WILEY | Hindawi

Review Article

Automation of Legal Precedents Retrieval: Findings from a Literature Review

Hugo Mentzingen (ID)**, Nuno António** (ID)**, and Fernando Bacao** (ID)

NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, Lisbon 1070-312, Portugal

Correspondence should be addressed to Hugo Mentzingen; hsilva@novaims.unl.pt

Judges frequently rely their reasoning on precedents. Courts must preserve uniformity in decisions while, depending on the legal system, previous cases compel rulings. The search for methods to accurately identify similar previous cases is not new and has been a vital input, for example, to case-based reasoning (CBR) methodologies. This literature review offers a comprehensive analysis of the advancements in automating the identification of legal precedents, primarily focusing on the paradigm shift from manual knowledge engineering to the incorporation of Artificial Intelligence (AI) technologies such as natural language processing (NLP) and machine learning (ML). While multiple approaches harnessing NLP and ML show promise, none has emerged as definitively superior, and further validation through statistically significant samples and expert-provided ground truth is imperative. Additionally, this review employs text-mining techniques to streamline the survey process, providing an accurate and holistic view of the current research landscape. By delineating extant research gaps and suggesting avenues for future exploration, this review serves as both a summation and a call for more targeted, empirical investigations.

1. Introduction

Civil, criminal, and administrative courts are challenged by the increasing need for justice systems' intervention in private and public affairs. Their actions must result in prompt and consistent judgments [1, 2]. Although the decision-making process of justice courts affects many facets of citizens' lives, these institutions have limited resources and strive to keep up with the rising caseload [3, 4].

The basis of the rationale of judges in national legal systems is precedent. Judges often follow precedent closely for legal certainty. Otherwise, their rulings could be appealed to higher instances [5].

In the same way, the Common Law system considers similar past cases to be precedents, implying that a result in a current issue is compelled by past issues [6]. Even in places that adopt Civil law, courts are required to regard former rulings when there is enough uniformity in case law. Typically, when consistent jurisprudence is formed, precedents become "soft" law, and courts consider them when making decisions [7].

Precedents are also fundamental to case-based reasoning (CBR). CBR considers similar previous cases to employ prior knowledge in answering new questions. CBR can clarify new situations by reasoning from precedents [8]. Artificial Intelligence (AI) and Law, a branch of AI, extensively utilize CBR to explore legal case-based reasoning [9].

Although research in CBR has been utilized in legal practice since the 1980s, techniques for identifying precedents are reasonably young and understudied in AI and Law. While methods for mining textual data and natural language processing (NLP) have evolved and provide promising opportunities, the number of papers that studied strategies for detecting similarity and recognizing such past cases is scant.

To our knowledge, no prior work has described the methodologies used to retrieve legal precedents (the present paper is an extended and updated version of our paper *A*

Rapid Semi-Automated Literature Review on Legal Precedents Retrieval, presented at the EPIA Conference on Artificial Intelligence, 2022 [10]; this paper incorporates feedback received, increases the search period, includes a new research question, and offers significantly more detailed results and discussion sections). In the same way, the research state on this subject needs clarification so that researchers and courts can consider such AI-based assistants. This paper identifies the most promising findings and the knowledge gaps about how legal practitioners can employ AI to retrieve similar cases. Moreover, we investigate the effectiveness of text mining (TM) and NLP in this semi-automated systematic review of the literature. An earlier version of this paper has been presented as a preprint [11].

Mainly, we concentrate on these research questions:

(i) RQ1: how did researchers address the challenge of automatically identifying prior relevant cases, and what methods have been used in the screened studies?

(ii) RQ2: what are the most promising methods for the automated search of legal precedents, and what research gaps exist?

(iii) RQ3: what is the taxonomy of existing methods, and what is their mainstream?

(iv) RQ4: what are the research domain's most influential journals and authors?

(v) RQ5: what data have been used in existing research?

(vi) RQ6: are there real-world applications of this topic?

2. Theoretical Background

2.1. Systematic Literature Reviews. We intend to present an unbiased literature assessment on automating legal precedent identification. To this purpose, we conducted a systematic literature review (SLR). An SLR is a method of synthesizing scientific data from explicitly defined research questions. It follows rigorous procedures to locate, select, and assess relevant scientific research. SLRs collect and critically analyze data from selected studies [12].

The systematic review process is based on predefined criteria and protocol [13], constituting evidence syntheses of high value used to inform decisions. However, it frequently takes one to two years to complete the process under the methodological rigor that renders SLR evidence reliable [14–17]. Garritty et al. [18] mentioned that this aspect reduces the usefulness of SLRs for not tackling the time constraints of stakeholders.

2.2. Rapid Literature Reviews. Various strategies exist to make reviews more time-efficient. These strategies can be employed separately or simultaneously. Review shortcuts are among these mechanisms, through which one or more steps in a systematic review can be simplified or omitted. Moreover, the typical systematic review approach is accelerated by automating review steps [19]. In this sense, Rapid Reviews (RRs) became an alternative method to save time and resources on literature reviews. At the same time, the core principles of knowledge synthesis are present [19].

Despite a surge in RR production, their process is still underdeveloped. The definition of an RR is still not universally recognized [16]. An extension is in progress to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for RRs [20]. However, until it is formally finalized, authors are encouraged to follow the generic PRISMA criteria as much as possible and adapt them appropriately [18].

PRISMA's four stages (identification, screening, eligibility, and inclusion) are used in this study to guide the critical components of an SLR.

2.3. Automating Systematic Literature Reviews. Numerous academics studied methods for automating the SLR procedure [21]. Conducting an SLR according to the best standards and with the required level of rigor is complex, developed in multiple stages, and considerably time-consuming [22]. While preserving the rigor expected from SLRs, text analytics and machine learning techniques could solve this scaling issue [23]. For this reason, automated literature reviews have been utilized for numerous subjects, including tourism and hospitality [24], climate change [25], and healthcare [26].

A recent review screened 41 articles focused on automating SLRs. It concluded that the primary election of studies was the most automated step. Even though various studies have proposed automation techniques for SLRs, none have automated the planning and reporting phases [21].

Four groups of applications were identified based on the SLR automation solutions proposed in 32 studies: automated document/text categorization, text mining based on visual methods (VTM) such as word clouds, federated search strategy, i.e., searching on multiple data sources at once, and document summarizing [22]. In this study, we used Python [27] and a combination of TM methods to identify primary studies for human screening.

VTM enabled the determination of the most pertinent terms and authors employing keywords and word clouds produced from the most frequent terms. Also, we used the vectorial representation of the studies to automate their classification. Documents were deduplicated based on their similarity and filtered based on keywords.

Lastly, Latent Dirichlet Allocation (LDA) was utilized to determine document admissibility. It was made by associating a topic's distribution to each paper and each topic with its corresponding probability. This way, documents were *softly* clustered according to the Topic with the most significant likelihood. The number of records still present following each phase is shown in Figure 1.

3. Materials and Methods

3.1. Keyword Identification. Precedents are influential in resolving any legal question [7]. According to the Legal Information Institute, based at Cornell Law School,
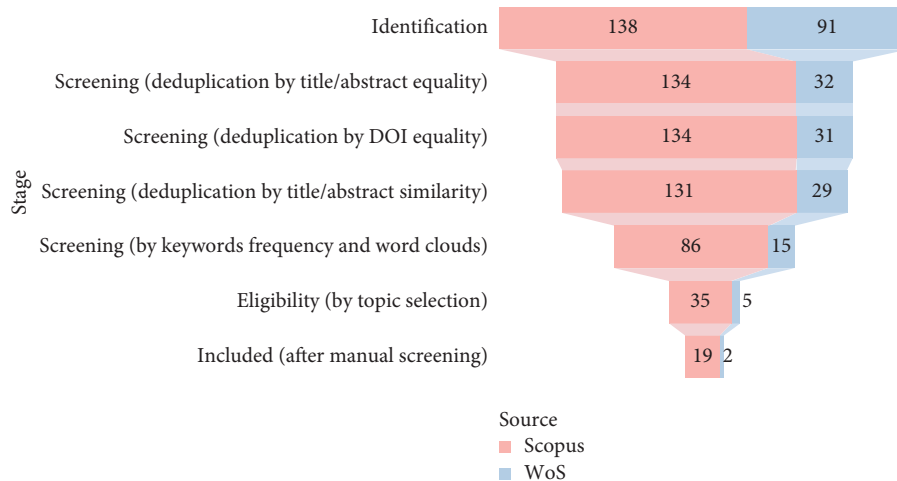
FIGURE 1: Documents chosen from each electronic database after the sources' selection stages.

a precedent is a decision that serves as a starting point for resolving later cases. These cases should involve similar facts or legal matters. However, a case cannot serve as a precedent if its facts or issues differ from the current case [28].

In this sense, the identification of legal precedents is anteceded by finding previous similar cases. In this sense, we derived central expressions from the research questions to allude to detecting similar cases. Also, we used "precedent" as the corresponding legal term and consulted the study from Mandal et al. [29] as a seminal reference on identifying precedents in justice courts. The resulting set was "precedents identification," "precedent retrieval," and "case similarity." Since CBR models are used in many disciplines, including engineering and medicine, we added the term "case-based reasoning" as a potential alternative. We made the keyword "legal" a required component to focus the search on the legal segment.

3.2. Electronic Databases and Eligibility Criteria.
We used Scopus, a publication of Elsevier, and Web of Science (WoS), a database managed by Thomson Reuters, as sources because of their extensive publications' coverage. Besides that, these are the most used electronic databases for bibliometric analyses [30].

Given that English is the language of most relevant research, the search was limited to English to reduce obstacles with text mining [24]. Additionally, to appraise the most recent methodologies, in line with advancements in TM and NLP, we restricted the search to research published from the 2000s onward ("Year" > 1999). We did not limit the publication categories provided by the search because precedent identification is still a relatively unexplored topic.

Figures 2 and 3 show the queries on Scopus and WoS, respectively.

The results from the search queries were combined into a single dataset that underwent preprocessing, including deduplication and feature engineering, before they could be employed for further studies' screening. The details about the extraction and preprocessing of the sources list are presented in Appendix A.

3.3. Screening Using the Keyword Frequency.
The queries for searching the electronic databases could retrieve any studies related to legal precedents, even those not involving computational methods. Thus, we wanted to apply a screening process so that the publications not mentioning computational techniques for identifying similar cases could be disregarded. We then chose to semi-automate the screening process of the remaining 160 documents.

This process started with tokenizing and stemming the abstracts, from which unique tokens (3256 unigrams and 14675 bigrams) were found. The number of occurrences of these tokens is shown in Figures 4 and 5.

Stemmed words (tokens) might be indicative of computer-aided precedent retrieval applications. We used word clouds for this purpose. The words appear more prominent in this visual representation the more frequently they occur in the corpus. The tokens shown in Figure 6 have been removed to prevent nondiscriminatory words from crowding word clouds.

When building the word cloud using unigrams, the tokens "reasoning," "cas," and "casebased" were discarded because they appeared in over half of the texts. Similarly, "casebased reasoning" exceeded 30% of the samples and was removed when the word cloud was built from bigrams. The word clouds were created using the WordCloud library [31] and are shown in Figures 7 and 8.

The terms' syst' ("support syst" included), "artific intelligenc," "machine learning," "natur language," and "computat model" pointed to the most promising literature segment. Following this, the only records kept were those containing such tokens in the title, abstract, or keywords, yielding 101 samples.

