

L'intérêt des méthodes computationnelles pour la recherche juridique empirique



Julien Bétaille

Maître de conférences en droit

Membre de l'*Institut universitaire de France*

S'adresser à la fois aux juristes et aux *computer scientists*.

Un point de vue naïf

Histoire de l'hybridation (du point de vue des juristes) : pionnier de l'informatique juridique en France (P. Catala, E. Serverin, D. Bourcier)

Focus sur la connaissance pure, pas sur les applications commerciales (justice prédictive, etc.)

Le droit de l'IA n'est pas du tout mon objet.

Objet : droit de l'environnement / épistémologie du droit

Méthodes computationnelles : approches qui utilisent l'informatique pour collecter, organiser et analyser des données. Elles permettent de traiter des données volumineuses et/ou complexes.

Recherche en droit : ça veut dire quoi ?

- Décrire le contenu des normes
- Systématisation du droit
- Constructions intellectuelles et théorisation
- Interprétation
- Argumentation

Recherche juridique empirique (#doctrinal, #analytique) : considère le droit comme une donnée susceptible d'être collectée de façon systématique et analysée en assurant la reproductibilité des résultats.

Usage opportuniste des méthodes computationnelles pour la connaissance du droit

Law as data : Au-delà d'un certain niveau, il devient impossible pour l'humain de traiter l'ensemble des données juridiques utiles, surtout dans le cadre d'une recherche individuelle.

Intérêt de l'informatique pour :

- explorer de vastes corpus de documents juridiques,
- identifier les points communs et les différences entre eux,
- rechercher des informations dans ces documents,
- automatiser leur annotation.

En pratique, cela décuple notre capacité d'analyse des documents juridiques.

Données juridiques :

- Traités internationaux,
- Constitutions,
- Actes législatifs et réglementaires,
- Contrats,
- Décisions de justice,
- Tout autre document juridique etc.

Trier des grandes masses de documents juridiques

Projet sur les données concernant les sanctions en matière d'installations industrielles

