> <li>* Ne présente aucune condition pour la valeur du prédicat</li> </ul>	<ul style="list-style-type: none"> <li>• Nécessite beaucoup d'espace mémoire</li> <li>• Devient moins performante en augmentant le nombre de TA</li> <li>• Ne traite le flou que pour la valeur du prédicat</li> </ul>
Approche à base de système expert	<ul style="list-style-type: none"> <li>• Peut être résolu en utilisant les langages standards (SQL)</li> <li>• Ne présente aucune condition pour la valeur du prédicat</li> </ul>	<ul style="list-style-type: none"> <li>• Décideur non expert</li> <li>• Ne traite le flou que pour la valeur du prédicat et le connecteur</li> <li>• Les résultats dépendent des paramètres fournis par l'expert</li> </ul>
Solution Ontologique	<ul style="list-style-type: none"> <li>• Permet d'évaluer le flou dans la valeur du prédicat, comme dans l'opérateur, le connecteur et autre</li> <li>• Robuste en raisonnement</li> </ul>	<ul style="list-style-type: none"> <li>• Dépend des algorithmes utilisés en plus des paramètres de l'expert</li> <li>• Ne concerne qu'un domaine bien déterminé</li> <li>• faible en termes de modélisation</li> </ul>

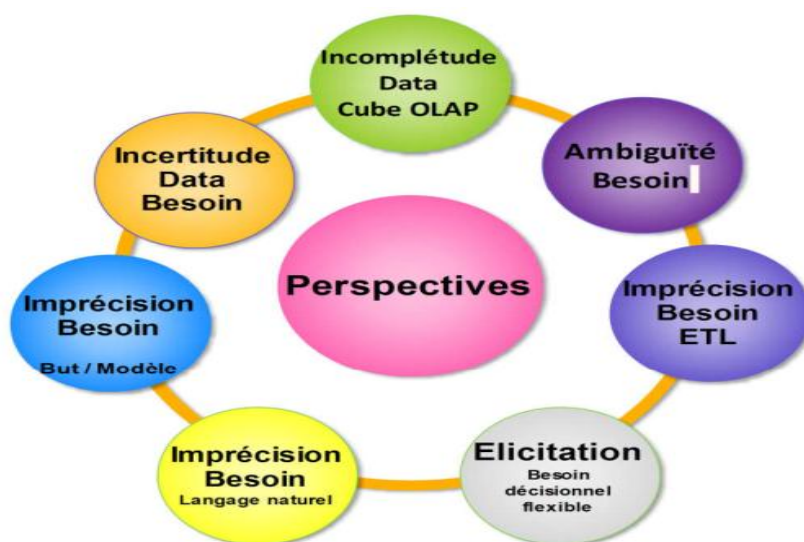
En résumé, voilà le résultat de la comparaison des trois approches :

- ❖ La solution à base de système expert : la principale difficulté rencontrée lors du développement d'un système expert est l'extraction et la formalisation, généralement sous forme de faits et de règles, des connaissances sur laquelle l'expert semble s'appuyer pour agir.
- ❖ La solution structurelle est une solution gourmande en espace mémoire puisque en plus de la grande taille qu'occupe le data warehouse, la solution nécessite une taille supplémentaire pour les tables additives. En plus de ce problème, on retrouve l'augmentation de temps de réponse de requête représentant les besoins dû au nombre important de jointures émanant de la nouvelle structure.
- ❖ La solution ontologique est une solution performante, robuste en raisonnement, faible en termes de modélisation.
- ❖ la difficulté de formaliser les connaissances et raisonnements susceptibles de les exploiter, ou à celle de maintenir ces représentations à jour

## 6.2 Perspectives

Comme perspectives, beaucoup de travail reste à faire dans ce domaine. Ce travail a fait un balayage des tâches qui suscitent beaucoup d'intérêt de la part des chercheurs et ce dans l'objectif d'améliorer la qualité du design de l'entrepôt de données d'une part et de satisfaire au mieux les besoins décisionnels émanant du décideur. On peut résumer les principales perspectives comme suit :

- Etude de l'incomplétude des données (Source ou cube OLAP) et son influence sur le besoin du décideur, sur la conception de l'entrepôt et sur l'analyse OLAP.
- Etude de l'incertitude du besoin décisionnel ou des données sources
- Etude de l'ambiguïté dans les besoins décisionnels (cas de décision collective)
- Etude de l'imprécision des besoins décisionnels à base de buts ou de modèles existants
- Etude de l'imprécision des besoins décisionnels dans la phase ETL
- Etude de l'imprécision des besoins décisionnels avec expression en langage naturel
- Etude de l'élicitation flexible des besoins décisionnels



Comme on peut prévoir d'étudier comme continuité de ce travail les opérateurs OLAP flous tels que slice, dice, etc., et utiliser la fonction d'agrégation floue FMOY avec la clause "Having" qui fournit la moyenne d'un attribut en utilisant un degré d'appartenance. En outre, nous allons d'étendre l'étude aux connecteurs flous tels que "et possible" et "surtout depuis". Ces connecteurs qui peuvent être présents dans la requête de prise de décision. En outre, nous prévoyons de traiter l'incertitude et l'incomplétude des expressions de besoins et les appliquer à d'autres domaines. Enfin, On considère aussi d'appliquer notre proposition à un vrai DW si ce n'est pas un grand ensemble de données ou un Big Data.

## Références bibliographiques

- [Abai et al., 2013] Abai N.H.Z, Yahaya J.H., Deraman A., User Requirement Analysis in Data Warehouse Design: A Review, *Procedia Technology*, 11 (2013) 801– 806
- [Abdelhedi & al., 2014] F. Abdelhedi & al. (2014) "Conception assistée d'entrepôts de données et de documents XML pour l'analyse OLAP", thèse.
- [Abran et al., 2004] A. Abran, J.W. Moore, P. Bourque, and R.E. Dupuis. Guide to the Software Engineering Body of Knowledge. [Http://www.swebok.org/](http://www.swebok.org/).IEEE Computer Society, 2004.
- [Agarwal, 2014] Mahak Agarwal, Expert System and it's Requirement Engineering Process 2014, CSED Thapar University Patiala
- [Agarwal & al., 2017] Ajay Agarwal, Apeksha Aggarwal, Apoorv Agarwal, An Approach for Augmenting Selection Operators of SQL Requêtes using Skyline and Fuzzy-Logic Operators, 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22- 24 August 2017, Cochin, India, Elsevier, *Procedia Computer Science* 115 (2017) 14–21
- [Ait Ameer et al., 2003] Y. Ait Ameer, M. Baron, and P. Girard. Formal validation of hci user tasks. In the International Conference on Software Engineering Research and Practice (SERP'03), pages 732–738. CSREA Press, 2003.
- [Amokrane et al., 2016] Nawel Amokrane, De l'Ingénierie des Besoins `a l'Ingénierie des Besoins : Vers une Démarche Méthodologique d'Ingénierie de Systèmes Complexes, de Vérification et de Validation Appliquée à l'Informatisation des PME, Thèse, 2016
- [Annoni, 2006] Annoni, Modélisation adaptée aux besoins utilisateurs dans le développement des systèmes d'information décisionnels, 2006
- [Annoni & al., 2007] Annoni E., (2007) Eléments méthodologiques pour le développement des systèmes décisionnels dans un contexte de réutilisation, Thèse, Computer Science. Université Paul Sabatier - Toulouse III. France.
- [Aqdas & al., 2015] Aqdas Ikram, Usman Qamar, Developing an expert system based on association rules and predicate logic for earthquake prediction, 2015
- [Arnott & al., 2008] Arnott, D., & Pervan, G. (2008). Eight key issues for the decision support systems discipline. *Decision Support Systems*, 44(3), 657–672.
- [Aude & al., 2007] B. Aude et G. Florian "Redesign prototype fuzzy query ISQLF", March 2007
- [Atigui, 2012] Atigui F.,(2012) Modélisation conjointe des données et des processus pour l'implantation de schémas d'entrepôts. Dans : *Journal of Decision Systems*, Taylor & Francis Group, Vol. 21, N. 1, p. 27-49
- [Baneyx, 2007] Baneyx, A. (2007). Construire Une Ontologie De La Pneumologie : Aspects Théoriques, Modèles Et Expérimentations. Thèse de doctorat, Université Pierre et Marie Curie - PARIS 6, Laboratoire INSERM UMR\_S 872 – Santé Publique et Informatique Médicale, France.
- [Bashiri & al., 2011] Bashiri, M., & Badri, H. (2011). A group decision making procedure for fuzzy interactive linear assignment programming. *Expert Systems with Applications*, 38(5), 5561–5568.
- [Bartel & al., 2006] Bartel S.« Interrogation floue de Bases de données : extension de isqlf » , 2006
- [Bellahsene & al., 2002] Bellahsene Z., Schema Evolution in Data Warehouses, *Knowledge and Information Systems*, (2002) – Springer
- [Ben Achour et al., 1999] Ben Achour, C. (1999). Extraction des Besoins par Analyse des Scénarios Textuels. Thèse de doctorat, Université de Paris 6, Paris, France.
- [Balmisse, 2002] Gilles Balmisse, La logique floue, Sept. 2002
- [Basili et al., 1992] V.R.Basili, G. Gianluigi, and H.D. Rombach. The goal question metric approach. Computer science technical report series cstr-2956. Technical report, 1992.

- [Bell et al., 1976] Bell, T.E. & Thayer, T.A., (1976). Software requirements: Are they really a problem? Proc. ICSE-2: 2nd International Conference on Software Engineering, San Francisco, pp.61–68.78
- [Bellatrèche et al., 2000] Ladjel Bellatreche, Utilisation des vues matérialisées, les indexes et de la fragmentation dans la conception logique et physique d'un entrepôt de données, thèse de doctorat, 18 décembre 2000.
- [Bellatrèche et al., 2000a] Ladjel Bellatreche, LISI/ENSMA, La conception physique des data warehouses, 2006.
- [Bellatrèche et al., 2007] L. Bellatreche, K. Boukhalfa, and M. K. Mo. Pruning search space of physical database design. In DEXA, pages 479–488, 2007.
- [Bellatrèche, 2003] Ladjel Bellatreche, LISI/ENSMA, Techniques d'optimisation des requêtes dans les data warehouses, 2003
- [Benkrid et al., 2009] Soumia Benkrid, Une méthodologie Complète de Conception d'un Entrepôt de Données Relationnel Parallèle, thèse de magistère, École nationale Supérieure en Informatique ouedsmar Alger, 2009.
- [Ben Mustapha, 2009] Nesrine Ben Mustapha | Recherche sémantique basée sur les ontologies modulaires et le raisonnement à base de cas, , Manouba, Tunisia, 2009)
- [Bensafi, 2013] Bensafi Abdeljawed, Réécriture de requêtes décisionnelles à l'aide d'ontologies mémoire présenté pour obtention le diplôme de Master option: système d'information avancé université de Béchar 2012-2013.
- [Blair & al., 1994] Andrew Blair, John Debenham and Jenny Edwards, Requirements analysis for intelligent decision support systems, 1994
- [Blazquez, 1998] Blazquez, M., Fernandez, M., Garcia-Pinar, J., & Gomez-Perez, A. (1998). Building Ontologies at the Knowledge Level using the Ontology Design Environment. In Proceedings of the Banff Workshop on Knowledge Acquisition for Knowledge-based Systems.
- [Bobillo & al., 2011] Bobillo F., Straccia, U. Fuzzy ontology representation using OWL 2. International Journal of Approximate Reasoning 52(7), 1073–1094. (2011)
- [Bobillo & al., 2012] Bobillo F. & al.; A Fuzzy Ontologies for Specialized Knowledge Representation in WordNet, Springer-Verlag Berlin Heidelberg; Volume 297, 187. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Italy, July 9-13, (2012)
- [Bosc & al., 1988] P. Bosc, M. Galibourg, and G. Hamon, “fuzzy querying with sql : extensions and implementation aspects,” *fuzzy sets and syst.*, vol. 28, pp. 333-349, 1988.
- [Bosc & al., 1994] Bosc P., Pivert O., Fuzzy Requêtes and relational databases, 1994
- [Bosc & al., 1995] Bosc P., Pivert O. “sqlf: a relational database language for fuzzy querying”. IEEE Transactions on Fuzzy Systems, vol(3) p.1–17, 1995.
- [Bosc et al., 1998] Bosc P., Pivert O. “Fuzzy databases”, Eds. Handbook of Fuzzy Computation, Chapter 4.1, Oxford University Press, 1998
- [Boukhalfa et al., 2008] Kamel Boukhalfa, Ladjel Bellatreche, Pascal Richard, Fragmentation Primaire et Dérivée: Étude de Complexité, Algorithmes de Sélection et Validation sous ORACLE10g 2008.
- [Boukhalfa et al., 2009] Kamel Boukhalfa, De la conception physique aux outils d'administration de tuning des entrepôts de données, thèse de doctorat, Ecole nationale supérieure de mécanique et d'aérotechnique, France 2009
- [Boukhari et al., 2014] Boukhari I., Intégration et exploitation de besoins en entreprise étendue fondées sur la sémantique, thèse, ENSMA Poitiers, France, (2014)
- [Boyadzhieva & al., 2014] Boyadzhieva D., & al., Intuitionistic Fuzzy Data Warehouse and Some Analytical operations, IEEE Transactions on Fuzzy Systems, VOL. 22, NO 4, AUGUST (2014)

