

# Les itemsets essentiels fermés : une nouvelle représentation concise

Tarek Hamrouni\*, Islem Denden\*  
Sadok Ben Yahia\*, Engelbert Mephu Nguifo\*\*, Yahya Slimani\*

\* Département des Sciences de l'Informatique  
Faculté des Sciences de Tunis  
Campus Universitaire 1060 Tunis, Tunisie  
{tarek.hamrouni, sadok.benyahia, yahya.slimani}@fst.rnu.tn  
\*\* CRIL CNRS FRE 2499  
Université d'Artois, IUT de Lens  
Rue Jean Souvraz, SP-18  
F-62307 Lens Cedex France  
mephu@cril.univ-artois.fr

**Résumé.** Devant l'accroissement constant des grandes bases de données, plusieurs travaux de recherche en fouille de données s'orientent vers le développement de techniques de représentation compacte. Ces recherches se développent suivant deux axes complémentaires : l'extraction de bases génériques de règles d'association et l'extraction de représentations concises d'itemsets fréquents.

Dans ce papier, nous introduisons une nouvelle représentation concise exacte des itemsets fréquents. Elle se situe au croisement de chemins de deux autres représentations concises, à savoir les itemsets fermés et ceux dits essentiels. L'idée intuitive est de profiter du fait que tout opérateur de fermeture induit une fonction surjective. Dans ce contexte, nous introduisons un nouvel opérateur de fermeture permettant de calculer les fermetures des itemsets essentiels. Ceci a pour but d'avoir une représentation concise de taille réduite tout en permettant l'extraction des supports négatif et disjonctif d'un itemset en plus de son support conjonctif. Un nouvel algorithme appelé D-CLOSURE permettant d'extraire les itemsets essentiels fermés est aussi présenté. L'étude expérimentale que nous avons menée a permis de confirmer que la nouvelle approche présente un bon taux de compacité comparativement aux autres représentations concises exactes.

## 1 Introduction

L'apparition de la "fouille de connaissances" a été un tournant dans les intérêts prioritaires de la communauté de la fouille de données. En effet, les efforts ne sont plus seulement déployés dans la réduction des temps d'extraction des motifs fréquents mais de plus en plus de travaux s'intéressent à l'extraction d'une connaissance de meilleure qualité tout en préservant la vertu de la compacité. Dans ce registre, nous relevons les travaux visant l'extraction des représentations concises. Ainsi, parmi les représentations exactes les plus connues, nous citons