A semantic analysis of a comprehensive corpus of court decisions

Gildas Tagny Ngompe[1,2], Sébastien Harispe[1], Jacky Montmain[1], Stéphane Mussard[2], Guillaume Zambrano[2]

[1] LGI2P, École des mines d'Alès, 69 Rue Georges Besse, Nîmes, France
[2] CHROME, Université de Nîmes, Rue du Dr Georges Salan, Nîmes, France

Abstract. We present our recent progress on the analysis of the French judicial corpus; we discuss in particular (i) results of our detailed study of HMM and CRF probabilistic models for detecting sections and entities in court decisions, and (ii) primilary work on the extraction of claim information through key-phase extraction and decision classification. We also discuss faced open challenges like dealing with multiple claims of similar or different types in the same decision, as well as lines of thought for solving these challenges based on discourse analysis.

Keywords: Natural Language Processing, probabilistic models, text classification, court decisions sectioning, entities and claims extraction

1 Introduction

A court decision is a document containing the description of a case, i.e. the decision of the judges as well as their motivations.

2 Extracting data about requests of parties

2.1 Localizing data using predefined and learned keyphrases

2.2 Indentify the polarity of the result by classifying documents

3 Categorizing the situations in which a type of request is formulated

4 Conclusion

References