- [Bresciani et al., 2004] P. Bresciani, A. Perini, P. Giorgini, F. Giunchiglia, and J. Mylopoulos. Tropos: An agent-oriented software development methodology. *Autonomous Agents and multiagent Systems*, 8:203–236, 2004
- [Carbonell & al., 2007] J.G. Carbonell et J. Siekmann. « Applications of Fuzzy Sets Theory ». *Proc. Of the 7th International Workshop on Fuzzy Logic and Applications*, Science. Pages 7–10. Camogli, Italy. Juillet 2007.
- [Carmè & al., 2010] Carmè A. & al., (2010) "A Model-Driven Heuristic Approach for Detecting Multidimensional Facts in Relational Data Sources", DaWaK, Springer-Verlag Berlin Heidelberg
- [Cavero & al., 2003] Cavero J.,(2003) A multidimensional DW development methodology. *Managing Data Mining Technologies in Organizations*
- [Chamberlin & al., 1976] D. Chamberlin *et al.*, "SEQUEL2: A unified approach to data definition, manipulation and control," *IBM J. Of Res. And Dev.*, vol. 20, pp. 56G575, 1976
- [Chen & al., 2009] Chen CT. &al, A comprehensive model for selecting information system project under fuzzy environment *Intern. Journal P.M.*, (2009) – Elsevier.
- [Chiu & al., 2014] Chiu CY. &al, Customer information system using fuzzy query and cluster analysis. *Journal of Industrial and ...*, (2014) - Taylor & Francis [Cornuéjols, 2013 ] Antoine Cornuéjols, "Introduction à la logique floue" 3<sup>e</sup> année, 2013, université Paris sud, France. [www.lri.fr/~antoine](http://www.lri.fr/~antoine)
- [Cox, 1992] E. Cox, Fuzzy fundamentals, *IEEE Spectrum*, Volume: 29 Issue: 10 , DOI: 10.1109/6.158640  
Page(s): 58 - 61, Oct. 1992
- [Cox, 1994] Earl D. Cox,. "The Fuzzy Systems Handbook", A Practitioner's Guide to Building Using and Maintaining Fuzzy Systems, Cambridge, MA. 1994
- [Cravero & al., 2014] Cravero A.; (2014) Goal oriented requirements engineering in DWs: A comparative study; *Ingeniería E Investigación* vol. 34 N<sup>o</sup>2.
- [Cox & al., 1998] Earl Cox , Michael O'Hagan, Rodman Taber, *The Fuzzy Systems Handkbook with Cdrom* , Book, 2<sup>nd</sup> Edition, Academic Press, Inc. Orlando, FL, USA, ISBN:0121944557, 1998,
- [Dardenne et al., 1993] A. Dardenne, A.van Lamsweerde, and S. Fickas. Goal-directed requirements acquisition. *Science. Computer. Program.*, 20:3–50,1993.
- [Delgado & al., 2015] Delgado A. & al.,Automating the process of building flexible Web Warehouses with BPM Systems, *Latin American Computing Conference (2015)*
- [Denger & al., 2003] Denger C. & al., (2003) "Higher Quality Requirements Specifications through Natural Langage Patterns", *Proceedings of the IEEE International Conference on Software—Science, Technology & Engineering*.
- [Di Tria & al., 2013] Di Tria F. & al., (2013) Ontological Approach to Data Warehouse Source Integration. In *Proceedings of the 28<sup>th</sup> International Symposium on Computer and Information Sciences*, Springer, 251-259.
- [Dubois & al., 1988] Dubois D. & H. Prade, (1988) "Representation and combination of uncertainty with belief functions and possibility measures". *Computational Intelligence*, 4:244–264.
- [Dieng, 2001] Dieng, R., Corby, O., Gandon, F., Giboin, A., GOLEBIEWSKA, J., MATTA, N., et al. (2001). *Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire du knowledge management*. Dunod Edition Informatiques Séries Systèmes d'Information (2<sup>ième</sup> édition).
- [Di Tria, 2012] Di Tria F.; (2012) *Academic DW Design Using a Hybrid Methodology*, *Computer Science and Information Systems*;
- [Di Tria & al., 2014] Di Tria & al., (2014) *Academic DW Design Using a Hybrid Methodology*, *Computer Science and Information Systems*
- [Djaaboub, 2009] DJAABOUB S, *Logique floue et SMA : Aide à la décision floue dans les systèmes multi-agents..* Magister. 2009
- [Dogac & al., 1989] Asuman Dogac, Birol Yuruten. Stefano Spaccapietra , A Generalized Expert System for Database Design SPACCAPIETRA, *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*. VOL. 15, NO. 4, APRIL 1989

- [Elamin, 2014] Elhaj elamin & al., (2014) "Toward an ontology based approach for data warehousing, State of the art and proposal",
- [El Golli, 2008] El Golli, I.G., (2008) «Ingénierie des besoins pour les Systèmes d'Information Décisionnels : Concepts, Modèles et Processus (la méthode CADWE) », Thèse de Doctorat, Université Paris-Panthéon-Sorbonne, France
- [El Sayed, 2007] A. El Sayed, H. Hacid et D. Zighed. « Recherche d'information par le contenu dans des données Hétérogènes ». Actes de la 3<sup>ème</sup> Rencontre Inter Association (RIA 2007). Toulouse, France. Mars 2007.
- [Fasel, 2012] Fasel D.; Concept and Implementation of a Fuzzy Data Warehouse; University of Fribourg; thesis; (2012)
- [Fasel & al., 2014] Fasel D. Fuzzy Data Warehouse, - Fuzzy Data Warehousing for Performance, (2014) Springer
- [Fernández-López & al., 1997] Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Méthontology: From Ontological Art Towards Ontological Engineering. Spring Symposium on Ontological Engineering of AAAI, 33–40.
- [Ferragu et al., 2013 ]Emmanuel Ferragu, Modélisation des systèmes d'information décisionnels,
- [Fred & al., 2001] Ana Fred, Joaquim Filipe, Syntax-directed translation schemes for multi-agent systems conversation modelling. Enterprise Information Systems II, 2001 – Springer
- [Fürst, 2002] F. Fürst. « L'ingénierie ontologique, Rapport de recherche » No 02-07 Octobre 2002.
- [Gaëlle, 2007] H. Gaëlle, P. Buche, J. Dibie-Barthélemy et H. Ollivier. « Annotation sémantique floue de tableaux guidée par une ontologie ». Actes de la conférence Extraction et Gestion des Connaissances (EGC 2007). Pages 587–598. Namur, Belgique. Janvier 2007.
- [Gagliardi, 2005] H. Gagliardi, O. Haemmerlé, N. Pernelle et F. Said. « An Automatic ontology-based approach to enrich tables semantically ». Proc. Of Context and Ontologies Workshop, Sponsored by the Association for the Advancement of Artificial Intelligence (AAAI). 2005.
- [Gagliardi & al., 2005] H. Gagliardi, O. Haemmerlé, N. Pernelle et F. Said. « An Automatic ontology-based approach to enrich tables semantically ». Proc. Of Context and Ontologies Workshop, Sponsored by the Association for the Advancement of Artificial Intelligence (AAAI). 2005.
- [Gam & al., 2006] Gam I., Camille Salinesi Analyse des Besoins pour la Conception d'Entrepôts de Données, Université de Paris 1, Panthéon Sorbonne, (2006)
- [Gandon, 2006] Gandon, F. (2006). Ontologies informatiques. Interstices, Journal en ligne de l'inria.
- [Garf, 1996] Garf, B. (1996). Lexique de philosophie. Paris: Edition du Seuil.
- [Ghorbel & al. , 2008] H. Ghorbel, A. Bahri et R. Bouaziz. « Les langages de description des ontologies : RDF & OWL ». Actes de la conférence Génie Electrique et Informatique (GEI 2008). Pages 597–606. Sousse, Tunisie. Mars 2008.
- [Ghorbel & al., 2009] Ghorbel,H.,Bahri,A., Bouaziz,R. Fuzzy Protégé for fuzzy ontology models. In: Proc. Of 11thInternational Conference IPC'2009. Academic Medical Center, University of Amsterdam, Amsterdam,Netherlands.(2009)
- [Ghozzi et al., 2005] Ghozzi F., Méthode de conception d'une base multidimensionnelle contrainte. In RNTI, Entrepôts de Données et l'Analyse en ligne (EDA'05), volume RNTI B-1 (2005)
- [Ghozzi & al., 2008] F. Ghozzi , (2008) Conception et manipulation de bases de Données dimensionnelles a contraintes, Thèse, université Toulouse III - Paul Sabatier
- [Ghozzi, 2010] Ghozzi Faiza, Conception et manipulation de bases de données dimensionnelles a` contraintes, Thèse, 2010
- [Giorgini & al., 2005] Giorgini P. & al. (2005) “Goal-oriented requirement analysis for DW design”, DOLAP'05. Proceedings of the 8th ACM international workshop on DWsing and OLAP, pages 47–56.

- [Giorgini et al., 2006] Giorgini, P., Rizzi, S., & Garzetti, M. (2006). Grand: A goal-oriented approach to requirement analysis in data warehouses. *Decision Support Systems*, 45(1), 4–21. DOI: 10.1016 /j.dss.2006.12.001
- [Giorgini & al., 2008] Giorgini P. & al., (2008) GRAnD: A goal-oriented approach to requirement analysis in data warehouses, *Decision Support Systems* 45 4 –21
- [Girard & al. , 1999] N. Girard , B. Hubert Modelling expert knowledge with knowledge-based systems to design decision aids, 1999. The example of a knowledge-based model on grazing management
- [Goh et al., 1999] C.H. Goh, S. Bressan, E. Madnick, and M.D.Siegel. Context interchange: New features and formalisms for the intelligent integration of information. *ACM Transactions on Information Systems*, 17(3):270–293, 1999.
- [Golfarelli & al., 2002] Golfarelli, M. & al., (2002) "WAND: A CASE Tool for Workload-Based Design of a Data Mart.- SEBD", Citeseer
- [Golfarelli & al., 2009] Golfarelli M. & al. ; (2009) "DW Design: Modern Principles and Methodologies". Book; USA
- [Golfarelli & al., 2009] Golfarelli M. & al., (2009), "A comprehensive approach to DW testing". Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP. ACM.
- [Golfarelli et al., 2009] Golfarelli M., A comprehensive approach to data warehouse testing. Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP. ACM, (2009)
- [Golfarelli et al., 2009] M. Golfarelli. From user requirements to conceptual design in data warehouse design- a survey. In *Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction*, pages 1–16. IGI Global, 2009.
- [Golfarelli et al., 2010] Matteo Golfarelli, Chapter 1, From User Requirements to Conceptual Design, *Data Warehouse Design. A Survey 2010 DEIS - University of Bologna, Italy*
- [Golfarelli et al., 2011] Golfarelli M., Rizzi S., Turricchia E., *Modern Software Engineering Methodologies Meet Data Warehouse Design: 4WD*, (2011)
- [Golfarelli & al., 2011a] Golfarelli M.& al. *Modern Software Engineering Methodologies Meet Data Warehouse Design: 4WD; DaWaK, LNCS 6862*, (2011), Springer
- [Gómez-Pérez et al., 2004] Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological Engineering (with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web)*. Springer.
- [Goncalves & al., 2001] M. Goncalves & al. “sqlf Flexible Querying Language Extension by means of the norm SQL2”, The 10th IEEE International Conference on Fuzzy Systems, 2001.
- [Gonzalez & al., 2009] Claudia González, Marlene Goncalves and Leonid Tineo, “A New Upgrade to sqlf : Towards a Standard in Fuzzy Databases”, 2009
- [González et al., 2014] Claudia González, Marlene Goncalves and Leonid Tineo, “A New Upgrade to sqlf: Towards a Standard in Fuzzy Databases”, 2009
- [Gruber, 2008] T. R. Gruber. « Ontology. Proc. Of the Encyclopedia of Database Systems », Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, ISBN 978-0-387-49616-0. 2008.
- [Guarino, 1998] N.Guarino “Formal ontology in information system “frontiers in artificial intelligence and application IOS press 1998.
- [Guergour, 2007] Habib-ellah Guergour Construction d’une ontologie d’application dans le cadre de l’eai MÉMOIRE Présenté en vue de l’obtention du diplôme de Magister en informatique Option : Génie logiciel & intelligence artificielle 16/06/2007.
- [Gupta et al., 1997] H. Gupta. Selection of views to materialize in a data warehouse. Proceedings of the 6th International Conference on Database Theory (ICDT '97), pages 98–112, 1997.
- [Gupta et al., 1997] H. Gupta, V. Harinarayan, A. Rajaraman, and J. Ullman. Index selection for olap. Proceedings of the International Conference on Data Engineering (ICDE), pages 208–219, April 1997.
- [Hachaichi & al., 2013] Hachaichi Y. (2013) "An automatic method for the design of multidimensional schemas from object oriented databases", *International Journal of Information Technology & Decision Making* Vol. 12, No. 6 1223–1259

