Enriching Legal Knowledge Through Intelligent Information Retrieval Techniques: A Review

Marco Gomes[1,2(✉)] ⬥, Bruno Oliveira[2] ⬥, and Cristóvão Sousa[3] ⬥

[1] CIICESI/ESTG, Polytechnic Institute of Porto, Felgueiras, Portugal
mfg@estg.ipp.pt
[2] ALGORITMI Centre, University of Minho, Guimaraes, Portugal
bmo@estg.ipp.pt
[3] INESC TEC, Porto, Portugal
cds@estg.ipp.pt

Abstract. This work aims to systematize the knowledge on emerging Intelligent Information Retrieval (IIR) practices in scenarios whose context is similar to the field of tax law. It is a part of a project that covers the emerging techniques of IIR and its applicability to the tax law domain. Furthermore, it presents an overview of different approaches for representing legal data and exposes the challenging task of providing quality insights to support decision-making in a dedicated legal environment. It also offers an overview of the related background and prior research referring to the techniques for information retrieval in legal documents, establishing the current state-of-the-art, and identifying its main drawbacks. A summary of the most appropriate technologies and research approaches of the technologies that apply artificial intelligence technology to help legal tasks is also depicted.

Keywords: Legal knowledge · Information Retrieval · Artificial intelligence · Legal domain

1 Introduction

Any technical domain, be it legal, tax, computational, or otherwise, is characterized by a highly specialized discourse using its terminology and style in the textual codification of the underlying themes. Thus, it is accepted that understanding a technical discourse requires a certain level of literacy suitable for the identification and analysis of the discourse. However, interpretation, as an act of inference about what is written from the perspective of its application to a given context, may require more than understanding the terminological and conceptual framework of the domain, implying a set of epistemic practices [7] from which new structures of knowledge emerge that tend to facilitate the process of interpretation. These practices contribute to meaning construction through collaborative activities, where experience and knowledge representation models

are crucial factors. Notwithstanding the importance of tacit knowledge and the models that facilitate its explanation, the interpretation process also depends on the quality of information and mechanisms of analysis and research of interrelated content, that is, identification of relationships between different types or categories of information.

The field of taxation and tax law is complex by nature, typically consisting of large volumes of highly technical textual information and categorized by codes (VAT, IRS, ...) and laws. Moreover, the different sources of information that make up this technical domain are characterized by a high degree of interdependence and dynamics, where inter and intratextual references and updates are frequent. Therefore, the correct and precise treatment of this information, its temporal location, and its availability in an organized and adequate way to the needs of different users can create added value for professionals and organizations, but, above all, it can serve to avoid legal disputes or promote necessary tools for support in resolving these, whether with the Tax Authority or with the Courts. In this work, the knowledge management process is assumed as a critical process, representing a key factor of organizational performance and an essential tool for competitiveness [26]. The availability and access to updated legislative information, complemented with objective and timely explanations, allows the users to comply with their tax obligations, promote tax efficiency, and reduce tax burdens. However, this process often results in querying data from multiple sources of information, where each document can handle updates and cross-references, which increases process complexity. In many cases, excessive, mismatched, and disorganized information can affect the entire data exploration process, which can lead to wrong decisions if improperly interpreted. For this, it is necessary to identify relevant information (Information Retrieval - IR), considering the existence of typically unstructured data and in large quantities [32]. However, considering the technological evolution in knowledge base management and new intelligent models of collection, processing, analysis, and representation of information, several of the existing gaps can be adequately explored. In this sense, the consortium presents this project, presenting complementary capabilities in the areas involved and proposing an intelligent solution to the problems encountered. This document contains four sections and is organized as follows. In the first section, the main problem is contextualized and described as the principal motivation for this work. Next, the following section (Sect. 2) presents an overview of different approaches for representing legal data and exposes the challenging task of providing quality insights to support decision-making in a dedicated legal domain. Section 3 proposes an overview of the related background and prior research referring to the techniques for information retrieval in legal documents, establishing the current state-of-the-art, and identifying its main drawbacks. A summary of the most appropriate technologies and research approaches for developing the research work is also depicted. Moreover, this section also gives an overview of the technologies that apply artificial intelligence technology to help legal tasks. Finally, the conclusion provides a summary of contributions to the main points of this work.