Index of /icpe/archives/

20190926T000000Z icpe.csv.gz	26-Sep-2019 08:31	1105316
20190926T000000Z icpe.geojson.gz	26-Sep-2019 21:56	4053279
20190927T000000Z icpe.csv.gz	27-Sep-2019 06:03	1105424
20190927T000000Z icpe.geojson.gz	27-Sep-2019 06:03	4049189
20190928T000000Z icpe.csv.gz	28-Sep-2019 06:03	1105472
20190928T000000Z icpe.geojson.gz	28-Sep-2019 06:03	4042581
20190928T000000Z icpe_detail.json.gz	28-Sep-2019 13:29	14956542
20191001T000000Z icpe.csv.gz	01-Oct-2019 06:40	1105561
20191001T000000Z icpe.geojson.gz	01-Oct-2019 06:41	4041179
20191002T000000Z icpe.csv.gz	02-Oct-2019 07:05	4055527
20191003T000000Z icpe.csv.gz	03-Oct-2019 06:47	1105716
20191003T000000Z icpe.geojson.gz	03-Oct-2019 06:47	4040503
20191004T000000Z icpe.csv.gz	04-Oct-2019 06:45	1105554
20191004T000000Z icpe.geojson.gz	04-Oct-2019 06:45	4039126
20191005T000000Z icpe.csv.gz	05-Oct-2019 09:47	1105571
20191005T000000Z icpe.geojson.gz	05-Oct-2019 09:47	4044735
20191008T000000Z icpe.csv.gz	08-Oct-2019 06:45	1105545
20191008T000000Z icpe.geojson.gz	08-Oct-2019 06:45	4056767
20191009T000000Z icpe.csv.gz	09-Oct-2019 06:44	4047680
20191010T000000Z icpe.csv.gz	10-Oct-2019 06:45	1105794
20191010T000000Z icpe.geojson.gz	10-Oct-2019 06:45	4056088
20191011T000000Z icpe.csv.gz	11-Oct-2019 06:45	4043157
20191012T000000Z icpe.csv.gz	12-Oct-2019 09:02	1105874
20191014T000000Z icpe.csv.gz	14-Oct-2019 07:08	1105925
20191015T000000Z icpe.csv.gz	15-Oct-2019 07:04	4044042
20191016T000000Z icpe.csv.gz	16-Oct-2019 07:09	1106320
20191016T000000Z icpe.geojson.gz	16-Oct-2019 07:09	4044397
20191017T000000Z icpe.csv.gz	17-Oct-2019 07:06	1106187
20191017T000000Z icpe.geojson.gz	17-Oct-2019 07:07	4075001
20191018T000000Z icpe.csv.gz	18-Oct-2019 07:28	1106342
20191018T000000Z icpe.geojson.gz	18-Oct-2019 07:28	4072097
20191019T000000Z icpe.csv.gz	19-Oct-2019 07:27	1106400
20191019T000000Z icpe.geojson.gz	19-Oct-2019 07:27	4075200
20191021T000000Z icpe.csv.gz	22-Oct-2019 06:39	1106537
20191022T000000Z icpe.csv.gz	22-Oct-2019 06:39	4067534
20191023T000000Z icpe.csv.gz	23-Oct-2019 07:06	4071462
20191024T000000Z icpe.csv.gz	24-Oct-2019 06:41	1106683
20191024T000000Z icpe.geojson.gz	24-Oct-2019 06:41	4072387
20191025T000000Z icpe.csv.gz	25-Oct-2019 06:40	1106511
20191025T000000Z icpe.geojson.gz	25-Oct-2019 06:40	4067312
20191028T000000Z icpe.csv.gz	28-Oct-2019 06:36	4075982
20191029T000000Z icpe.csv.gz	29-Oct-2019 06:36	1106519
20191029T000000Z icpe.geojson.gz	29-Oct-2019 06:36	4067376
20191030T000000Z icpe.csv.gz	30-Oct-2019 06:43	1106597
20191030T000000Z icpe.geojson.gz	30-Oct-2019 06:43	4060424
20191031T000000Z icpe.csv.gz	31-Oct-2019 06:38	1106568
20191031T000000Z icpe.geojson.gz	31-Oct-2019 06:38	4068239
20191102T000000Z icpe.csv.gz	02-Nov-2019 09:32	4067866
20191104T000000Z icpe.csv.gz	04-Nov-2019 06:36	1106627
20191105T000000Z icpe.csv.gz	05-Nov-2019 06:41	4061865
20191105T000000Z icpe.geojson.gz	05-Nov-2019 06:41	1106784
20191106T000000Z icpe.csv.gz	06-Nov-2019 06:41	4087488
20191107T000000Z icpe.csv.gz	07-Nov-2019 06:38	4075077
20191108T000000Z icpe.csv.gz	08-Nov-2019 06:49	1106779
20191108T000000Z icpe.geojson.gz	08-Nov-2019 06:49	4065724
20191109T000000Z icpe.csv.gz	09-Nov-2019 06:46	1106796
20191109T000000Z icpe.geojson.gz	09-Nov-2019 06:46	4069605
20191110T000000Z icpe.csv.gz	12-Nov-2019 06:39	1106570
20191113T000000Z icpe.csv.gz	13-Nov-2019 06:41	1106756
20191113T000000Z icpe.geojson.gz	13-Nov-2019 06:41	4071347
20191114T000000Z icpe.csv.gz	14-Nov-2019 06:44	4057999
20191115T000000Z icpe.csv.gz	15-Nov-2019 06:40	1106891
20191115T000000Z icpe.geojson.gz	15-Nov-2019 06:40	4058065
20191118T000000Z icpe.csv.gz	18-Nov-2019 07:05	4062177
20191119T000000Z icpe.csv.gz	19-Nov-2019 06:46	1106855
20191119T000000Z icpe.geojson.gz	19-Nov-2019 06:46	4062616
20191120T000000Z icpe.csv.gz	20-Nov-2019 06:42	1106956
20191120T000000Z icpe.geojson.gz	20-Nov-2019 06:42	4082279
20191121T000000Z icpe.csv.gz	21-Nov-2019 06:42	4066835
20191122T000000Z icpe.csv.gz	22-Nov-2019 06:38	4062202
20191123T000000Z icpe.csv.gz	23-Nov-2019 06:40	1107073
20191123T000000Z icpe.geojson.gz	23-Nov-2019 06:41	4067832
20191125T000000Z icpe.csv.gz	25-Nov-2019 06:42	1107093
20191126T000000Z icpe.csv.gz	26-Nov-2019 08:11	4057286
20191127T000000Z icpe.csv.gz	27-Nov-2019 06:43	1107065
20191127T000000Z icpe.geojson.gz	27-Nov-2019 06:43	4069888
20191128T000000Z icpe.csv.gz	28-Nov-2019 06:41	4058941
20191129T000000Z icpe.csv.gz	29-Nov-2019 06:42	1107202
20191129T000000Z icpe.geojson.gz	29-Nov-2019 06:43	4067745
20191130T000000Z icpe.csv.gz	30-Nov-2019 06:44	4067245
20191202T000000Z icpe.csv.gz	02-Dec-2019 06:42	1107010