- [Hammouche & al., 2017] D. Hammouche, M. Loukam, Fuzzy MDX Requêtes for taking into account the ambiguity in querying the baccalaureate data warehouse , 4th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE November 2017
- [Huang & al., 2006] Liang Huang, Kevin Knight, Aravind Josh, A Syntax-Directed Translator with Extended Domain of Locality., Workshop on Computationally Hard Problems and Joint Inference in Speech and Language Processing, pages 1–8, New York City, New York, June 2006.
- [Hull et al., 2010] Hull, E., Jackson, K., & Dick, J. (2010). Requirements engineering. Springer Science & Business Media.
- [Inmon et al., 2001] Inmon, W.H., Joyce Montanari, Bob Terdeman, and Dan Meers. Data Warehousing for E-Business. New York: John Wiley & Sons. 2001.
- Inmon, W.H. Building the datawarehouse. Third Edition, New York: John Wiley & Sons. 2002.
- [Jarzabek & al., 1998] S. Jarzabek , “Design of Flexible Static Program Analyzers with PQL”, IEEE Computer Society 1998
- [Jean, 2007] Jean S., OntoQL, un langage d’exploitation des bases de données à base ontologique, Thèse, France, (2007)
- [Jensen & al., 2004] Jensen C.S. (2004) "Multidimensional data modeling for location-based services". The VLDB Journal.
- [Jiang & al., 2011] Jiang, Y.& al., Reasoning and change management in modular fuzzy ontologies. Expert Systems with Applications 38 (11), 13975–13986. (2011)
- [Jilani, 2010] Zouhir Djilani, Donner une autre vie à vos besoins fonctionnels : une approche dirigée par l’entreposage et l’analyse en ligne. Thèse, ENSMA, 2017
- [Jindal & al., 2012] Jindal R. & al., (2012) "Comparative study of DW Design approaches: a survey", IJDMS
- [Joseph & al., 1998] Joseph J., James D. Palmer Romano TBRIM: Decision Support for Validation Verification of Requirements 1998, Fairfax, VA 22030 USA
- [Jovanovic & al., 2014] P. Jovanovic et al. (2014) A requirement-driven approach to the design and evolution of data warehouses. Information Systems 44 94–119
- [Kahramana & al., 2015] Cengiz Kahramana, Sezi Cevik Onara & Basar Oztaysia Fuzzy Multicriteria Decision-Making: A Literature Review 2015 Istanbul, Turkey May 2015.
- [Khalfi, 2009] Souheila Khalfi ; construction d’une ontologie pour la prise en charge des patients à domicile mémoire de magister discipline : informatique, option : systèmes distribués université Mentouri Constantine 2008/2009.
- [Kimball et al., 1996] Ralph Kimball, The data warehouse toolkit: Practical Techniques for Building Dimensional Data Warehouses, John Wiley and Sons, 1996
- [Kimball et al., 2002] Ralph Kimball, Margaery Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd Edition, John Wiley and Sons, 2002.
- [Kimball, 2008 & al.] R. Kimball, M. Ross & al. (2008) “The DW Lifecycle Toolkit”
- [Khouri & al., 2012] Khouri, S.;Boukhari I.;Bellatreche L.;Sardet E.; Jean S.; Baron M.; Ontology-based structured web data warehouses for sustainable interoperability :requirement modeling, design methodology and tool, Computers in industry; Elsevier (2012)
- [Khouri et al., 2014] Selma Khouri, Cycle de vie sémantique de conception de systèmes de stockage et manipulation de données. Thèse 2014.
- [Khouri et al., 2011] S. Khouri and L. Bellatreche. Dwobs: Data warehouse design from ontology-based sources. In 16th International Conference on Database Systems for Advanced Applications (DASFAA’ 2011), pages 438–441, 2011.

- [Klai, 2009], Sihem magistère «construction d'une ontologie a partir de bases de données pour l'aide a la maintenance industrielle application », 2009 .
- [Kumar & al., 2011] Manoj Kumar, Quality-Oriented Requirements Engineering for a Data Warehouse, ACM SIGSOFT Software Engineering, (2011)
- [Lamsweerde et al., 2009] Lamsweerde, A. Van, (2009). Requirements engineering: from system Buts to UML models to software specifications.
- [Larbi & al., 2013] Larbi A.; Malki M.; Boukhalfa K.; Layachi h.; Modeling the Imprecision of Flexible Requêtes Using a Fuzzy SQL Langage; 2nd International Conference on Software Engineering and New Technologies; ICSSENT'13. Tunis. ISBN: 978-9938-12-094-3, (2013)
- [Larbi & al., 2017] Larbi A., MALKI M., Boukhalfa K., Merzoug A., A Survey of Decisional Requirements: Imprecision study; JFSE 2017; Avril 2017 Proceeding, ISSN 1613-0073, pp 21-26
- [Larbi & al., 2017a] Larbi Abdelmadjid, Malki Mimoun, B. Seddiki, I. Larbi, A semi-automatic solution for XML query response enrichment using a terminological domain ontology, Proceedings of the International Conference on Computing for Engineering and Sciences (ICCES 2017), Istanbul, Turkey — July 22 - 24, 2017, ACM New York, NY, USA ©2017, Pages: 76-81, doi>10.1145/3129186.3131941; ISBN: 978-1-4503-5309-0
- [Larbi & al., 2018] Larbi Abdelmadjid, Malki M., Requêtes-based requirements Imprecision Study for Data Warehouse Update: Structural Approach. 8th International Conference on Information Systems and Technologies Istanbul, Turkey 16 - 18 March 2018, ACM ISBN 978-1-4503-6404-1, <https://doi.org/10.1145/3200842.3200851>
- [Larbi & al., 2018a] Larbi Abdelmadjid, Malki M., Boukhalfa K., Fuzzy Ontology-based Approach for the Requirements Query Imprecision Assessment in DW Design Process. NEAR Negative Fuzzy Operator ; International Journal of Information Technology and Computer Science (IJITCS). Volume 10 N°2, ISSN: 2074-9007 (Print), ISSN: 2074-9015 (Online). DOI: 10.5815/ijitcs.2018.02.03/IJITCS
- [Lebraty et al., 2006] Lebraty Jean-Fabrice, Les systèmes décisionnels. Encyclopédie de l'informatique et des systèmes d'information, Vuibert, pp.1338-1349, 2006.
- [Liu, 1986] Chen-Ching Liu, Kevin Tomsovic, An expert system assisting decision-making of reactive power/voltage control, 1986
- [Liu & al., 2014] Shaofeng Liu, Pascale Zaraté, Knowledge Based Decision Support Systems: A Survey on Technologies and Application Domains – 2014, International Conference on Group Decision and Negotiation
- [Loucopoulos et al., 1989] Loucopoulos, P., & Champion, R. E. M. (1989). Knowledge-based support for requirements engineering. *Information and software technology*, 31(3), 123-135.
- [Lubars et al., 1993] M. Lubars, C. Potts, and C. Richter. A review of the state of the practice in requirements modeling. In Proceedings of IEEE International Symposium on Requirements Engineering (RE'93), pages 2–14, 1993.
- [Lamsweerde et al., 2009] Lamsweerde, A. Van, (2009). Requirements engineering: from system Buts to UML models to software specifications.
- [Lee & al., 1999] J. Lee & al., (1999) "Modeling imprecise requirements with fuzzy objects", *Information Sciences* 118 101–119.
- [Lee & al., 2003] J. Lee & al. (2003) "Modeling imprecise requirements with XML", *Information and Software Technology* 45 445–460
- [Legros, 2009] Legros D., "Matrise des risques dans les systmes de transport", Thèse, (2009).
- [Lujan-Mora & al., 2006] Lujan-Mora S. & al., "Physical modeling of DWs using UML component and deployment diagrams: design and implementation issues". *Journal of Database Management (JDM)* 17.2 (2006): 12-42.
- [Marhavilas & al., 2011] Marhavilas P.K, Risk analysis and assessment methodologies in the work sites: On a review, classification and comparative study of the scientific literature of the period 2000–2009, (2011)

- [Martinez-Cruz et al., 2012] Carmen Martínez-Cruz ,Albertvan der Heide, Daniel Sánchez, bc Gracian Trivino, An approximation to the computational theory of perceptions using ontologies, Expert Systems with Applications, Volume 39, Issue 10, August 2012, Pages 9494-9503
- [Martinez-Cruz & al., 2014] Martinez-Cruz, C. ; An Ontology to represent Requêtes in Fuzzy Relational Databases, ISDA, ,Page(s): 1317-1322 ISSN : 2164-7143 , (2014)
- [Mayer & al., 2006] Mayer N., Humbert P., la gestion des risques pour les systèmes d'information, Revue MISC, Luxembourg. Avril-Mai (2006)
- [Maté et al., 2012] Maté A., Trujillo J., de Gregorio e., improving the Maintainability of Data Warehouse Designs: Modeling Relationships between Sources and User Concepts, International Workshop on Data Warehousing and OLAP, DOLAP'12, (2012)
- [Mazón & al., 2007] Mazón J.N.& al. (2007) A Model-Driven Goal-Oriented Requirement Engineering Approach for DWs. International Workshop on Ontologies and Information Systems for the Semantic Web.
- [Mazon & al., 2008] Mazon J.N. & al., (2008) Applying the i\* framework to the development of DWs. Proceedings of the 3rd Intern.I\* Workshop
- [Mazón & al., 2009] Mazón, J.N. & al. (2009) "A hybrid model driven development framework for the multidimensional modeling of data warehouses". ACM SIGMOD Record 38.2: 12-17.
- [Mazziéri, 2004] M. Mazziéri. « A fuzzy RDF Semantics to represent trust Metat-Data ». Proc. Of the 1 st Workshop on Semantic Web Application and Perspectives (SWAP 2004). Pages 83–89. Ancona, Italy. 2004.
- [Memon & al., 2007] Tasneem Memon ; Mian Muhammad Usman ghani ; Nadeem Iftikhar , Category of fuzzy operators in SQL, International Conference on Emerging Technologies, 2007.
- [Mendel et al., 2010] What Computing with Words Means to Me. Jerry M. Mendel, Lotfi A. Zadeh, Enric Trillas IEEE Computational Intelligence Magazine, Volume: 5, Issue: 1, Feb. 2010, DOI: 10.1109/MCI.2009.934561, Print ISSN: 1556-603X
- [Mendel & al., 2010] Mendel, J. M., Lawry, J., Zadeh, L. Foreword to the special section on computing with words. IEEE T. Fuzzy Systems 18 (3), 437–440. (2010)
- [Messlem et al., 2017] Conception et réalisation d'un Dataware house « Métier ressources humaines », Mémoire Master, Messelem Mohamed Merouane, Mebtouche Lilia, Université de Boumerdes, Décembre 2017
- [Miloud & al., 2009] Miloud D. & al., A fuzzy logic model for identifying spatial degrees of exposure to the risk of road accidents, 978-1-4799-0313-9/13 (2013) IEEE 193.
- [Miloud, 2013] Miloud D. al., "A fuzzy logic model for identifying spatial degrees of exposure to the risk of road accidents", 978-1-4799-0313-9/13 (2013) IEEE.
- [Morente-Molinera & al., 2017] Juan Antonio Morente-Molinera, Ignacio Javier Pérez, Francisco Javier Cabrerizo, Sergio Alonso, and Enrique Herrera-Viedma, Using Group Decision Making Methods to Extract Experts Knowledge, 2017
- [Morgan, 2014] M. Granger Morgan Use (and abuse) of expert elicitation in support of decision making for public policy 2014, Carnegie Mellon University, Pittsburgh
- [Musa & al., 2015] Mustafa Musa J. Flexible Data Warehouse Parameters: Toward Building an Integrated Architecture, Intern. Journal of Computer Theory and Eng.(2015)
- [Mylopoulos et al., 1992] Mylopoulos, J., Chung, L., & Nixon, B. (1992). Representing and using non - functional requirements: A process-oriented approach. IEEE Transactions on Software Engineering, 18(6), 483–497. Doi:10.1109 / 32. 142871
- [Navathe et al., 1989] S. Navathe and M. Ra. «Vertical partitioning for database design: a graphical algorithm. ACM SIGMOD, pages 440–450, 1989.