3.4. Eligibility Screening Based on Topic Modeling.
During the eligibility assessment stage of the SLR, the publications are assessed for eligibility based on prespecified methods. Therefore, the review team must meet the criteria for including and excluding publications. The result of this stage is the reduction of the studies under evaluation, keeping the

```
(TITLE-ABS-KEY ({precedents identification} OR {precedent
  identification} OR {precedent retrieval} OR {precedents
retrieval} OR {case similarity} OR {cases similarity} OR {case-
  based reasoning}) AND TITLE-ABS-KEY (legal)) AND (LIMIT-
        TO(LANGUAGE, "English")) AND (PUBYEAR > 1999)
```

Figure 2: The search query used on Scopus.

```
(((TI=("precedent retrieval" OR "precedents retrieval" OR "case
similarity" OR "cases similarity" OR "case-based reasoning") OR
  AB=("precedent retrieval" OR "precedents retrieval" OR "case
similarity" OR "cases similarity" OR "case-based reasoning") OR
  AK=("precedent retrieval" OR "precedents retrieval" OR "case
  similarity" OR "cases similarity" OR "case-based reasoning"))
AND (TI=(legal) OR AB=(legal) OR AK=(legal))) AND LA=(English)
              Timespan: 2000-01-01 to 2022-12-31
```

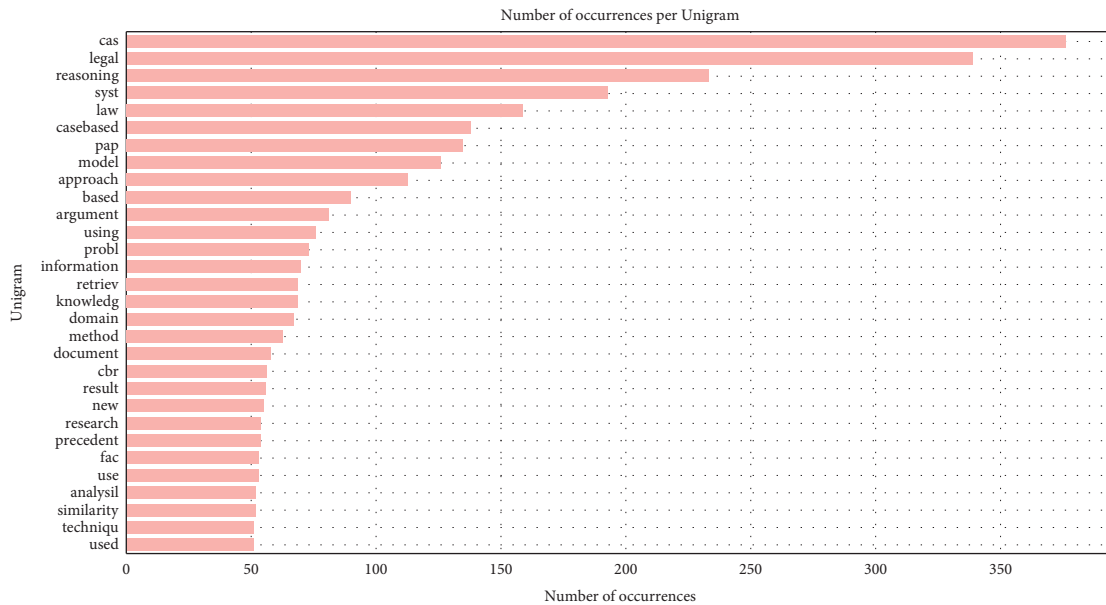Figure 3: The search query used on Web of Science.

Figure 4: Abstracts' unigram frequency distribution.

evidence set that can provide answers to the research questions [32].

We used topic modeling to cluster the papers and choose the studies with a higher probability of effectively responding to the research questions. Although topic modeling is not a recent concept, it is remarkable that just a few publications use this method to cluster research papers [33]. We decided to employ LDA [34]. It is usually the preferred approach for topic modeling and is considered state of the art [33], using the Gensim [35] library. LDA is a topic model that uses a generative statistics approach to unveil semantic topics in extensive text collections and classifies documents into these topics. The documents are categorized according to their distance from a given topic [24, 36].

We classified the documents according to four topics. The one with the most explicit link to the research subject was chosen as the eligibility criterion for including a document in the literature review. It resulted in forty eligible documents. The procedure for obtaining the optimal number of topics and the topics' description is presented in Appendix B.

3.5. Full-Text Screening for Inclusion. After clustering the documents according to their most relevant topics and selecting the set of publications mainly related to Topic 2, the remaining forty studies had their full texts examined to eliminate those irrelevant to the automation of similar case identification or precedent retrieval. The nineteen
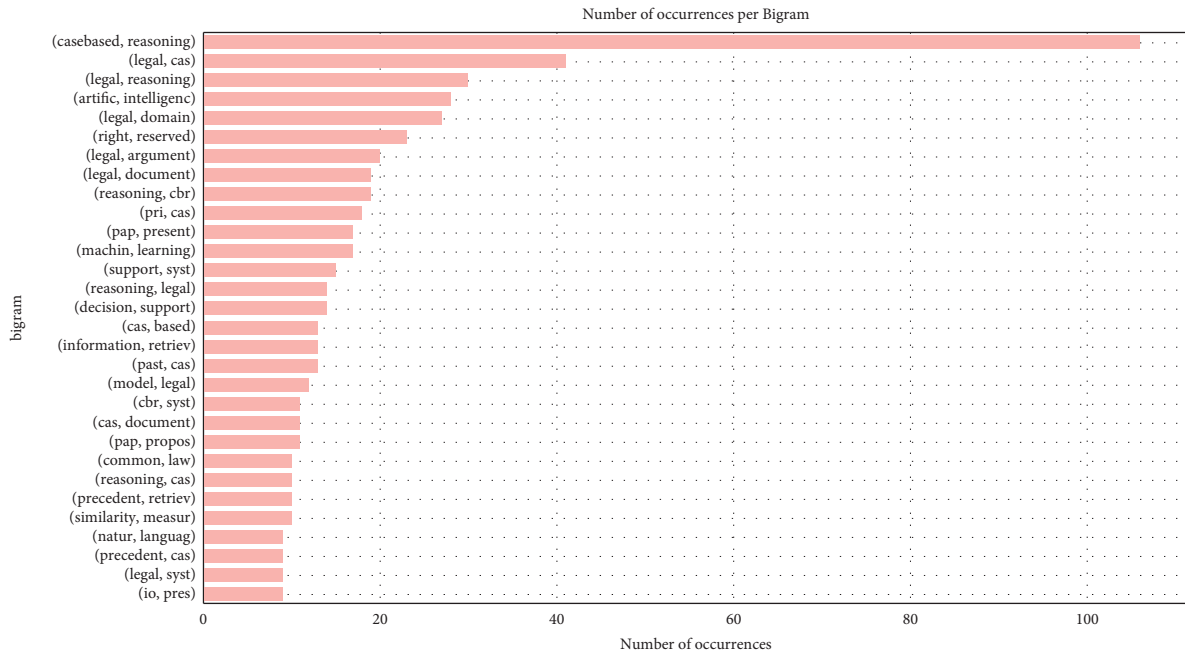
Figure 5: Abstracts' bigram frequency distribution.

```
    'legal', 'law', 'pap', 'approach', 'based', 'using',
'knowledg', 'domain', 'document', 'result', 'new', 'research',
  'use', 'analysil', 'used', 'work', 'two', 'dat', 'provid',
   'proposed', 'one', 'present', 'term', 'text', 'outcom',
'expert', 'exampl', 'show', 'study', 'issu', 'propos', 'task',
  'first', 'theory', 'describ', 'reserved', 'developed', 'way',
                    'given', 'formal'
```

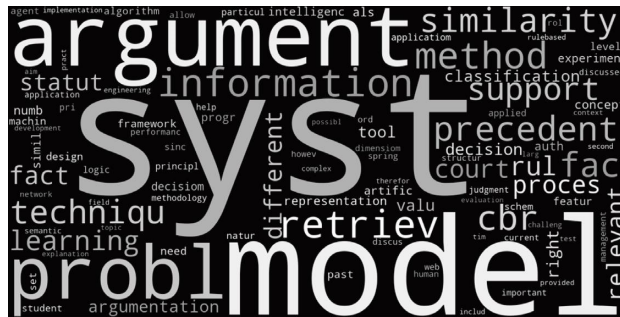Figure 6: The set of tokens with low document discrimination capacity.

Figure 7: Abstracts' unigram word cloud.

Figure 8: Abstracts' bigram word cloud.

publications that focus on subjects unrelated to the research questions are synthesized in Appendix C (Table 1). It contains the excluded document titles and respective research topics. In Appendix D, we included the method used for validating the clustering through topic modeling as an effective method to assess document eligibility.

In conclusion, the ultimate analysis included the twenty-one papers in Table 2.

4. Results

4.1. Descriptive Analytics. Examining Figure 9 demonstrates that the studies' publications were dispersed in numerous sources, including conference proceedings. The peer-reviewed articles were found in six journals, one per journal.

When we examine the number of publications per year and geographical origin (Figure 10), we observe a few studies, approximately one per year, in the first decade of the 2000s. After the initial studies, there was a long period in which there were virtually no publications in the field. After 2016, a growing interest in this field is observable, with Indian researchers contributing the most to this subject. This growth is possibly a result of the developments observed in TM and NLP in the latter half of the 2010s: word embeddings and neural network (NN) applications to NLP [57], recurrent neural networks (RNNs) [58], long short-term memory networks (LSTM) [59], attention mechanisms and transformers [60], and language models pretrained through transfer learning [61, 62].

4.2. Content Analysis. The two first relevant studies [55, 56] presented a simplified model to store and retrieve information in the legal context. The model approximated the human cognitive process and combined keywords with related scenarios to facilitate understanding. The scheme matched individual cases to story patterns, clustering cases with some similarity. However, the patterns, keywords, and importance were manually crafted according to the specific domain, e.g., bankruptcy legislation.

The following paper proposed a model using content vectors to recover principles and previous cases. In this work, the author assessed similarities in each case's actions and events [54]. Content vectors summarize the information included in intricate relational structures. A case description's content vector identifies the functors (a functor is defined as a function that converts items of one set into those of another set) that were used in that description and how frequently they occurred, including connectives, relations, object properties, and functions [63].

A 2004 paper presented an algorithm for obtaining valuable legal information, constructing case examples from prior lawsuits, recognizing comparable samples, and refining them by combining cases and deleting irrelevant data. It required encoding the texts as ordered sets of keywords and evaluating the similarity between pairs of case scenarios. To measure similarity, the authors applied word count-based metrics. Using predetermined crime types, the nearest neighbors created clusters [53].

The subsequent research introduced a function based on nonlinear nearest-neighbor (NNN) for finding similar accident compensation cases. This method used "dimensions" to compare cases. The dimensions can be interpreted as factors representing, analogizing, and differentiating legal cases. From data observation, groupings of variables were set to represent clusters of cases. The authors employed four groupings of variables [52].

Nouaouria et al. [51] presented a prototype of a tool for applying interpretative case-based reasoning to verdicts involving alcohol consumption and smoking under Islamic legislation. The representation of cases was based on attributes inferred as relevant to historical decisions, such as the type of fact and the product used. The similarity was addressed by grouping cases with similar attributes.

Between 2008 and 2009, a methodology inspired by CBR for retrieving legal precedents depicted lawsuits as pairs of attributes and their respective values. Knowledge from experts and critical legal circumstances were the sources for attributes and values. The process involved a case-similarity calculation, and cases with strong similarities were selected according to the value distribution of each attribute [49, 50].

In all previously analyzed studies, cases were represented using a predetermined set of dimensions corresponding to the attributes of each sample. Consequently, the factors or dimensions describing each case were subject-specific. Indeed, there were two broad approaches to applying information retrieval (IR): methods built on manual knowledge engineering (KE) and other methods based on NLP. The existing technology and scientific knowledge limited the former methods. So far, the studies have emphasized KE-based retrieval, even though these methods were not viable in the long run [64].

McLaren and Ashley [48] evaluated the influence of temporal orderings of facts in distinguishing among cases in ethical case issues. Each case was represented by a predefined set of actors, objects, actions, and events that appeared in the narrative. Also, temporal knowledge was expressed through a time qualifier: an association between time and how a fact relates to other facts. The study did not detail the logic employed to obtain the time qualifier for each case. The authors could not confirm the hypothesis that introducing temporal knowledge into a computational model would increase the accuracy of the model's predictions.

