

# Méthodes d'analyse sémantique de corpus de décisions jurisprudentielles

Soutenance de thèse de doctorat en informatique de l'IMT Mines Alès

---

Gildas TAGNY NGOMPÉ

24 janvier 2020

Jury:

- Stéphane MUSSARD, Professeur, Université de Nîmes (Directeur de thèse)
- Jacky MONTMAIN, Professeur, IMT Mines Alès (Co-directeur de thèse)
- Sandra BRINGAY, Professeur, Université Paul Valéry Montpellier (Rapporteur)
- Mohand BOUGHANEM, Professeur, Université Toulouse III Paul Sabatier (Rapporteur)
- Françoise SEYTE, Maître de Conférences (HDR), Université de Montpellier (Examineur)
- Fabrice MUHLENBACH, Maître de Conférences, Université Jean Monnet de Saint-Étienne (Examineur)
- Guillaume ZAMBRANO, Maître de Conférences, Université de Nîmes (Encadrant de proximité)
- Sébastien HARISPE, Maître Assistant, IMT Mines Alès (Encadrant de proximité)



1. Introduction
2. Annotation des sections et entités judiciaires
3. Identification des demandes
4. Identification du sens du résultat
5. Découverte des circonstances factuelles
6. Conclusions

# Introduction

---

# Expression du besoin

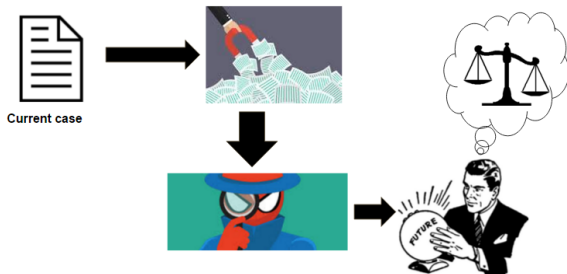


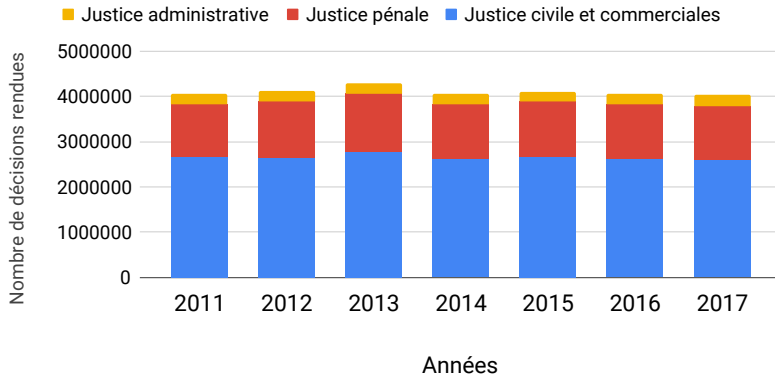
FIGURE — Les juristes analysent les décisions

## Pourquoi?

- comprendre et comparer l'application de la loi (contentieux, ville, ...)
- estimer le risque judiciaire
- ...

# Motivation : gros volume de décisions

## Plus de 4 millions de décisions prononcées / an



Source : <http://www.justice.gouv.fr/statistiques-10054/chiffres-cles-de-la-justice-10303/>

FIGURE — Nombre de décisions prononcées en France par an de 2011 à 2017.

# Motivation : recherches et analyses sémantiques difficiles

Moteurs de recherche juridique à mots-clés

Aucune analyse synthétique des décisions

The image shows a screenshot of the Légifrance search interface. A dropdown menu is open for the 'Nom de la juridiction' field, displaying options: '-- Toutes les juridictions --', 'Cour de cassation', 'Juridictions d'appel', 'Juridictions du premier degré', and 'Tribunal des conflits'. The 'Cour de cassation' option is currently selected. To the right of the dropdown, the 'Numéro d'affaire' field contains '06-81968'. Below the dropdown, the 'Date de décision' field is visible with a date picker set to 'Ex: 2019'. On the right side of the form, there are two checkboxes: 'Arrêts publiés au bulletin (Cour de cassation)' and 'Arrêts non publiés au bulletin (Cour de cassation)'. Below these are two text input fields labeled 'Mots recherchés' and 'Autres mots recherchés'. At the bottom left, there is a checkbox for 'Période de (1) à (2)' followed by date selection fields for 'Jour', 'Mois', and 'Année' (with 'Ex: 2019' as an example).

FIGURE — Formulaire de Légifrance.