- [Navathe et al., 1995] S. Navathe, K. Karlapalem, M. Ra. A mixed fragmentation methodology for initial distributed database design, JCSE, 1995
- [Navathe et al., 1993] S.B.Navathe, Evolution of data modeling for databases. Communications of the ACM, 35(9) : 112–123, 1992.
- [Nebot et al., 2009] Nebot V. And Berlanga R., Efficient retrieval of ontology fragments using an interval labeling scheme, information Sciences, 2009 –Elsevier
- [Nordin & al , 2016] Ariza Nordin and Suriyani Ariffin; Conceptualizing Data Driven Decision , Chapter 8; Support Component in Academic Information Management System 2016
- [Nuseibeh et al., 2000] Bashar Nuseibeh, Steve Easterbrook, Requirements engineering: a roadmap, Proceeding ICSE '00 Proceedings of the Conference on The Future of Software Engineering, Pages 35-46, ACM New York, NY, USA 2000, ISBN:1-58113-253-0 doi>10.1145/336512.336523
- [Olteanu, 2008] Ana-Maria Olteanu , Fusion de connaissances imparfaites pour l'appariement de données géographiques : proposition d'une approche s'appuyant sur la théorie des fonctions de croyance, Thèse, 2008
- [Paim & al., 2003] Paim R. F. S. & al. (2003) "DWARF: An Approach for Requirements Definition and Management of DW Systems". Proceedings of the 11<sup>th</sup> IEEE International Requirements Engineering Conference.
- [Pardillo et al., 2011] Pardillo J. , Mazón J.N., Using ontologies for the Design of data warehouses , International Journal of Database Management Systems ( IJDMS ), Vol.3, No.2, May (2011)
- [Pardillo et al., 2011] Pardillo J. , Mazón J.N., Using ontologies for the Design of data warehouses , International Journal of Database Management Systems ( IJDMS ), Vol.3, No.2, May (2011)
- [Park & al., 2013] Taeil Park, Hyoungkwan Kim, A data warehouse-based decision support system for sewer infrastructure management, 2013
- [Prakash & al., 2005] Prakash N., Gosain A., Requirements Driven Data Warehouse Development, CAISE Short Paper Proceedings, India. (2005)
- [Pinet & al., 2009] Pinet F. & al.,(2009)"A Unified Object Constraint Model for Designing and Implementing Multidimensional Systems", Journal on Data Semantics XIII, Volume 5530 pp37-71,
- [Pivert & al., 2012] Pivert O.& al., Fuzzy preference Requêtes to relational databases, Book, (2012) - World Scientific
- [Phipps & al., 2002] Phipps C. and Davis K.C. (2002) Automating DW Conceptual Schema Design and Evaluation. In Proceedings of the 4th International Workshop Design and Management of DWs, 23-32.
- [Prat et al., 2006] Prat, N., Akoka J., Comyn Wattiau I., A UML-based data warehouse design method. Decision Support Systems 42.3, (2006)
- [Quintana-Amate & al., 2015] S. Quintana-Amate, Pablo Bermell-Garcia, A. Tiwari, Transforming expertise into Knowledge-Based Engineering tools: A survey of knowledge sourcing in the context of engineering design, 2015
- [Ravat , 2007] 23. Ravat F., OLAP Aggregation function for textual data warehouse. ICEIS, 2007
- [Reix et al., 1995] Robert Reix, Systèmes d'information et management des organisations, Éditions Vuibert, First Edition ; 1995, 367 pages
- [Rivière et al., 2014] Guillaume Rivière, Informatisation du Système d'Information, ESTIA 2ème Année –, Dernière révision : Avril 2014 ESTA, Pays Basque
- [Rizzi & al., 2006] Rizzi S. & al., (2006) "Research in DW modeling and design: dead or alive?" Proceedings of the 9th ACM international workshop on Data warehousing and OLAP.
- [Rochfeld et al., 1983] A. Rochfeld and H. Tardieu. Merise: An information system design and development methodology. Information & Management, 6:143–159, 1983.
- [Rodríguez & al., 2014] Rodríguez ND. & al., A fuzzy ontology for semantic modelling and recognition of human behavior - Knowledge-Based ..., (2014) – Elsevier
- [Rolland et al., 1998] C. Rolland, C. Souveyet, and C.B. Achour. Guiding goal modeling using scenarios. IEEE Transactions on Software Engineering, 24(12):1055–1071, 1998

- [Rolland et al., 1999] C. Rolland, G. Grosz, and R. Kla. Experience with goal-scenario coupling in requirements engineering. In Proceedings. IEEE International Symposium on Requirements Engineering, (RE 1999)., pages 74–81, 1999.
- [Rolland, 2003] Colette Rolland, Ingénierie des Besoins : L'Approche L'Ecritoire, Journal Techniques de l'Ingénieur, 2003, pp.1.
- [Romero & al., 2007] Romero, O&al. (2007) "Automating multidimensional design from ontologies". Proceedings of the ACM 10<sup>o</sup> International workshop on Data warehousing and OLAP. ACM.
- [Romero & al., 2009] Romero O. Abelló; A Survey of Multidimensional Modeling Methodologies, International Journal of Data Warehousing & Mining, (2009)
- [Romero & al., 2010] Romero, O.; Abelló, A.; (2010) A framework for multidimensional design of DWs from ontologies; Data & Knowledge Engineering; Volume 69, Issue 11, Pages 1138–1157– Elsevier.
- [Romero & al., 2011] Romero, O. &al. (2011) "GEM: requirement-driven generation of ETL and multidimensional conceptual designs". Data Warehousing and Knowledge Discovery. Springer, 80-95.
- [Sabri & al., 2012] Sabri, A. ; Kjiri L. (2012) "Processus de traitement et de formalisation des besoins métier d'un SI Décisionnel"
- [Sabri & al., 2012] Sabri Aziza, Kjiri Laila ; Patterns to analyze requirements of a Decisional Information System, Journal of Computer Application (IJCA), Special Issue On Software Engineering Databases and EXpert Systems (SEDEXS), Number 2, ISBN: 973-93-80870-26-8, 17, 2012
- [Sabri et al., 2013] Aziza Sabri, Vers une approche d'analyse des besoins métier d'un Système d'Information Décisionnel, ETI, Revue électronique en Technologies de l'Information. [Http://www.revue-eti.net](http://www.revue-eti.net), Numéro 7. 2013. ISSN 1114-8802
- [Sanchez, 2006] E. Sanchez et T. Yamanoi. « Fuzzy Ontologies for the Semantic Web ». Proc. Of the 7th International Conference on Flexible Query Answering Systems (FQAS 2006). Pages 691–699. Milan, Italy. 2006.
- [Sanjay et al., 2004] A. Sanjay, V. R. Narasayya, and B. Yang. Integrating vertical and horizontal partitioning into automated physical database design. Proceedings of the ACM SIGMOD International Conference on Management of Data, pages 359–370, June 2004.
- [Sapir & al., 2008] Sapir L. & al., A methodology for the design of FDW 4th Intern. IEEE Conference "Intelligent Systems", (2008)
- [Scukanec et al., 2007] Scukanec, M. & Gaasbeek, J. Van, (2007). A day in the life of a verification requirement. USAir Force T&E Days.
- [Sequeira & al., 2014] Siqueira &al, Modeling vague spatial data warehouses using the VSCube conceptual model TLL Geo Informatica, (2014) – Springer
- [Seo & al., 2009] Seo, K. & al.. Ontology-based fuzzy support agent for ship steering control. Expert Systems with Applications 36 (1), 755–765. (2009)
- [Sfafi et al., 2014] Lilia Sfafi, Business Intelligence, Institut National des Sciences Appliquées et de Techn, Tunis, 2014
- [Siu-Kwong et al., 1996] E. Siu-Kwong. Yu. Modelling strategic relationships for process reengineering. PHD thesis, University of Toronto, 1996.
- [Siu-Kwong et al., 1997] E. Siu-Kwong. Yu. Towards modeling and reasoning support for early-phase requirements engineering. In Proceedings of the 3<sup>rd</sup> IEEE International Symposium on Requirements Engineering, (RE'97), page 226. IEEE Computer Society, 1997.
- [Smits & al., 2012] Gregory Smits, "Postgre sqlf: A Fuzzy Query System", December 2012
- [Soler, 2008] Soler E., (2008) Towards Comprehensive Requirement Analysis for DWs: Considering Security Requirements Third Intern. Conference on Availability, Reliability and Security, Spain.

- [Song et al., 2007] Song Yeol, Ritu Khare, and Bing Dai. "SAMSTAR: a semi-automated lexical method for generating star schemas from an entity-relationship diagram". Proceedings of the ACM tenth international workshop on Data warehousing and OLAP. ACM, (2007)
- [Srivastava et al., 1996] D. Srivastava, S. Dar, H. Jagadish and A. Y. Levy. Answering Requetes with aggregation using views». Proceedings of the International Conference on Very Large Databases, pages 318{329, September 1996.
- [Soussi & al., 2005] Soussi A. & al. (2005) « Approche semi automatisée de conception de schémas multidimensionnels valides ». Revue des Nouvelles Technologies de l'Information. EDA'05
- [Straccia, 2007] U. Straccia. « A Fuzzy Description Logic for the Semantic Web ». Elie Sanchez, editor, Fuzzy Logic and the Semantic Web, Capturing Intelligence (chapter 4). Pages 73–90, Elsevier. 2006.
- [Stohr et al., 2000] T. Stöhr, H. Märten and E. Rahm. Multi-dimensional database allocation for parallel data warehouses. Proceedings of the International Conference on Very Large Databases, pages 273–284, 2000.
- [Supriadi, 2018] Leni Sagita Riantini Supriadi, Low Sui Pheng, Knowledge Based Decision Support System (KBDSS), Chapter ; Business Continuity Management in Construction pp 155-174 |; ISBN 978-981-10-5487-7; Springer Nature Singapore Pte Ltd. 2018
- [Tahani & al., 1977] Tahani V., "A conceptual framework for fuzzy query processing; a step toward very intelligent database systems", Inf. Processing and Management.13, 1977
- [Tamani & al., 2011] N. Tamani, "An Extension of a Fuzzy Ontology for Flexible Querying", 2011
- [Tamani & al., 2013] Tamani N., A Fuzzy Ontology for Database Querying with Bipolar Preferences, International Journal of Intelligent Systems, (2013)
- [Tebourski & al., 2013]Tebourski W., (2013) "Semi-automatic DW Design methodologies: a survey"
- [Thakur & al., 2011] Garima Thakur & al., DWEVOLVE: A Requirement Based Framework for Data Warehouse Evolution , ACM SIGSOFT Software Engineering Notes, Volume 36 Number 6, (2011)
- [Thenmozhi & al., 2012] Thenmozhi M& al., (2012) An Ontology based Hybrid Approach to Derive Multidimensional Schema for DW, International Journal of Computer Applications
- [Thenmozhi & al., 2013] Thenmozhi, M. & al. (2013) "A Tool for Data Warehouse Multidimensional Schema Design using Ontology". International Journal of Computer Science Issues (IJCSI) 10.2.
- [Thomsen et al., 2002] Erik Thomsen. OLAP Solutions: Building Multidimensional Information Systems, 2nd Edition, John Wiley and Sons, ISBN: 0-471-15337 0, Apr 2002
- [Trujillo & al., 2004] Trujillo J.& al., (2004) "Applying UML and XML for designing and interchanging, information for DWs and OLAP applications", Journal of Database Management, Vol. 15, No.1.
- [Turban, 1986] Efraim Turban Paul R., Integrating Expert Systems and Decision Support Systems, 1986, Watkins Univ. of Southern California Los Angeles, California
- [Uschold et al., 1995] Uschold, M., & King, M. (1995). Towards a Methodology for Building Ontologies. In D. Skuce (Ed.), IJCAI'95 Workshop on Basic Ontological Issues in Knowledge Sharing, (pp. 6.1-6.10). Montreal, Canada.
- [Winter, 2003] Winter R., (2003) A Method for Demand-driven Information Requirements Analysis in Data Warehousing Projects
- [Worboys & al., 1998] Worboys M., Imprecision in Finite Resolution Spatial Data, Keele University, UK, GeoInformatica, (1998) – Springer
- [Yue et al., 1987] K. Yue. What does it mean to say that a specification is complete? In Fourth International Workshop on Software Specification and Design (IWSSD-4)., 1987.
- [Zadeh, 1965] L. Zadeh, « Fuzzy sets », Information and Control 8, 1965, p. 338–353.
- [Zadeh, 1975] L. Zadeh Calculus of fuzzy restrictions, Fuzzy sets and their applications to cognitive and ..., 1975
- [Zadeh, 1979] L. Zadeh, Approximate reasoning based on fuzzy logic, Proceedings of the 6th international joint conference on Artificial, 1979

