Curation Technologies for the Construction and Utilisation of Legal Knowledge Graphs

Julian Moreno Schneider, Georg Rehm
Language Technology Lab, DFKI GmbH
Alt-Moabit 91c, 10559 Berlin, Germany
julian.moreno_schneider@dfki.de, georg.rehm@dfki.de

Abstract

Under the umbrella of the European project LYNX we currently develop technologies for the construction of a legal knowledge graph and a corresponding system that makes use of this legal knowledge graph. The final platform will eventually bundle a set of semantic services into a curation technology system, which is meant to assist users to process legal and regulatory content and data more efficiently and more effectively. In this paper we present an overview of the current state of the art with regard to semantic technologies and natural language processing approaches applied to the legal domain.

Keywords: Curation Technologies, Natural Language Processing, Legal Domain

1. Introduction

The ever growing amount of digital information not only offers immense opportunities but also makes it necessary, in practically all professional areas and also niches, to develop new, efficient and effective approaches for processing digital content in order to make the information available in a way that fits the users' specific use cases as adequately as possible (Rehm et al., 2018). In a general sense, these professional users are the *curators* of digital content, for example, a journalist, producer of a television programme, a knowledge worker, a scholar, or someone who is collecting information to put together a report. The processes involved in digital curation include, among others, sorting, analysing, summarising, translating and paraphrasing digital content in terms of large amounts of incoming data and producing some kind of output, e. g., a study that relies on facts and figures found in a large data collection.

This data collection is typically a combination of publicly available sources (e. g., Wikipedia and other websites) and in-house collections owned by the respective organisation or data sets that the organisation has access to. The amount of time digital curators have so that they can familiarise themselves with new topics depends on the respective sector and is typically not a lot, ranging from a couple of hours or days to a few weeks at most. Working under intense time pressue, digital curators may not be able to identify and locate all relevant information contained in a sizable document collection (Neudecker and Rehm, 2016).

Ideally, digital curators should be able to explore, handle, analyse, summarise, translate, curate their data collections as quickly and efficiently as possible, enabling them to concentrate on producing the required output document or piece of information (Schneider et al., 2016; Bourgonje et al., 2016; Srivastava et al., 2016). The brief description given above is what we perceive as the core of any content curation system. In the highly dynamic legal domain we face the additional challenge of new decisions and dynamic updates where the law, or its interpretation, can change significantly with every court case.

Legal documents have a certain set of key characteristics (van Opijnen and Santos, 2017): high volume, extensive document length, very specific (internal) structure, heterogeneity of types, self-contained documents, hierarchy, temporal aspects, legal terminology, multilingualism, multi-jurisdictionality and, crucially, importance and abundance of citations and cross-references. All of these features make documents from the legal domain highly interesting and also challenging objects for a digital curation system. The idea is to analyse the documents automatically in order to provide added value based, among others, on semantically enriched documents – an important prerequisite for providing suitable curation services in different use cases. This is one of the objectives of the project LYNX ("Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe"), a three-year project, funded by the European Union, that consists of a consortium of ten partners.[1] Lynx aims to create a knowledge graph of legal and regulatory data towards compliance, in which heterogeneous data sources from different jurisdictions, languages and orders are aggregated and interlinked by a collection of advanced analysis and curation services.

In this article we provide an overview of the current state-of-the-art in curation technologies in the legal domain, concentrating on the following three questions:

1. What kind of (semantic) technologies are currently being used in production systems and research prototypes in the field of smart digital services for legal data, legal documents, etc.?
2. What kind of features and functionalities are currently explored in research labs and what is actually being used in terms of novel technologies?
3. What are the most important open research questions in Natural Language Processing and Language Technology for the legal domain, NLP for legal documents, processing legal information, automatically understanding and machine-reading the law, etc.?

