



NO d'ordre :

UNIVERSITE DE M'SILA

Faculté des Mathématiques et de l'informatique
Département d'informatique

MEMOIRE

Présenté pour l'obtention du diplôme de Magistère
Spécialité : Informatique
Option : Systèmes d'Information et de Connaissance

Par :

Hamani Mohamed Said

SUJET

***Une approche à base d'ontologies
floues pour la fouille de données***

Soutenu publiquement le 10/06/2010 devant le jury composé de :

Président : Pr. Boufaida Mahmoud, Université de Constantine

Rapporteur : Dr. Maamri Ramdane, Université de Constantine

Examineur : Pr. Bouderah Brahim, Université de M'sila

Examineur : Pr. Boufaida Zaizette, Université de Constantine

Examineur : Pr. Benmohammed Mohamed, Université de Constantine

Promotion : 2009/2010

Remerciement

Mes remerciements vont tout d'abord à Mr Maamri Ramdane, Maître de conférence de l'université de Constantine pour avoir accepté de diriger ce travail, pour ses précieux conseils, ses orientations sa compréhension et sa patience.

Je remercie Mr Boufaida Mahmoud, professeur de l'université de Constantine, qui me fait l'honneur de présider ce jury.

Je remercie Mr Bouderah Brahim, professeur de l'université de M'sila, qui me fait l'honneur d'avoir accepté d'examiner ce travail.

Je remercie Mme Boufaida Zizette, professeur de l'université de Constantine, qui me fait l'honneur d'avoir accepté d'examiner ce travail.

Je remercie Mr Benmohammed Mohamed, professeur de l'université de Constantine, qui me fait l'honneur d'avoir accepté d'examiner ce travail.

Je remercie, autre fois, Mr Guemra Mohamed pour avoir relu ce mémoire et avoir enlevé les fautes d'orthographe que j'avais laissées.

Mes remerciements vont également à tous ceux qui ont contribué à la réalisation de ce travail.

TABLE DES MATIERES

Introduction Générale

1. Contexte de l'étude et problématique	1
2. Plan de mémoire	3

Partie 1 : Etat de l'art

Chapitre 1: Fouille de données

1. Introduction	7
2. Extraction de connaissances à partir de données (ECD)	8
2.1. Le Processus de l'ECD	9
2.1.1. Préparation des données	9
2.1.2. Fouille de données	10
2.1.2.1 Tâches de fouille de données	10
2.1.2.2 Classification des méthodes de fouille de données ...	12
2.1.3. Interprétation des résultats	12
2.3. Modèle CRISP-DM	13
3. Règles d'association	15
3.1. L'extraction des motifs fréquents	15
3.2. Génération des règles d'associations	16
3.3 L'algorithme d'extraction des règles 'Apriori'	17
3.4. Mesures d'intérêts des règles d'associations	18
3.4.1. Mesures d'intérêts objectives	19
3.4.2. Mesures d'intérêts subjectives	24
3.5. L'intérêt de classement des règles d'associations	24
4. Conclusion	26

Chapitre 2: Représentation de la connaissance et ontologie

1. Introduction	28
2. Représentation de la connaissance	29
2.1. Langages de représentation des connaissances	29
3. Ontologie	31

3.1. Composantes d'une ontologie	31
3.2. Le cycle de vie des ontologies	33
3.3. Construction d'ontologies	33
3.4. Typologie des ontologies	34
3.5. Langages de représentation des ontologies	35
3.6. Ontologie et représentation des connaissances	37
3.7. Ontologie floue	38
4. Similarités entre concepts	40
4.1. Mesures reposant sur la distance	41
4.2. Mesures reposant sur le contenu en information	43
4.3. Mesures Mixtes	43
5. Logique floue	44
5.1. Variables linguistiques floues	44
5.2. Les ensembles flous et la fonction d'appartenance	44
5.3. Opérations logiques sur les ensembles flous	46
6. Conclusion	48

Chapitre 3: L'utilisation des ontologies dans le processus de fouille de données

1. Introduction	50
2. Ontologies en fouille de données	51
2.1. Les ontologies dans le cycle de vie CRISP-DM	52
2.2. Travaux existants sur l'intégration des ontologies dans l'ECD	53
3. Travaux existants sur les mesures d'intérêt et évaluation	54
4. Conclusion	59

Partie 2 : Contributions

Chapitre 4: Utilisation des ontologies floues dans le classement des règles intéressantes

1. Introduction	62
2. Présentation de l'approche	63
2.1. Principe de l'approche	63
2.2. Description de l'approche	63
2.3. Notre approche dans le modèle CRISP-DM	64

3. Présentation du système	65
3.1. Processus de classement des règles d'association	65
3.2. Représentation des connaissances du système	68
3.2.1. Représentation des règles d'associations	68
3.2.2. Représentation d'une ontologie floue	69
4. Présentation de l'algorithme de classement des règles	71
4.1. Calcul de la distance sémantique entre concepts	71
4.2. L'ontologie floue et le calcul de distance	72
4.3. Pondération des relations taxonomique	72
4.4. Calcul de la distance d'une règle d'association	73
4.5. Présentation des résultats	73
4.6. Algorithme Pseudo code	73
5. Conclusion	75

Chapitre 5: Etude de cas

1. Introduction	77
2. Présentation du cas d'étude théorique	78
2.1. Application de l'algorithme de classement des règles	78
2.1.1. Construction de la matrice des relations pondérées	79
2.1.2. Construction de la matrice du plus court chemin	80
2.1.3. Calcul des distances des règles	81
2.1.4. Présentation des résultats	83
2.2. Interprétation des résultats	83
3. Présentation du cas d'étude réel	84
3.1. Extraction des règles d'associations	88
3.2 Application de l'algorithme de classement des règles	89
3.2.1 Cas d'ontologie ordinaire	89
3.2.2. Cas d'ontologie Floue	92
4. Conclusion	97

Conclusion générale	98
----------------------------------	----

Références	100
-------------------------	-----

Annexe	109
---------------------	-----