- [Zadeh, 1983] L. Zadeh, Commonsense knowledge representation based on fuzzy logic, CUMINCAD, 1983.
- [Zadeh, 1997] L. Zadeh, The roles of fuzzy logic and soft computing in the conception, design and deployment of intelligent systems, Software Agents and Soft Computing Towards Enhancing Machine Intelligence, 1997
- [Zadeh, 2008] L. Zadeh, Is there a need for fuzzy logic? Information sciences 178 (13), 2751-2779 , 2008
- [Zadeh, 2009] L. Zadeh, Computing with words and perceptions—a paradigm shift Information Reuse & Integration, 2009. IRI'09. IEEE International Conference, 2009
- [Zapeda & al., 2008] Zepeda L., (2008) A Mixed Approach for Data Warehouse Conceptual Design with MDA, ICCSA '08 Proceedings of the international conference on Computational Science and Its Applications
- [Zhai & al., 2008] J. Zhai, L. Yiduo, J. Jiatao et Y. Yi. « Fuzzy Ontology Models Based on Fuzzy Linguistic Variable for Knowledge Management and Information Retrieval ». Proc. Of the 5th international conference on Intelligent Information Processing (IIP2008). Pages 58–67. Beijing, China. 2008.
- [Zoghliami et al., 2013] Asma Zoghliami, Modélisation et conception de systèmes d'information géographique gérant l'imprécision, Thèse, 2013,

### **Netographie**

- [Net01] <http://www.piloter.org> , Piloter la Performance: Méthodes, Outils et Techniques, Consulté 01/2018
- [Net02] <http://eduscol.education.fr/math>, Portail national de ressources EDUSOL, Consulté 01/2018

# ANNEXE

## Annexe A

**Requête SQL flou :** Nous allons maintenant considérer l'utilisation de la logique floue dans la formulation de requêtes sur des bases de données relationnelles standard. Pour illustrer les différences entre requêtes conventionnelles et requêtes floues, nous allons nous servir d'une petite base de données de "profils" de sociétés. Le tableau suivant montre le contenu de cette base [Cox & al., 1998]

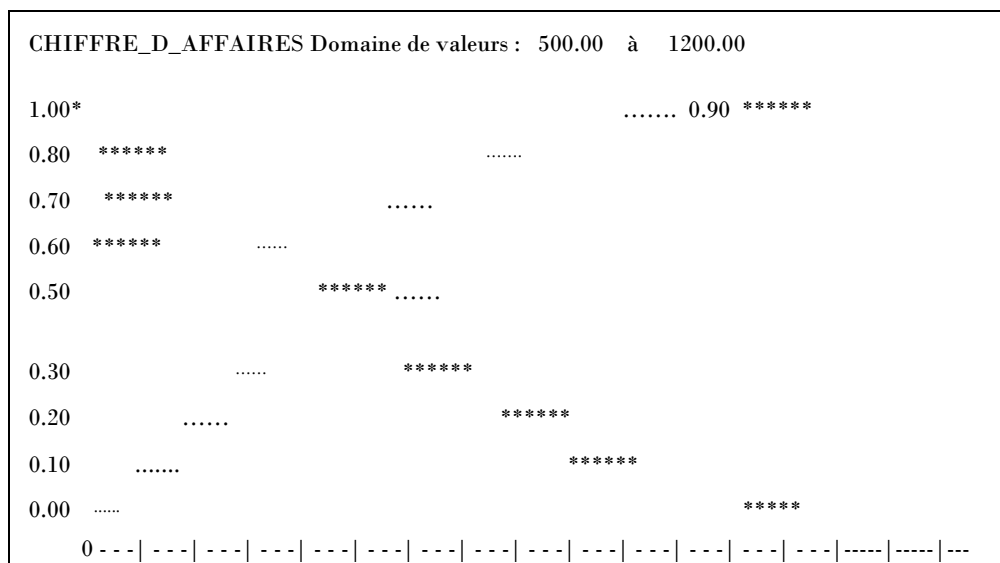
Société	Année de création	C.A annuel	Nombre de produits	Nombre d'employés	Bénéfice ou perte	Bénéfice par action
SocA	1948	570	1	16	3	0,2
SocB	1972	478	3	27	243	1,77
SocC	1982	401	9	58	17	1,02
SocD	1987	321	3	8	25	3,11
SocE	1988	650	5	33	-87	0,44
SocF	1994	550	16	9	99	0,79
SocG	1992	597	4	11	-13	0,04
SocH	1990	602	7	41	33	2,14
SocI	1991	498	5	15	207	1,33
SocK	1994	501	2	57	133	2,83
SocL	1985	920	8	5	201	1,55
SocM	1983	555	2	9	509	1,05
SocN	1988	1050	10	25	427	2,94
SocO	1990	530	14	12	22	0,3
SocP	1988	522	4	21	62	1,28
SocR	1959	900	3	5	-8	0
SocS	1964	659	2	7	0	0,1
SocT	1979	619	5	17	280	3,17
SocU	1967	1233	2	28	322	2,4
SocW	1964	973	11	48	422	2,88

Tableau Base de données "société"

### a. Une requête simple en SQL flou

Pour exprimer l'intention d'une requête, plutôt que son mécanisme, nous voudrions arriver à ce que les instructions SQL qui sélectionnent les enregistrements de la base, le fassent en fonction d'une interprétation plus large de l'instruction WHERE. Les ensembles flous sont une solution à ce problème, puisque, en tant que variables linguistiques, ils permettent de représenter le domaine conceptuel de la requête. Ainsi, afin de pouvoir exprimer des requêtes floues sur une base de données, nous devons d'abord définir les domaines de valeurs à l'aide de descripteurs flous [92, 93]. La figure suivante montre ces descripteurs pour

### CHIFFRE\_D\_AFFAIRES.élevé et CHIFFRE\_D\_AFFAIRES.bas.



• Ensemble flou : ÉLEVÉ  
 Support : 500.00 1200.00  
 \* ensemble flou : BAS  
 Support : 500.00 1200.00

Figure Ensembles flous pour CHIFFRE\_D\_AFFAIRES.élevé et CHIFFRE\_D\_AFFAIRES.bas

L'ensemble flou CHIFFRE\_D\_AFFAIRES.élevé a une valeur d'appartenance non nulle à partir d'environ 500 M\$, qui croît ensuite linéairement jusqu'à 1200 M\$. L'ensemble flou CHIFFRE\_D\_AFFAIRES.bas est son complémentaire. Cette nouvelle description sémantique du terme ÉLEVÉ commence à une valeur légèrement supérieure à 600 M\$ (par le choix d'un seuil de prise en compte de niveau 0,1) et attribue des valeurs de vérité de plus en plus significatives aux montants supérieurs, ce qui correspond à l'intuition. L'ensemble flou CHIFFRE\_D\_AFFAIRES.élevé devient un terme du vocabulaire de notre SQL. La figure montre comment il est utilisé dans une requête approximative.

Seuil de prise en compte: 0,1	
SELECT société FROM BDRSoc WHERE chiffre_d_affaires IS ÉLEVÉ;	
Société	Indice de compatibilité
SocU	1,000
SocN	0,782
SocW	0,669
SocL	0,595
SocR	0,568
SocS	0,226
SocE	0,210
SocT	0,167

SocH	0,144
SocG	0,136
Enregistrements sélectionnés : 10 Rejets dus au seuil de prise en compte : 5	
Pourcentage (%) de la base complète : 50,00	

Tableau Requête simple en SQL flou pour "chiffre d'affaires élevé"

La requête en SQL flou donne comme résultat une liste de sociétés classée par ordre décroissant d'indice de compatibilité (qui est ici simplement la valeur d'appartenance de la valeur `chiffre_d_affaires_annuel` à l'ensemble flou `CHIFFRE_D_AFFAIRES.élevé`). Nous pouvons constater que notre intention réelle a été mieux prise en compte. Les sociétés sont distribuées le long de l'ensemble flou, nous indiquant dans quelle mesure elles satisfont à notre notion de ce qu'est un chiffre d'affaires "élevé". En particulier les sociétés SocU, SocN et SocW sont bien classées comme très représentatives. En queue de liste, la société SocG n'obtient qu'un score assez faible.

**b. Seuils de prise en compte :** On peut aussi remarquer que certaines sociétés ayant un chiffre d'affaires entre 500 M\$ et 1200 M\$ n'ont pas été incluses dans la liste (par exemple la société SocA). Ceci est dû au seuil de prise en compte spécifié dans la requête. Toute valeur d'appartenance inférieure à ce seuil est en fait considérée comme nulle.

La liste de la figure montre toutes les valeurs d'appartenance de `chiffre_d_affaires_annuel` à l'ensemble flou `CHIFFRE_D_AFFAIRES.élevé`, pour toutes les sociétés de la base. Les valeurs du chiffre d'affaires et de l'indice de compatibilité sont indiquées entre parenthèses [92, 93].

- 1. SocA chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 570,00;0,097) **Rejet sur seuil : 0,097 < 0,100**
- 2. SocB chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 478,00;0,000)
- 3. SocC chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 401,00;0,000}
- 4. SocD chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 321,00;0,000)
- 5. SocE chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 650,00;0,210)
- 6. SocF chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 550,00;0,070) **Rejet sur seuil : 0,070 < 0,100**
- 7. SocG chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 597,00;0,136)
- 8. SocH chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 602,00;0,144}
- 9. SocI chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 498,00;0,000)
- 10. SocK chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 501,00;0,000)
- 11. SocL chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 920,00;0,595)
- 12. SocM chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 555,00;0,078} **Rejet sur seuil: 0,078 < 0,100**
- 13- SocN chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' (1050,00;0,782)
- 14. SocO chiffre\_d\_affaires\_annuel dans 'ÉLEVÉ' ( 530,00;0,039) **Rejet sur seuil: 0,039 < 0,100**

Figure Analyse des degrés d'appartenance pour la requête en SQL flou  
 Cinq enregistrements sont éliminés du fait de leur **trop faible degré d'appartenance**. Le message de rejet apparaît en gras après chacun des enregistrements concernés. Ce seuil est un élément important d'une requête en SQL flou. Un seuil est appliqué à la valeur de vérité composite de l'instruction WHERE complète (qui peut contenir un grand nombre de propositions floues) et se comporte comme une barrière en ce qui concerne l'admission dans la liste résultat. Le fait, par exemple, de hausser le seuil à la valeur 0,5, conduit à la liste beaucoup plus réduite visible en figure

Seuil de prise en compte: 0,5 -		SELECT société
FROM BDRSoc		
WHERE chiffre_d_affaires IS ÉLEVÉ;		
<u>Société</u>		<u>Indice de compatibilité</u>
SocU		1,000
SocN		0,782
SocW		0,669
SocL		0,595
SocR		0,568
Enregistrements sélectionnés		: 5
Rejets dus au seuil de prise en compte		: 10
Pourcentage (%) de la base complète		: 25,00

Tableau : requête en SQL flou avec un seuil de prise en compte de 0,5

## Annexe B : Evaluation du risque

Le risque est tout événement susceptible de se produire et de compromettre la réussite du projet. Le **type de risques des projets** dépendent de plusieurs caractéristiques telles que : la taille du projet, difficulté technique, degré d'intégration, configuration organisationnelle, changement et Instabilité de l'équipe projet

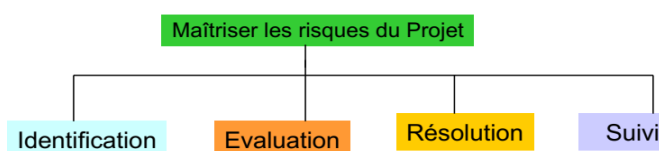


Figure Risque de projet

**La phase d'analyse :** Elle permet d'identifier les risques puis d'évaluer les risques selon leur criticité

**La phase de gestion des risques :** Permet d'assurer un suivi régulier du plan d'action et de ses résultats, gérer l'évolution des risques retenus (statut et résultats des actions entreprises) et s'assurer que les risques non retenus n'ont pas évolués

### Cas d'étude 1 : Evaluation du risque dans la gestion de projet intelligent

Cette modélisation interactive d'entreprise associe l'évaluation de risque d'un projet, budgétisation du capital, et aide à la décision stratégique, dans un même système utilisant sa propre gestion de bases de données relationnelles. La figure 5.2 montre le schéma général des composants du modèle de gestion de projet.