Graphs

Romain Boulet, Pierre Mazzega et Danièle Bourcier, « Network Analysis of the French Environmental Code », in P. Casanovas et al. (dir.), AICOL Workshops 2009, LNAI 6237, Springer, p. 39.

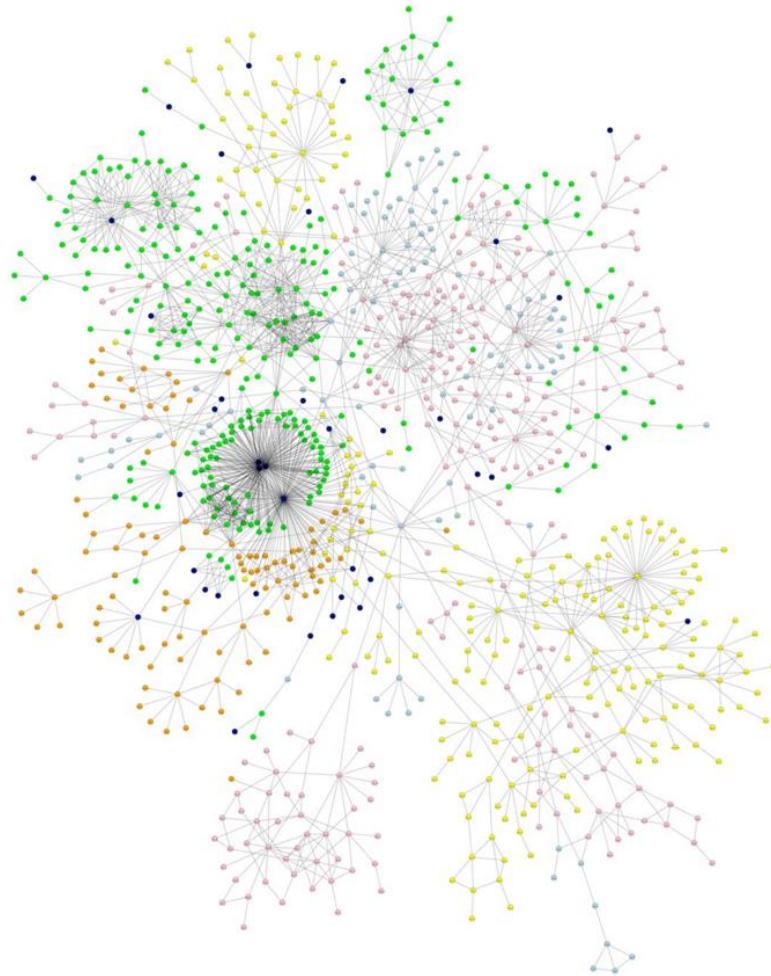


Figure 1. The graph G associated to the main connected component (980 vertices, 2186 edges) of the Environmental Code (legislative part only). The colors given to the vertices indicate the Book they are belonging to (hereafter only short names are given): (blue) Common provisions; (green) Physical environments; (orange) Natural space; (yellow) Flora and fauna; (pink) Prevention of pollutions, Risks nuisances; (dark blue) Provisions applicable in New Caledonia etc.; (grey) Environmental protection in Antarctica.

Pierre Mazzega, Danièle Bourcier et Romain Boulet, « The Network of French Legal Codes », in *ICAAIL '09: Proceedings of the 12th Int. Conference on AI and Law*, Assoc. for Computing Machinery, 2009, p. 236.