Eyorokon et al. [47] presented Kyudo, a system that used conversational CBR to support knowledge discovery. Cases were represented in Kyudo as sequences of questions or knowledge goals represented by TF-IDF vectors. The similarity between answers or new goals with the existing knowledge base was calculated as a dialogue between the user and the system. By expressing knowledge goals as multidimensional vectors of semantic attributes, the system could recognize similarities with other knowledge goals and alert the user to other pertinent goals as the number of examples increased.

Later, a business intelligence (BI) solution was proposed by Oconitrillo and De La Ossa Osegueda [45] to support judges' decision making. The authors advocated a new formal representation of legal cases, depending on facts and

TABLE 1: Publications that were excluded during the full-text screening phase.

Reference	Year	Research topic
[78]	2021	This paper proposed a formal model that derived factors from linguistic terms and expert knowledge and used a neutrosophic pairwise matrix of factors for determining case similarity
[79]	2021	This paper defined legal debugging and used logic programming to anticipate unexpected consequences in the legal domain
[80]	2021	This paper proposed an approach to classify legal documents employing convolutional neural networks (CNNs)
[81]	2021	This conference paper proposed a model for predicting and explaining European Court of Human Rights (ECHR) application admissibility results based on a set of questions
[82]	2019	Instead of retrieving precedents, this conference paper concentrated on retrieving legal statutes from existing cases
[83]	2019	This conference paper presented a web-based platform for developing legal case-based reasoners
[84]	2016	This proceedings paper presented a model for an automotive design system based on CBR
[85]	2014	This paper outlined a model to assess the constitutional admissibility of legal changes
[86]	2014	This conference paper reported preliminary results of a CBR system designed to support decision making on the sustainable development of regions
[87]	2014	This proceedings paper provided a prototype for using decision support approaches to function as a "virtual courtroom" for plaintiffs and defendants to resolve their legal cases without a trial
[88]	2013	This paper proposed a method for retrieving similar construction accident cases for dispute resolution
[89]	2010	This proceedings paper proposed a method to produce an annotated text that supports information extraction
[90]	2010	This paper presented a framework to suggest techniques for solving risk management problems. The proposal involved matching new cases to similar previous cases based on their attributes
[91]	2010	This paper examined text mining algorithms for automating profiling and extracting arguments from court cases
[92]	2005	This conference paper introduced the conversational case-based reasoning (CCBR) methodology, described its evolution, examined its various manifestations and motivations, highlighted recent work, and described unsolved issues that could be the dissertations' focus
[93]	2003	This review summarized the conference content without further details on the studies
[94]	2003	The entire text of this conference paper could not be accessed from electronic databases
[95]	2003	This conference paper evaluated the lazy learning implemented by the HIPO family of legal expert systems in helping courts' decision making
[96]	2002	This paper discussed the evolution of a system to support court decisions and its transition to a hybrid system (based on rules and cases), evolving from a system based solely on rules

attributes. It considered how the law is applied during a judge's reasoning process to decide on each case and the relations that a judge develops among such regulations. Nevertheless, the authors did not incorporate a solution for automating the retrieval of attributes and their values.

With a study authored by Kulkarni et al. [43] proposing the detection of precedents using regular expression rules, cosine similarity between Doc2Vec embeddings [65], and topic modeling [34], we notice a move towards TM and NLP. Another research published in 2017 detected precedents integrating genetic algorithms (GAs) with k-nearest neighbors (KNNs) [46].

Additionally, the performance of legal catchphrases for precedent retrieval was studied. Thuma and Motlogelwa [44] isolated legal catchphrases from new cases, primarily bigrams and trigrams represented by TF-IDF vectors, and compared them to gold standard catchphrases extracted from previous cases. The results indicated that the technique needed improvements.

In a 2018 article, the citations of a document were used to perform association-rule mining as an alternative method of identifying similarities. This time, cases with matching citations were considered comparable [42].

TABLE 2: Documents contained in the final selection.

Reference	Year	Type	Cited	Country
[29]	2021	A	11	India
[37]	2020	CP	0	India
[38]	2020	CP	1	Italy
[39]	2019	CP	4	India
[40]	2019	CP	1	Germany
[41]	2019	CP	6	India
[42]	2018	A	1	India
[43]	2017	CP	2	India
[44]	2017	CP	2	Botswana
[45]	2017	CP	1	Costa Rica
[46]	2017	A	22	China
[47]	2016	CP	3	United States
[48]	2011	CP	0	United States
[49]	2009	PP	0	Malaysia
[50]	2008	PP	0	Malaysia
[51]	2006	A	3	Algeria
[52]	2005	CP	2	New Zealand
[53]	2004	A	40	Taiwan
[54]	2003	CP	67	United States
[55]	2001	A	2	Hungary
[56]	2000	A	30	Hungary

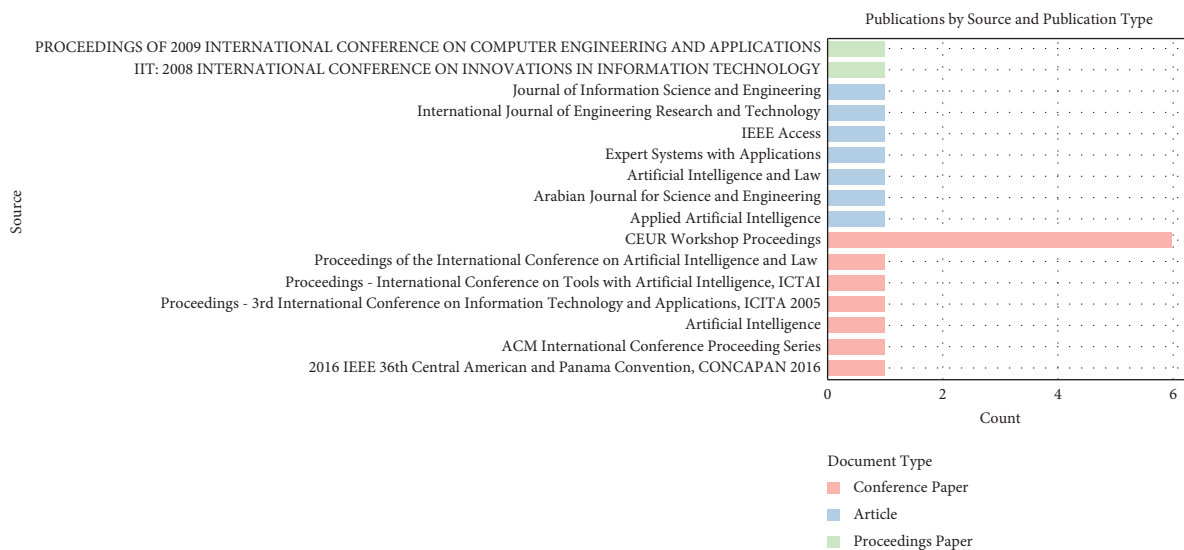(A = article, CP = conference paper, and PP = proceedings paper)

FIGURE 9: The number of publications per source and type.

To identify similar cases from unstructured text, an unsupervised Autoencoder [66] was used as a substitute for word embeddings based on neural networks, such as Word2Vec [57]. The Autoencoder was utilized with LSTM to retrieve similar cases from unstructured text. It reportedly led to quicker training and more accurate results [40].

Using Named Entity Recognition (NER) [67] to preprocess documents and the input query, More et al. [39] reported the extraction of data from legal texts. First, the vectorization used TF-IDF, while the comparison of documents used BM25 [68]. This algorithm also won the Artificial Intelligence for Legal Assistance (AILA) track at the 2019 FIRE Conference [41], in which the task involved identifying legal precedents. In the subsequent edition of the FIRE Conference, Di Nunzio [38] explored techniques to reduce the dimensionality of vectorized texts by employing lemmatization and stemming. The author compared the techniques for retrieving precedents with no outstanding results for any method.

Recently, the novel text embedding technique Top2Vec [69] was used to retrieve precedents in combination with BM25 in a paper published by Arora et al. [37]. The authors evaluated the similarity obtained from Top2Vec embeddings with BM25, outperforming BM25-only measures.

The most recent publication in this rapid literature review was released in 2021. Mandal et al. [29] presented a comprehensive study of fifty-six distinct combinations of
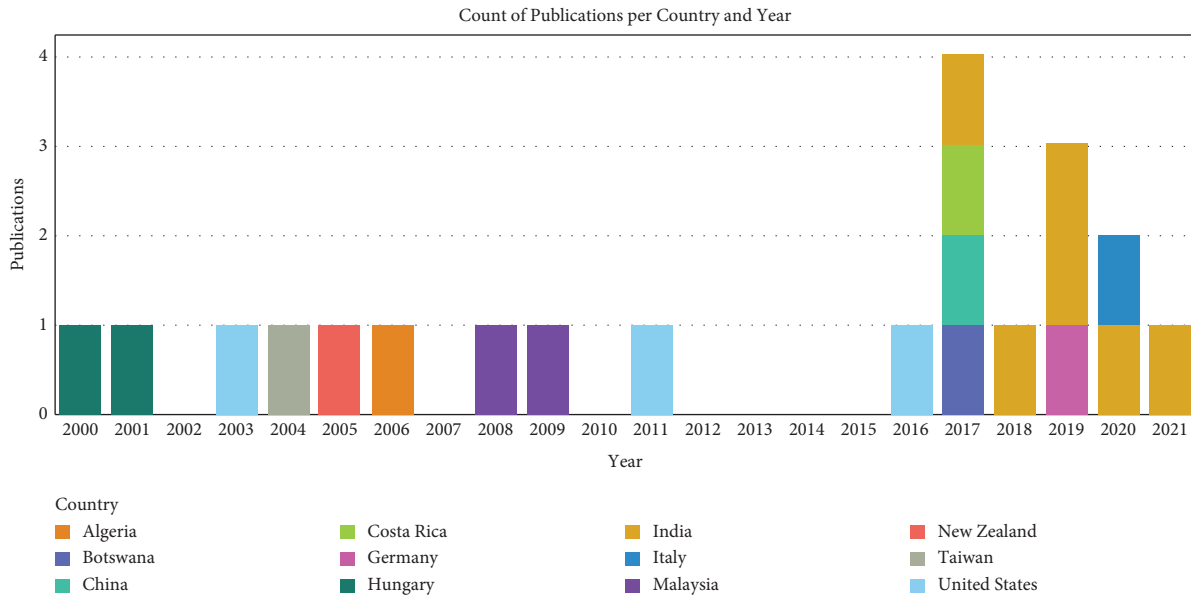
FIGURE 10: Publications per year and country.

document representation techniques (eight) and similarity measures (seven) to identify similarities between case report texts. Their methods included author-designed methodologies, BERT, and Law2Vec [70], a set of Word2Vec embeddings trained on legal corpora.

When comparing the various methods for similarity measurement, the authors noticed a similar performance between neural network-based embeddings (Word2Vec, Doc2Vec, and Law2Vec) and conventional embeddings, whereas BERT produced unsatisfactory results. They also noted that the conventional vectorization techniques that represent text using bag-of-words, for instance, TF-IDF, outperformed the more sophisticated methods that consider the context (such as Law2Vec and BERT).

4.2.1. The Era of Manual Knowledge Engineering (2000–2009). As identified in the content analysis, in eight papers published in 2009 and before, cases were represented using a predetermined set of dimensions corresponding to the attributes of each sample that may be grouped as methods built on manual knowledge engineering (KE). Early computational models, like Elhadi's [55, 56] studies, aimed to store and retrieve information by manually matching cases to story patterns. The efforts during this period primarily relied on predetermined sets of attributes derived from keywords and factual aspects, which were crafted to the specifics of the domain, such as bankruptcy legislation. Such dimensions were handcrafted from intricate cognitive structures like content vectors [54] or text encoding as ordered sets of keywords [53].