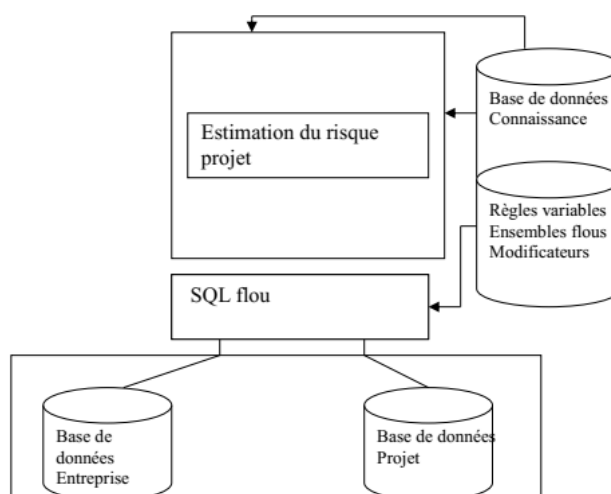


Figure Système gestion de projet et estimation de risque [Bosc, 95]

Ce modèle comporte trois modules principaux :

- ❖ Estimation de risque, qu'est un petit système expert indépendant qui contient un jugement d'experts cohérent, au sujet du risque à lancer un projet, en fonction d'un certain nombre de caractéristique,
- ❖ SQL flou et
- ❖ Base de données de l'entreprise

### **Justification du choix du système expert pour l'évaluation du risque**

Vu l'absence de méthodes ou d'algorithmes pour la gestion de risque, on a opté pour la résolution du problème par le biais du système expert où on a pu accueillir l'expertise de quelques cadres et chefs de projets dans le domaine gestion de projet. L'évolution rapide de la technologie en informatique et dans l'industrie, la difficulté à tirer parti des résultats de la recherche, et le taux élevé d'«usure» des employés techniquement compétents, incite à la prudence en ce qui concerne les projets à haute technologie. De fait, les échecs et les dérapages importants sont devenus fréquents, et presque attendus. Les chefs de projet expérimentés, les analystes techniques et les cadres moyens ont tous quelques idées sur ce que sont des projets à haut ou à faible risque. Coder cette connaissance dans un système d'estimation de risque d'un projet devrait permettre aux cadres supérieurs et aux dirigeants (sans revenir au chef de projet lui-même), de peser les avantages et les inconvénients respectifs de différents projets proposés

**Description de monde réel :** On présente dans cette partie la description de la BD Entreprise [Bosc, 1995] pour qu'elle puisse supporter le concept de requêtes flexibles. On a ajouté des nouveaux objets flous en particulier les étiquettes linguistiques, etc. Cette description est comme suit :

- ✓ Un département possède un nom, un numéro unique et un directeur dont la date d'entrée est mémorisée.
- ✓ Un département possède un local. Il peut contrôler plusieurs projets mais un projet est contrôlé par un seul département.
- ✓ Chaque projet est caractérisé par un nom, un numéro, un budget, une durée de réalisation, un effectif projet, un nombre de phases.
- ✓ Un employé est décrit par un nom, un matricule unique, une adresse, un salaire, un sexe et une date de naissance.
- ✓ Un employé possède un seul supérieur hiérarchique immédiat. Il possède aussi un niveau d'études et un rendement.
- ✓ Un employé est affecté à un seul département, mais il peut travailler sur plusieurs projets non nécessairement rattachés à son département. Le nombre d'heures par semaine passée par un employé sur chaque projet est mémorisé.

- ✓ Un poste est décrit par un code unique, un nom (profession) et exige un niveau d'études et une expérience.
- ✓ Chaque employé possède des personnes qui sont à sa charge (conjoint, enfants, etc.) ou dépendants.

Ces derniers sont décrits par : un prénom, un sexe, une date de naissance et un lien de parenté avec l'employé.

Chaque projet à des risques, le niveau de risque d'un projet est souvent un facteur pris en compte, ce risque dépend de:

- (1) l'expérience du chef projet,
- (2) l'effectif projet,
- (3) l'effectif budget,
- (4) la complexité,
- (5) la durée,
- (6) le rendement d'employés

Le système expert est composé de deux parties indépendantes [Bosc, 1995]:

- une base de connaissances elle-même composée d'une base de règles qui modélise la connaissance du domaine considéré et d'une base de faits qui contient les informations concernant le cas que l'on est en train de traiter. Pour notre cas le modèle estimation de risque, il y a deux experts, qui déterminent le risque associé à un projet, compte tenu des valeurs prévues pour la durée, le budget et l'effectif. La première pile établie par Bill Smith, un chef de projet confirmé, contient quatre règles :

- si durée\_projet est longue alors risque est acceptable.
- si durée\_projet est assez\_longue alors risque est important.
- si budget\_projet est élevé alors risque est important.
- si effectif\_budget est confortable alors risque est élevé.

La seconde pile, proposée par Mary Williams, directrice de recherche de la division Marketing, contient trois règles :

- si durée\_projet est longue alors risque est important
- si budget\_projet est élevé alors risque est acceptable
- si effectif\_budget est confortable alors risque important

- Un moteur d'inférences capable de raisonner à partir des informations contenues dans la base de connaissance, de faire des déductions, etc.

Le système expert est souvent complété par des interfaces plus ou moins riches permettant un dialogue avec les utilisateurs, l'idéal étant une interface en langage naturel.

Certains attributs décrits précédemment (en gras) peuvent être interrogés par des requêtes flexibles (floues). Ils possèdent les caractéristiques estimées suivantes :

**Salaire** : possède les étiquettes linguistiques définies sur les distributions de possibilité trapézoïdales suivantes :

**BAS**(50, 80, 120,180), **MOYEN**(150, 300, 400,550), **ÉLEVÉ**(400,600,800,1000).

Ces valeurs sont en Dinars Algériens.

**Nombre\_heures** : possède les étiquettes linguistiques :

**PETIT**(0, 2, 7, 8), **MOYEN**(5, 10, 15, 20), **ÉLEVÉ**(15, 20, 25, 30).

Ces valeurs représentent le nombre d'heures.

**Budget** : possède les étiquettes linguistiques : **PETIT**(1000, 5000, 10000, 18000), **MOYEN**(15000, 50000, 100000, 150000), **GROS**(120000, 200000, 500000, 1000000).

Ces valeurs sont en Dinars Algériens.

**Effectif\_budget** : possède les étiquettes linguistiques :

**PETIT**(1000, 5000, 10000, 18000), **MOYEN**(15000, 50000, 100000, 150000), **Confortable**(120000, 200000, 500000, 1000000).

Ces valeurs sont en Dinars Algériens

**Durée** : possède les étiquettes linguistiques :

**PETITE**(0, 10, 50, 100), **ASSEZ\_LONGUE**(80, 120, 180, 220), **LONGUE**(210, 270, 310, 380).

Ces valeurs sont en jours.

**Effectif\_projet** : possède les étiquettes linguistiques définies sur les distributions de possibilité trapézoïdales suivantes :

**PAUVRE**(1, 6, 10, 16), **MOYEN**(14, 18, 22, 28), **CONFORTABLE**(25, 30, 36, 43).

Ces valeurs sont, nombre équipe.

**Complexité** : c'est le nombre de phases du projet, elle possède les étiquettes linguistiques définies sur les distributions de possibilité trapézoïdales suivantes :

**BAS**(1, 2, 4, 6), **MOYEN**(5, 8, 10, 12), **ÉLEVÉ**(11, 13, 15, 17).

Ces valeurs expriment le nombre de phases du projet.

**Âge** : on a ajouté cet attribut à la place de `date_naissance` pour faciliter la représentation et pour ne pas répéter le calcul `DATE_SYSTEME -`

`DATE_NAISSANCE` chaque fois que nous avons besoin de l'âge d'une personne. Il possède les étiquettes linguistiques définies sur les distributions de possibilité trapézoïdales suivantes :

**JEUNE**(18, 22, 30, 35), **ADULTE**(25, 32, 45, 50), **VIEUX**(50, 55, 62, 70)

**Expérience** : possède les étiquettes linguistiques :

**PETITE** (2, 3, 5, 6), **BONNE**(5, 7, 10, 12), **SUFFISANTE**(7, 8, 15, 20), **GRANDE** (12, 15, 50, 50).

Ces valeurs dépendent des nombres d'années travaillés par un employé.

**Niveau de risque** : possède les étiquettes linguistiques : **FAIBLE** (0,5, 10, 15, 20)

**ACCEPTABLE** (15, 20, 25, 40), **ÉLEVÉ** (35, 40, 75, 100). Ces valeurs décrivent le niveau de risque (pourcentage).

## Cas d'étude 2 : Evaluation de risque dans le transport routier (Solution structurelle)

**Le risque dans le secteur des transports:** Les modèles routiers de sécurité mettent l'accent sur le processus de génération de l'accident et les victimes qui surviennent et se situent à l'intérieur du système de transport routier. Pour concevoir et estimer des modèles de surveillance de la sécurité routière, le modélisateur dispose d'une grande liberté, le résultat dépendra des objectifs [Mayer, 2006]. Le tableau ci-dessous décrit la fréquence et la sévérité des critères de qualité et de gravité couramment utilisés dans le chemin de fer [Legros, 2009] et [Miloud, 2013]. Ces critères sont utilisés pour produire un tableau bidimensionnel "Occurrence-gravité", comme dans l'exemple ci-dessous, pour assigner un niveau de risque correspondant à un niveau d'acceptabilité.

Fréquence de Situation dangeureuse	Gravité des conséquences d'une situation dangeureuse			
	Insignifiante	Marginale	Critique	Catastrophique
Frequente	Indésirable	Inacceptable	Inacceptable	Inacceptable
Probable	Acceptable	Indésirable	Unacceptable	Unacceptable
Occasionnel	Acceptable	Indésirable	Indésirable	Unacceptable
Rare	Négligeable	Acceptable	Indésirable	Indésirable
Improbable	Négligeable	Négligeable	Acceptable	Acceptable
Incroyable	Négligeable	Négligeable	Négligeable	Négligeable

Tableau Matrix-type occurrence-gravity [Mayer, 2006]

Notre exemple d'illustration traite le risque du secteur des transports en particulier l'accident qui est le premier danger vécu dans le transport et le plus dangereux. Les différents acteurs de l'accident sont: Le conducteur, le bus et la ligne. Il y a d'autres acteurs comme le climat dont nous ne parlerons pas dans ce document. Ces critères sont inclus dans les normes à titre indicatif seulement pour limiter les critères de décision. Les indicateurs d'accidents sont à la base des programmes de sécurité routière. Les données sur les accidents sont collectées depuis de nombreuses années comme point de départ pour améliorer la sécurité routière.

Les significations des situations dangereuses sont: Inacceptable: Doit être éliminé; Non désiré: Le risque n'est acceptable que lorsque la réduction de ce dernier est impossible; dans ce cas, l'accord de l'opérateur ou des régulateurs de réseau est impératif; Acceptable: Le risque est acceptable avec la supervision appropriée et l'accord de l'exploitant ou de l'organisme de réglementation; Négligeable: Acceptable sans condition. Alors, appliquons cette matrice dans notre cas de transport routier. Ces règles ont été validées avec des experts en ingénierie du trafic. À travers ce

contexte, nous voyons que ce travail représente l'étude d'imprécision dans les besoins décisionnels seulement présentes dans la valeur de prédicat de requête des besoins.

## **Annexe C : Modélisation du flou**

Pour modéliser notre cas avant de l'implémenter, on a fait appel à une variété spécifique d'UML appelé : UML Flou ou Fuzzy UML

**Fuzzy UML** : L'un des axes importants de recherche dans les bases de données est l'effort continu d'enrichir les modèles de bases de données existants d'une collection plus vaste de concepts sémantiques. C'est dans ce contexte qu'une méthodologie de modélisation conceptuelle orientée objet a été développée pour la modélisation floue de l'information. Les diagrammes de classe UML peuvent modéliser des informations floues à travers le concept du FUZZY UML [Klai, 2009]

### **Diagramme de classes floues**

**Classe floue** : Une classe peut être floue à cause des raisons suivantes :

1. Certains objets qui ont des propriétés semblables sont considérés comme flous. Une classe définie par ces objets flous sera donc floue. Les objets appartiennent à la classe avec un degré d'appartenance entre  $[0,1]$ .
2. Quand une classe est intentionnellement définie (classe définie par un ensemble d'attributs et les valeurs admissibles des attributs), le domaine d'un attribut peut être flou et une classe floue est alors formée [Zoghلامي, 2013]
3. Une sous-classe produite par une super-classe floue au moyen d'une relation de spécialisation est considérée comme une classe floue. Suivant ce même principe, une super-classe produite par des sous-classes, dans lesquelles il y a au moins une classe qui est floue au moyen de la relation de généralisation, est considérée comme une classe floue. Donc dans la classe UML, il y a trois niveaux d'imprécision définis comme suit :

Niveau 1: l'imprécision dans l'appartenance de la classe au modèle de données ou l'imprécision dans l'appartenance de l'attribut à la classe. Pour modéliser ce niveau d'imprécision, l'attribut ou le nom de classe devrait être suivi par les mots : « WITH d DEGREE » (avec degré d'appartenance) où  $0 < d \leq 1$ . Le degré d'appartenance est utilisé pour indiquer le degré auquel l'attribut appartient à la classe ou la classe appartient au modèle de données. Généralement, une classe ou un attribut ne sera pas déclaré si son degré est 0. De plus, "WITH 1.0 DEGREE" peut être omise quand le degré d'un attribut ou d'une classe est 1.

Niveau 2 : Il est possible qu'une instance de la classe appartienne à la classe avec un degré d'appartenance. Pour la modélisation de ce deuxième type d'imprécision, nous devons indiquer le degré d'appartenance auquel une instance appartient à la classe. Pour cela, un attribut supplémentaire est introduit dans la classe pour représenter le

degré d'appartenance à la classe avec un domaine d'attribut dans l'intervalle [0,1]. Cet attribut spécial est représenté par «  $\mu$  ».