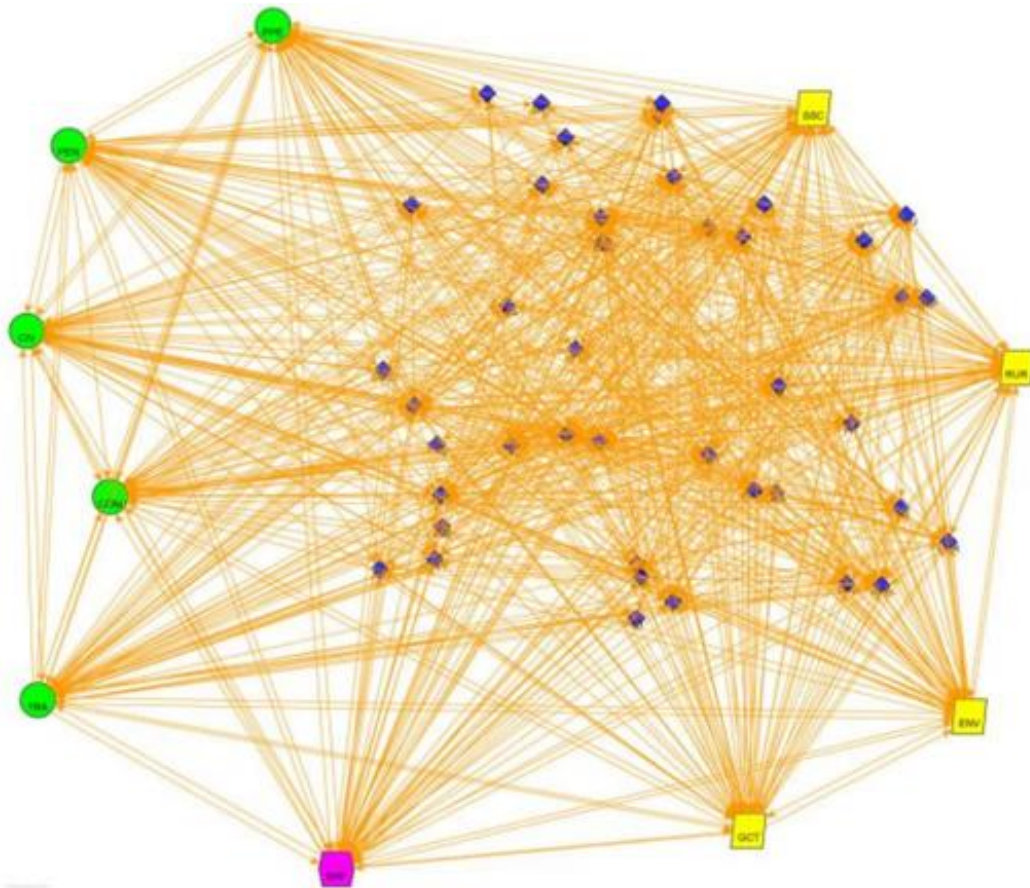


Figure. The graph associated with the network of French Legal Codes. Each vertex (blue diamonds, yellow squares, green circles, cyan hexagon) represents a Code. The yellow squares stand for the 4 codes most citing other codes. The 5 green circles are the most cited codes. The cyan hexagon is in both sets of the most cited and most citing codes. All these codes together (not including the diamond shaped symbols) form the rich club.



Goodbye *van Gend en Loos*, Hello *Bosman*? Using Network Analysis to Measure the Importance of Individual CJEU Judgments

*Mattias Derlén and Johan Lindholm**

Table 3. Top Judgments Based on CJEU Citations (In-Degree, 2011)

Rank	Case	In-degree
1	C-415/93 <i>Bosman</i>	134
2	8/74 <i>Dassonville</i>	112
3	C-379/98 <i>PreussenElektra</i>	89
T4	120/78 <i>Cassis de Dijon</i>	68
T4	C-168/03 <i>Commission v. Spain</i>	68
6	C-23/05 <i>Commission v. Luxembourg</i>	67
T7	C-279/93 <i>Schumacker</i>	57
T7	C-136/92 <i>Augusto Brazzelli</i>	57
9	C-55/94 <i>Gebhard</i>	56
10	8/81 <i>Ursula Becker</i>	54