2   Legal Knowledge Representation

Legal data is typically represented using natural language framed within a specific domain and context. For that reason, expressing and sharing legal knowledge so that computers can explore is a challenging task. Therefore, the emergence of disruptive techniques for handling, modeling, and using data became very popular in the last decade, particularly with the advent of Artificial intelligence techniques used to implement Machine Learning algorithms that have enhanced the development of expert systems. However, systems cannot analyze and provide quality insights to support decision-making in a dedicated legal domain without a proper representation of knowledge. As stated by Ramakrishna and Paschke [24], knowledge represents a relation between a knower and a proposition, expressed by a declarative sentence. In [3], three types of knowledge are discussed: experiential knowledge is acquired based not only on experiences but also connects to the environment through the sensory before being processed by the individual, which means that the same experience may result in different knowledge (since it is associated with previous experiences and knowledge); the skills that represent the know-how resulting from doing specific actions; and claims that define what is known based on explicit knowledge (provided by, for example, by books or legislations). Each type is interconnected and is particularly relevant in the legislation domain since the same explicit knowledge (law) and its application by judiciary entities is influenced by experimental and know-how knowledge. In the last years, several research works have proposed technologies, methods, and languages to identify requirements and represent the specificities of the legal domain. The primary purpose is to capture information that can be processed and shared by computers. In [33] a categorization based on generations for describing the efforts to provide access to legal electronic data is presented. The first generation refers to a representation closer to word processing and database models; the second generation refers to the adoption of metadata for structuring and modeling, the third generation focuses on grammar for preserving consistency over time and the ability to share and integrate new knowledge to existing one (e.g., using ontologies). Finally, the four-generation provides prescriptiveness using constraint-based grammar. Some approaches have been used for legal knowledge representation during the last decades. For example, the LEXML [33] is a European model to promote interoperability for legal and legislative data. The LEXML was born from several countries' initiatives to find similarities between the national and international legal systems. However, LEXML is currently only implemented in Brazil and can be framed within the second generation of legal data representation. The EUR-Lex [33] provides European legal documents with Formex, an XML standard used for managing legal (not for representation) data in the EUR-Lex service that provides access to EU legislation such as the case law of the Court of Justice of the European Union, and other EU public documents. EUR-Lex supports multi-language to cover several European languages, supporting law or international agreements. Formex is widely adopted in the European community and defines the logical markup for legal documents. It can be framed in the first

generation of initiatives to standardize legal documents. The CEN MetaLex [2] is an XML standard used to represent sources of law and references to sources of law as CEN MetaLex documents. It provides the interchange of data in a standardized way. In addition, it provides mechanisms to link legal information from various levels of authority, supporting different countries and languages and information exchange and interoperability. These characteristics allow for the classification of this standard as the third generation. Akoma Ntoso [27] is an international technical standard for representing executive, legislative, and judicial documents in a structured way using a domain-specific XML vocabulary. It provides a framework for exchanging parliamentary, legislative, and judiciary documents. In addition, Akoma Ntoso maintains connected standards and languages that provide: document format (for open documents that cover areas such as Parliamentary Debates, Primary Legislation, or Judgements); a model for document interchange (supporting the generation, presentation, accessibility, and description of documents), data schema (all document types share the same basic structures), metadata schema and ontology (the ontology is designed to be extensible to accommodate extra elements and qualifiers to meet specific requirements), and a schema for citation and cross-referencing (relying on a name convention and a reference mechanism to connect a distributed document corpus). Due to these characteristics, Akoma Ntoso was originally a third-generation Initiative for legal knowledge representation. LegalXML[1] approach produces technical standards for structuring legal documents and information using XML, enabling the adoption and convergence of e-business standards for the legal domain. LegalXML has standards to support court documents, legal citations, or transcripts. In addition, it includes the LegalDocML, a legal rule representation language based on Akoma Ntoso for structuring legal content, the Electronic Court Filing for supporting interoperability among electronic courts, the LegalRuleML for representing legal norms and rules, among other technical specifications such as the LegalRuleML for supporting legal arguments representation and evaluation.

