

# Text2Geo : des données textuelles aux informations géospatiales

Sabiha Tahrat \*, Eric Kergosien \*,\*\*,\*\*\*, Sandra Bringay \*,  
Mathieu Roche \*, Maguelonne Teisseire \*\*

\* LIRMM, UMR 5506, 161 rue Ada, 34392 Montpellier - France  
{prénom.nom}@lirmm.fr

\*\* UMR TETIS, 500 rue Jean-François Breton, 34093 Montpellier - France  
{prénom.nom}@teledetection.fr

\*\*\* LIUPPA, UPPA, Avenue de l'Université, BP 576, 64012 PAU cedex - France  
{prénom.nom}@univ-pau.fr

**Résumé.** Dans cet article, nous nous intéressons aux méthodes d'extraction d'informations spatiales dans des documents textuels. Nous présentons la méthode hybride Text2Geo qui combine une approche d'extraction d'informations, fondée sur des patrons avec une approche de classification supervisée permettant d'explorer le contexte associé. Nous discutons des résultats expérimentaux obtenus sur le jeu de données de l'étang de Thau.

## 1 Introduction

Au-delà de sa stricte définition d'entité administrative et politique, le territoire, selon Guy Di Méo, témoigne d'une *"appropriation à la fois économique, idéologique et politique de l'espace par des groupes qui se donnent une représentation particulière d'eux-mêmes, de leur histoire, de leur singularité"* (Di Méo (1998)). Dans ce contexte éminemment subjectif, la caractérisation et la compréhension des perceptions d'un même territoire par les différents acteurs est difficile, mais néanmoins particulièrement intéressante dans une perspective d'aménagement du territoire et de politique publique territoriale. Le travail présenté s'inscrit dans le cadre du projet Senterritoire<sup>1</sup>, qui adopte une démarche pluridisciplinaire, initiée à partir d'une méthode automatique et visant à fournir aux géographes et aux environnementalistes, une aide à la découverte de connaissances. Nos contributions portent, dans cette publication, sur l'accès à l'information spatiale et proposent (1) d'affiner et d'enrichir les patrons d'extraction d'informations existants dans la littérature afin d'améliorer l'identification du sens de l'entité spatiale extraite et (2) de définir une approche originale utilisant différentes techniques de fouille de textes afin de distinguer une entité spatiale d'une entité d'organisation.

La suite de l'article est organisée de la façon suivante. En section 2, nous présentons les définitions préliminaires et les travaux du domaine. En section 3, nous décrivons la méthode hybride Text2Geo. En section 4, nous présentons les expérimentations réalisées sur le jeu de données du bassin de Thau et concluons dans la section 5.

1. <http://www.msh-m.fr/programmes/senterritoire/>