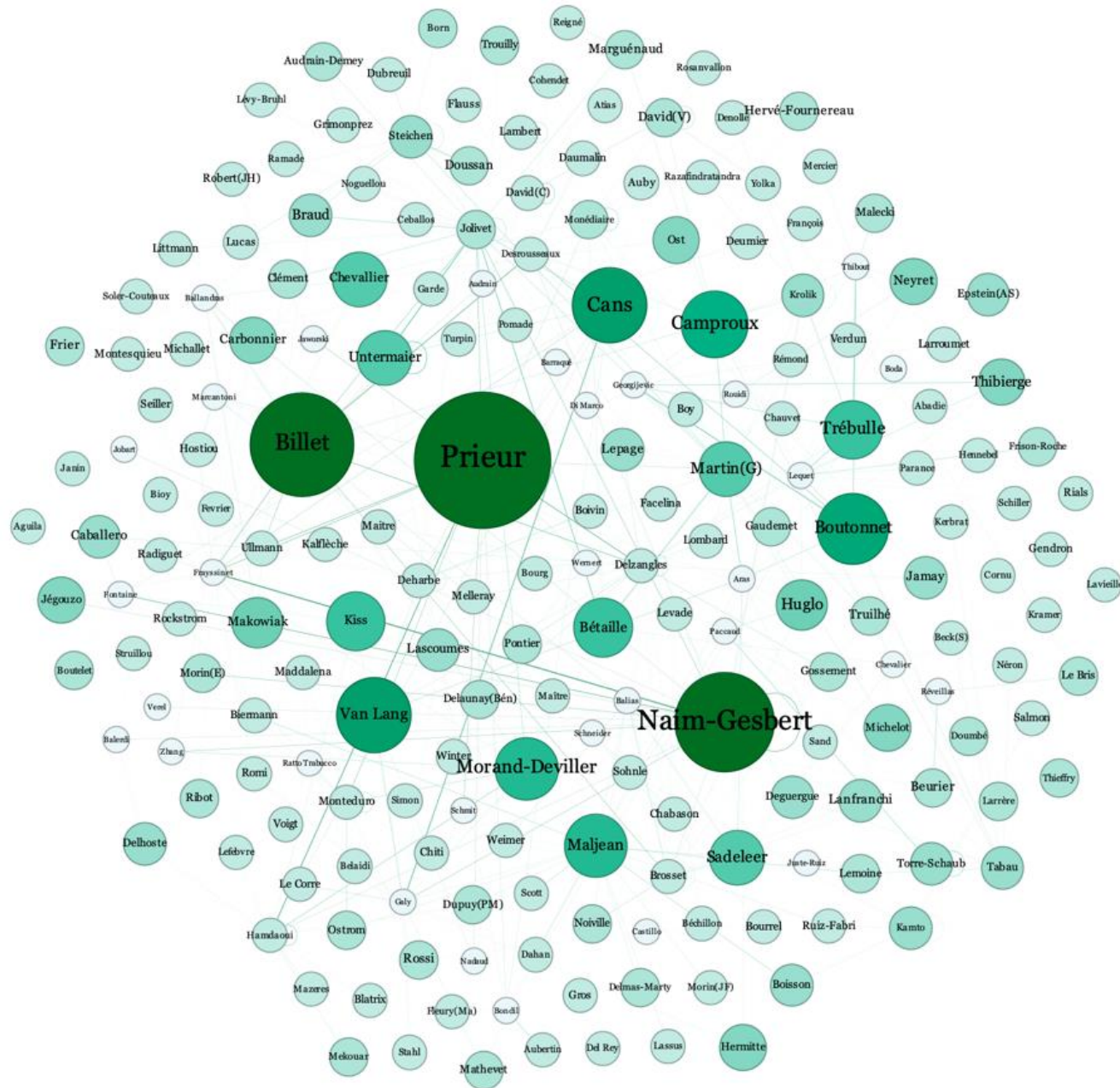
Table 2. Top Judgment Based on Advocates General Citations

Rank	Case	Citations
1	8/74 <i>Dassonville</i>	174
2	120/78 <i>Cassis de Dijon</i>	162
3	C-415/93 <i>Bosman</i>	145
4	C-279/93 <i>Schumacker</i>	106
5	C-55/94 <i>Gebhard</i>	100
6	222/84 <i>Royal Ulster Constabulary</i>	96
7	43/75 <i>Defrenne II</i>	94
8	C-6/90 <i>Francovich</i>	91
9	6/64 <i>Costa v. ENEL</i>	90
T10	152/84 <i>Marshall</i>	89
T10	33/76 <i>Rewe Zentralfinanz</i>	89

Table 6. Betweenness Centrality

Rank	Case	Betweenness
1	C-415/93 <i>Bosman</i>	234395
2	C-344/04 <i>IATA</i>	203989
3	C-156/98 <i>New Lander</i>	148818
4	C-210/03 <i>Swedish Match</i>	143393
5	C-491/01 <i>British American Tobacco</i>	139459
6	C-379/98 <i>PreussenElektra</i>	132705
7	C-191/95 <i>Commission v. Germany</i>	130763
8	C-280/93 <i>Germany v. Council</i>	123354
9	C-46/93 & 48/93 <i>Brasserie du Pêcheur</i>	117118
10	C-495/03 <i>Intermodal Transport</i>	111843