The main contribution of this paper is a detailed description of previous research efforts and commercial tools, such as

[1] http://www.lynx-project.eu

curation systems and technologies, the application of curation technologies to a new, concrete and specific domain, i. e., legal information systems. The remainder of this article is structured as follows. Section 2. provides a summary of curation systems in the legal domain, both commercial and prototypical. In Section 3., an overview of important research areas in the legal domain with regard to Natural Language Processing (NLP, Section 3.1.) and semantic technologies (Section 3.2.) is presented. Section 4. concludes the article.

2. Curation Systems in the Legal Domain

Even if they do not use this specific name, curation technologies have been in use in several different domains, including the legal area. Here, the uptake has been a bit slower than in other domains because, in many countries, collections of legal documents and data sets have been monopolised by commercial enterprises which means that there are access restrictions on multiple levels.

2.1. Commercial Systems and Services

One of the most visible companies in the area of semantic technologies and services for the legal domain is LexisNexis.[2] Their system is the market leader and offers services for the legal domain, such as legal research, practical guidance, company research and media-monitoring solutions, intellectual property, litigation strategy and discovery, practice and legal department management as well as compliance and due diligence among others.

Also visible in the legal area is WestLaw, an online service that allows legal professionals to find and consult the needed legal information.[3] Developed by Thomson Reuters, one of the goals of Westlaw is to enable professionals to put together the strongest argument possible.

Apart from these two providers, there are other smaller companies and services that offer legal research solutions and analytic environments, such as RavelLax,[4] which "provides services designed to help legal professionals draw insights and connections using advanced analytical algorithms", or Lereto[5], offering tools for legal document processing. A commercial search engine for legal documents, iSearch, is a service offered by LegitQuest.[6]

The Casetext CARA Research Suite allows uploading a brief and then retrieving, based on its contents, useful case law.[7] In its own words, CARA is an AI-backed automated research assistant, empowering litigators to better serve their clients through advanced information services, supported through technology, and expert analysis from the legal community. CARA's contextual search to help litigators get the answers they need fast so they can spend more time on higher-value work. The company is comprised of lawyers, data scientists, engineers, and designers helping attorneys better to represent their clients.

In addition to these products and services there is also a growing number of startup companies active in the legal domain – from applying AI techniques for automatically analysing large amounts of documents to better supporting communication among law firms, clients and stakeholders.

2.2. Research Prototypes

While there are several research prototypes that can be considered "curation systems for the legal domain", most of the documented systems were developed in the 1990s under the umbrella of Computer Assisted Legal Research – CALR (Span, 1994). In the following we briefly review several of these systems.

Most prototypes we have been able to find in the literature are not curation systems per se (meaning, in the sense described above) but systems that offer a very specific functionality that a legal document curation system (in our sense) would offer together with many other functionalities. One example is the open source software for the analysis and visualisation of networks of Dutch case law presented by (van Kuppevelt and van Dijck, 2017). This technology assists in answering legal research questions by means of determining relevant precedents (analysing the citation network of case law), comparing them with those identified in the literature, and determining clusters of related cases. Another prototype extracting references from legal documents is described by (Agnoloni et al., 2017). They introduce a framework for the extraction of legal references from case-law of European Member States based on an approach applicable to multiple languages and jurisdictions, helping national data providers to reduce the effort needed to develop their own extraction solution. (Gifford, 2017) propose a search engine for legal documents where arguments are extracted from appellate cases and are accessible either through selecting nodes in a litigation issue ontology or through relational keyword search.

A relevant curation prototype is Lucem (Bhullar et al., 2016), a web-based system that provides a solution for obtaining legal information in an accessible and intuitive way. The system tries to mirror the way lawyers approach legal research, developing visualisations that provide lawyers with an additional tool to approach their research results. Eunomos is a curation prototype that uses NLP techniques to semi-automate the construction and analysis of knowledge. This legal knowledge management service enables users to view legislation from various sources and to find the relevant definitions and explanations of legal concepts in a given context (Boella et al., 2012). Functionalities included are the ability to view legislation at European, national and regional level, links between different parts of legislation, lists of similar legislation, a mechanism for classifying norms in user-defined categories and a notification service that alerts users when a newly downloaded legislation appears.