The commonality in all these methods was their reliance on domain expertise and manual screening of cases to create such attributes. Such methods made scalability a persistent challenge, proving them unfeasible when dealing with vast numbers of legal documents. Even as computational power grew, the emphasis remained on KE-based retrieval methods [49, 50].

4.2.2. The Wave of Artificial Intelligence (2016 Onwards). The next significant evolution in the domain came almost a decade later. A renewed interest in legal precedent retrieval was marked by a notable shift towards utilizing NLP and ML techniques to identify precedents through textual similarity. Studies like Eyorokon et al. [47] and Kulkarni et al. [43] marked the beginning of this transition. The authors of the first study used TF-IDF to represent documents as vectors of important words. The approach of Kulkarni et al. [43] combined regular expression rules with sophisticated techniques like Doc2Vec embeddings [65] and topic modeling [34].

More contemporary methods, like the unsupervised Autoencoder paired with LSTM [40] and Named Entity Recognition [39], were employed as research evolved. The later part of this era saw researchers evaluating many algorithms and representations to optimize precedent retrieval. Techniques like Top2Vec [69] combined with BM25, a ranking function, began outperforming legacy methods. A comprehensive assessment by Mandal et al. [29] encompassed fifty-six unique document representation techniques and similarity measure combinations. Their observations testify to the potential of conventional vectorization techniques in identifying legal precedents while also indicating areas of improvement for methods that heavily rely on context.

4.2.3. The Pipeline of Legal Precedents Retrieval. In the scholarly landscape focused on automating the identification of legal precedents, a noticeable structure has emerged, as depicted in Figure 11. The initial phase of this process, here termed "representation," involves encoding each legal case based on a predefined set of attributes, thereby preparing it for subsequent "similarity assessment." Following this second step, most studies incorporate "evaluation" as the final stage. This phase is
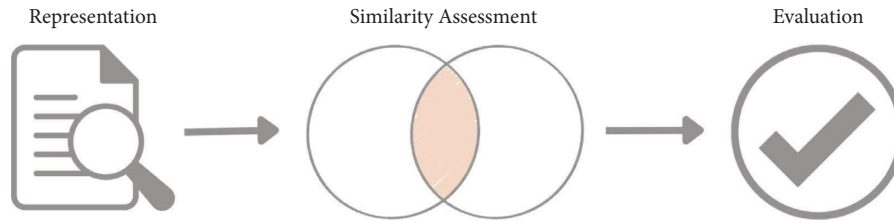
FIGURE 11: The pipeline for automating the identification of legal precedents.

dedicated to assessing the effectiveness of the proposed techniques for legal precedent retrieval.

Concerning the representation of legal cases, the existing body of research can be categorized based on the specific attributes employed to characterize each document. It is noteworthy that a singular study may incorporate methodologies from multiple categories. The delineated groups are as follows:

Keywords: in this approach, legal cases were characterized through the utilization of specific keywords or sets of keywords.

Facts: this category concentrated on representing cases based on factual elements, encompassing scenarios, actions, events, and the content of judicial decisions.

Time: this group of studies considered the temporal sequencing of the cases or the facts within each case.

Text-based vectors: Here, vectorial representations of text are employed. These may be generated through various means, such as regular expressions or vectorization techniques, and may or may not incorporate semantic considerations.

References: this category pertains to representations that account for the statutes cited in each case or cross-references to other legal cases.

The similarity assessment phase constitutes a critical component of the legal precedents' retrieval pipeline, serving as the linchpin that connects the representation of legal cases with the subsequent evaluation of a methodology's efficacy. Following the encoding of cases based on specific attributes as previously described—from keywords and factual elements to text-based vectors and references—the similarity assessment phase employs computational techniques to determine the resemblance between cases. The methods of similarity assessment identified among the studies are multifaceted, encompassing clustering, pairwise distance calculations, attribute matching, association rules, and ranking functions. These techniques measure how closely a target case aligns with one or more source cases, thereby enabling the accurate retrieval of relevant legal precedents. Therefore, the efficacy of similarity assessment methodologies plays an instrumental role in enhancing the precision and utility of automated legal precedent retrieval systems. The studies can be classified into the following categories according to the type of similarity assessment:

Clustering: this method for identifying similar cases involves retrieving sets of documents based on document clustering.

Pairwise distance: the potential precedents were identified by calculating a distance measure, mainly cosine similarity, to all other cases in the dataset.

Attribute matching: in such studies, cases sharing one or more attributes or facts were considered similar.

Association rules: similar cases were identified by mining frequent item sets and calculating association rules metrics such as support and confidence.

Ranking function: These studies used a ranking function to compare documents to a given query or another document. Most studies in this category employed BM25 to score documents, while divergence from randomness (DFR) (the DFR framework aims to rank documents based on the idea that terms that diverge significantly from their expected random distribution in a corpus are informative and thus useful for determining the relevance of a document to a query; essentially, the more a term's distribution in the document set deviates from a random distribution, the more "useful" or "informative" that term is for distinguishing relevant from nonrelevant documents) [71] was also observed.

We designated the terminal phase of the legal precedents' retrieval pipeline as the "evaluation" stage, which substantiates the employed methodologies' utility and precision. This section gauges the effectiveness of the techniques demonstrated in the studies in automating the identification of relevant similar cases. A well-constructed evaluation, therefore, not only confirms the validity of the similarity assessment techniques but also serves as a benchmark for future studies seeking to contribute to this growing field of research. Notably, the evaluative methods employed by researchers in this domain can be broadly classified into three categories:

Document citation: This approach entails juxtaposing the results produced by the computational model against the cases cited in the target document. Such a comparison provides an empirical assessment of the model's ability to identify precedents already recognized and cited in the legal literature.

Expert evaluation: Some studies opt for a more comprehensive approach by comparing the model's results with cases appraised similarly by legal experts. This

method adds a layer of professional scrutiny, offering insights into how well the computational methods align with human expertise in the field. The comparison of a target document is made with all the documents included in the corpus and is not limited to the documents cited by the target document.

Authors' appraisal: A subset of studies adopts a somewhat subjective methodology, wherein the authors manually compare the results generated by their model against a small set of case pairs. Typically, this involves examining one to three pairs of cases. Although less rigorous, this evaluation form is an initial test for the model's performance.

Ultimately, a handful of studies forego the evaluation phase entirely, either due to the exploratory nature of the research or other constraints. In these instances, the absence of an evaluative component leaves the model's effectiveness untested, limiting the findings' generalizability.

4.2.4. The Taxonomy of Legal Precedents Retrieval. Based on our review of the existing literature and findings detailed in Sections 4.2.1 to 4.2.3, we propose a taxonomy to systematize the field of legal precedents retrieval in Figure 12. The aim is to provide a structured framework that categorizes the studies according to the technological context and techniques employed, facilitating a more comprehensive understanding of the field's evolution and current trends.

We also classified the existing studies under this taxonomy according to the characteristics described in Table 3.

4.2.5. Data Sources. In the topic of legal precedent retrieval, academic research has utilized a diversity of data sources from varied geographical and legal contexts. An observation that stands out is the wide-ranging geographical representation of data sources, lending a global perspective to the research. Early studies, such as those by Elhadi [55, 56] and McLaren [54], mainly used US-based data focusing on bankruptcy law and professional ethics cases. These were followed by research that expanded the geographical scope to include criminal summary judgments from Taiwan [53] and Islamic legislation [51]. There were also more focused datasets, such as those involving the Law of Negligence [50] and judicial declarations of abandonment in Costa Rican Juvenile Courts [45].

Notably, there has been a prominent utilization of Indian court data in recent studies. Beginning with Thuma and Motlogelwa [44] who used case documents from the Indian Supreme Court, subsequent research such as that by Kulkarni et al. [43] and Nair and Wagh [42] further delved into various facets of the Indian justice system. Kulkarni et al. [43] analyzed a broad dataset comprising both court cases and statutes. Nair and Wagh [42] used cases under the Information Technology Act 2000 conducted in different high courts in India. This trend continued in recent years with work by More et al. [39]; Bhattacharya et al. [41]; Di Nunzio [38]; Arora et al. [37]; and Mandal et al. [29], all of whom engaged with case documents adjudicated by the Supreme Court of India. The emphasis on Indian court data enriches the field by incorporating the complexities and idiosyncrasies of a legal system influenced by a rich tapestry of cultural, historical, and social factors.

Adding another layer of complexity are studies like that of Amin et al. [40]; which employed a mixed-language customer support tickets dataset from an automotive company in Germany, and Zhang et al. [46]; who investigated Chinese statutes and judicial cases. These data types, though unusual, pave the way for exploring the adaptability of legal precedent retrieval algorithms to varied data formats and languages. The diversity in data sources and geographical settings reveals the demand from multiple justice systems and jurisdictions for legal precedent retrieval methods. It also embeds a challenge for the adaptability of such models and raises questions about their universal applicability, thereby serving as a compelling avenue for future research.

5. Discussion

5.1. RQ1: How Did Researchers Address the Challenge of Automatically Identifying Prior Relevant Cases, and What Methods Have Been Used in the Screened Studies? Researchers have employed various approaches to identify relevant prior cases automatically. Such methodologies were extensively described in Section 4.2. Two major paradigms emerge.

5.1.1. Manual Knowledge Engineering (2000–2009). Initially, the research relied on domain expertise to create manually defined attributes and dimensions representing each case. The first techniques included matching individual cases to story patterns [55, 56] and content vectors [54]. Keywords, facts, and other elements were manually chosen, which made these methods less scalable.

5.1.2. Artificial Intelligence Wave (2016 Onwards). The subsequent wave saw the use of ML and NLP techniques to handle the task. Methods like TF-IDF vectors [47], regular expression combined with Doc2Vec embeddings [43], Named Entity Recognition [39], and Autoencoders [40] were employed. These methods significantly improved scalability and accuracy.

5.2. RQ2: What Are the Most Promising Methods for the Automated Search of Legal Precedents, and What Research Gaps Exist? Based on the review, the most promising automated legal precedent retrieval techniques seem to lie within NLP and ML. While early efforts were primarily built on manual knowledge engineering, the shift towards NLP and ML has heralded promising advances. Document embeddings such as Doc2Vec and topic modeling [29, 43] have effectively represented textual data. Additionally, methods such as Top2Vec combined with BM25 ranking functions have been highlighted as outperforming approaches in a recent precedent retrieval challenge [37, 41].
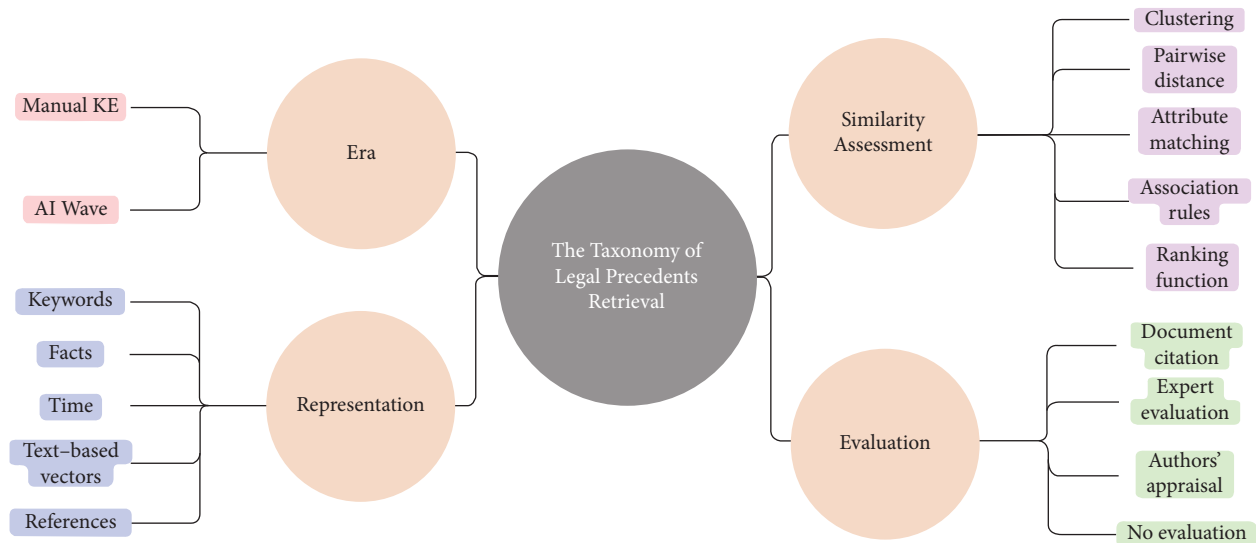
FIGURE 12: Taxonomy of the legal precedents retrieval field.

