

# Fouille de textes pour orienter la construction d'une ressource terminologique

Valentina CEAUSU, Sylvie DESPRES

Université René Descartes  
CRIP5 – Equipe IAA – Groupe SBC  
UFR Mathématiques et Informatique  
45 rue des Saints-Pères  
75006 PARIS  
valentina.ceausu@math-info.univ-paris5.fr  
sd@math-info.univ-paris5.fr

**Résumé.** La finalité de ce papier est d'analyser l'apport de techniques de fouille de données textuelles à une méthodologie de construction d'ontologie à partir de textes. Le domaine d'application de cette expérimentation est celui de l'accidentologie routière. Dans ce contexte, les résultats des techniques de fouille de données textuelles sont utilisés pour orienter la construction d'une ressource terminologique à partir de procès-verbaux d'accidents. La méthode TERMINAE et l'outil du même nom offrent le cadre général pour la modélisation de la ressource. Le papier présente les techniques de fouille employées et l'intégration des résultats des fouilles dans les différentes étapes du processus de construction de la ressource.

## 1 Introduction

La finalité de ce papier est d'analyser l'apport des techniques de fouille de données textuelles à une méthodologie de construction d'ontologie à partir de textes. Le domaine d'application de cette expérimentation est celui de l'accidentologie routière. Une ontologie du domaine a été élaborée à partir de connaissances expertes (Després, 2002). Le travail présenté concerne la construction d'une ressource terminologique à partir de procès verbaux d'accidents (PV) rédigés par les forces de l'ordre. Les deux ressources (ontologique et terminologique) seront exploitées dans un système de raisonnement à partir de cas ayant comme cas cible des procès verbaux et comme cas source des scénarios d'accidents.

Dans ce contexte sont utilisés simultanément : (a) un algorithme de reconnaissance de motifs qui engendre un ensemble de syntagmes nominaux et verbaux ; (b) l'algorithme Apriori pour affiner les syntagmes nominaux identifiés à l'aide de motifs ; (c) l'ontologie de l'accidentologie pour affiner les syntagmes verbaux et (d) la méthodologie TERMINAE de construction de la ressource terminologique (Biébow, Szulman, 2000). Après avoir présenté les techniques de fouille de textes utilisées, leur apport à l'élaboration de la ressource terminologique est discuté. En conclusion, les améliorations à apporter aux différentes techniques sont discutées.