# État de l'art : analyse automatique de décisions judiciaires

- Extraction d'information dans les décisions
  - entités juridiques [Waltl et al., 2016, Andrew and Tannier, 2018]
  - faits [Wyner, 2010, Wyner and Peters, 2010, Shulayeva et al., 2017]
  - définitions de concept juridiques [Waltl et al., 2016, Waltl et al., 2017]
  - arguments [Moens et al., 2007]
- Classification de décisions
  - Prédiction des décisions de justice [Ashley and Brüninghaus, 2009, Aletras et al., 2016]
  - identification de la formation et la période [Șulea et al., 2017b, Șulea et al., 2017a]
  - identifier la sentence prononcée (Chine) [Ma et al., 2018]
- Similarité entre décisions
  - décisions qui citent les mêmes lois et précédents [Nair and Wagh, 2018]
  - recherche d'affaires antérieures pertinentes [Thenmozhi et al., 2017]
  - identifier la sentence prononcée (Chine) [Ma et al., 2018]
  - similarité basée sur la question discutée et les faits sous-jacents (Inde) [Kumar et al., 2011]
  - regroupement non-supervisé [Ravi Kumar and Raghuv eer, 2012]

# Objectif : automatiser des tâches d'analyse de décisions

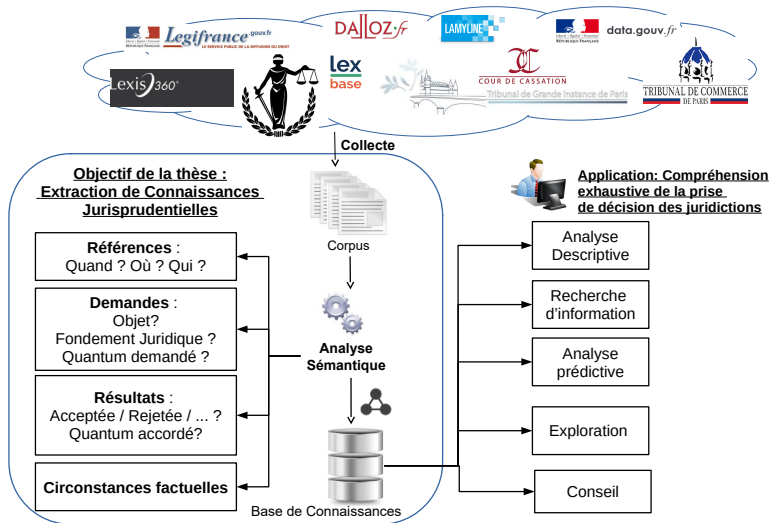


FIGURE — Objectifs et exemples d'application de la thèse.



# Problème : annotation de sections, métadonnées, normes

Cour d'appel  
Lyon  
6e chambre  
17 Mars 2016  
Répertoire Général : 14/06777  
APPELANTE :  
Mme Monique V. ...  
Représentée par Me Chrystelle P. , avocat au ...  
INTIMES :  
Mme Sylvianne C. ...  
Composition de la Cour ... :  
- Claude VIEILLARD , président ...

FAITS, PROCÉDURE, MOYENS ET ...  
Suite à un prêt de 10.000 € ...  
Par jugement en date du 4 avril 2013, ...  
Dans leurs conclusions ..., Mme Sylvianne C. , M. ...  
demandent à la cour de :  
- condamner Mme V. à leur payer ... au titre de l'article 700 du  
code ... , ...

MOTIFS DE LA DÉCISION  
La cour constate au préalable que le jugement n' est pas remis  
en causes ...  
...  
La Cour estime par contre que ... application  
de l' article 700 du code de procédure civile en cause d' appel  
au profit des  
intimés et il convient de leur allouer à ce titre la somme de 1.000  
€ .

PAR CES MOTIFS  
La Cour , ...

▼ Original markups

- ☒ appellant
- ☒ avocat
- ☒ corps
- ☒ date
- ☐ decision
- ☐ dispositif
- ☐ entete
- ☒ fonction
- ☒ formation
- ☒ intime
- ☒ juge
- ☒ juridiction
- ☐ litige
- ☐ motifs
- ☒ norme
- ☒ rg
- ☒ ville

# Problème : extraction des demandes

Cibles : sens du résultat, montant demandé, montant accordé

## EXPRESSION DE DEMANDE ET RESULTAT

Jennifer M. et Catherine M. ... demandent à la Cour de :

- **infirmen le dit jugement** en **toutes ses dispositions**; ...

Statuant à nouveau ...

- les condamner au paiement d' une somme de 3 000,00 € **pour procédure abusive** et aux entiers dépens; ...

La cour ...

CONFIRME **le jugement entrepris** en **toutes ses dispositions**.