Autoencoders, especially when coupled with LSTMs, have been claimed to be effective for training on unstructured text, offering both speed and accuracy [40]. NER has also been employed as a preprocessing step to extract meaningful data from legal documents [39]. Conventional vectorization techniques, such as TF-IDF, still maintain robust performance in similarity measurements, as evidenced by the findings of Mandal et al. [29].

Despite the progress, there are some glaring research gaps in the field. One significant gap is the continuation of studies employing AI to retrieve similar cases. The lack of consensus about the most effective methodology is evident. No definitive technique stands out as the best for all types of legal documents and jurisdictions, indicating room for more comparative studies. Additionally, no work has analyzed the effect of text preprocessing on the results, which may prove to be decisive considering that, so far, no methodology presents superior performance.

Moreover, there is a lack of uniform benchmarks for comparing different methodologies. While Mandal et al. [29] conducted a comprehensive study on various document representation techniques and similarity measures, such exhaustive evaluations are not commonly found in the existing literature, and some studies did not evaluate the proposed methods. Larger corpora and expert-supplied ground truth should be incorporated into new studies, preferably in new legal contexts. It is crucial to notice that the mentioned works were conducted on very small corpora, with the most comprehensive study comparing similarity measurement methods using only 50 pairs of documents as the gold standard.

Temporal elements in cases, although explored, have not been comprehensively studied or incorporated into existing models, leaving a gap in our understanding of how temporal relationships between events might affect the relevance of legal precedents. Another aspect that has not yet been thoroughly evaluated is the applicability of models across different legal domains and jurisdictions, considering studies focused only on a single corpus each.

Moreover, while there has been a move towards utilizing sophisticated techniques like NLP and ML, it is worth noting that conventional methods like TF-IDF are still highly effective. This effectiveness leaves an open question about the actual incremental benefits of using more complex methods, which remains underexplored.

Finally, there is plenty of room for assessing the effects of contextual understanding on this topic. While techniques considering semantics have been effective, there is still a gap in understanding the deep context of legal language, which neural embeddings like BERT and other transformers may fulfill.

In summary, the field has seen a promising shift towards automation using AI techniques, but many questions and gaps remain, suggesting avenues for further research.

5.3. RQ3: What Is the Taxonomy of Existing Methods, and What Is Their Mainstream? The taxonomy of methods for automating the identification of legal precedents has evolved significantly over the years, with two distinct eras emerging, as described in Sections 4.2.1 and 4.2.2. In terms of operational pipeline, the taxonomy can be broadly divided into three main phases: representation, which involves encoding legal cases based on predefined attributes; similarity assessment, the techniques employed to measure the resemblance between cases; and evaluation, the final stage of measuring the effectiveness of the retrieval techniques.

As per the existing literature, there has been a mainstream shift towards AI-based methods in recent years, with a particular emphasis on vectorial text representations and advanced similarity assessment techniques like ranking functions and pairwise distance measures. The comprehensive assessment by Mandal et al. [29] stood out as a seminal work that spanned over fifty-six unique combinations of document representation techniques and similarity measures, validating the efficacy of conventional vectorization techniques while indicating areas for improvement.

TABLE 3: Studies are classified according to the proposed taxonomy.

Reference	Era	Representation	Similarity assessment	Evaluation
[29]	AI wave	Text-based vectors	Pairwise distance	Expert evaluation
[37]	AI wave	Text-based vectors	Ranking function	Document citation
[38]	AI wave	Text-based vectors	Ranking function	Document citation
[39]	AI wave	Text-based vectors	Ranking function	Document citation
[40]	AI wave	Text-based vectors	Attribute matching	No evaluation
[41]	AI wave	This study described the Artificial Intelligence for Legal Assistance (AILA) track at the 2019 FIRE Conference and did not include representation or similarity assessment stages		Document citation
[42]	AI wave	References	Association rules	No evaluation
[43]	AI wave	Text-based vectors	Pairwise distance	No evaluation
[44]	AI wave	Text-based vectors	Ranking function	Authors' appraisal
[45]	Manual KE	Facts	The authors did not describe the similarity assessment method in detail	No evaluation
[46]	AI wave	Facts, time, references	Clustering	Expert evaluation
[47]	AI wave	Facts, text-based vectors	Clustering	No evaluation
[48]	Manual KE	Facts, time	Attribute matching	Document citation
[49]	Manual KE	Facts	The authors did not describe the similarity assessment method in detail	No evaluation
[50]	Manual KE	Facts	Attribute matching	Authors' appraisal
[51]	Manual KE	Facts	Clustering	No evaluation
[52]	Manual KE	Facts	Clustering	Authors' appraisal
[53]	Manual KE	Keywords	Clustering	Expert evaluation
[54]	Manual KE	Facts	Pairwise distance	Document citation
[55]	Manual KE	Keywords, facts	Clustering	Authors' appraisal
[56]	Manual KE	Keywords, facts	Clustering	Authors' appraisal

TABLE 4: Variables resulting from merging of datasets.

| Scopus variable name	WoS variable name	Resulting variable
Authors	Authors	Authors
Author(s) ID	Researcher ID	Author(s) ID
Title	Article title	Title
Year	Publication year	Year
Source title	Source title	Source title
Volume	Volume	Volume
Issue	Issue	Issue
Cited by	Times cited, all databases	Cited by
DOI	DOI	DOI
Affiliations	Affiliations	Affiliations
Abstract	Abstract	Abstract
Author keywords	Author keywords	Author keywords
Correspondence address	Addresses	Correspondence address
Publisher	Publisher	Publisher
ISSN	ISSN	ISSN
ISBN	ISBN	ISBN
Document type	Document type	Document type
—	—	Source
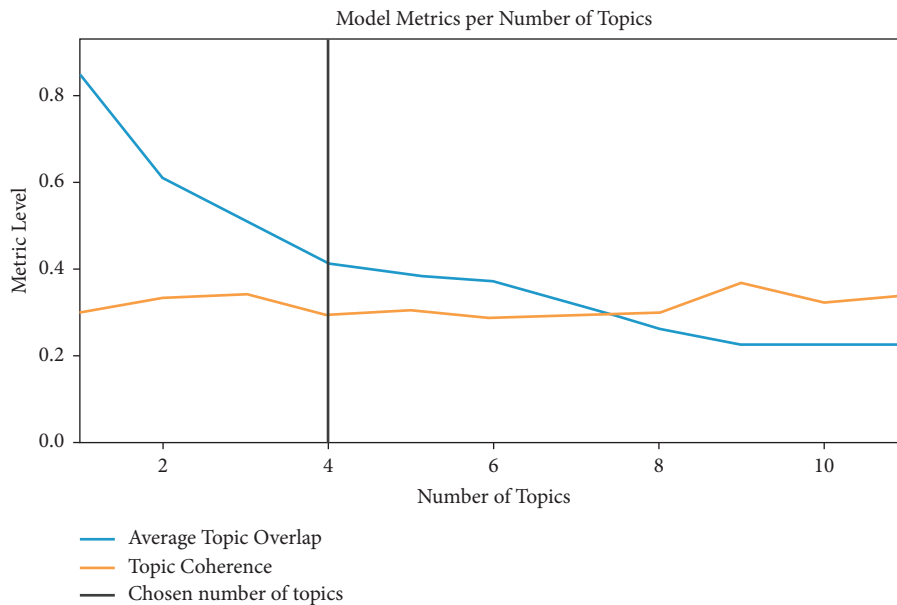
FIGURE 13: A comparison between overlap and coherence for different number of topics.

In terms of evaluation, the most rigorous methods involve either document citation or expert evaluation, with a less rigorous yet prevalent approach involving authors' appraisal of a handful of case pairs. However, it is noteworthy that some studies have foregone the evaluation phase entirely, possibly due to the lack of an annotated corpus or other constraints, which leaves a gap in validating the effectiveness of the proposed methods.

The trend has moved towards more automated, scalable, and data-driven approaches, leaving behind the past's cumbersome, manual, and less scalable methods. Nevertheless, there remains considerable diversity in both representation and similarity assessment techniques, suggesting that the field is still in a state of active exploration and development.

5.4. RQ4: What Are the Research Domain's Most Influential Journals and Authors? As described in Table 2, the most cited paper had sixty-seven citations. It was published in *Artificial Intelligence*. The following two papers with more citations were published in the *Journal of Information Science and Engineering* and *Expert Systems with Applications*. This set of publications is concentrated between the years 2000 and 2004. In line with the time gap we observe in the literature associated with this field, the following three most cited papers were published between 2017 and 2021 on *IEEE Access, Artificial Intelligence and Law,* and *ACM International Conference Proceeding Series*.

Except *CEUR Workshop Proceedings*, all sources had only one paper considered relevant to this literature review (Figure 9). Consequently, it is still not possible to say that there is
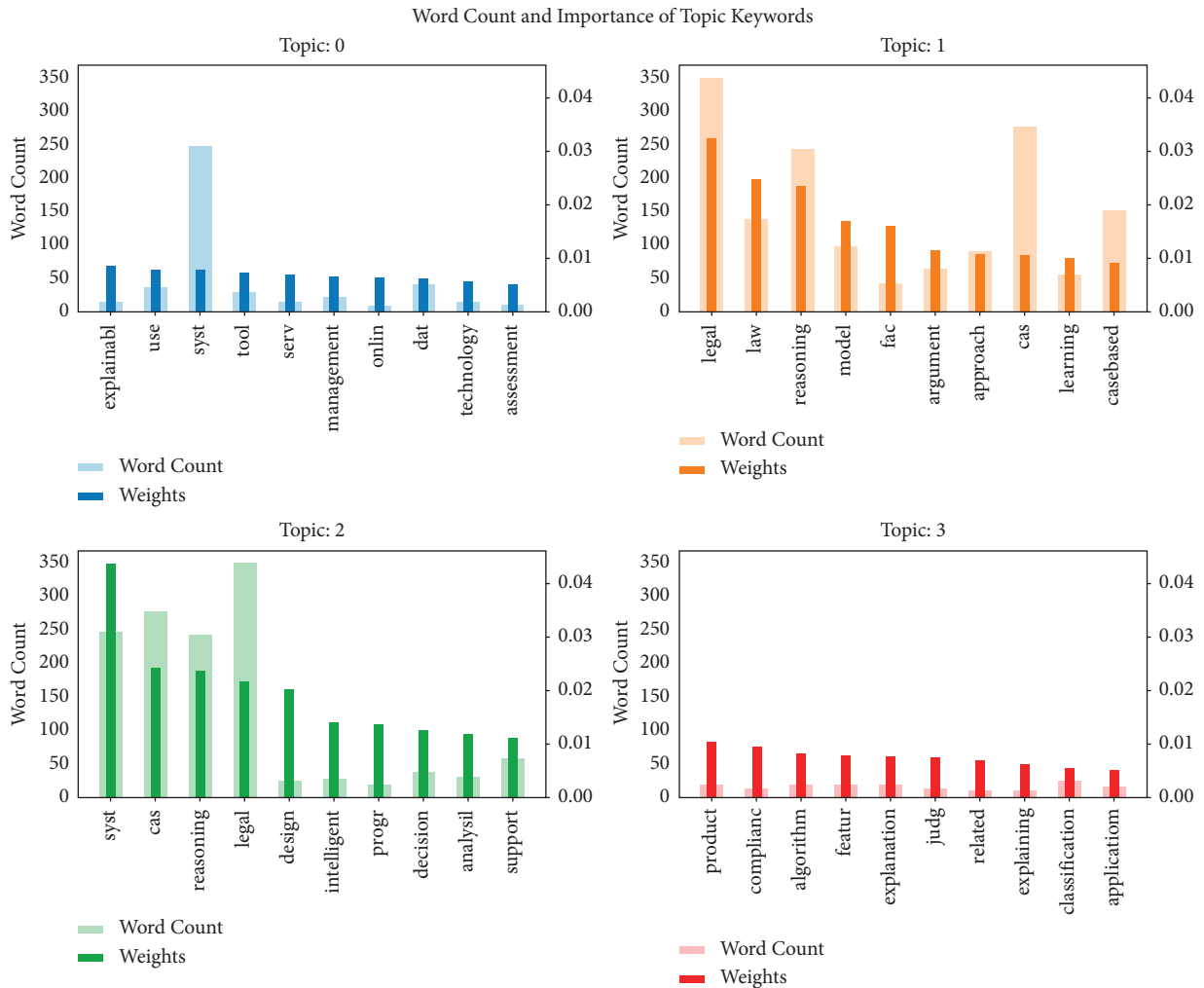
Figure 14: Word count and importance of keywords for each topic.

a source considered more influential than the others in the field of automation of identification of legal precedents.

