

# Extraction d'opinions appliquée à des critères

Benjamin Duthil\*, François Troussel\*  
G rard Dray\*, Pascal Poncelet\*\*, Jacky Montmain\*

\*EMA-LGI2P, Parc Scientifique Georges Besse, 30035 N mes Cedex, France  
pr nom.nom@mines-ales.fr

\*\*LIRMM UM2 - CNRS 5506, 161 Rue Ada, 34095 Montpellier Cedex 5, France  
Pascal.Poncelet@lirmm.fr

**R sum .** Les technologies de l'information et le succ s des services associ s (e.g., blogs, forums,...) ont ouvert la voie   un mode d'expression massive d'opinions sur les sujets les plus vari s. R cemment, de nouvelles techniques de d tection automatique d'opinions (*opinion mining*) ont fait leur apparition et via des analyses statistiques des avis exprim s, tendent   d gager une tendance globale des opinions exprim es par les internautes. N anmoins une analyse plus fine de celle-ci montre que les arguments avanc s par les internautes rel vent de crit res de jugement distincts. Ici, un film sera d cri  pour un sc nario d cousu, l  il sera encens  pour une bande son  poustouflante. Dans cet article, nous proposons, apr s avoir caract ris  automatiquement des crit res dans un document, d'en extraire l'opinion relative. A partir d'un ensemble restreint de mots cl s d'opinions, notre approche construit automatiquement une base d'apprentissage de documents issus du web et en d duit un lexique de mots ou d'expressions d'opinions sp cifiques au domaine d'application. Des exp riences men es sur des jeux de donn es r elles illustrent l'efficacit  de l'approche.

## 1 Introduction

Avec le d veloppement du Web, de plus en plus de documents textuels sont disponibles et de plus en plus d'outils permettent de rechercher de l'information pertinente. Conna tre l'opinion des personnes sur un produit, rechercher et classer des documents, indexer de mani re automatique des documents sont des probl matiques d'actualit  (e.g. Xu et al. (2011); Bai (2011); Morinaga et al. (2002)). Par exemple, dans le cas de l'opinion de cin philes, de nombreux outils sont disponibles pour conna tre l'avis g n ral des spectateurs d'un film. Traditionnellement, pour extraire ces opinions, deux grandes approches existent : (i) celles bas es sur un corpus d'apprentissage qui s'appuient g n ralement sur une analyse syntaxique et de co-occurrence des mots; (ii) et celles utilisant un "dictionnaire" sp cifique (e.g. *SenticNet* Cambria et al. (2010)) pour obtenir des orientations s mantiques d'un mot (Xu et al. (2011); Kamps et al. (2004)). M me si ces approches sont tr s efficaces, elles souffrent pour les premi res de la n cessit  de devoir constituer un corpus d'apprentissage. Dans des travaux pr c dents (Harb et al. (2008)), nous avons mis en avant le fait que l'expression d'opinions