Analyse de la doctrine environmentaliste



Données : environ 1500 références citées dans les 120 articles parus dans la *Revue juridique de l'environnement* entre 2013 et 2022

Machine Learning / Topic Modeling

Data: 200,000 legislative acts, 55,000 court rulings and opinions, and 4,000 articles from a leading EU law journal

ARTICLE

The Voices of European Law: Legislators, Judges and Law Professors

Arthur Dyevre¹, Monika Glavina^{1,*} and Michal Ovrádek¹

Unsupervised Machine Learning / Probabilistic topic modelling
Results: « We find that economic integration remains the focus of EU law, but that scholars tend to emphasize rights issues more and ignore certain topics, such as farming regulations, almost entirely ».

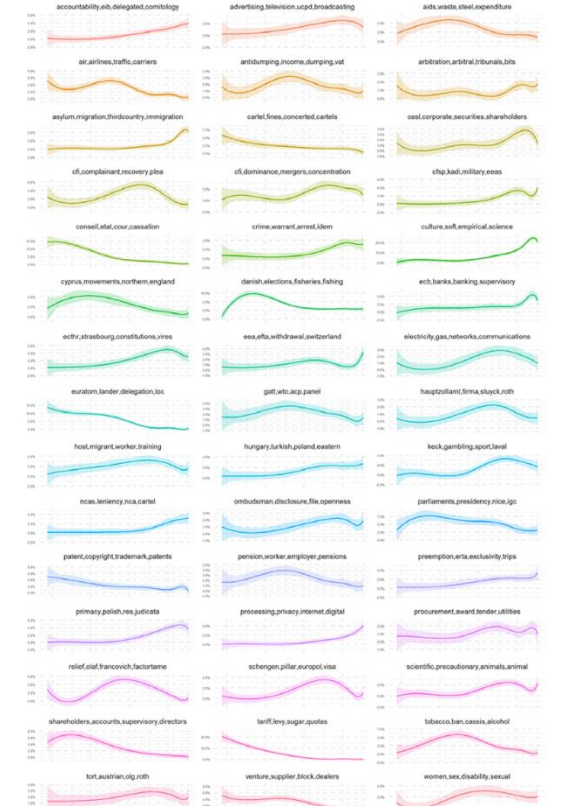
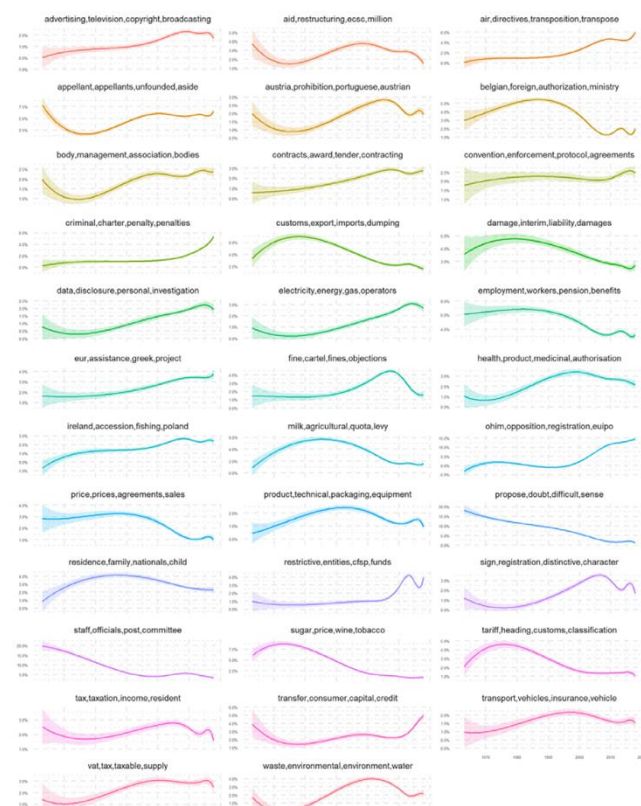


Figure A1. Topic proportion over time for legislative topics.

Figure A2. Topic proportion over time for CJEU topics.