Bruce M. McLaren (sixty-seven citations) is the author with the most citations. However, this author cannot be considered influential in this field as the paper to which the citations correspond was published in 2003. The same applies to Chao L. Liu, Cheng T. Chang, and Jim H. Ho, with forty citations, and Mohamed T. Elhadi, cited thirty times. The authors who published papers more recently and that obtained the most citations are Kripabandhu Ghosh (seventeen citations), Saptarshi Ghosh (seventeen citations), and Arpan Mandal (eleven citations). The literature in this field is also very sparse regarding influential authors, with no author with more than two relevant publications.

5.5. RQ5: What Data Have Been Used in Existing Research? In the academic research included in this literature review, a wide variety of data sources have been utilized, spanning multiple geographical and legal contexts. The diversity in data is indicative of the field's global scope and its attempts to adapt to various legal systems. Below is a breakdown of the geographical scopes and corresponding legal contexts used.

5.5.1. United States. Early studies primarily used US-based data [54–56]. These studies focused on areas such as bankruptcy law and professional ethics cases.

5.5.2. Taiwan. The research includes Taiwanese data in criminal summary judgments [53].

5.5.3. Algeria. One study has delved into Islamic legislation [51].

5.5.4. Costa Rica. Research has been conducted on judicial declarations of abandonment in Costa Rican Juvenile Courts [45].

5.5.5. India. A significant emphasis has been placed on Indian court data in recent studies. Research in this context includes various aspects of the Indian justice system, including the Supreme Court and different high courts [29, 37–39, 41–44].

TABLE 5: Relevant publications excluded during the eligibility screening phase (by topic selection).

LDA topic	Reference	Year	Research topic
0	[97]	2018	This conference paper proposed a precedent retrieval system using ontology. Ontology is a set of concepts and categories in a domain representing their properties and relations.
0	[98]	2017	This study summarizes technological advancements to law practitioners, including AI and non-AI tools
1	[99]	2015	This paper proposed a system that responded to a user's questions by evaluating semantical similarity with previously answered inquiries
1	[100]	2009	This conference paper introduced an agent-based architecture for a tool that provides users with a list of potential case outcomes based on legal precedents
1	[64]	2008	This conference paper examined existing approaches to legal information retrieval, namely, manual KE and NLP
1	[101]	2003	This conference paper modeled legal precedents based on their defining properties: the relevance of a precedent is dictated by the legal support under which it was decided
3	[102]	2003	This conference paper used criminal judgments to investigate the efficacy of artificially created criteria for classification problems
1	[103]	2001	This proceedings paper proposed a factor-based similarity measure between legal cases
1	[104]	2000	This conference paper detailed a CBR system based on fuzzy logic developed for the United Nations Convention on Contracts for the International Sale of Goods

TABLE 6: Relevant and irrelevant documents identified in each cluster.

Cluster	Type		
	Relevant	Irrelevant	Total
Topic 0	2	12	14
Topic 1	6	33	39
Topic 2	21	19	40
Topic 3	1	7	8

TABLE 7: Confusion matrix resulting from the eligibility screening method.

Eligibility screening		Identified	
		Relevant	Irrelevant
Actual	Relevant	21	9
	Irrelevant	19	52

5.5.6. China. One study has also investigated Chinese statutes and judicial cases [46].

5.5.7. Germany. One unusual dataset employed was a mixed-language customer support ticket dataset from an automotive company in Germany [40].

Finally, multiple types of documents have been employed by researchers. Most studies used case documents from courts, while some research incorporated statutes [43, 46] or even customer support tickets [40], a non-traditional source for this field.

Using data from multiple justice systems and jurisdictions embeds a challenge for the adaptability of legal precedent retrieval models and raises questions about their universal applicability. This diversity serves as a compelling avenue for future research.

5.6. RQ6: Are There Real-World Applications of This Topic? In synthesizing the range of research approaches discussed in Section 4.2, it is notable that none of the studies in this literature review report real-world implementations of models designed for legal precedents retrieval. This gap presents a central direction for future research. While various models have been proposed, tested, and adapted to

tackle different aspects of precedent retrieval—from early KE systems to more recent systems based on neural network embeddings—their applications have remained mainly theoretical. The absence of real-world case studies in applying these models to actual legal systems raises pertinent questions: why have these models not yet been adopted in judicial settings? Are there inherent limitations in the current models that deter their application, or are there external factors such as ethical, legal, or operational constraints?

This situation underscores the need for the following research phase to focus on refining algorithms and methods and transitioning from theoretical frameworks to applied solutions. Implementing these models in real-world scenarios could provide insights into their efficacy, scalability, and limitations. Similarly, in-depth investigations into the challenges preventing the implementation can yield valuable lessons. Both avenues would significantly contribute to the field's maturity, ensuring that future advancements are theoretically robust and practically applicable.

6. Conclusions

This literature review aimed to synthesize the state of research on automating the identification of legal precedents, focusing on the techniques employed, influential journals, and authors while assessing the effectiveness of these

techniques and existing research gaps. Using textual mining methods for semi-automating the review process, this study identified 70% of relevant publications, thereby reducing the number of studies that required in-depth analysis by 82.5%. This highlights the transformative potential of automation in expediting the creation of literature reviews, particularly in a field as intricate as legal studies.

The findings suggest that the automation of legal precedent retrieval is increasingly leveraging text processing and machine learning techniques. However, the field is nascent, with techniques often tested in isolated studies and specific contexts. Most alarmingly, the limited validation samples used in these studies and the lack of a noteworthy advantage from advanced text embeddings underscore the challenges in establishing universally applicable models. This leads to the realization that automating legal precedent retrieval is not a one-size-fits-all proposition but requires nuanced approaches tested across different legal contexts and datasets.

Given these complexities, there is an urgent call for more comprehensive research. Future work should focus on employing more extensive and diverse corpora, including expert-provided ground truths, to accurately evaluate the efficacy and adaptability of various strategies accurately. The application of emerging technologies, such as LSTM networks and transformers, also merits rigorous exploration to understand their potential in legal precedent retrieval.

However, the limitations of this literature review must be acknowledged. The insights were drawn from a relatively small set of 21 papers, and the human cognition involved in selecting databases, queries, and eligibility criteria could have influenced the findings. Notably, the lack of snowballing in related work identification due to varied citation structures could also be a limitation. As such, future research could benefit from investigating citation networks or employing automated methods for selecting studies based on objective criteria for literature reviews.

In summary, while this review marks a significant step in understanding the possibilities and limitations of automating legal precedent retrieval, it also serves as a clarion call for further, more expansive research to grasp the opportunities and challenges that lie ahead fully.

Appendix

A. Extraction and Preprocessing of Data

A comma-separated values (CSV) file was created from the Scopus results, and WoS results were stored in an Excel spreadsheet for analysis. The relevance of the search terms was confirmed by validating well-known studies on the topic of interest.

We combined both datasets into a single table. Given the difference between variables in the two sources, only the coincident variables in both databases were kept. A "Source" feature was engineered to specify the database (WoS or Scopus). The set of variables that resulted from this procedure is found in Table 4.

We could determine the country from the writers' affiliations and mailing addresses for one hundred eighty-one entries. It represented 79% of the dataset. When it was impossible to separate the nation from the affiliation of the initial writers, the correspondence address was utilized as a substitute. However, the country of origin could not be determined for forty-eight of the initial samples.

The next stage eliminated duplicate samples, focusing primarily on the document titles. These were preprocessed to remove such elements to prevent capitalization, punctuation, and white spaces from harming the identification of publications with the same titles. Sixty-three duplicates were eliminated due to this operation, and the Scopus samples were kept whenever possible.

The Digital Object Identifiers (DOIs) and the abstracts underwent the same process. In the case of DOIs, a duplicate sample was found, which was removed. The comparison between abstracts found no duplicates. The titles and abstracts later underwent a more sophisticated process comprising tokenization and stemming. Tokenization is the text segmentation into basic units (tokens) such as words and punctuation, while stemming accounts for reducing words to their root form based on preestablished rules. For example, a rule may state that any expression with -ing as a suffix will be condensed by suffix removal [72].

The reason for deepening the title and abstract similarity check was that two publications in the dataset could represent iterations of the same research, such as a journal article originating from a conference paper. For this reason, a method was used to filter out articles with titles or abstracts that were remarkably similar.

The abstracts and titles were compared separately. Each of these variables was initially vectorized into bigrams and unigrams. Every publication represented an element of a bag-of-words model. This text embedding model assumes that a text is nothing more than a histogram of the words it contains. Thus, words' order or context is not considered [73]. In the vectorial representation, the term frequency-inverse document frequency (TF-IDF) was used to normalize word occurrence frequency based on the number of publications where each word is present [74, 75]. As a rule for removing recurring or infrequent words, ones appearing in more than 80% of the corpus or appearing in one document were ignored.

After that, the cosine similarity between pairs of titles and abstracts was determined. One of the publications was discarded when the pair had a similarity above 0.8. Indeed, two duplicate documents were found when titles were evaluated, and three duplicates were found using abstracts. Three duplicates were updates of the same paper released in the subsequent year, and the two remaining were studies that appeared in different publishers.

B. Topic Selection Process

We compared coherence and overlap measures to obtain the optimal number of topics. Additionally, it should be recognized that model interpretation becomes increasingly challenging as the number of topics increases. Coherence evaluates the level of semantic similarity among words with the highest scores in a topic [76]. In contrast, the occurrence

of a term in several topics and overlapping topics due to a few unique terms may point to less valuable (less coherent) models [77]. Figure 13 compares topic coherence and topic overlap for different number of topics.

The coherence remains approximately constant while the average topic overlaps continuously decay. There is also a reduction in the gradient ("elbow") of the topic overlap for two or four topics and after nine topics. Considering the problematic interpretability of many topics and the significant reduction of topic overlap from two to four topics, the authors opted to use four topics. The preprocessed versions of each document's abstract, title, and keywords were then utilized to construct a corpus, which was then subjected to LDA.

The weights of the top ten terms discovered in the four topics are shown in Figure 14. Topic 0 was characterized by a higher weight attributed to the "explainabl" token. It means that documents more likely to be associated with Topic 0 are possibly related to the explainability of models' outcomes.

Topic 1 was associated with case-based reasoning. Still, the prevalence of general terms such as "legal" and "law" associated the topic with the legal sector while hampering more specific intuitions. Topic 2's most important word was "syst," and the top ten words included "legal," "decision," and "support," indicating some association with decision support systems in the legal environment. Finally, Topic 3's most essential words included "product," "complianc," "algorithm," "explanation," "explaining," "classification," and "application." It indicated some degree of overlapping with Topic 0.

