Automated Legal Research for German Law

Automation in law can have far reaching advantages in providing direct access to justice. Simple to use applications could help in querying legal concerns and obtaining a preliminary analysis. Legal professionals could use those applications for help with case research and for detecting edge conditions, inequality and loopholes (Ashley, 2017). The idea of adding automation to (some of) legal processes is not new. Techniques such as using first order logic for modeling cases have been shown to be useful. Unfortunately, they were not practical solutions as modeling the law as logic programs is a difficult task. New machine learning techniques and more accessible computation power has led to improved NLP and also legal automation. These machine learning models for NLP are trained and evaluated over large datasets (Mahfouz and Kandil, 2012).

The existing techniques are usually applicable to *Case Law* in which the results of previous cases become part of the law. Obtaining good accuracy of the systems requires a large enough dataset. *Civil Law* or rule-based law, such as the one practiced in Germany is different from case law mainly because in such systems the laws change only when the definitions of the laws are amended. Datasets with case texts under Civil Law are difficult to find.

We have built a system for performing *automated legal research* of Civil Law. The dataset has all the legal text organized according to legal code, sections, paragraphs and sentence numbers. Relevant links connecting related laws are present. Supporting information such as POS tags, parse-trees, synonyms and similar words (found using Wikipedia word embeddings) are used to enrich the dataset with features. The user can input a simple sentence(s) describing the case, according to which the legal case is classified to a specific part of the law. Interactive fact collection is then performed (Iyer et al., 2017). Once enough facts are collected and particular legal texts can be matched with sufficient confidence, judgment prediction is performed. All the collected facts, matching legal text with justification (Silva et al., 2018) and predictions are compiled into a report for the user. Future work on this system will include an argument mining system based on rhetorical relations and figures in law text (Mitrović et al., 2017).

References

Kevin D Ashley. Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge University Press, 2017.

Srinivasan Iyer, Ioannis Konstas, Alvin Cheung, Jayant Krishnamurthy, and Luke Zettlemoyer. Learning a neural semantic parser from user feedback. Proceedings of the 55th Annual Meeting of ACL (Vol 1: Long Papers), 2017.

Tarek Mahfouz and Amr Kandil. Litigation outcome prediction of differing site condition disputes through machine learning models. Journal of Computing in Civil Engineering, 26(3):298–308, may 2012.

Vivian S Silva, André Freitas, and Siegfried Handschuh. Recognizing and justifying text entailment through distributional navigation on definition graphs. 2018.

Jelena Mitrović, Cliff, O'Reilly, Miljana Mladenović, Siegfried Handschuh. Ontological Representations of Rhetorical Figures for Argument Mining. Argument and Computation. 7(3), 2017.