

# Reconnaissance de sections et d'entités dans les décisions de justice : application des modèles probabilistes HMM et CRF

Gildas Tagny Ngompé<sup>\*,\*\*</sup>, Sébastien Harispe<sup>\*</sup>, Guillaume Zambrano<sup>\*\*</sup>, Jacky Montmain<sup>\*</sup>,  
Stéphane Mussard<sup>\*\*</sup>

<sup>\*</sup>Laboratoire LGI2P, École des mines d'Alès  
{gildas.tagny-ngompe, sebastien.harispe}@mines-ales.fr  
<sup>\*\*</sup>Equipe CHROME, Université de Nîmes

**Résumé.** Une décision de justice est un document textuel rapportant le déroulement d'une affaire judiciaire. Les juristes s'en servent régulièrement comme source d'interprétation de la loi et de compréhension de l'opinion des juges. La masse disponible de décisions exige des solutions automatiques pour aider les acteurs du droit. Nous proposons d'adresser certains des défis liés à la recherche et l'analyse du volume croissant de décisions de justice en France dans un projet plus global. La première phase de ce projet porte sur l'extraction d'information des décisions dans l'objectif de construire une base de connaissances jurisprudentielles structurant et organisant les décisions. Une telle base facilite l'analyse descriptive et prédictive de corpus de décisions. Cet article présente une application des modèles probabilistes pour la segmentation des décisions et la reconnaissance d'entités dans leur contenu (lieu, date, participants, règles de loi, ...). Nos tests montrent l'avantage d'approches basées sur les champs aléatoires conditionnels (CRF) par rapport à des modèles plus simples et rapides basés sur les modèles cachés de Markov (HMM). Nous présentons ici les aspects techniques de la sélection et l'annotation du corpus d'apprentissage, et la définition de descripteurs discriminants. La spécificité des textes est importante et doit être prise en compte lors de l'application de méthodes d'extraction d'information dans un domaine spécifique.

## 1 Introduction

Une décision de justice est soit le résultat rendu par des juges à l'issue d'un procès, soit un document contenant la description de l'affaire, le résultat des juges et les motifs qui ont conduit à ce résultat. Cet article présente une approche de reconnaissance de sections (entête, exposé de l'affaire, et dispositif) et d'entités (date, ville, nom des juges, ...) dans ces documents. Plus précisément, nous évaluons l'application de deux approches de reconnaissance d'information à base de deux modèles markoviens HMM (*Hidden Markov Model*) et CRF (*Conditional Random Fields*). Les décisions jurisprudentielles sont essentielles pour les juristes parce qu'elles sont des sources d'interprétation de la loi. Les juristes doivent rassembler et analyser des décisions pertinentes pour résoudre les problèmes auxquels ils s'intéressent afin de mieux