

# Hierarchisation des règles d'association en fouille de textes

Rokia BENDAOUD\*, Yannick TOUSSAINT\*  
Amedeo NAPOLI\*

\*LORIA Campus Scientifique - BP 239  
54506 VANDOEUVRE-lès-NANCY CEDEX  
{bendaoud,toussaint,napoli}@loria.fr,

**Résumé.** L'extraction de règles d'association est souvent exploitée comme méthode de fouille de données. Cependant, une des limites de cette approche vient du très grand nombre de règles extraites et de la difficulté pour l'analyste à appréhender la totalité de ces règles. Nous proposons donc de pallier ce problème en structurant l'ensemble des règles d'association en hiérarchies. La structuration des règles se fait à deux niveaux. Un niveau global qui a pour objectif de construire une hiérarchie structurant les règles extraites des données. Nous définissons donc un premier type de subsomption entre règles issue de la subsomption dans les treillis de Galois. Le second niveau correspond à une analyse locale des règles et génère pour une règle donnée une hiérarchie de généralisation de cette règle qui repose sur des connaissances complémentaires exprimées dans un modèle terminologique. Ce niveau fait appel à un second type de subsomption inspiré de la subsomption en programmation logique inductive. Nous définissons ces deux types de subsomptions, développons un exemple montrant l'intérêt de l'approche pour l'analyste et étudions les propriétés formelles des hiérarchies ainsi proposées.

## 1 Introduction

L'extraction des règles d'association appliquée à des textes est une méthode de fouille de données qui permet de mettre en valeur des liens entre les termes des textes. Ces liens peuvent alors être interprétés par des experts en vue, par exemple, de la construction d'une ontologie.

Que ce soit à partir de textes où à partir de base de données, le nombre de règles extraites est souvent très grand et difficile à appréhender par un expert humain. De nombreux travaux se sont intéressés à élaguer l'ensemble des règles et à les classer soit par rapport à des critères statistiques, soit par rapport à une base de connaissances (Janetzko et al. 2004). Nous proposons dans cet article une approche visant à structurer les règles sous forme hiérarchique afin de permettre à l'expert une approche descendante de la lecture de l'ensemble des règles. En réalité, nous proposons à l'expert deux approches d'analyse, un niveau global et un niveau local, tous deux reposant sur une structuration hiérarchique des règles. Ces deux types de structuration hiérarchique nous ont conduit à définir deux types de subsomption qui, au final, peuvent être combinés.