Figure A3. Topic proportion over time for CMLR topics.

Annotation de documents juridiques

Claws & Laws | Litigation database

A database covering all species protection administrative court cases from all national court in all Member States of the European Union. Read more information about the database or search cases below.



The database contains **6827** court rulings. Last updated on **2024-06-17**

Protected species

Search interface:

The database can be searched through different fields, or combination of fields. Choosing a field on the left search menu updates the corresponding options on the right search menu.

Criteria to search the database:

You can search court cases by the protected species (for Habitats Directive cases) or bird (for Birds Directive cases) they deal with, the country where the cases took place and the case language, the year of the cases, the activity involved, the types of plaintiff or defendant, the types of court and litigation as well as the judicial decision.

Base de données européennes : $n = 6827$
Base de données française : codage manuel ($n = 2114$)

PREVENTING ENVIRONMENTAL HARM THROUGH LITIGATION

THE EFFECTIVENESS OF FRENCH ADMINISTRATIVE
SUMMARY PROCEEDINGS

Données : plus de 2500 décisions des
juridictions administratives françaises

Objectif : éviter le codage manuel !

DATA ANALYSIS



Court decisions
(heterogeneous data)

- Case number
- Court : type and location
- Plaintiff
- Complaint date
- Decision date

- Date of the administrative
decision for which suspension
is requested

Script with pattern recognition



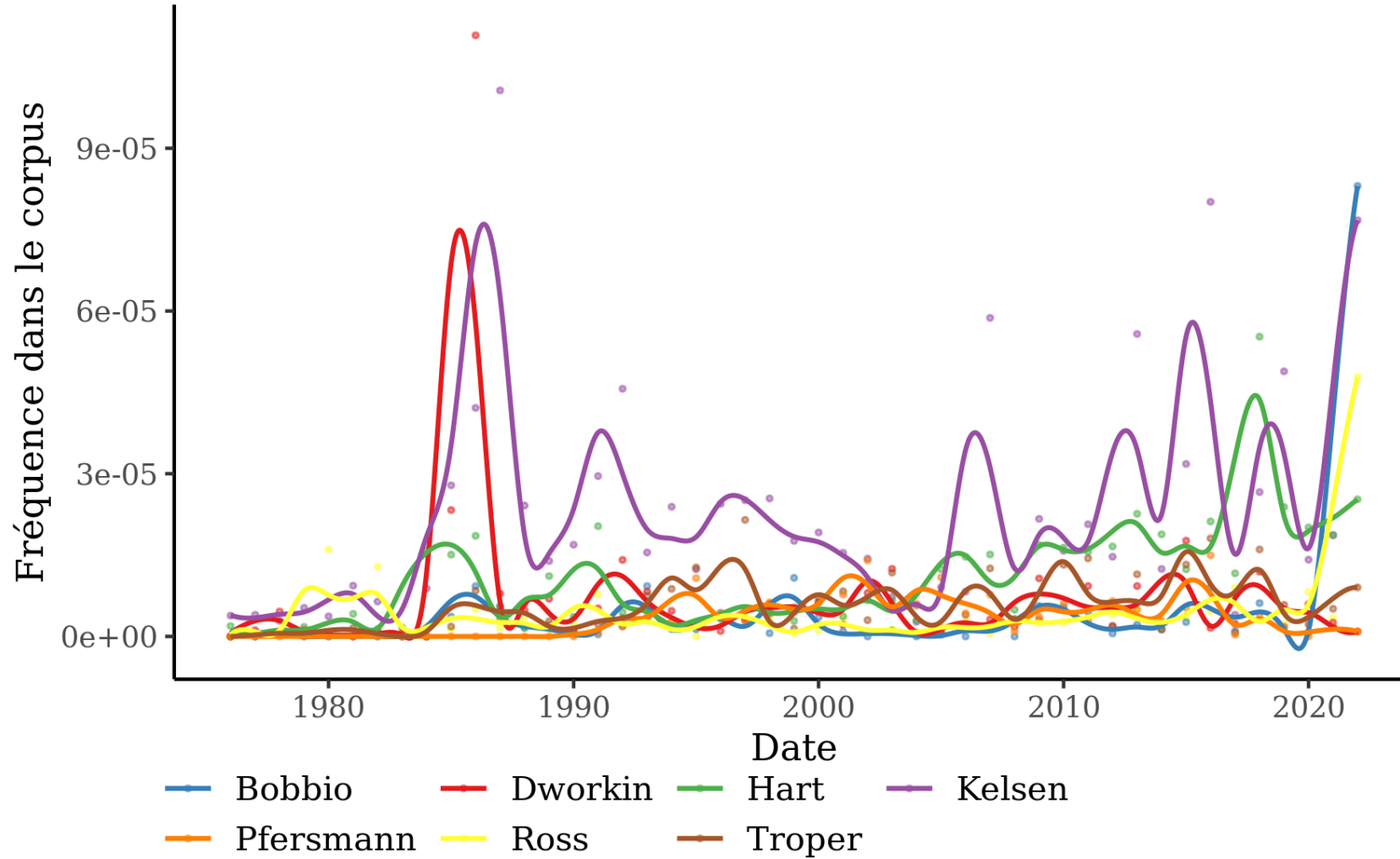
Formatted data
(usable)