Légende : *référence au jugement antérieur, agrégation*

IDENTIFICATION DE LA DECISION ▼			DESCRIPTION DE LA PRETENTION ▼			DESCRIPTION DU RESULTAT ▼	
Type	Ressort	RG	OBJET	NORME	QUANTUM	RESULTAT	QUANTUM RESULTAT (obtenu)
CA	Saint Denis	14/01082	dommages-intérêts	1382 code civil + 32-1 code de procédure civile : en procédure abusve	3,000.00 €	rejette	0.00 €

TABLE – Informations à extraire (dommages-intérêts pour procédure abusive)

# Problème : découverte des circonstances factuelles

Déterminer les situations distinctes où sont formulées les demandes d'une catégorie données.

## CATEGORIE : ACTION EN RESPONSABILITE CIVILE PROFESSIONNELLE CONTRE LES AVOCATS

- cas *a* (46 documents) : il s'agit d'un avocat négligent qui envoie son assignation de manière tardive ;
- cas *b* (20 documents) : il s'agit d'un avocat qui n'a pas donné un conseil opportun, qui n'a pas soulevé le bon argument ;
- cas *c* (18 documents) : un avocat qui n'a pas rédigé un acte valide ou réussi à obtenir un avantage fiscal ;
- cas *d* (3 documents) : il s'agit d'un avocat attaqué par son adversaire et non par son propre client.

# Positionnement en fouille de texte

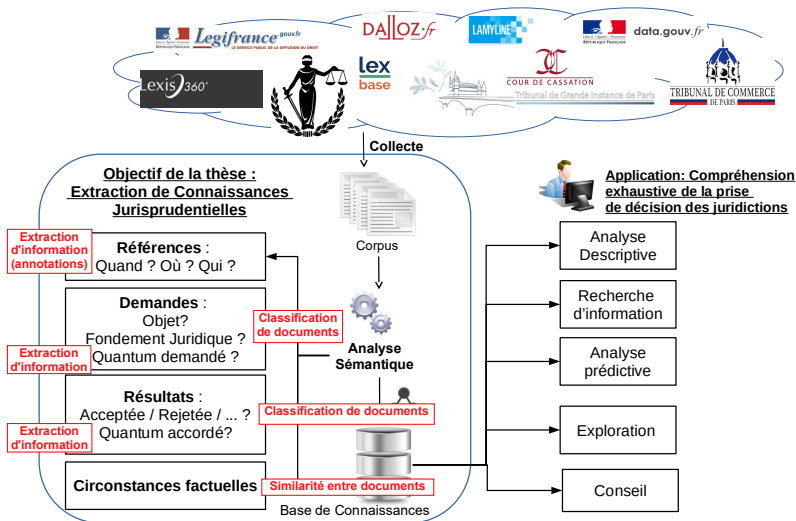


FIGURE – Tâches abordées en analyse de données textuelles.

# Difficultés rencontrées par l'automatisation de ces tâches

- Les décisions sont des textes non-structurés
- Le langage juridique est complexe

ARRÊT N°

R.G : 11/03924

...

COUR D'APPEL DE NÎMES

CHAMBRE CIVILE

1ère Chambre A

ARRÊT DU 20 MARS 2012

APPELANTE :

Madame Michèle A. ...

assistée de la SELARL VAJOU, ...

INTIMES :

Monsieur Martial B ...

assisté de la SCP MARION GUIZARD PATRICIA

SERVAIS, ...

COMPOSITION DE LA COUR LORS DU DÉLIBÉRÉ :

M. Dominique BRUZY, Président

M. Serge BERTHET, Conseiller

...

FAITS, PROCEDURE, ...

Madame Michèle A. demande :

...

- de condamner Madame JONES-B. à lui payer la somme de 2.500 euros au titre de l'article 700 du Code de Procédure Civile,

PAR CES MOTIFS, LA COUR :

...

Vu l'article 809 du Code de Procédure Civile,

...

Déboute Madame A. de sa demande de provision sur dommages-intérêts.

...

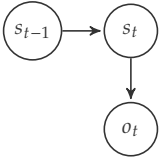
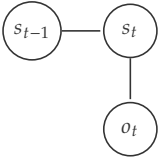
Vu l'article 700 du Code de Procédure Civile,  
Condamne Madame JONES-B. à verser à Madame A. la somme de 2.500 euros.