Among the presented approaches for representing legal data, the LegalXML initiative covers several aspects of legal knowledge representation [33]:

– Supporting the ability to represent different knowledge aspects with clear and expressive semantics.
– Provide mechanisms for knowledge sharing, reusability, and extensibility.
– Provide support for reasoning and inference over the legal content.
– The capacity to reference other legal documents, which is very useful in the jurisdiction context;
– Defining rules for legally checking
– Capacity to support authoring and to link the law context according to a specific temporal occurrence
– Support to track changes and amendments
– Support directions or injunctions indicate how a language should be used in specific contexts.

---

[1] http://www.legalxml.org/.

The main difference between LegalXML, CEN-Metalex, and Akoma Ntoso is related to the lack of prescriptiveness in CEN-Metalex and the original Akoma Ntoso version. With LegalRuleML, prescriptive statements are modeled by "If" conditions, describing rules application. Not only baseline rules but also the exceptions to the baseline.

3   Approaches for Information Retrieval in Legal Documents

Legal documents represent a complex knowledge composed of lengthy texts that need to be analyzed and interpreted by domain personnel (e.g., lawyers or judges) to extract meaningful information. In several cases, data is extracted with other related documents to select the right documents and for understanding the document content.

A legal document can involve several uses cases (such as contracts, regulations, or privacy documents). They typically involve several entities, relationships between, and the relationships between external documents, such as laws, amendments, or revocations. Thus, the complexity and time spent extracting relevant data in legal documents are challenging and error-prone. Since legal documents are subject to different interpretations, misinterpretations or precision loss are common problems related to text interpretation.

To analyze and reason over the documents, users need expert systems to support decision-making requirements. Due to their nature, extracting, organizing, and interpreting legal documents requires the application of several advanced techniques and algorithms. Techniques such as semantic web, text mining, and NLP (Natural Language Processing) techniques can be used to reveal understandings and patterns that can be used to support decision-making. Moreover, the knowledge extracted from legal documents will be used to support critical decisions related to its applications in judgments or legal decisions. This means the inappropriate legal data handling can lead to disastrous results and can be seen with mistrust from the decision-making personnel [5].

Legal search queries can be framed according to several dimensions (e.g., legal issues, jurisdictions), which imposes the evaluation of the proper algorithm, retrieval and ranking models to effectively extract meaningful data. Legal Information Retrieval (IR) and legal argument data mining represent two typical strategies to extract knowledge from legal documents. Despite focusing on different perspectives, several approaches can be used together [1]. The legal argument retrieval (AR) [1,36,37] use these two techniques for returning arguments and not just documents. In this context, Xu [39] addressed the possibility to automatically generate succinct summaries of legal documents through the identification of legal arguments.

3.1   Information Science

In the field of legal knowledge, there are several contributions not only for knowledge representation, as stated in the previous section, but also for rule

interchange [13]. Furthermore, for judicial interpretation based on domain, conceptualization [8].

The approach presented in [8] connects the knowledge coming from different decisions and highlights similarities and differences between them. The authors introduce JudO, an OWL2 ontology library of legal knowledge that relies on the metadata contained in judicial documents. JudO represents the interpretations performed by a judge while conducting legal reasoning toward adjudicating a case. JudO provides meaningful legal semantics while retaining a strong connection to source documents (fragments of legal texts). This approach detects and models jurisprudence-related information directly from the text and performs shallow reasoning on the resulting knowledge base.

In [14], the authors present a formal model of legal norms modeled in OWL. It is intended for semiautomatic drafting, semantic retrieval, and browsing legislation. Most existing solutions model legal norms by formal logic, rules, or ontologies. The proposed model formally defines legal norms using the elements of legal relations they regulate. The paper presents a formal model of legal norms used to develop expert systems for semiautomatic drafting and semantic retrieval and browsing of legislation.

