

# Vers une fouille sémantique des brevets : Application au domaine biomédical

Nizar Ghoula\*, Khaled Khelif\*, Rose Dieng-Kuntz\*

\*INRIA Sophia Antipolis 2004 route des Lucioles 06902, BP93  
Sophia Antipolis - France  
{Nizar.ghoula, Khaled.Khelif, Rose.Dieng}@sophia.inria.fr

**Résumé.** Les brevets sont une source d'information très riche puisque ce sont des documents qui servent à décrire les inventions. L'accès aux documents de brevets en ligne est possible grâce aux efforts des offices nationaux de la propriété intellectuelle. Par ailleurs, ayant des objectifs différents, la présentation de ces documents a pris des formes variées loin d'être unifiées. Ce papier présente une méthode et un système permettant l'analyse de brevets "Patent Mining" pour générer des annotations sémantiques. L'idée principale est de pouvoir prendre en considération la structure des brevets pour pouvoir trouver un lien entre le contenu du brevet et les concepts des différentes ontologies.

## 1 Introduction

### 1.1 Contexte

Le traitement des documents de propriété intellectuelle, tels que les brevets, est important pour l'industrie, les affaires et les communautés juridiques. Récemment, les communautés de recherche académiques et en particulier, les chercheurs de traitement automatique de la langue naturelle et de la recherche documentaire ont reconnu l'importance du traitement des brevets. En fouillant les brevets scientifiques, nous pouvons remarquer un volume important d'informations sur la biologie, les substances et les procédures médicales. En effet, l'extraction des informations de ces brevets permet de donner une idée précise sur : (i) par exemple les interactions biomédicales et l'effet pharmacologique résultant, et (ii) la propriété intellectuelle dans un certain contexte biologique.

Durant ces dernières années, de grands efforts ont été exercés pour mettre les données relatives aux brevets sous une forme électronique et les présenter au public via les services en ligne. De nos jours, nous remarquons que ces services présentent et fournissent des structures de données hétérogènes, ce qui rend difficile à mettre en œuvre une analyse automatique des brevets.

Dans ce papier, nous présentons l'approche **PatAnnot** fondée sur les principes du web sémantique et qui se réfère aux notions de métadonnée et ontologies pour faciliter l'extraction des connaissances et la recherche d'informations relatives aux brevets.

Ce travail rentre dans le cadre du projet européen **Sealife** (Schroeder et al, 2006) qui a pour objectif la réalisation d'un navigateur Web sémantique pour le domaine des sciences de la vie, qui exploitera les ressources du Web en les rendant partageables, accessibles et manipulables par plusieurs utilisateurs dans différents domaines biomédicaux et ce afin de favoriser le partage des connaissances.