3. Important Research Areas

To identify current research strands and trends we checked the scientific programme of the most relevant conferences in the area to identify common topics. The conferences in

[2] https://www.lexisnexis.com

[3] http://legalsolutions.thomsonreuters.com/law-products/westlaw-legal-research/

[4] http://ravellaw.com

[5] https://www.lereto.at

[6] https://www.legitquest.com

[7] https://casetext.com

question are AI4J[8], JURIX[9], JURISIN[10] and ICAIL (International Conference on AI and Law)[11].
There are several overarching topics that are recurrent among the conferences listed above. These are reasoning and inference, argumentation extraction, evidential reasoning, legal interpretation, decision making, extraction of structure and connections of legal texts and rules, annotation, information retrieval and discovery, text classification, summarisation, translation, linked data and open data, knowledge acquisition, natural language processing, legal knowledge representation, including legal ontologies and common sense knowledge.

3.1. Natural Language Processing for the Legal Domain

Within the broad field of NLP, research currently focuses upon the topics briefly reviewed below.

3.1.1. Citation Analysis

Almost all types of documents that belong to the legal domain refer to laws, paragraphs, rules, correspondence or arbitrary other documents, which is why citation and cross-reference analysis is an almost mandatory step in any processing pipeline. There is a multitude of approaches focused on citation analysis, addressing the challenge from different perspectives and with different methods. There appear to be two major directions, i. e., applying network analysis to citations (Zhang and Koppaka, 2007), (Winkels et al., 2011), (Lupu and Voeten, 2012), (Neale, 2013) and classification systems estimating the status of the cited case (Galgani et al., 2015). (Zhang and Koppaka, 2007) develop a semantics-based legal citation network, which is a tool that extracts and summarises citation information into a network, allowing the users to navigate the citation network and to study how citations are interrelated and how legal issues have evolved in the past. LEXA (Galgani et al., 2015) is a system that relies on Ripple Down rules approach to identify citations within the "distinguished" class. This category is generally best linguistically signaled and is therefore suitable for achieving high precision and recall.

3.1.2. Argument Extraction and Mining

Like citation analysis, argument extraction is an important part in the understanding of legal documents. Recognising the arguments used in case law is vital for identifying similar arguments in other documents and to predict possible outcomes of a specific case. Many different approaches have been applied, such as statistical methods over annotated corpora, used by (Moens et al., 2007) to automatically detect sentences that are part of a legal argument. (Cabrio et al., 2016) summarise current trends in argumentation mining and discuss future challenges.

3.1.3. Reasoning

Logical reasoning is, naturally, an important part of a legal expert's day-to-day work, which is why there have been several attempts at performing automatic reasoning techniques based on a specific set of information or knowledge provided. As stated by (Vlek et al., 2014), there are three main approaches to performing reasoning with or over evidence: argumentative, narrative and probabilistic approaches. (Vlek et al., 2014) combine these approaches to form a design method for constructing a Bayesian network based on narratives. An extension of this work, (Vlek et al., 2016), proposes a method combining a probabilistic approach with a narrative approach to reasoning with legal evidence. Whereas a Bayesian network is a popular tool for analysing parts of a case, the advantage of a narrative approach is that it provides the global perspective on the case as a whole. (Verheij, 2017) use a different approach, in which they propose a formalism, in which the validity of arguments is defined in terms of case models.