Semantic web techniques are also used for modeling legal information and reason about related data. These networks represent the relevant entities, their properties, and their relationship considering the legal domain [38]. The research work presented in [9] describes the implementation of a semantic network. The authors implemented an entity recognition task using a NER Tagging tool to identify victims, places, or organizations as entities involved in the related legal case to produce nodes for the knowledge graph. To identify words and their context and their relationship with associated words, Part-of-speech [18] tagging was used for identifying edges (mainly verbs) between the entities previously identified. Additionally, they used an information Extraction tool to identify the relationships between entities from plain text. In [16], the authors present ALDA, a legal cognitive assistant to analyze digital legal documents. They addressed several components, including the development of ontological representation: the extraction of data to create knowledge bases (using text-mining and natural language processing), the cross-referencing between related documents (which results in the development of subgraphs), and the use of deep learning to extract semantically similar legal entities and terms.

3.2   Artificial Intelligence

Almost since 1970, it has been noticed the "information crisis in law" (an ever-increasing amount of legal data that is being generated and not plenty or properly used) encouraging the development of legal Information Retrieval systems [28]. Also, almost since 2007, the "natural language barrier" [23] has been discussed as a barrier that hinders artificial intelligence in the legal domain. However, even using artificial intelligence and other computing approaches almost since the 1970 s,s, there has been no breakthrough in such matter [25]. Most of the research performed on the use of Artificial Intelligence in the legal

domain appears to relapse into three main categories: Computer-Aided Reasoning (CAR), Knowledge-Based Systems (KBS), and Legal Language Processing (LLP). Those categories are highly coupled since CAR needs KBS, which relies on LLP, so those groups are a matter of research focus. This is suggested to be called a LIIS. Considering a legal system stack, CAR appears to be performed most by Case-based Reasoning. KBS seems to be built in most cases using one or another ontology strategy; on the other hand, on LLP, a commonly used approach appears not yet to exist. Regarding application, most of the research seems to aim one way or another at Legal Information Retrieval. Datastore and querying are ancient human needs performed by cataloging and retrieving systems [25,31]. Cataloging systems are used to handle structured data and retrieve the system's unstructured data. The first is related to SQL (or relational) within a computer science scope, yet the second is NoSQL (or non-relational) technology. Shall highlight that it is not a matter of choice but the impossibility to structure some types of information [22]. Moreover, Computer-aided Reasoning cannot only rely on formal logic [6]. It must also rely on a knowledge-based system that needs a legal language processor to be composed [31]. The stack sits on legal language processing. While it is not properly settled, it will not be possible to reach the expected breakthrough, called the "natural language barrier" [31]. The stack sits on legal language processing. While it is not properly settled, it will not be possible to reach the expected breakthrough, called the "natural language barrier" [23]. In this sense, it was realized that most natural language processing approaches do not adequately suit legal texts due to their idiosyncrasies, and legal language processing comes into use [6].

Specific issues of legal texts include sentences twice longer, and prepositional chaining is a third deeper than those used in newspapers [6]. Also, legal terms may present proper semantics being different from the regular use. Moreover, the law of each country is written in its official language and considers a particular legal structure [19]. Because of that, those concerns must be dealt with locally and hinders international cooperation. In other words, the "natural language barrier" in the legal field includes lexical, syntactic, semantic, pragmatic, and idiom obstacles. But once surpassed, it is believed that conventional approaches of Computer-aided Reasoning and knowledge-based systems are feasible to be used [23]. Meanwhile, researchers' efforts over the years led to tremendous advances in applying artificial intelligence (especially NPL) technology to help legal tasks. Nowadays, the Legal Information Retrieval datasets, including COLIEE, CaseLaw, and CM. Both COLIEE and CaseLaw are involved in retrieving the most relevant articles from a large corpus, while data examples in CM give three legal documents for calculating similarity. Moreover, these datasets provide benchmarks for the studies of LegalIR.

The difficulties are not restricted to legal language processing, although it typifies a barrier. Information retrieval itself presents several issues to be handled. The index formation (that can be static or dynamic), according to the complexity of data and the index itself, may lead to computational complexity (time and space) issues requiring some cluster processing and other big-data

approaches [11,22]. A significant data source may return more information than the user can handle, requiring a scoring system to rank the fetched documents. Due to the diversity of possible approaches, shall build an information retriever as specific as possible given a domain and user [25] within a user-centered design. Also another peculiarity of legal retrieving is the need to encompass juridical dynamics and context. CBR is a problem-solving method that addresses new problems by remembering and adapting solutions previously used to solve similar issues [17]. CBR is based on two tenets to understand intelligence: problems tend to occur repeatedly, and similar problems have similar solutions [20]. Also, the answer to each new problem in CBR becomes the basis for a new case, being learned and stored for potential reuse in the future [35].