The abstracts of the documents were then checked to validate these initial hypotheses. Topic 3, for instance, included papers on electric motor compliance verification, new product development, and the prediction of case-based legal arguments. Topic 2, in turn, included the paper treated as a referential work authored by Mandal et al. [29]. As a result, the highest probability associated with Topic 2, the topic with the most explicit link with the research subject, was chosen as the eligibility criterion for each document. It resulted in forty eligible documents.

C. Publications Excluded during the Full-Text Screening Phase

D. Eligibility Screening Method Validation

Although screening resulted in recognizing twenty-one publications eligible for the literature review, we wanted to validate the clustering through topic modeling as an effective method to assess document eligibility.

Thus, the abstracts of documents associated with the other topics were also evaluated to identify false negatives, i.e., publications relevant to the review objectives, which could be considered in a systematic literature review.

Among the sixty-one publications related to Topic 0, Topic 1, or Topic 3, it was possible to identify nine additional publications relevant to clarifying our research questions.

They are listed in Table 5, and Table 6 details the number of publications (relevant or irrelevant) by topic. The relevant publications in Topic 2 are true positives identified with the clustering method. On the other hand, the relevant publications associated with other topics are considered false negatives.

Next, it was possible to formulate the confusion matrix of Table 7 and calculate the recall value, respectively, 0.70. Recall was the performance metric because it measures the ability to identify the maximum number of relevant studies. Recall (recall is also known as sensitivity or true positive rate) is the proportion of true positives to all positive samples (sum of true positives and false negatives).

To maintain consistency with the method used in this rapid semi-automated review, only studies associated with the chosen topic (Topic 2) were considered when answering the research questions.

Data Availability

The data used in this study are results of searches in electronic scientific research databases. Such data are available in the databases mentioned in the study.

Disclosure

An earlier version of this paper has been presented as a preprint in the following link: https://www.researchsquare.com/article/rs-2292464/v1.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] A. R. Gomez, "Demand side justice," *Georgetown Journal on Poverty Law and Policy*, no. 3, pp. 411–436, 2021.

[2] D. L. Rhode, *Access to Justice*, Oxford University Press, Oxford, UK, 2004.

[3] O. Popova, R. Maroz, and M. A. Q. Gámez, *The Undeniable Benefits of Court Automation*, Let's Talk Development, Dubai, United Arab Emirates, 2021.

[4] R. Susskind, "The future of courts," *Practice*, vol. 6, no. 5, 2020.

[5] G. Guillaume, "The use of precedent by international judges and arbitrators," *Journal of International Dispute Settlement*, vol. 2, no. 1, pp. 5–23, 2011.

[6] A. Rigoni, "Common-law judicial reasoning and analogy," *Legal Theory*, vol. 20, no. 2, pp. 133–156, 2014.

[7] V. Fon and F. Parisi, "Judicial precedents in civil law systems: a dynamic analysis," *International Review of Law and Economics*, vol. 26, no. 4, pp. 519–535, 2006.

[8] J. L. Kolodner, "An introduction to case-based reasoning," *Artificial Intelligence Review*, vol. 6, no. 1, pp. 3–34, 1992.

[9] A. Roth, "Case-based reasoning in the law: a formal theory of reasoning by case comparison," Doctoral Thesis, Universiteit Maastricht, Maastricht, Netherlands, 2003.

[10] H. Silva, N. António, and F. Bacao, "A rapid semi-automated literature review on legal precedents retrieval," in *Progress in Artificial Intelligence*, G. Marreiros, B. Martins, A. Paiva, B. Ribeiro, and A. Sardinha, Eds., vol. 13566, pp. 53–65, Springer International Publishing, Berlin, Germany, 2022.

[11] H. Mentzingen, N. António, and F. Bacao, "Automation of legal precedents retrieval: findings from a rapid literature review," 2022, https://www.researchsquare.com/article/rs-2292464/v1.

[12] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *Journal of Clinical Epidemiology*, vol. 62, no. 10, pp. 1006–1012, 2009.

[13] N. Jahan, S. Naveed, M. Zeshan, and M. A. Tahir, "How to conduct a systematic review: a narrative literature review," *Cureus*, vol. 8, no. 11, 2016.

[14] L. Bero, G. Busuttil, C. Farquhar et al., "Measuring the performance of the cochrane library," *Cochrane Database of Systematic Reviews*, vol. 12, Article ID ED000048, 2012.

[15] R. Ganann, D. Ciliska, and H. Thomas, "Expediting systematic reviews: methods and implications of rapid reviews," *Implementation Science*, vol. 5, no. 1, p. 56, 2010.

[16] S. Khangura, K. Konnyu, R. Cushman, J. Grimshaw, and D. Moher, "Evidence summaries: the evolution of a rapid review approach," *Systematic Reviews*, vol. 1, no. 1, pp. 10–19, 2012.

[17] G. Tsafnat, P. Glasziou, M. K. Choong, A. Dunn, F. Galgani, and E. Coiera, "Systematic review automation technologies," *Systematic Reviews*, vol. 3, no. 1, p. 74, 2014.

[18] C. Garritty, G. Gartlehner, B. Nussbaumer-Streit et al., "Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews," *Journal of Clinical Epidemiology*, vol. 130, pp. 13–22, 2021.

[19] A. C. Tricco, E. V. Langlois, and S. E. Straus, *Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide*, vol. 119, World Health Organisation, Geneva, Switzerland, 2017.

[20] A. Stevens, C. Garritty, M. Hersi, and D. Moher, "Developing prisma-rr, a reporting guideline for rapid reviews of primary studies (protocol)," 2018, http://www.equator-network.org/wp-content/uploads/2018/02/PRISMA-RR-protocol.pdf.

[21] R. van Dinter, B. Tekinerdogan, and C. Catal, "Automation of systematic literature reviews: a systematic literature review," *Information and Software Technology*, vol. 136, Article ID 106589, 2021.

[22] L. Feng, Y. K. Chiam, and S. K. Lo, "Text-mining techniques and tools for systematic literature reviews: a systematic literature review," in *Proceedings of the Asia-Pacific Software Engineering Conference, APSEC*, pp. 41–50, Nara, Japan, December, 2018.

[23] J. Zimmerman, R. E. Soler, J. Lavinder et al., "Iterative guided machine learning-assisted systematic literature reviews: a diabetes case study," *Systematic Reviews*, vol. 10, no. 1, p. 97, 2021.

[24] N. António, A. D. Almeida, and L. Nunes, "Predictive models for hotel booking cancellation: a semi-automated analysis of the literature," *Tourism and Management Studies*, vol. 15, no. 1, pp. 7–21, 2019.

[25] V. J. Schweizer, J. H. Kurniawan, and A. Power, "Semi-automated literature review for scientific assessment of socioeconomic climate change scenarios," in *Proceedings of the Companion Proceedings of the Web Conference 2022*, pp. 789–799, Lyon, France, August, 2022.

[26] D. Golinelli, A. G. Nuzzolese, F. Sanmarchi et al., "Semi-automatic systematic literature reviews and information extraction of COVID-19 scientific evidence: description and preliminary results of the COKE project," *Information*, vol. 13, no. 3, p. 117, 2022.

[27] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA, USA, 2009.

[28] Cornell University Law School, "Precedent," Legal Information Institute, New York, NY, USA, 2020, https://www.law.cornell.edu/wex/precedent.

[29] A. Mandal, K. Ghosh, S. Ghosh, and S. Mandal, "Unsupervised approaches for measuring textual similarity between legal court case reports," *Artificial Intelligence and Law*, vol. 29, no. 3, pp. 417–451, 2021.

[30] P. Mongeon and A. Paul-Hus, "The journal coverage of Web of Science and Scopus: a comparative analysis," *Scientometrics*, vol. 106, no. 1, pp. 213–228, 2016.

[31] L. Oesper, D. Merico, R. Isserlin, and G. D. Bader, "WordCloud: a Cytoscape plugin to create a visual semantic summary of networks," *Source Code for Biology and Medicine*, vol. 6, no. 1, p. 7, 2011.

[32] G. K. Frampton, B. Livoreil, and G. Petrokofsky, "Eligibility screening in evidence synthesis of environmental management topics," *Environmental Evidence*, vol. 6, no. 1, pp. 27–13, 2017.

[33] C. B. Asmussen and C. Møller, "Smart literature review: a practical topic modelling approach to exploratory literature review," *Journal of Big Data*, vol. 6, no. 1, pp. 93–18, 2019.

[34] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. 4–5, pp. 993–1022, 2003.

[35] R. Rehurek and P. Sojka, *Gensim-python Framework for Vector Space Modelling*, vol. 3, NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, 2011.

[36] D. J. Hu, *Latent Dirichlet Allocation for Text, Images, and Music*, University of California, San Diego, CA, USA, 2009.

[37] J. Arora, T. Patankar, A. Shah, and S. Joshi, "Artificial intelligence as legal research assistant," in *Proceedings of the CEUR Workshop Proceedings, 2826*, pp. 60–65, Hyderabad, India, December, 2020.

[38] G. M. G. M. Di Nunzio, "A study on lemma vs stem for legal information retrieval using R tidyverse. IMS UniPD @ AILA 2020 Task 1," in *Proceedings of the CEUR Workshop Proceedings, 2826*, pp. 54–59, Hyderabad, India, December, 2020.

[39] R. More, J. Patil, A. Palaskar, and A. Pawde, "Removing named entities to find precedent legal cases," in *Proceedings of the CEUR Workshop Proceedings, 2517*, pp. 13–18, Kolkata, India, December 2019.

[40] K. Amin, S. Kapetanakis, K. D. Althoff, A. Dengel, and M. Petridis, "Cases without borders: automating knowledge acquisition approach using deep autoencoders and siamese networks in case-based reasoning," in *Proceedings of the International Conference on Tools with Artificial Intelligence, ICTAI*, pp. 133–140, Portland, OR, USA, November, 2019.

[41] P. Bhattacharya, K. Ghosh, S. Ghosh et al., "FIRE 2019 AILA track: artificial intelligence for legal assistance," in *Proceedings of the ACM International Conference Proceeding Series, 2517*, pp. 4–6, New York, NY, USA, February, 2018.

[42] A. M. A. M. Nair and R. S. R. S. Wagh, "Similarity analysis of court judgements using association rule mining on case citation data-a case study," *International Journal of Engineering Research and Technology*, vol. 11, no. 3, pp. 373–381, 2018.

[43] Y. H. Y. H. Kulkarni, R. Patil, and S. Shridharan, "Detection of catchphrases and precedence in legal documents," in *Proceedings of the CEUR Workshop Proceedings, 2036*, pp. 86–89, Kolkata, India, December, 2017.

[44] E. Thuma and N. P. N. P. Motlogelwa, "On the importance of legal catchphrases in precedence retrieval," in *Proceedings of the CEUR Workshop Proceedings, 2036*, pp. 92–94, Kolkata, India, December, 2017.

[45] L. R. R. Oconitrillo and A. De La Ossa Osegueda, "Business intelligence model to support a judge's decision making about legal situations," in *Proceedings of the 2016 IEEE 36th Central American and Panama Convention, CONCAPAN 2016*, pp. 1–5, San Jose, Costa Rica, November, 2017.

[46] N. N. Zhang, Y. F. Y.-F. Pu, S. Q. S.-Q. Yang, J.-L. J. L. Zhou, and J. K. J.-K. J. K. J.-K. Gao, "An ontological Chinese legal consultation system," *IEEE Access*, vol. 5, pp. 18250–18261, 2017.

[47] V. B. Eyorokon, B. Bengfort, U. S. Panjala, and M. T. Cox, "Goal trajectories for knowledge investigations," in *Proceedings of the CEUR Workshop Proceedings, 1815*, pp. 202–211, Atlanta, GA, USA, December, 2016.