3.1.4. Summarisation

Many researchers emphasise that the average length of documents in the legal domain is rather extensive, plus, one case usually comprises many different documents of several different types (van Opijnen and Santos, 2017). This is why it is, for legal experts, a difficult challenge to acquire first an overview and then detailed knowledge of the content of all of these documents. Automatic summarisation could help lawyers to familiarise themselves quickly and efficiently with a new set of documents on a specific case.
A common approach in automatic summarisation, also used in the legal domain, is sentence classification and sentence ranking. The SUM project (Grover et al., 2003) applied automatic summarisation to the legal domain by means of sentence classification based on the sentences' rhetorical roles. They explored the relationship between linguistic features and argumentative roles in order to classify sentences. Another approach using sentence classification is the prototypical summarisation system, LetSum (Legal text Summarizer) (Farzindar and Lapalme, 2004). It classifies sentences into four themes: introduction, context, juridical analysis and conclusion. Summaries are generated in four steps: thematic segmentation, filtering to eliminate unimportant quotations and noise, selection of candidate units and generation of the summary.
(Polsley et al., 2016) use a sentence classification method, that is based on word frequency augmented with domain-specific knowledge. They implemented a tool called CaseSummarizer, whose processing pipeline consists of three steps: preprocessing, scoring of sentence relevance, and domain processing. They present summaries to the user through a multi-faceted interface with abbreviations, significance heat maps, and other flexible controls.
(Yousfi-Monod et al., 2010) use supervised machine learning for summarising legal documents based on a Naive Bayes classifier. They use a set of surface, emphasis, and content features. For the training of these machine learning based approaches, annotated data is needed. For data acquisition, a corpus of UK House of Lords judgments[12] is created (Grover et al., 2004). It contains three layers: rhetorical status annotation, detailed linguistic markup, and relevance annotation.

[8] http://www.ai.rug.nl/~verheij/AI4J/
[9] https://jurix2017.gforge.uni.lu
[10] http://research.nii.ac.jp/~ksatoh/jurisin2017/
[11] https://nms.kcl.ac.uk/icail2017/
[12] http://www.ltg.ed.ac.uk/SUM/

3.1.5. Information Retrieval

Given the large amount of information handled in legal cases, it is essential to have good search and retrieval capabilities. Many researchers focus on improving search engines in this domain. Two approaches to legal IR, based on manual knowledge engineering (KE) and NLP, are presented and compared in (Schafer and Maxwell, 2008). They concluded that IR based solely on KE is not sustainable in the long run.

The ontology-based IR system EgoIR is presented by (Gómez-Pérez et al., 2006). It aims to retrieve government documents in a timely and accurate manner. Ontologies are used for two purposes: to guide users to the legal terms, enabling them to avoid mistakes at constructing a query and to improve interoperability in legal applications.

Apart from the topics mentioned above, there are many additional questions being investigated. A sentence classification approach in the legal domain is presented by (van Opijnen and Santos, 2017; Shulayeva et al., 2017), where a set of linguistic features (part of speech tags, unigrams, dependency pairs, length of the sentence, position in the text and cita, which indicates whether there is a citation instance in the sentence) is extracted using NLTK (Loper and Bird, 2002) and CoreNLP (Manning et al., 2014), later on to classify the sentence with WEKA (Hall et al., 2009).

3.2. Semantic Technologies

The legal domain is characterised by having an incredibly large number of established terms. There have been several attempts to organise these terms in ontologies and semantic systems, which is why there is a lot of research related to semantic technologies including ontology bootstrapping and generation, ontology population and the use of ontologies for IR and semantic annotation.

Some common approaches for the population of ontologies use standard NLP tools (such as TreeTagger, GATE, YaTeA, etc.) or ontology learning tools (Lehmann and Voelker, 2014) (such as OntoGen, ASIUM, Text-To-Onto, Text2Onto and TERMINAE). (El Ghosh et al., 2017) use the methodology Terminae (Aussenac-Gilles et al., 2000) for legal ontology population based on two approaches: top-down and botton-up. The bottom-up approach uses linguistic information (using YaTeA) for extracting features (concepts and relations) and to convert them into domain-specific ontologies. The top-down approach is based on the definition and (partial) reuse of existent ontologies.