The typical CBR cycle is composed of [17,34]: Retrieve, Reuse, Revise and Retain. The retrieving step is responsible for determining which case is most similar to the new problem. Two different approaches can calculate cases' similarities: the K-Nearest neighbours approach - which uses a weighted sum of features to identify the similarities - and the template retrieval approach - which returns all cases that fit within specific parameters. During the CBR cycle, some stages usually have human interaction. For example, while it may automate case retrieval and reuse, case revision and retention are performed by human experts. Reuse is the process that receives the retrieved cases and makes the necessary adapts to solve the new problem. There are two methods used to perform this action: transformational and derivational. While derivational methods modify the previous solution using domain-specific transformation operators, the transformational methods reuse the algorithms, techniques, or rules that generated the original solution to produce a new solution to the current problem.

Traditionally, intelligent systems procedures were described through rules called RBR. For Hayes [15], expressing knowledge using this model is difficult and time-consuming in real situations. In contrast, in CBR systems, the cases are knowledge, leading to automatic maintenance and updating of knowledge, while in RBR systems, new rules are created. This type of system proposes a solution to a given problem, from knowledge not wholly defined, low structured, or unknown, as well as allows it to pay attention during the process of constructing a solution to the aspects/characteristics of the problem considered a determinant for the construction the same solution.

In the legal context, CBR-based systems have found vast opportunities to develop their application methodologies and with satisfactory results [29]. The SCALIR is an example of a hybrid symbolic/sub-symbolic system that uses legal network knowledge to perform retrieval through spreading activation to perform the task [30], considering legal decisions as complex networks [21]. Indeed, the law may thus be thought of as a giant network containing information embedded in cases (nodes) and relationship information called citations (arcs) going from node to node. Measures such as betweenness, closeness, and Markov Centrality can help find the causes at the core of a judicial system. Likewise, measures such as clustering enable understanding the degree of interdependence of cases that comprise a jurisprudence database.

Meanwhile, as pointed out by Carneiro and Gomes [4,10,12], both RBR and CBR approaches to negotiation face criticism. Thus the main drawbacks can be briefly enumerated as follows [ibid]:

– Laws constantly change, thus implying updates to the rules that establish how solutions are generated in RBR. This may result in inconsistencies and/or redundancy. Moreover, this might be quite a complex task (depending on the complexity of the legal domain) that must be performed manually, despite the use of some supporting tool;
– The quality of an RBR tool is directly dependent on the quality of the work of the humans, translating the legal norms into rules. The quality of information of the rules may be hard to determine;
– RBR are static and will not shape changes in the legal domain unless these are coded manually by a human expert;
– The quality of a CBR tool is directly dependent on the quality and amount of past cases known;
– The fact that legal norms change frequently also has a negative impact on CBR approaches, rendering past cases potentially useless under the light of the new norms;
– Both CBR and RBR approaches are domain-dependent. This implies that rules are defined independently for each legal domain and that cases from a specific field can hardly be reused.

3.3   An Overview of the Related Background and Prior Research

According to the most recent literature, AI solutions in legal services can be grouped into document analysis, legal research, and practice automation. While the first two categories correspond to tools that support lawyers in their work, practice automation refers to the automation of a lawyer's work. Practice automation via AI tools might bring considerable gains in productivity and a significant change in the legal profession, with the automation of discovery (e-discovery) and the redaction of court briefs. However, datasets are essential for AI systems, both as training material for developing AI algorithms and as input material for its actual use. The data (or its lack) might constitute a barrier to entry for small law firms or solo practitioners who want to create their own AI systems. Data has been considered a bottleneck: In its decision v. Google (Shopping), the European Commission stated that the search data held by Google constituted a barrier to entry for other prospective market players. In the legal context, it has been described that most law firms are "document rich and data-poor", and public data such as judicial decisions and opinions are either not available or so varied in the format as to be challenging to use effectively.