Local LLM (LLaMA 3) pretrained
with court decisions



Lexicométrie

Influence des principaux auteurs de théorie du droit sur les travaux des juristes universitaires



Corpus : revues juridiques
présentes sur Persée
Analyse : Gallicagram

Extraction et classification de normes (norms mining, LLM)

Once Upon a Time: Introducing the Historical Values Survey¹

VERY FIRST DRAFT – COMMENTS MOST WELCOME

MAHDI KHESALI‡, STEFAN VOIGT* and NADIA VON JACOBI‡

Abstract

This article introduces the Historical Values Survey. It contains information on historically prevalent values and norms encouraging cooperation, altruism, work ethic, entrepreneurship, but also discrimination and the challenging of authority. Information on values and norms is elicited from more than 15,000 folktales relying on artificial intelligence. The Survey covers 156 countries. Folktales, such as legends and fairy tales, are transmitted virtually unchanged from generation to generation and thus constitute an excellent source for historically prevalent values and norms. Some historically prevalent values and norms correlate with contemporaneous values and economic outcomes. Including the values and norms into a growth model based on long-term determinants of economic growth improves model fit and documents the relevance of culture for economic development.

Keywords: folktales, fairy tales, culture, values, norms, economic development, values survey

JEL classification: A13, D90, K00, N30, O10, P51, Z10

Extraction de valeurs et de normes à partir des histoires racontées aux enfants

Classification des textes selon qu'ils encouragent le respect de la norme ou non

Difficultés:

- Sensibilité des données : ChatGPT vs RGPD, nécessité de travailler en local avec Llama ou Mistral, limité par la puissance de nos ordinateurs
- Codage Python, un fardeau ?
- Travailler à plusieurs ?
- Des logiciels simples ? Gephi.
- Formations peu adaptées.

Environnement de la recherche

- Conférences ICAIL, JURIX
- NLLP: <https://sites.google.com/view/nllp/home>

2nd Workshop on Natural Legal
Language Processing (NLLP)

Des initiatives concrètes

- Toulouse Summer School 2023, 2024
- Legal Networks Analysis :
<https://maastrichtlawtech.github.io/LegalNetworkAnalysis/#/>
- Annotation de données juridiques : <https://www.lawnotation.org/>
- Gallicagram : <https://shiny.ens-paris-saclay.fr/app/gallicagram>



TOULOUSE CAPITOLE
UNIVERSITY

SUMMER SCHOOL

**ARTIFICIAL INTELLIGENCE &
QUANTITATIVE METHODS FOR
EMPIRICAL LEGAL RESEARCH**

JUNE • 10 - 18 • 2024

Avenir

- L'intérêt de ces méthodes pour les juristes est indéniable
- Mais quid de l'intérêt du terrain d'application juridique pour les *computer scientists* ?

Idées pour avancer

- Un cours sur l'application des méthodes computationnelles en droit
- Exemple du *Law & Tech Lab* Maastricht (dir. Gijs van Dijck) : <https://www.maastrichtuniversity.nl/about-um/faculties/law/research/law-and-tech-lab>



Faculty of Law >

About the Faculty >

Education >

Research >

Law and Tech Lab

At the Maastricht Law and Tech Lab, we believe that we can create a better society by combining expertise in law and computer science to study digital transformation. Our interdisciplinary lab fosters a community of researchers who work collaboratively to integrate law and computer science methods.

Maastricht University
Maastricht Law and Tech Lab

julienbetaille@gmail.com