(Francesconi et al., 2010) perform legal knowledge acquisition based on top-down and bottom-up approaches. They present a methodology for multilingual legal knowledge acquisition and modeling. The top-down approach is the definition of the conceptual structure of the legal domain on the basis of expert judgments. This structure is language-independent, modeled as an ontology, and can be aligned with other ontologies that capture similar or complementary knowledge, in order to provide a wider conceptual embedding. The bottom-up approach is a linguistic text-based population of conceptual structures using semi-automatic NLP techniques, which maximise the completeness and domain-specificity of the resulting knowledge. A different approach using semantic information is SALEM, the automatic enrichment of legal texts with semantic annotations (Biagioli et al., 2005). SALEM is an NLP system for the classification and semantic enrichment of articles of law. The enrichment helps effectively index and retrieve legal documents. It classifies paragraphs according to their regulatory content and extracts relevant text fragments corresponding to specific semantic roles. The ontology distinguishes three categories: obligations, definitions and modifications.

In addition, knowledge representation is an important topic in the legal domain. (Perinan-Pascual and Arcas-TÃ, 2014) define a knowledge representation model inside the frame of FunGramKB, a lexical-conceptual multilingual knowledge graph. The generation of the knowledge graph is divided into five steps: (1) definition of filters, (2) corpus indexing, (3) n-gram and statistics extraction, (4) terms identification and (5) corpus validation. A proof of concept for the ontological representation of normative requirements as Linked Data on the Web is proposed by (Gandon et al., 2017), who present an extension of the LegalRuleML ontology to model normative requirements and rules.

Furthermore, there are multiple available ontologies for the legal domain. (Breuker et al., 2009) provide a corresponding list. Several examples are FOLaw – Functional Ontology of Law (Valente et al., 1994), OPLK – Ontology of the Professional Legal Knowledge (Benjamins et al., 2004), Jur-Wordnet (Gangemi et al., 2003), DALOS (Francesconi and Tiscornia, 2008) and OPJK – Ontology of Professional Judicial Knowledge (Casanovas et al., 2009).

An ontology learning system (T2K) that includes NLP tools, statistical text analysis and machine learning is used by (Lenci et al., 2007). Their approach allows the dynamic integration of new modules to provide an incremental representation of the content of vast repositories of unstructured documents. They also include bootstrapping techniques to develop more sophisticated levels of content representation starting from knowledge-poor language tools.

(Casanovas et al., 2016) provide an overview of a special issue of the journal Semantic Web, aimed at the legal domain, summarising research carried out in the legal domain in the last 15 years. They emphasise five ontology definition and generation approaches: (1) an OWL ontology making it possible to describe a judge's interpretations of the law while engaging in the legal reasoning on which basis a case is adjudicated (Ceci and Gangemi, 2016). (2) an OWL ontology, framed in CELLAR, for describing normative provisions to enable advanced access to legal documents (Francesconi, 2016). (3) the LOTED2 ontology for the representation of European public procurement notices, enabling legal reasoning (Distinto et al., 2016). (4) the PPROC ontology, which enables the description of procurement processes and contracts (Muñoz-Soro et al., 2016). (5) the MPEG-21 Media Contract Ontology, which enables the description of contracts dealing with rights to multimedia assets and with any content protected by intellectual property (Rodríguez-Doncel et al., 2016).

4. Conclusions

This article presents an overview of approaches that are highly relevant for the development of a system for the

curation of content and documents from the legal domain, aimed at the construction and utilisation of legal knowledge graphs. The article is structured into two parts. The first part presents existing curation systems, both research prototypes and commercial systems. The commercial market is currently dominated by two major players and several smaller companies (including several startups) that try to penetrate the market.

In our desk research we have found only very few non-commercial and/or free systems – prototypes are rarely used outside the laboratory. Among the main reasons for this situation is the fact that many important data and document collections are controlled by commercial companies and because of privacy and data protection issues. These issues are so severe that the situation is unlikely to change. Legal document collections contain very high numbers of names and events that many would not want and proabably also cannot be made public – the publication of a collection in the legal domain that has previously been anonymised has little or no value for the development of functional technologies. Despite a very high amount of research activity in the legal domain, this effort does not immediately translate into prototypes or free systems that are in widespread use.

