

# JEN : un algorithme efficace de construction de générateurs pour l'identification des règles d'association

Amélie Le Floch<sup>\*</sup>, Christian Fiset<sup>\*</sup>, Rokia Missaoui<sup>\*\*</sup>  
Petko Valtchev<sup>\*\*\*</sup>, Robert Godin<sup>\*</sup>

<sup>\*</sup> Département d'informatique, Université du Québec à Montréal

C.P. 8888, succursale Centre-Ville, Montréal, Québec, Canada, H3C 3P8

<sup>\*\*</sup> Département d'informatique et d'ingénierie, Université du Québec en Outaouais  
rokia.missaoui@uqo.ca

<sup>\*\*\*</sup> Département d'informatique et recherche opérationnelle, Université de Montréal

**Résumé.** L'article décrit un algorithme, appelé *JEN*, d'identification efficace des générateurs à partir du treillis de concepts (Galois) pour l'extraction des règles d'association. Cette dernière est immédiate dans notre approche : les règles exactes sont obtenues à partir des concepts individuels en exploitant leurs générateurs et l'*itemset* fermé correspondant, tandis que les règles approximatives sont identifiées en consultant les générateurs d'un concept et l'*itemset* fermé des prédécesseurs immédiats de ce concept. Une analyse comparative empirique illustre la supériorité de *JEN* sur trois autres procédures de génération de règles et de générateurs, particulièrement lors de l'analyse de données fortement corrélées.

## 1 Introduction

La fouille de données (*data mining*) est une discipline récente, issue de la confluence des statistiques, de l'intelligence artificielle et des bases de données. Elle vise la découverte de régularités (ex. concepts, liens, règles) dans de grands volumes de données. La dernière décennie a enrichi la discipline d'un large spectre d'approches visant à améliorer la performance des méthodes de fouille et à accroître leur capacité de produire un ensemble réduit mais pertinent de concepts et règles.

Cet article traite du problème de l'extraction des connaissances sous forme de règles d'association (Agrawal 1994) en mettant davantage l'accent sur les travaux opérant dans le contexte de l'analyse formelle des concepts (Ganter 1999). Plusieurs études (Godin 1994), (Zaki 2002), (Pasquier 2000) ont notamment souligné le nombre prohibitif de règles d'association extraites et la nécessité de définir des bases pour les règles d'association. Ces bases constituent des ensembles réduits de règles informatives (prémisse minimale, conséquence maximale) permettant de ne conserver que les règles les plus pertinentes, sans perte d'information. La génération de ces règles d'association informatives peut se faire par une extraction efficace des concepts et de leurs générateurs associés.

Des algorithmes produisant l'ensemble des concepts ainsi que leurs générateurs associés ont été développés avec succès (Godin 1994), (Pasquier 2000), (Pfaltz 2002), (Bastide et al. 2002). Cependant, peu de travaux font état de l'efficacité et des performances de l'extraction des générateurs dans le processus de production des règles.