**Master 2 MIAGE CFA**

2010-2011

**Base de données multi-dimensionnelles et OLAP**

**Projet n°2 : Règles d’associations**

**Florian GOUIN et Arnaud BRETON**

Sommaire

[Introduction 3](#_Toc283225853)

[Présentation du projet 4](#_Toc283225854)

[Conception générale 4](#_Toc283225855)

[Choix technologiques 4](#_Toc283225856)

[Répartition du travail 5](#_Toc283225857)

# Introduction

L’objet de projet de dernière année de master 2 MIAGE est d’appréhender les règles d’associations de manière pratique via une implémentation informatique.

Le but de cette implémentation est de permettre à un utilisateur de connaître l’ensemble des règles d’associations intéressantes, c'est-à-dire celle dépassant une certaine confiance, afin d’établir des liens entre des éléments à priori sans aucune connexion.

Cette méthode est utilisé pour analyser les comportements des consommateurs et d’ainsi dégager des tendances (par exemple, que les consommateurs achetant du lait, achètent aussi de la bière).

Dans notre implémentation, nous partons d’un tableau disjonctif complet, à savoir une table binaire (ne contenant donc que des 0 et des 1) pour calculer, dans un premier temps les ensembles dits fréquents (ceux qui dépassent un seuil donné) puis, à partir de ces ensembles, nous calculons les règles d’associations intéressantes (celles qui dépassent un seuil donné).

# Présentation du projet

Comme nous l’avons vu ci-dessus, l’application permet à l’utilisateur d’obtenir les règles d’associations à partir d’une table binaire (un tableau disjonctif complet).

Pour cela, l’utilisateur doit saisir un seuil de confiance et de support, le nom de la table à exploiter (parmi celles présentes dans la base de données).

Le lancement des calculs recherche alors via l’algorithme A-Priori l’ensemble des attributs fréquents de la table.

Ce premier retour intermédiaire est affiché par l’IHM qui distingue par un jeu de couleur les ensembles fréquents des autres.

A partir de cet ensemble, le programme génère l’ensemble des règles possibles (au nombre de ) puis calcule leur confiance.

Le retour est donné à l’IHM qui affiche l’ensemble des règles générés et distingue celles intéressantes des autres, via un jeu de couleur.

Le programme permet aussi à l’utilisateur de générer des tables aléatoires à n colonnes et m lignes (n et m étant définis par l’utilisateur) peuplés de données aléatoires. Ces tables permettent à l’utilisateur de tester les performances de l’algorithme en faisant varier la taille de la table.

## Conception générale

L’application est divisée en quatre couches :

1. L’Interface graphique, qui permet à l’utilisateur d’entrer les seuils de support et de confiance et de calculer les règles d’associations d’une table donnée et qui réalise les contrôles de surface.
2. Le modèle, qui calcule les règles d’associations
3. La couche d’accès aux données qui interroge la base de données
4. La base de données qui contient les tables sur lesquelles on veut extraire les règles d’associations

Nous avons donc mis en place, côté applicatif (IHM, modèle, CAD), un découpage du type Modèle-Vue-Contrôleur (MVC).

## Choix technologiques

Le projet a entièrement été réalisé en Java (implémentation de l’algorithme A-Priori, l’interface graphique et JDBC (*Java DataBase Connectivity,* qui est une interface de programmation standard pour l’accès aux bases de données) pour l’accès à la base de données *MySQL*).

L’interface graphique a été réalisée grâce aux librairies graphiques SWT (*Standard Widget Toolkit,* <http://www.eclipse.org/swt/>)

## Répartition du travail

L’ensemble de l’interface graphique a été réalisée par Florian GOUIN.

Le moteur de calcul des règles d’associations a été entièrement réalisé par Arnaud BRETON.

Bien sûr, chacun d’entre nous est intervenu de manière ponctuelle sur la partie de l’autre afin de donner son point de vue ou d’apporter son aide sur un sujet délicat (point de conception, écriture d’algorithmes performants, etc.).

# Conclusion

Ce projet nous a permis de mettre en pratique un algorithme vu en cours, en l’implémentant complètement sur machine.

Il nous a permis de comprendre l’intérêt économique qui est induit par les règles d’associations qui vont permettre aux analystes de déduire des règles qui ne paraissent par évidente lors de l’observation des tables d’une base, surtout quand celle-ci comporte un grand nombre d’entrées.