The second part summarises current research strands in this area and analyses the main conferences on the topic. These research lines are divided into two main groups: NLP and semantic approaches. Regarding NLP, there are several interesting topics for the legal domain, such as reasoning, argument mining, summarisation and document linking. In the case of semantic approaches, the variety of topics is not as rich, and there are mainly three main topics: knowledge base construction, mainly based on existing ontologies, knowledge base population, mainly based on (semi-)automatic NLP and ontology learning, and semantic enrichment of documents.

This contribution has been prepared under the umbrella of the EU project LYNX, which has started in December 2017. The analysis of available technologies and current research strands will inform the design and development of a system, which makes use of curation technologies for the construction and utilisation of a legal knowledge graph.

5. Acknowledgements

The project LYNX has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 780602. More information is available online at http://www.lynx-project.eu.

6. Bibliographical References

Agnoloni, T., Bacci, L., Peruginelli, G., van Opijnen, M., van den Oever, J., Palmirani, M., Cervone, L., Bujor, O., Lecuona, A. A., García, A. B., Caro, L. D., and Siragusa, G. (2017). Linking european case law: BO-ECLI parser, an open framework for the automatic extraction of legal links. In Wyner and Casini (Wyner and Casini, 2017), pages 113–118.

Aussenac-Gilles, N., Biébow, B., and Szulman, S. (2000). Corpus analysis for conceptual modelling.

Benjamins, V. R., Contreras, J., Casanovas, P., Ayuso, M., Becue, M., Lemus, L., and Urios, C. (2004). Ontologies of professional legal knowledge as the basis for intelligent it support for judges. *Artificial Intelligence and Law*, 12(4):359–378.

Bhullar, J., Lam, N., Pham, K., Prabhakaran, A., and Santillano, A. J., (2016). *Lucem: A Legal Research Tool*. Number 63. Computer Engineering Senior Theses.

Biagioli, C., Francesconi, E., Passerini, A., Montemagni, S., and Soria, C. (2005). Automatic semantics extraction in law documents. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, ICAIL '05, pages 133–140, New York, NY, USA. ACM.

Boella, G., di Caro, L., Humphreys, L., Robaldo, L., and van der Torre, L. (2012). Nlp challenges for eunomos, a tool to build and manage legal knowledge.

Bourgonje, P., Schneider, J. M., Rehm, G., and Sasaki, F. (2016). Processing Document Collections to Automatically Extract Linked Data: Semantic Storytelling Technologies for Smart Curation Workflows. In Aldo Gangemi et al., editors, *Proceedings of the 2nd International Workshop on Natural Language Generation and the Semantic Web (WebNLG 2016)*, pages 13–16, Edinburgh, UK, September. The Association for Computational Linguistics.

Breuker, J., Casanovas, P., Klein, M. C. A., and Francesconi, E. (2009). *Law, ontologies and the Semantic Web: Channelling the legal information flood*. IOS Press, Amsterdam.

Cabrio, E., Hirst, G., Villata, S., and Wyner, A. (2016). Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments (Dagstuhl Seminar 16161). *Dagstuhl Reports*, 6(4):80–109.

Casanovas, P., Casellas, N., and Vallbé, J.-J. (2009). An ontology-based decision support system for judges. In *Proceedings of the 2009 conference on Law, Ontologies and the Semantic Web: Channelling the Legal Information Flood*, pages 165–175. IOS Press.

Casanovas, P., Palmirani, M., Peroni, S., van Engers, T. M., and Vitali, F. (2016). Semantic web for the legal domain: The next step. *Semantic Web*, 7(3):213–227.

Ceci, M. and Gangemi, A. (2016). An owl ontology library representing judicial interpretations. *Semantic Web*, 7(3):229–253.

Distinto, I., d'Aquin, M., and Motta, E. (2016). Loted2: An ontology of european public procurement notices. *Semantic Web*, 7(3):267–293.