Niveau 3 : L'imprécision sur les valeurs d'attribut des instances de la classe. Un attribut dans une classe définit un domaine de valeurs. Quand ce domaine définit un sous ensemble flou, l'imprécision d'une valeur d'attribut apparaîtra. Pour modéliser ce niveau d'imprécision, le mot clé fuzzy est introduit et placé devant l'attribut.

La figure suivante illustre un exemple d'une classe floue. Le premier niveau d'imprécision est présent dans la classe à travers l'attribut bureau qui appartient à la classe avec un degré de 0.8. Le deuxième niveau d'imprécision est modélisé par l'insertion de l'attribut  $\mu$  dans la classe. Finalement, le troisième niveau d'imprécision se manifeste dans la classe à travers l'attribut flou âge

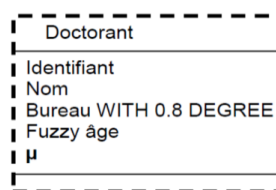


Figure Exemple d'une classe floue

Les différents types de relations entre les classes UML (association, relation de généralisation, composition, agrégation, etc.) peuvent être flous.

### Généralisation floue

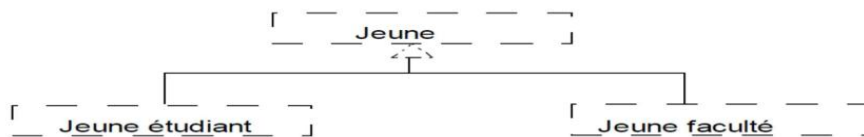


Figure Exemple de relation de généralisation floue

### Association floue

Deux niveaux d'imprécision peuvent être identifiés dans la relation d'association. Niveau 1: une relation d'association existe mais d'une manière imprécise entre deux classes associées.

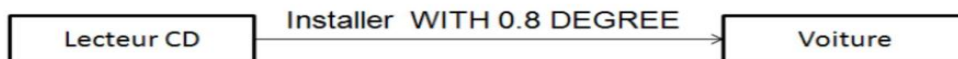


Figure Exemple de relation UML imprécise de premier niveau

Niveau 2 : il est possible qu'on ignore si deux instances de classes, qui appartiennent à deux classes associées auront une relation d'association donnée, bien que cette relation d'association existe entre ces deux classes.

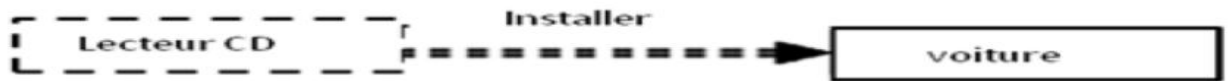


Figure Exemple de relation UML imprécise de deuxième niveau

Niveau 3 : Les deux niveaux de flou mentionnés ci-dessus peuvent se produire simultanément dans une relation d'association.

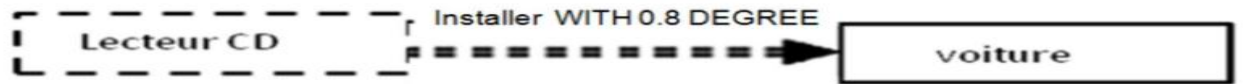


Figure Exemple de relation UML imprécise de troisième niveau

### Agrégation floue

Une agrégation représente une relation de type «ensemble/élément» dans laquelle l'une des classes décrit un tout, alors que la classe associée décrit des parties. Le tout est appelé «composé» et la partie est appelée «composant». Ces composants peuvent exister indépendamment. Ainsi, chaque objet d'un agrégat peut être projeté dans l'ensemble d'objets des éléments constitutifs.

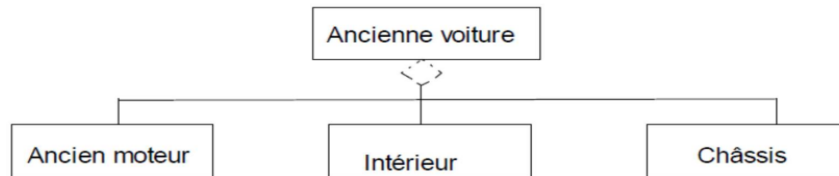


Figure Exemple de relation d'agrégation floue dans Fuzzy UML

### Dépendance floue

Une relation de dépendance floue est une relation de dépendance avec un degré de possibilité, qui peut être indiqué d'une manière explicite par les concepteurs ou implicite par la classe source, en se basant sur le fait que la classe cible est déterminée par la classe source.

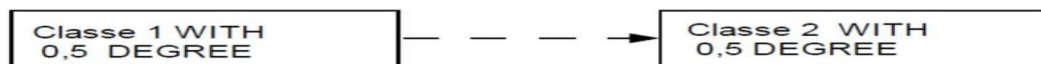


Figure Relation de dépendance floue

### Cas d'utilisation flou

Un cas d'utilisation décrit l'ensemble de séquences d'actions qu'un système effectue pour fournir un résultat observable ayant une valeur pour un acteur. Fuzzy UML permet de spécifier les besoins incertains d'un système.



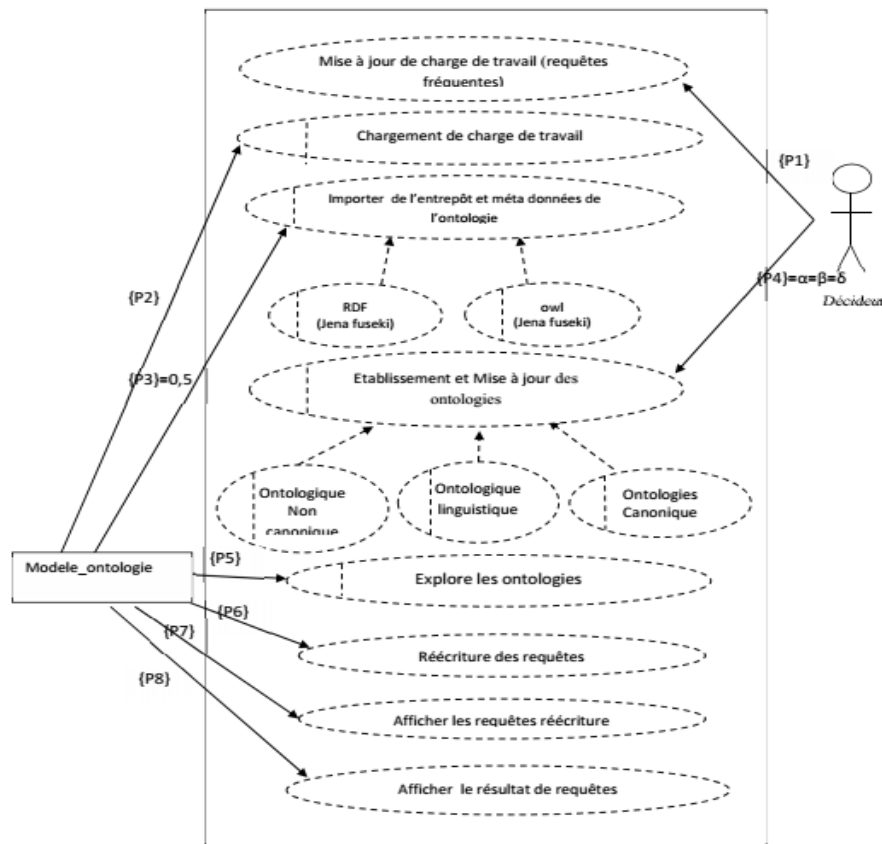
Figure Symbole d'un cas d'utilisation floue

## Diagramme de séquence flou

Dans le diagramme de séquence, les messages expliquent les méthodes. L'incertitude de la méthode a deux niveaux de flou. Le premier niveau est le degré d'appartenance de la méthode à un objet. Le deuxième niveau du flou est exprimé quand la méthode elle-même est floue.



Figure 5.11 Symbole d'un message flou en Fuzzy UML



$p_i$  ( $i = 1, \dots, m$ ) utilisation dans le système par utilisateur.  
 $p_j$  ( $j = 1, \dots, n$ ) les probabilités sont utilisés dans le uses case.

$$\sum_{i=1}^n P_{ij} = 1 \text{ et } \sum_{j=1}^m P_i = 1$$

$$\sum \alpha + \beta + \gamma = 1$$

Figure Cas d'utilisation floue

## Diagramme de séquence flou

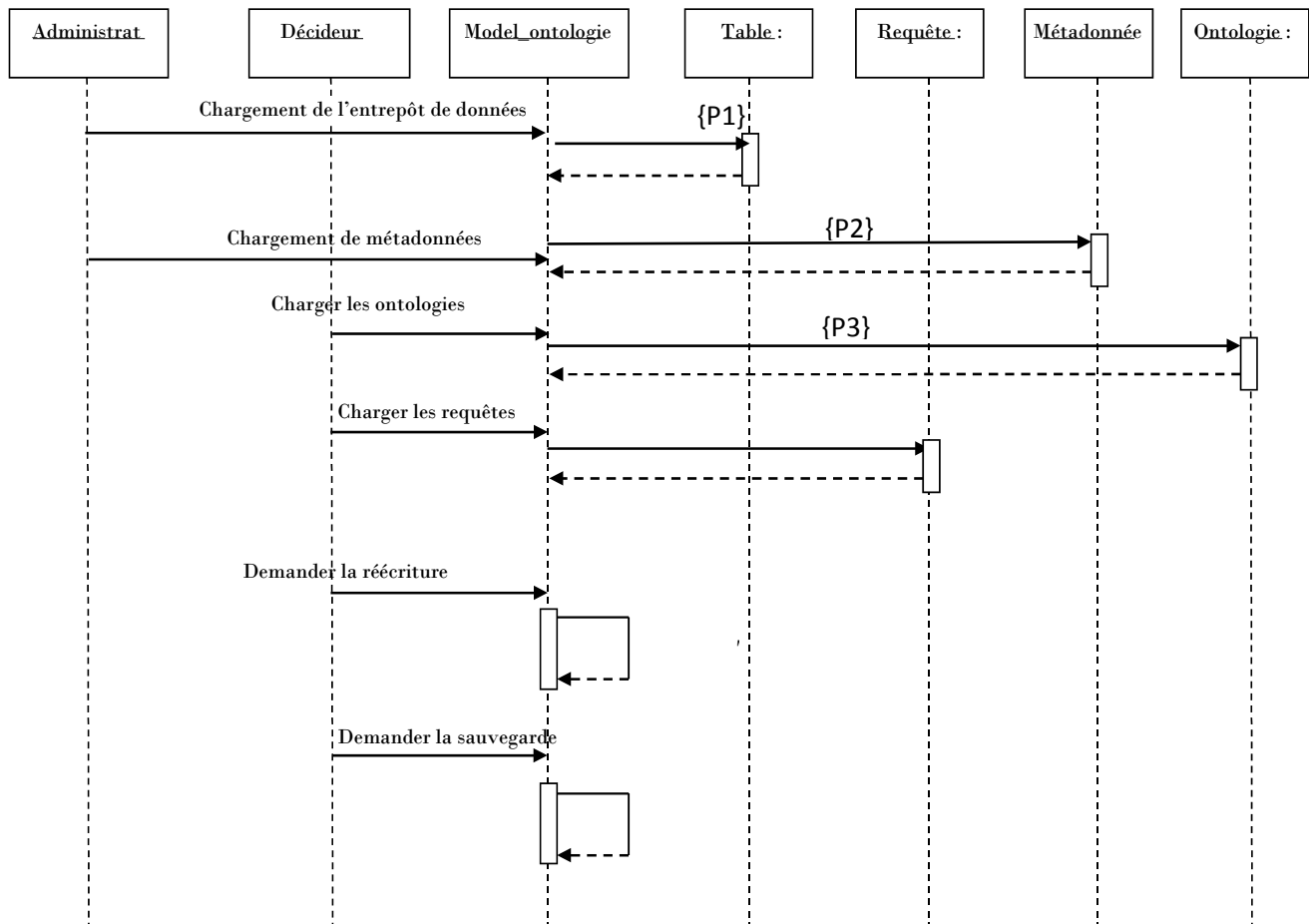


Figure Diagramme de séquence flou

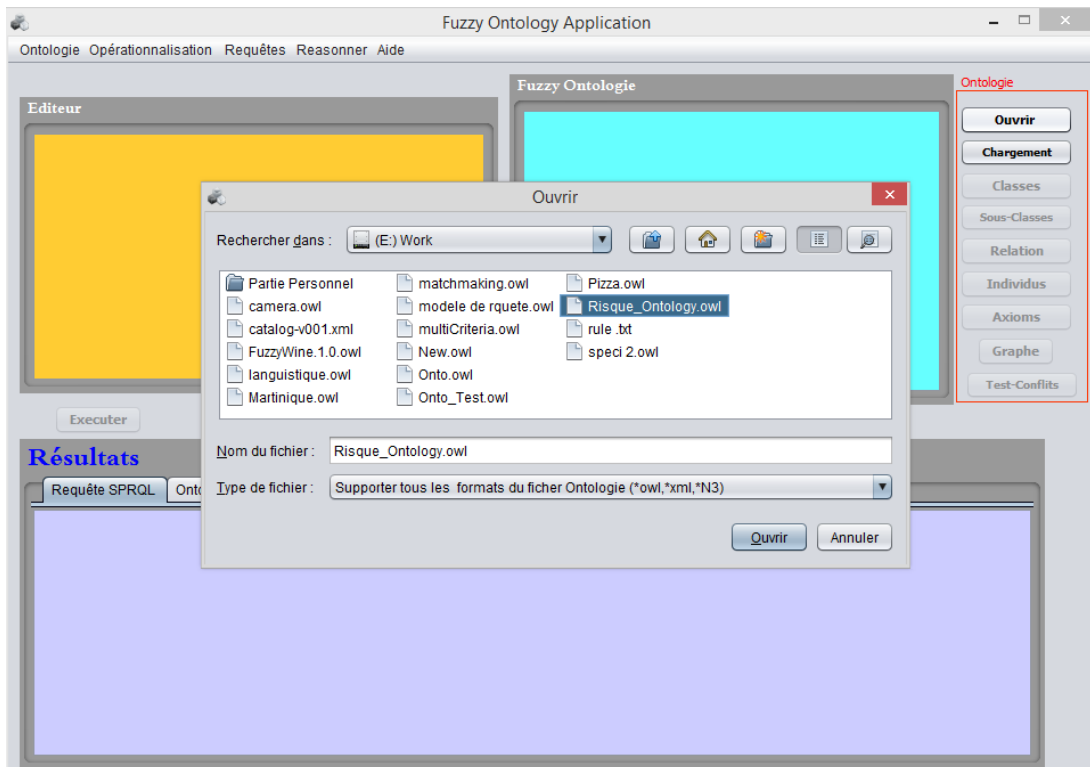
$p_i$  ( $i = 1, \dots, m$ ) utilisation dans le système par utilisateur.