[48] B. M. McLaren and K. D. Ashley, "Can temporal representation and reasoning make a difference in automated legal reasoning? Lessons from an AI-based ethical reasoner," in *Proceedings of the International Conference on Artificial Intelligence and Law*, pp. 229–238, Pittsburgh, PA, USA, June, 2011.

[49] V. Raman and A. Palanissamy, "Computer aided legal support systems: methodology to automatically convert the legal text into legal cases for building case base reasoning," in *Proceedings of the 2009 International Conference On Computer Engineering And Applications Issue International Conference on Computer Engineering and Applications*, P. S. Sandhu and Y. Xie, Eds., pp. 283–287, Washington, DC, USA, January, 2009.

[50] V. Raman and A. Palanissamy, "Computer aided legal support system: an initial framework retrieving legal cases by case base reasoning approach," in *Proceedings of the 2008 International Conference on Innovations in Information Technology, IIT 2008*, pp. 317–321, Al Ain, United Arab Emirates, December, 2008.

[51] N. Nouaouria, F. Atil, M. T. T. Laskri, D. Bouyaya, and A. H. A. H. Amari, "A case based tool as intelligent assistance to Mufti," *Arabian Journal for Science and Engineering*, vol. 31, no. 1 B, pp. 75–87, 2006.

[52] R. Wang and Y. Zeng, "Nonlinear nearest-neighbour matching and its application in legal precedent retrieval," in *Proceedings of the 3rd International Conference on Information Technology and Applications, ICITA 2005*, pp. 341–346, Sydney, Australia, July, 2005.

[53] C.-L. Liu, C.-T. Chang, and J.-H. Ho, "Case instance generation and refinement for case-based criminal summary judgments in Chinese," *Journal of Information Science and Engineering*, vol. 20, no. 4, pp. 783–800, 2004.

[54] B. M. McLaren, "Extensionally defining principles and cases in ethics: an AI model," *Artificial Intelligence*, vol. 150, no. 1–2, pp. 145–181, 2003.

[55] M. T. Elhadi, "Using statutes-based ir drive legal cbr," *Applied Artificial Intelligence*, vol. 15, no. 6, pp. 587–600, 2001.

[56] M. T. Elhadi, "Bankruptcy support system: taking advantage of information retrieval and case-based reasoning," *Expert Systems with Applications*, vol. 18, no. 3, pp. 215–219, 2000.

[57] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems*, vol. 2, pp. 3111–3119, Lake Tahoe NV, USA, December, 2013.

[58] K. Cho, B. Van Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the Conference EMNLP 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, Doha, Qatar, October, 2014.

[59] L. Yao, Z. Pan, and H. Ning, "Unlabeled Short text similarity with LSTM encoder," *IEEE Access*, vol. 7, pp. 3430–3437, 2019.

[60] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, pp. 5999–6009, 2017.

[61] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proceedings of the ACL 2018 56th Annual Meeting of the Association for Computational Linguistics*, vol. 1, Melbourne, Australia, July, 2018.

[62] X. P. Qiu, T. X. Sun, Y. G. Xu, Y. F. Shao, N. Dai, and X. J. Huang, "Pre-trained models for natural language processing: a survey," *Science China Technological Sciences*, vol. 63, no. 10, pp. 1872–1897, 2020.

[63] K. D. Forbus, D. Gentner, and K. Law, "Mac fac: a model of similarity-based retrieval," *Cognitive Science*, vol. 19, no. 2, pp. 141–205, 1995.

[64] K. T. T. Maxwell and B. Schafer, "Concept and context in legal information retrieval," *Frontiers in Artificial Intelligence and Applications*, vol. 189, no. 1, pp. 63–72, 2008.

[65] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *. Proceedings of the 31st International Conference on Machine Learning, ICML 2014*, vol. 4, pp. 2931–2939, Beijing, China, June, 2014.

[66] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning workshop*, pp. 37–50, Washington, DC, USA, July, 2012.

[67] A. Mansouri, L. S. Affendey, and A. Mamat, "Named entity recognition approaches," *Journal of Computer Science*, vol. 8, no. 2, pp. 339–344, 2008.

[68] S. Robertson and H. Zaragoza, "The probabilistic relevance framework: BM25 and beyond," *Foundations and Trends in Information Retrieval*, vol. 3, no. 4, pp. 333–389, 2009.

[69] D. Angelov, "Top2vec: distributed representations of topics," pp. 1–25, 2020, http://arxiv.org/abs/2008.09470.

[70] I. Chalkidis, "Law2vec: legal word embeddings," 2018, https://archive.org/details/Law2Vec.

[71] G. Amati and C. J. Van Rijsbergen, "Probabilistic models of information retrieval based on measuring the divergence from randomness," *ACM Transactions on Information Systems*, vol. 20, no. 4, pp. 357–389, 2002.

[72] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media, Sebastopol, CA, USA, 2009.

[73] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, Elsevier, Amsterdam, The Netherlands, 2nd edition, 2020.

[74] H. P. Luhn, "A statistical approach to mechanized encoding and searching of literary information," *IBM Journal of Research and Development*, vol. 1, no. 4, pp. 309–317, 1957.

[75] K. Spärck Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, no. 1, pp. 11–21, 1972.

[76] D. O'Callaghan, D. Greene, J. Carthy, and P. Cunningham, "An analysis of the coherence of descriptors in topic modeling," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5645–5657, 2015.

[77] S. Arora, R. Ge, and A. Moitra, "Learning topic models-going beyond SVD," in *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pp. 1–10, New Brunswick, NJ, USA, October, 2012.

[78] E. B. P. Calderon, M. A. Martinez, T. T. C. Martinez, and L. X. C. Ona, "Case-based legal reasoning and inductive reasoning: determination of the input parameters using neutrosophic AHP," *Neutrosophic Sets and Systems*, vol. 44, pp. 71–81, 2021.

[79] W. Fungwacharakorn, K. Tsushima, and K. Satoh, "Resolving counterintuitive consequences in law using legal debugging," *Artificial Intelligence and Law*, vol. 29, no. 4, pp. 541–557, 2021.

[80] D. Mohan and L. R. Nair, "Probit regressive tversky indexed rocchio convolutive deep neural learning for legal document data analytics," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 9, no. 3, pp. 109–115, 2021.

[81] K. Atkinson, J. Collenette, T. Bench-Capon, and K. Dzehtsiarou, "Practical tools from formal models: the ECHR as a case study," in *Proceedings of the 18th International Conference on Artificial Intelligence and Law, ICAIL 2021*, pp. 170–174, São Paulo, Brazil, June, 2021.

[82] M. Lefoane, T. Koboyatshwene, G. Rammidi, and V. Lakshmi Narasimham, "Legal statutes retrieval: a comparative approach on performance of title and statutes descriptive text," *CEUR Workshop Proceedings*, vol. 2517, pp. 52–57, 2019.

[83] J. Morris, "User-friendly open-source case-based legal reasoning," in *Proceedings of the 17th International Conference on Artificial Intelligence and Law, ICAIL 2019*, pp. 270-271, Montreal, Canada, June, 2019.

[84] Z. Yuan, "Intelligent decision support system development technology of automotive mechanical system user expert admin man-machine interactive interface control and management module case library save case part feature information extraction case match success case," in *Proceedings of the 3rd International Conference on Education, Management and Computing Technology (ICEMCT 2016)*, pp. 1373–1377, Hangzhou, China, June, 2016.

[85] M. Araszkiewicz, "Time, trust and normative change. On certain sources of complexity in judicial decision-making," *Lecture Notes in Computer Science*, vol. 8929, pp. 100–114, 2014.

[86] A. Nagata, T. Kobayashi, K. Hasegawa, K. Moroga, Y. Kuriyama, and T. Ohno, "Development of the case-based reasoning system for regional science and technology policy: an interim report," in *Proceedings of the PICMET 2014-Portland International Center for Management of Engineering and Technology, Proceedings: Infrastructure and Service Integration*, pp. 135–139, Kanazawa, Japan, July, 2014.

[87] T. K. Wah and M. Muniandy, "Courtroom decision support system using case based reasoning," *Procedia Social and Behavioral Sciences*, vol. 129, pp. 489–495, 2014.

[88] H. Fan and H. Li, "Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques," *Automation in Construction*, vol. 34, pp. 85–91, 2013.

[89] A. Wyner and W. Peters, "Towards annotating and extracting textual legal case factors," in *Proceedings of The Lrec 2010 Seventh International Conference On Language Resources And Evaluation*, N. Calzolari, K. Choukri, B. Maegaard et al., Eds., pp. I36–I45, Valletta, Malta, May, 2010.

[90] D. R. D. R. Forbes, S. D. S. D. Smith, and R. M. W. M. W. Horner, "The selection of risk management techniques using case-based reasoning," *Civil Engineering and Environmental Systems*, vol. 27, no. 2, pp. 107–121, 2010.

[91] A. Wyner, R. Mochales-Palau, M. F. Moens, and D. Milward, "Approaches to text mining arguments from legal cases," in *Semantic Processing of Legal Texts*, vol. 6036, Berlin, Germany, Springer, 2010.

[92] D. W. Aha, "Conversational case-based reasoning," *Applied Intelligence*, vol. 14, 2005.

[93] K. D. Ashley, "Case-based reasoning research and development," in *Proceedings of the 5th International Conference on Case-Based Reasoning, ICCBR 2003*, vol. 2689, Trondheim Norway, June, 2003.

[94] H. C. Hoeschl, T. C. D. Bueno, A. Bortolon et al., "Structured contextual search for the un security council," in *ICEIS 2003 Proceedings of the 5th International Conference on Enterprise Information Systems*, vol. 2, pp. 100–107, Angers, France, April, 2003.

[95] K. D. K. D. Ashley and E. L. E. L. Rissland, "Law, learning and representation," *Artificial Intelligence*, vol. 150, no. 1–2, pp. 17–58, 2003.

[96] J. Zeleznikow, "An Australian perspective on research and development required for the construction of applied legal decision support systems," *Artificial Intelligence and Law*, vol. 10, no. 4, pp. 237–260, 2002.

[97] Y. Kiryu, A. Ito, T. Kasahara, H. Hatano, and M. Fujii, "A study of precedent retrieval system for civil trial," in *Computer Aided Systems Theory – EUROCAST 2017*, pp. 151–158, Springer, Berlin, Germany, 2018.

[98] E. Nissan, "Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement," *AI and Society*, vol. 32, no. 3, pp. 441–464, 2017.

[99] I. Elhalwany, A. Mohammed, K. Wassif, and H. Hefny, "Using textual case-based Reasoning in intelligent Fatawa QA system," *The International Arab Journal of Information Technology*, vol. 12, no. 5, pp. 503–509, 2015.

[100] D. Carneiro, P. Novais, F. Andrade, J. Zeleznikow, and J. Neves, "The legal precedent in online dispute resolution," *Frontiers in Artificial Intelligence and Applications*, vol. 205, 2009.

[101] L. K. K. Branting, "A reduction-graph model of precedent in legal analysis," *Artificial Intelligence*, vol. 150, no. 1–2, pp. 59–95, 2003.

[102] C.-L. C. L. Liu, C. T. C.-T. Chang, and J.-H. J. H. Ho, "Classification and clustering for case-based criminal summary judgments," in *Proceedings of the International Conference on Artificial Intelligence and Law*, pp. 252–261, São Paulo, Brazil, June, 2003.

[103] M. Xu, K. Hirota, and H. Yoshino, "Knowledge representation and similarity measure in learning a vague legal concept," in *A New Paradigm of Knowledge Engineering By Soft Computing*, L. Ding, Ed., vol. 5,pp. 189–212, 2001.

[104] K. Hirota, Y. Takama, M. Xu, and H. Yoshino, "Legal argument in fuzzy legal expert system (FLES)," in *Proceedings of the 2000 26th Annual Conference of the IEEE Industrial Electronics Society. IECON 2000. 2000 IEEE International Conference on Industrial Electronics, Control and Instrumentation. 21st Century Technologies*, vol. 1, pp. 207–210, Nagoya, Japan, October, 2000.