El Ghosh, M., Naja, H., Abdulrab, H., and Khalil, M. (2017). Towards a legal rule-based system grounded on the integration of criminal domain ontology and rules. *Procedia Computer Science*, 112:632–642.

Farzindar, A. and Lapalme, G., (2004). *LetSum, an automatic Legal Text Summarizing system*, pages 11–18. IOS Press, Berlin, dec.

Francesconi, E. and Tiscornia, D. (2008). Building semantic resources for legislative drafting: The dalos project.

Francesconi, E., Montemagni, S., Peters, W., and Tiscornia, D., (2010). *Integrating a Bottom–Up and Top–Down Methodology for Building Semantic Resources for*

the Multilingual Legal Domain, pages 95–121. Springer Berlin Heidelberg, Berlin, Heidelberg.

Francesconi, E. (2016). Semantic model for legal resources: annotation and reasoning over normative provisions. *Semantic Web*, 7(3):255–265.

Galgani, F., Compton, P., and Hoffmann, A. (2015). Lexa: Building knowledge bases for automatic legal citation classification. *Expert Systems with Applications*, 42(17):6391–6407.

Gandon, F., Governatori, G., and Villata, S. (2017). Normative requirements as linked data. In Wyner and Casini (Wyner and Casini, 2017), pages 1–10.

Gangemi, A., Sagri, M.-T., and Tiscornia, D. (2003). Metadata for content description in legal information. In *Procs. of LegOnt Workshop on Legal Ontologies*.

Gifford, M. (2017). Lexridelaw: an argument based legal search engine. In *ICAIL '17*.

Gómez-Pérez, A., Ortiz-Rodriguez, F., and Villazón-Terrazas, B. (2006). Ontology-based legal information retrieval to improve the information access in e-government. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, pages 1007–1008, New York, NY, USA. ACM.

Grover, C., Hachey, B., Hughson, I., and Korycinski, C. (2003). Automatic summarisation of legal documents. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, ICAIL '03, pages 243–251, New York, NY, USA. ACM.

Grover, C., Hachey, B., Hughson, I., and Place, B. (2004). The holj corpus: supporting summarisation of legal texts. In *In Proceedings of the 5th International Workshop on Linguistically Interpreted Corpora*.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November.

Lehmann, J. and Voelker, J. (2014). An introduction to ontology learning. *Perspectives on Ontology Learning. IOS Press, Amsterdam, The Netherlands*.

Lenci, A., Montemagni, S., Pirrelli, V., and Venturi, G. (2007). Nlp-based ontology learning from legal texts. a case study. In Pompeu Casanovas, et al., editors, *LOAIT*, volume 321 of *CEUR Workshop Proceedings*, pages 113–129. CEUR-WS.org.

Loper, E. and Bird, S. (2002). Nltk: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*, ETMTNLP '02, pages 63–70, Stroudsburg, PA, USA. Association for Computational Linguistics.

Lupu, Y. and Voeten, E. (2012). Precedent in international courts: A network analysis of case citations by the european court of human rights. *British Journal of Political Science*, 42(2):413–439.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

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.

Muñoz-Soro, J. F., Esteban, G., Corcho, O., and Serón, F. (2016). Pproc, an ontology for transparency in public procurement. *Semantic Web*, 7(3):295–309.

Neale, T. (2013). Citation analysis of canadian case law. *J. Open Access L.*, 1:1.

Neudecker, C. and Rehm, G. (2016). Digitale Kuratierungstechnologien für Bibliotheken. *Zeitschrift für Bibliothekskultur 027.7*, 4(2), November.

Perinan-Pascual, C. and Arcas-TÃ, F. (2014). La ingenierÃa del conocimiento en el dominio legal: La construcciÃ de una OntologÃa SatÃ©en FunGramKB. *Revista signos*, 47:113 – 139, 03.

Polsley, S., Jhunjhunwala, P., and Huang, R. (2016). Casesummarizer: A system for automated summarization of legal texts. In *COLING*.