4   Conclusion

The field of taxation and tax law is complex by nature, typically consisting of large volumes of highly technical textual information and categorized by codes

(VAT, IRS, ...) and laws. The development and deployment of techniques and approaches using AI and Law can help this field of knowledge to assist the legal professionals in interpreting this type of information, providing added value in a collaborative context that enriches the information and the professionals who use it. Bearing this in mind, this work intended to characterize the existing technological infrastructure and present the most relevant literature and understanding gathered related to these topics.

References

1. Ashley, K.D., Walker, V.R.: From information retrieval (IR) to argument retrieval (AR) for legal cases: report on a baseline study. Front. Artif. intell. Appl. **259**, 29–38 (2013). https://doi.org/10.3233/978-1-61499-359-9-29
2. Boer, A., Hoekstra, R., Winkels, R.: MetaLex: legislation in XML (2002)
3. Bolisani, E., Bratianu, C.: The elusive definition of knowledge. In: Emergent Knowledge Strategies. KMOL, vol. 4, pp. 1–22. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-60657-6_1
4. Carneiro, D.: An agent-based architecture for online dispute resolution services. Ph.D. thesis, University of Minho (2013). https://hdl.handle.net/1822/28773
5. Devins, C., Felin, T., Kauffman, S., Koppl, R.: The law and big data. Cornell J. Law Public Policy **27**(2), 357–413 (2017)
6. Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D. (eds.): Semantic Processing of Legal Texts. LNCS (LNAI), vol. 6036. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12837-0
7. Frické, M.: Big data and its epistemology. J. Am. Soc. Inf. Sci. **66**(4), 651–661 (2015). https://doi.org/10.1002/asi.23212
8. Gangemi, A., Sagri, M.-T., Tiscornia, D.: A constructive framework for legal ontologies. In: Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A. (eds.) Law and the Semantic Web. LNCS (LNAI), vol. 3369, pp. 97–124. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-32253-5_7
9. Giri, R., Porwal, Y., Shukla, V., Chadha, P., Kaushal, R.: Approaches for information retrieval in legal documents. In: 2017 10th International Conference on Contemporary Computing, IC3 2017 2018-January(November 2019), pp. 1–6 (2018). https://doi.org/10.1109/IC3.2017.8284324
10. Gomes, M., Carneiro, D., Novais, P., Neves, J.: Modelling stress recognition in conflict resolution scenarios. In: Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B. (eds.) HAIS 2012. LNCS (LNAI), vol. 7208, pp. 533–544. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-28942-2_48
11. Gomes, M., Silva, F., Ferraz, F., Silva, A., Analide, C., Novais, P.: Developing an ambient intelligent-based decision support system for production and control planning. In: Madureira, A.M., Abraham, A., Gamboa, D., Novais, P. (eds.) ISDA 2016. AISC, vol. 557, pp. 984–994. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53480-0_97
12. Gomes, M.F.V.: A conflict management environment to support decision-making process (2021)

13. Gordon, T.F., Governatori, G., Rotolo, A.: Rules and norms: requirements for rule interchange languages in the legal domain. In: Governatori, G., Hall, J., Paschke, A. (eds.) RuleML 2009. LNCS, vol. 5858, pp. 282–296. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04985-9_26

14. Gostojić, S., Milosavljević, B., Konjović, Z.: Ontological model of legal norms for creating and using legislation. Comput. Sci. Inf. Syst. **10**(1), 151–171 (2013). https://doi.org/10.2298/CSIS110804035G

15. Hayes-Roth, F.: Rule-based systems. Commun. ACM **28**(9), 921–932 (1985). https://doi.org/10.1145/4284.4286

16. Joshi, K.P., Gupta, A., Mittal, S., Pearce, C., Joshi, A., Finin, T.: ALDA: cognitive assistant for legal document analytics. In: AAAI Fall Symposium - Technical Report FS-16-01-September, pp. 149–152 (2016)

17. Kolodner, J.: Case-Based Reasoning. Morgan Kaufmann, Burlington (2014)

18. Koniaris, M., Anagnostopoulos, I., Vassiliou, Y.: Evaluation of diversification techniques for legal information retrieval. Algorithms **10**(1), 1–24 (2017). https://doi.org/10.3390/a10010022

19. Lame, G.: Using NLP techniques to identify legal ontology components: concepts and relations. Artif. Intell. Law **12**(4), 379–396 (2004). https://doi.org/10.1007/s10506-005-4160-3

