

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

INSTITUT NATIONAL DE FORMATION EN INFORMATIQUE
OUED-SMAR

MEMOIRE POUR L'OBTENTION DU DIPLOME
D'INGENIEUR D'ETAT EN INFORMATIQUE

Option

SYSTEMES INFORMATIQUES

Thème

EXTRACTION AUTOMATIQUE DE LA CONNAISSANCE
UNE APPROCHE CENTREE-OBJET

Organisme d'accueil :

Université de Pau
Département d'Informatique

Encadré par :

- Teste Alain, Université de Pau
- Bouzidi Zair, Université de Béjaia

Réalisé par: Smadhi Samia

Promotion 1995-1996

Résumé

Le travail que nous avons réalisé au sein du groupe de recherche ICOG (Département d'Informatique de l'Université de Pau) s'inscrit autour du domaine des systèmes de découverte de connaissances dans les bases de données.

Nous avons contribué à améliorer une méthode d'extraction automatique des connaissances. Ces connaissances extraites sont représentées dans le formalisme des règles de production dont la compréhension est plus facile à l'utilisateur.

Notre travail porte sur:

- l'intégration des connaissances du domaine sous la forme de hiérarchies de concepts fournies par l'expert du domaine. Ces connaissances supplémentaires ont été utilisées dans le processus de généralisation.
- la modélisation de ces hiérarchies de concepts et des données initiales à l'aide d'un sous ensemble de RECOS (une couche centrée-objet au dessus de Prolog).
- la mise en œuvre d'un algorithme d'extraction de règles caractéristiques.
- l'extension de l'algorithme de caractérisation logique d'une partition au contexte d'une base de données, et ce pour découvrir des règles décisionnelles.

Ces algorithmes ont été programmés en Prolog.

Ils feront partie du système EKDB développé par le groupe ICOG. Nous présentons, dans ce mémoire une version appelée EKDB-P, qui a été mise au point pour une utilisation autonome.

Sommaire

Chapitre1 : Les systèmes de découverte de connaissances dans les bases de données

1. Introduction	5
2. Aperçu de quelques systèmes de découvertes de connaissances (KDD)	6
2.1 - Un modèle idéal de système KDD	6
2.2 - Le système KID3	9
2.3 - Le système DBLearn.....	10
3. Quelques rappels sur les bases de données relationnelles	12
3.1. Concepts de base du modèle relationnel.....	13
3.2. Contraintes d'intégrité.....	15
3.3. Dépendances fonctionnelles et normalisation des relations	16
3.4 - Le langage de requête SQL	17
4. Génération de règles: quelques algorithmes basés sur les arbres de décision	19
4.1. Acquisition des connaissances et apprentissage.....	19
4.2. Les techniques d'apprentissage.....	22
4.3. L'apprentissage inductif	23
4.4. ID3	25
4.5. SIPINA.....	29
4.6. La caractérisation logique d'une partition	31

Chapitre 2 : Intégration de la connaissance du domaine par les hiérarchies de concepts : une approche centrée-objet

1. Introduction	38
2. Hiérarchies de concepts et principe de la généralisation	38
3. Le problème des ambiguïtés.....	44
4. Une généralisation basée sur une modélisation centrée-objet des hiérarchies des concepts.....	47
4.1. Caractéristiques du langage de schémas utilisé.....	47
4.2. Modélisation des hiérarchies de concepts pour un attribut généralisable	51
4.3. Modélisation des données de la table à généraliser	52

Chapitre 3: Extraction automatique de règles caractéristiques et de règles décisionnelles

1. Introduction	55
2. Le système ERSQL	57
2.1. Le processus d'extraction de règles à l'aide d'objets symboliques	57
2.2. Le système ERSQL	59
3. Position de notre travail par rapport à la démarche ERSQL	60
4. L'extraction de règles caractéristiques	62
4.1. Le processus d'extraction de règles caractéristiques	62
4.2. Description de l'algorithme proposé	63
5. L'extraction de règles décisionnelles: une approche fondée sur la caractérisation logique d'une partition	64
5.1. Définition d'une règle décisionnelle	64
5.2. Extension de l'algorithme de caractérisation pour l'exploration d'une base de données relationnelle	64

Chapitre 4: Implémentation des algorithmes: le logiciel EKDB-P

1. Introduction	69
2. Architecture du logiciel	70
3. Structuration des informations	71
3.1. Les informations initiales	71
3.2. La représentation des données lors de leur traitement	73
3.3. Représentation des résultats obtenus	75
4. Dialogues avec l'utilisateur	75
4.1. Chargement d'une table de données	75
4.2 Généralisation	75
5. Un exemple de fonctionnement	76