Rehm, G., Schneider, J. M., Bourgonje, P., Srivastava, A., Fricke, R., Thomsen, J., He, J., Quantz, J., Berger, A., König, L., Räuchle, S., Gerth, J., and Wabnitz, D. (2018). Different Types of Automated and Semi-Automated Semantic Storytelling: Curation Technologies for Different Sectors. In Georg Rehm et al., editors, *Language Technologies for the Challenges of the Digital Age: 27th International Conference, GSCL 2017, Berlin, Germany, September 13-14, 2017, Proceedings*, number 10713 in Lecture Notes in Artificial Intelligence (LNAI), pages 232–247, Cham, Switzerland, January. Gesellschaft für Sprachtechnologie und Computerlinguistik e.V., Springer. 13/14 September 2017.

Rodríguez-Doncel, V., Delgado, J., Llorente, S., Rodríguez, E., and Boch, L. (2016). Overview of the mpeg-21 media contract ontology. *Semantic Web*, 7(3):311–332.

Schafer, B. and Maxwell, T., (2008). *Concept and Context in Legal Information Retrieval*, pages 63–72. Frontiers in Artificial Intelligence and Applications. IOS Press.

Schneider, J. M., Bourgonje, P., Nehring, J., Rehm, G., Sasaki, F., and Srivastava, A. (2016). Towards Semantic Story Telling with Digital Curation Technologies. In Larry Birnbaum, et al., editors, *Proceedings of Natural Language Processing meets Journalism – IJCAI-16 Workshop (NLPMJ 2016)*, New York, July.

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

Span, G. (1994). Lites: An intelligent tutoring system shell for legal education. *International Review of Law, Computers & Technology*, 8(1):103–113.

Srivastava, A., Sasaki, F., Bourgonje, P., Moreno-Schneider, J., Nehring, J., and Rehm, G. (2016). How to Configure Statistical Machine Translation with Linked Open Data Resources. In Joao Esteves-Ferreira, et al., editors, *Proceedings of Translating and the Computer 38 (TC38)*, pages 138–148, London, UK, November. Editions Tradulex.

Valente, A., Breuker, J., et al. (1994). Ontologies: The missing link between legal theory and ai & law. *Legal knowledge based systems JURIX*, 94:138–150.

van Kuppevelt, D. and van Dijck, G. (2017). Answering legal research questions about dutch case law with network analysis and visualization. In Wyner and Casini (Wyner and Casini, 2017), pages 95–100.

van Opijnen, M. and Santos, C. (2017). On the concept of relevance in legal information retrieval. *Artificial Intelligence and Law*, 25(1):65–87, Mar.

Verheij, B. (2017). Proof with and without probabilities. *Artificial Intelligence and Law*, 25(1):127–154, Mar.

Vlek, C. S., Prakken, H., Renooij, S., and Verheij, B. (2014). Building bayesian networks for legal evidence with narratives: a case study evaluation. *Artificial Intelligence and Law*, 22(4):375–421, Dec.

Vlek, C. S., Prakken, H., Renooij, S., and Verheij, B. (2016). A method for explaining bayesian networks for legal evidence with scenarios. *Artificial Intelligence and Law*, 24(3):285–324, Sep.

Winkels, R., Ruyter, J. d., and Kroese, H. (2011). Determining authority of dutch case law.

Adam Z. Wyner et al., editors. (2017). *Legal Knowledge and Information Systems - JURIX 2017: The Thirtieth Annual Conference, Luxembourg, 13-15 December 2017*, volume 302 of *Frontiers in Artificial Intelligence and Applications*. IOS Press.

Yousfi-Monod, M., Farzindar, A., and Lapalme, G. (2010). Supervised machine learning for summarizing legal documents. In Atefeh Farzindar et al., editors, *Advances in Artificial Intelligence*, pages 51–62, Berlin, Heidelberg. Springer Berlin Heidelberg.

Zhang, P. and Koppaka, L. (2007). Semantics-based legal citation network. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, ICAIL '07, pages 123–130, New York, NY, USA. ACM.