

A list of selected resources, methods, and tools dedicated to Legal Text Analytics.

 CC-BY-SA-4.0 License

 7 stars    2 forks


 Star


 Watch ▾

-  **Code**
-  Issues
-  Pull requests
-  Actions
-  Projects
-  Wiki
-  Security

 main ▾



 **bwaltl** Update README.md ...

2 days ago  72

[View code](#)

README.md

# Legal Text Analytics

CLP **Ecosystem**

CLP **Code**

CLP **Community**

 **awesome**

License **CC BY-SA 4.0**

A list of selected resources, methods, and tools dedicated to Legal Text Analytics.



Please read the [contribution guidelines](#) before contributing. Please add a resource by raising a [pull request](#). We also seek for discussion and proposal of new ideas (including additional content sections) as [issues](#).

## Contents

- [Selected Use Cases](#)
- [Methods](#)
- [Libraries](#)
- [Datasets and Data](#)

- [Annotation and Data Schemes](#)
- [Annotation Tools](#)
- [Research Groups and Labs](#)
- [Tutorials](#)
  
- [Credits](#)

## Selected Use Cases

---

[Back to Top](#)

- [Optical Character Recognition](#) (find more information [here](#))
- [Legal Document Pre-processing](#) (find more information [here](#))
- [Clause Segmentation](#) and [Sentence Boundary Detection](#)
- [Information Extraction](#) and [Named Entity Recognition](#) (find more information [here](#))
- [Legal Norm Classification](#)
- [Machine Translation](#)
- [Document Comparison](#) and [Semantic Matching](#)
- [Text Summarization](#)
- [Argument Mining](#)
- [Question Answering](#)
- [Legal Case Outcome Prediction](#)
- [Reference and Coreference Extraction](#)
- [Document Assembling and Generation](#)
- [Voice Transcription](#)
- [Anomaly Detection](#)
- [Data Anonymization](#)

## Methods

---

[Back to Top issues](#)

- [NLP Overview](#)
- [NLP Progress](#)
- [Text Visualizations](#)
- [Optical Character Recognition](#)
- [Rule-based methods for NLP](#), [Apache Ruta](#), [Jape Grammar](#)
- [Statistical NLP](#)
- [Machine Learning Frameworks](#)
- [Neural networks and deep learning for NLP Tutorial](#)

## Libraries

---

[Back to Top](#)

- [Spacy - Industrial-Strength Natural Language Processing](#)
- [Scikit - machine learning in python](#)
- [NLTK - Natural Language Toolkit](#)

- [Apache UIMA](#)
- [Gate - General Architecture for Text Engineering](#)
- [Transformers and re-trained embedding models including LegalBERT](#)
- [German Bert Model: Deepset AI](#)
- [Flair - SOTA NLP \(incl. biomedical and legal data\)](#)
- [Blackstone - Legal Named Entity Recognition and Text Categorizer](#)
- [Legal Reference Detection, Legal Reference Detection II](#)
- [Haystack - Transformers at scale for question answering & neural search](#)

## Datasets and Data

---

[Back to Top](#)

- [NLP Datasets](#)
- [OpenLegalData](#)
- [Legal Entity Recognition](#)
- [Legal Text Summarization](#)
- [Legal Text Translation](#)
- [Legal Document Classification](#)
- [Legal Sentence Classification \(German\)](#)
- [100k German Court Decisions](#)
- [Legal Paper Datasets](#)
- [Awesome Legal Data](#)
- [Germany: Gesetze im Internet, Rechtsprechung im Internet, Verwaltungsvorschriften im Internet, Annotated German Court Decisions \(Judgment style\), German Federal Courts Dataset](#)
- [Switzerland: Swiss Legislation Corpus](#) French and German
- [German NLP Ressources: Awesome German NLP](#)
- [ECtHR: Judicial Decisions of the European Court of Human Rights](#)
- [EU: European Union Law \(eurlex R Package\), Digital Corpus of the European Parliament \(DCEP\)](#)
- [Canada: Federal Laws and Regulations \(ftp://205.193.86.89/\)](#)
- [UK: UK Law Reports & Case Law Search](#)
- [USA: Caselaw Access Project](#)
- [USA: Supreme Court Database](#)
- [Israel: The Israeli Supreme Court Database](#)
- [United Nations: United Nations General Debate Corpus, United Nations Parallel Corpus](#)
- [International Law: Text of Trade Agreements \(ToTA\), Electronic Database on Investment Treaties \(EDIT\)](#)
- [Meta Search: Google Dataset Search](#)
- [Overview of Political Science Datasets: PolData](#)

## Annotation and Data Schemes

---

[Back to Top](#)

- [Annotation guidelines for Legal Entity Recognition \(Germany\)](#)

- [Semantic Types of Legal Norms](#)

## Annotation Tools

---

[Back to Top](#)

- [Awesome data annotation](#)
- [Prodigy](#)
- [Doccano](#)
- [Brat](#)

## Research Groups and Labs

---

[Back to Top](#)

- [Stanford University - CodeX: The Stanford Center for Legal Informatics](#)
- [Technical University of Munich](#)
- [Bucerius Center on the Legal Profession](#)
- [Suffolk Law School - Legal Innovation & Technology \(LIT\) Lab](#)
- [University of Ottawa - Legal Technology Lab](#)
- [University of Vienna - Department of Innovation and Digitalisation in Law](#)
- [University of Amsterdam - Leibniz Center for Law](#)
- [University of Helsinki - LegalTech Research Lab](#)
- [Hofstra University - Law, Logic & Technology Research Laboratory](#)
- [Computational Legal Studies](#)

## Tutorials

---

[Back to Top](#)

- [Monkey Learn - Text Analysis](#)
- [Using NLP to understand laws](#)
- [Document Representation for Legal Texts](#)
- [Data Science for Lawyers - Learning Resources](#)
- [Coding for Lawyers \(discontinued\)](#)
- [Custom NLP Approaches to Data Anonymization](#)
- [Information Extraction in legal documents](#)

## Credits

---

[Back to Top](#)

See contributors and committers (and many more).

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

### Releases

No releases published

No releases published

Contributors 4



**bwaltl** Bernhard Waltl



**pdufter** Philipp Dufter



**bnczk** Georg Bonczek



**WAlschner**