$p_{ij}$  ( $j = 1, \dots, n$ ) les probabilités sont utilisés dans le uses case.

# Annexe D : Ontologie de Risque

## Sélectionnement du fichier OWL de « Onto-Risque »

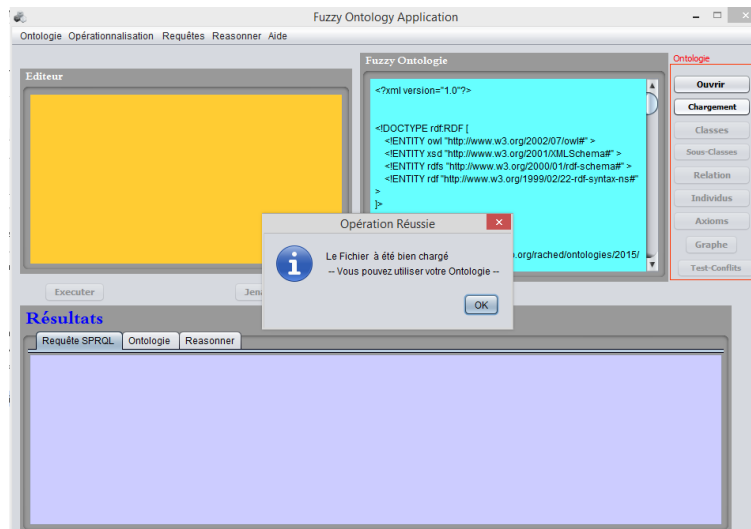
L'écran ci-dessous montre comment sélectionner un fichier OWL de notre ontologie floue « Onto-Risque sous notre application.



*Selection du fichier OWL« Onto-Risque ».*

## Chargement du fichier OWL de « Onto-Risque »

L'écran ci-dessous montre comment charger un fichier OWL de notre ontologie floue « Onto-Risque sous l'application

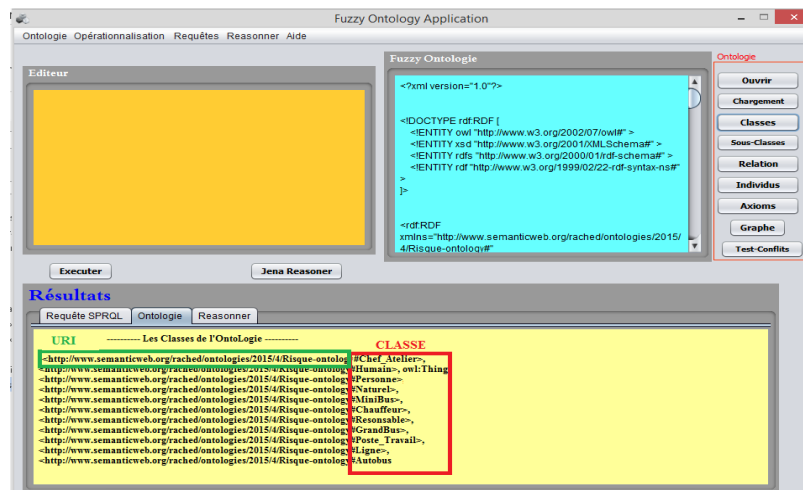


*Chargement du fichier OWL « Onto-Risque ».*

Donc l'opération de chargement du fichier owl consiste un test de vérification l'extension du fichier, Si le fichier est de type owl donc l'opération sera réalisée,

### **Extraction des classes de « Onto-Risque »**

L'écran ci-dessous représente une opération d'extraction des classes à partir de fichier OWL de notre ontologie floue « Onto-Risque ».



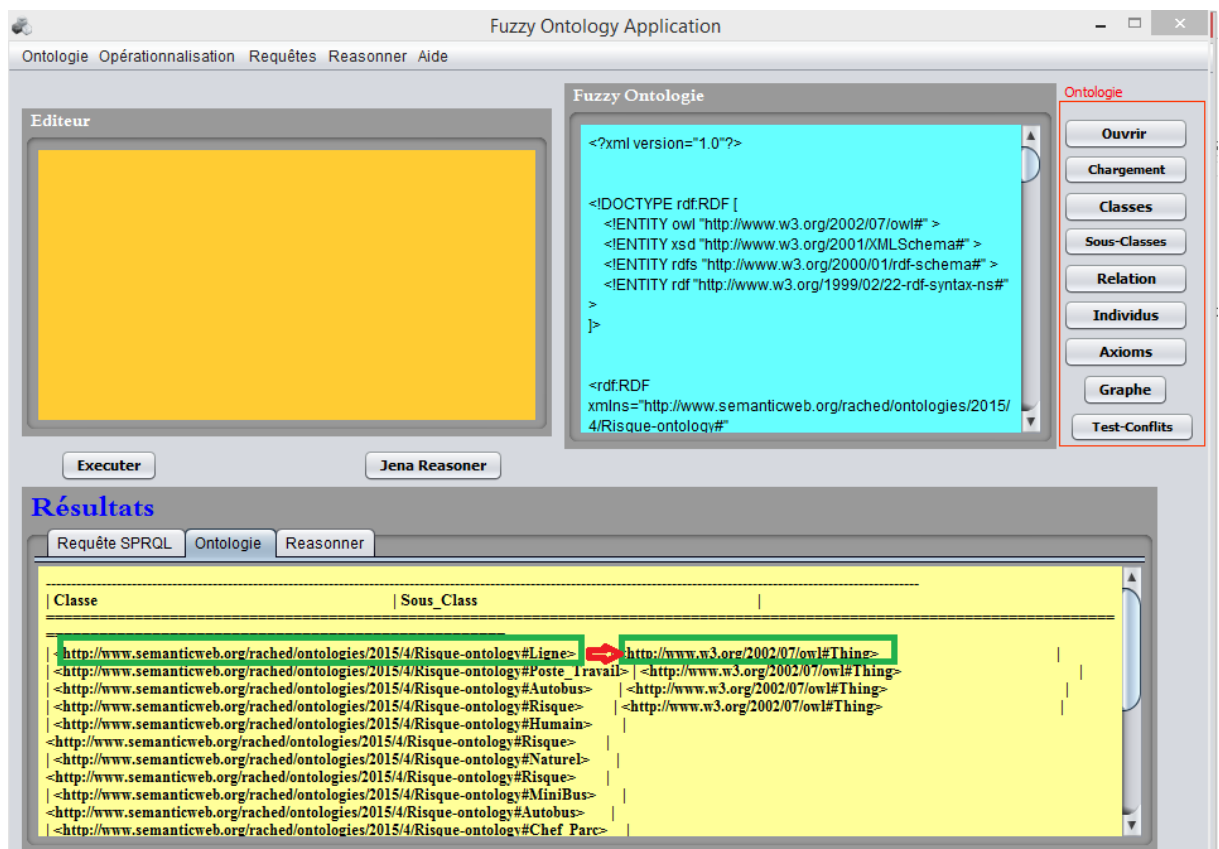
*Extraction des classes « Onto-Risque ».*

Après l'extraction les résultats sont affichent par la syntaxe suivante :

<URI # Name Of Classe>

### Extraction des Sous-classes de « Onto-Risque »

La figure ci-dessous représente une opération d'extraction des sous classes à partir de fichier OWL de notre ontologie floue « Onto-Risque ».



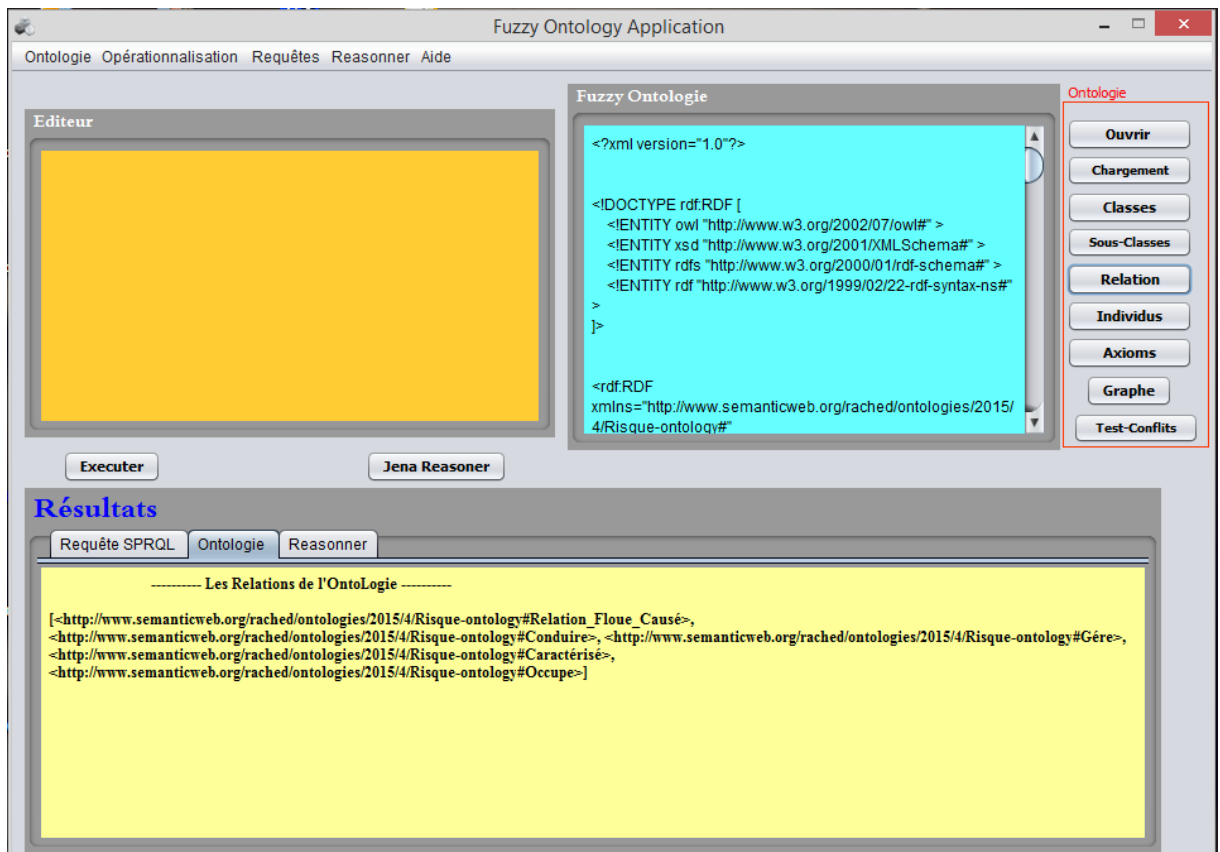
### Extraction des Sous -classes « Onto-Risque ».

Après l'extraction les résultats sont affichent par la syntaxe suivante :

<URI # Name Of SubClasse> <URI # Name Of Classe>

## Extraction des relations de « Onto-Risque »

L'écran ci-dessous représente une opération d'extraction des relations à partir de fichier OWL de notre ontologie floue « Onto-Risque ».



### *Extraction des relations « Onto-Risque ».*

Après l'extraction les résultats sont affichent par la syntaxe suivante :

<URI # Name Of relationship>

## Extraction des individus de « Onto-Risque »

L'écran ci-dessous représente une opération d'extraction des instances à partir de fichier OWL de notre ontologie floue « Onto-Risque ».