20. Leake, D.B.: Case-based reasoning: experiences, lessons, and future directions (1996)

21. Lupu, Y., Voeten, E.: Precedent in international courts: a network analysis of case citations by the European court of human rights. Br. J. Polit. Sci. **42**(2), 413–439 (2012). https://doi.org/10.1017/S0007123411000433

22. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008). https://doi.org/10.1017/CBO9780511809071

23. McCarty, L.T.: Deep semantic interpretations of legal texts. In: Proceedings of the 11th International Conference on Artificial Intelligence and Law, pp. 217–224 (2007)

24. Merkl, D., Schweighoffer, E., Winiwarter, W.: Exploratory analysis of concept and document spaces with connectionist networks. Artif. Intell. Law **7**(2), 185–209 (1999). https://doi.org/10.1023/A:1008365524782

25. Nissan, E.: Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement. Ai Soc. **32**(3), 441–464 (2015). https://doi.org/10.1007/s00146-015-0596-5

26. Omotayo, F.O.: Knowledge management as an important tool in organisational management: a review of literature. Libr. Philos. Pract. **1**(2015), 1–23 (2015)

27. Palmirani, M., Vitali, F.: Akoma-Ntoso for legal documents. In: Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, M. (eds.) Legislative XML for the Semantic Web. Law, Governance and Technology Series, vol. 4. Springer, Dordrecht (2011). https://doi.org/10.1007/978-94-007-1887-6_6

28. Ramakrishna, S., Paschke, A.: Bridging the gap between Legal Practitioners and Knowledge Engineers using semi-formal KR. CoRR abs/1406.0 (2014). https://arxiv.org/abs/1406.0079

29. Rissland, E.L., Daniels, J.J.: A hybrid cbr-ir approach to legal information retrieval. In: Proceedings of the 5th International Conference on Artificial Intelligence and Law, pp. 52–61 (1995). https://doi.org/10.1145/222092.2221250

30. Rose, D.E.: A Symbolic and Connectionist Approach to Legal Information Retrieval. Psychology Press, London (2013)

31. Sanderson, M., Croft, W.B.: The history of information retrieval research. In: Proceedings of the IEEE 100(Special Centennial Issue), pp. 1444–1451 (2012). https://doi.org/10.1109/JPROC.2012.2189916

32. Schütze, H., Manning, C.D., Raghavan, P.: Introduction to Information Retrieval, vol. 39, p. 2. Cambridge University Press, Cambridge (2008)

33. Shelar, A., Moharir, M.: A comparative study to determine a suitable legal knowledge representation format. In: 2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), pp. 514–519 (2018). https://doi.org/10.1109/ICEECCOT43722.2018.9001363

34. Slade, S.: Case-based reasoning: a research paradigm. AI Mag. **12**(1), 42–42 (1991). https://doi.org/10.1609/aimag.v12i1.883

35. Smyth, B., Keane, M.T.: Remembering to forget. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, pp. 377–382. Citeseer (1995)

36. Walker, V.R., Han, J.H., Ni, X., Yoseda, K.: Semantic types for computational legal reasoning: propositional connectives and sentence roles in the veterans' claims dataset. In: Proceedings of the International Conference on Artificial Intelligence and Law, pp. 217–226 (2017). https://doi.org/10.1145/3086512.3086535

37. Walker, V.R., Pillaipakkamnatt, K., Davidson, A.M., Linares, M., Pesce, D.J.: Automatic classification of rhetorical roles for sentences: comparing rule-based scripts with machine learning. In: CEUR Workshop Proceedings, vol. 2385 (2019)

38. Winkels, R., Boer, A., De Maat, E., Van Engers, T., Breebaart, M., Melger, H.: Constructing a semantic network for legal content. In: Artificial Intelligence Conference, Belgian/Netherlands, pp. 405–406(2005)

39. Xu, H., Savelka, J., Ashley, K.D.: Toward summarizing case decisions via extracting argument issues, reasons, and conclusions. In: Proceedings of the 18th International Conference on Artificial Intelligence and Law, pp. 250–254 (2021). https://doi.org/10.1145/3462757.3466098