# Annotation des sections et entités judiciaires

---

# Appliquer le HMM ou le CRF pour l'annotation

## 1. Méthodes : Modèles probabilistes à états et observations

HMM	CRF
un seul descripteur par observation	plusieurs descripteurs complexes par observation
	
$P_{\lambda}(S O) = \prod_{t=1}^T P(s_t s_{t-1}) * P(o_t s_t)$ <p>[Seymore et al., 1999]</p>	$P_{\lambda}(S O) = \frac{1}{Z(O)} \exp \left( \sum_{t=1}^T \sum_k \lambda_k f_k(s_{t-1}, s_t, o_t) \right)$ <p>[Peng and McCallum, 2006]</p>

Objectif : Trouver la séquence la plus probable d'étiquetage pour l'ensemble du texte

**Entrainement fait sur des séquences préalablement étiquetées**

## 2. Données



## 3. Résultats

## Identification des demandes

---

## 1. Méthodes

## 2. Résultats

## Identification du sens du résultat

---

# Etendre le Gini-PLS pour la classification de textes

## 1. Méthodes

## 2. Résultats

## Découverte des circonstances factuelles

---



# Apprendre la similarité par transformation de document

## 1. Méthodes

## 2. Résultats

## Conclusions

---



# Conclusions : perspectives

---

## Questions

# References I



Aletras, N., Tsarapatsanis, D., Preoțiuc-Pietro, D., and Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights : A Natural Language Processing perspective.  
*PeerJ Computer Science*, 2 :e93.



Andrew, J. J. and Tannier, X. (2018). Automatic Extraction of Entities and Relation from Legal Documents.  
*In Proceedings of the Seventh Named Entities Workshop*, pages 1–8.



Ashley, K. D. and Brüninghaus, S. (2009). Automatically classifying case texts and predicting outcomes.  
*Artificial Intelligence and Law*, 17(2) :125–165.



Kumar, S., Reddy, P. K., Reddy, V. B., and Singh, A. (2011). Similarity analysis of legal judgments.  
*In Proceedings of Compute 2011 - Fourth Annual ACM Bangalore Conference*, page 17. ACM.



Ma, Y., Zhang, P., and Ma, J. (2018). An Efficient Approach to Learning Chinese Judgment Document Similarity Based on Knowledge Summarization.  
arXiv preprint arXiv :1808.01843 [cs.AI].



Moens, M.-F., Boiy, E., Palau, R. M., and Reed, C. (2007). Automatic detection of arguments in legal texts.  
*In Proceedings of the 11th international conference on Artificial intelligence and law*, pages 225–230. ACM.

# References II



Nair, A. M. and Wagh, R. S. (2018).  
Similarity Analysis of Court Judgements Using Association Rule Mining on Case Citation Data - A Case Study.

*International Journal of Engineering Research and Technology*, 11(3) :373–381.



Peng, F. and McCallum, A. (2006).  
Information extraction from research papers using conditional random fields.  
*Information processing & management*, 42(4) :963–979.



Ravi Kumar, V. and Raghuveer, K. (2012).  
Legal documents clustering using latent dirichlet allocation.  
*International Journal of Applied Information Systems (IJ AIS)*, 2(6) :34–37.



Seymore, K., McCallum, A., and Rosenfeld, R. (1999).  
Learning hidden Markov model structure for information extraction.  
*AAAI-99 workshop on machine learning for information extraction*.



Shulayeva, O., Siddharthan, A., and Wyner, A. (2017).  
Recognizing cited facts and principles in legal judgements.  
*Artificial Intelligence and Law*, 25(1) :107–126.



Şulea, O.-M., Zampieri, M., Malmasi, S., Vela, M., P. Dinu, L., and van Genabith, J. (2017a).  
Exploring the Use of Text Classification in the Legal Domain.  
In *Proceedings of 2nd Workshop on Automated Semantic Analysis of Information in Legal Texts*, page 5, London, United Kingdom. ASAIL'2017.



# References III



Șulea, O.-M., Zampieri, M., Vela, M., and van Genabith, J. (2017b).  
Predicting the Law Area and Decisions of French Supreme Court Cases.  
In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP 2017*,  
pages 716–722.



Thenmozhi, D., Kannan, K., and Aravindan, C. (2017).  
A Text Similarity Approach for Precedence Retrieval from Legal Documents.  
In *Proceedings of Forum for Information Retrieval Evaluation - FIRE (Working Notes)*, pages 90–91.



Waltl, B., Landthaler, J., Scepankova, E., Matthes, F., Geiger, T., Stocker, C., and Schneider, C. (2017).  
Automated extraction of semantic information from German legal documents.  
In *IRIS : Internationales Rechtsinformatik Symposium. Association for Computational Linguistics*.



Waltl, B., Matthes, F., Waltl, T., and Grass, T. (2016).  
LEXIA - A Data Science Environment for Semantic Analysis of German Legal Texts.  
In *IRIS : Internationales Rechtsinformatik Symposium*.  
Salzburg, Austria.



Wyner, A. and Peters, W. (2010).  
Lexical Semantics and Expert Legal Knowledge towards the Identification of Legal Case Factors.  
In *JURIX*, volume 10, pages 127–136.



Wyner, A. Z. (2010).  
Towards annotating and extracting textual legal case elements.  
*Informatica e Diritto : special issue on legal ontologies and artificial intelligent techniques*, 19(1-2) :9–18.