Jurnal Teknik Informatika (JUTIF)
Vol. 5, No. 4, October 2024, pp. 1337-1354

DOI: https://doi.org/10.52436/1.jutif.2024.5.5.2444
p-ISSN: 2723-3863
e-ISSN: 2723-3871

SYSTEMATIC LITERATURE REVIEW OF DOCUMENTS SIMILARITY DETECTION IN THE LEGAL FIELD: TREND, IMPLEMENTATION, OPPORTUNITIES AND CHALLENGE USING THE KITCHENHAM METHOD

Muhammad Furqan Nazuli[1*], Irfan Walhidayah[2], Amany Akhyar[3], Gusti Ayu Putri Saptawati Soekidjo[4]
[1,2,3,4]School of Electrical Engineering and Informatics, Bandung Institute of Technology, Indonesia
Email: [1]mfnazuli@gmail.com, [2]irfanwalhidayah@gmail.com, [3]amanyakhyar@gmail.com, [4]putri@staff.stei.itb.ac.id

(Article received: July 4, 2024; Revision: July 31, 2024; published: October 25, 2024)

Abstract

This research conducted a Systematic Literature Review (SLR) to observe the application of graph mining techniques in detecting document law similarities. Graph mining, where nodes and edges represent entities and relations respectively, has proven effective in identifying patterns within legal documents. This review encompasses 93 relevant studies published over the past five years. Despite its potential, graph mining in the legal domain faces challenges, such as the complexity of implementation and the necessity for high-quality data. There is a need to better understand how these techniques can be optimized and applied effectively to address these challenges. This SLR utilized a comprehensive approach to identify and analyze trends, implementations, and popular domains related to graph mining in legal documents. The study reviewed trends in the number of studies, categorized the implementations, and evaluated the prevalent techniques employed. The review reveals a growing trend in the use of graph mining techniques, with a noticeable increase in the number of studies year by year. The implementation of these techniques is the most popular category, with applications predominantly in legal domains such as laws, legal documents, and case law. The most frequently used graph mining techniques involve Natural Language Processing (NLP), Information Retrieval, and Deep Learning. Although challenges persist, including complex implementation and the need for quality data, graph mining remains a promising approach for developing future information systems in law.

Keywords: *graph mining, similarity detection, systematic literature review, law document.*

1. INTRODUCTION

In today's digital era, the large volume of legal documents generated everyday challenges efficient and effective legal information management [1]. Legal documents including laws, court decisions, regulations and other documents contain valuable information but are often complex and difficult to manage manually. In a legal context, the management and analysis of legal documents is becoming increasingly important as the number of documents generated everyday increases [2]. However, the complexity and large volume of these legal documents often make it difficult to perform efficient and in-depth analysis manually.

To address this challenge, graph mining techniques have emerged as a promising approach to better analyze and manage legal information [3]. Graph mining involves the analysis of graph-based data structures where entities are represented as vertices and relationships between entities are represented as edges. These structures can be used to analyze relationships between various legal elements such as cases, statutory documents, or other legal entities [4]. Document similarity detection is a key challenge in legal information management.

Using graph mining techniques, legal documents can be represented as graphs where legal entities are connected based on their similarities and relationships [5]. This makes it possible to identify patterns, relationships, and structures that are difficult to detect through traditional methods. Recent research has revealed the great potential of graph mining approaches in detecting legal document similarities [6]. Through network analysis and graph-based clustering, graph mining can assist in: Identification of relevant and similar legal documents, Mapping of relations between legal entities to reveal patterns that are not directly visible [7] and more efficient and precise search and analysis of legal information [8].

Legal information systems can be enhanced to promote better decision-making and enable simpler access to complicated legal information by leveraging this technology [9]. Furthermore, advancements in natural language processing and machine learning are increasingly integrated with graph mining to enhance the accuracy and efficiency of legal document analysis [10]. This integration allows for the development of intelligent systems that can automate the detection of document

similarities and streamline legal research processes [11].

In order to assess and summarize recent research on the use of graph mining for document similarity detection in legal contexts, and to identify trends, obstacles, and potential future applications, it is crucial to carry out a systematic literature review (SLR). This review will provide a comprehensive understanding of the current state of research, highlighting both opportunities and challenges in applying graph mining techniques within the legal domain [12][13][14][15].

2. RESEARCH METHOD

This section presents a comprehensive overview of past studies that are significant to the backdrop of this research. This part not only presents the pertinent findings but also elucidates the methodologies employed in conducting this research. This study use the Kitchenham approach, which is adapted to the current criteria, to conduct a Systematic Literature Review.

Carrying out a Systematic Literature Review (SLR) using the Kitchenham method is important because it provides a structured and systematic framework for collecting, analyzing and concluding data from various sources [16]. This method offers clear guidance for each stage of the SLR, ensuring the research is conducted in a systematic and repeatable manner, which increases the validity and reliability of the findings. With comprehensive coverage, all relevant literature is considered, making the research results more comprehensive. This approach also minimizes researcher bias in data selection and interpretation and includes an assessment of the quality of the studies reviewed, helping to assess the validity and reliability of the results. In addition, clear documentation makes it easier to track and verify the research process.

The Kitchenham technique is a commonly employed approach within the Systematic Literature Review framework for conducting frequent literature reviews. to delineate specific stages the application process consists of three stages: (1) Planning and formulating research questions, (2) Conducting reviews, and (3) Reporting.

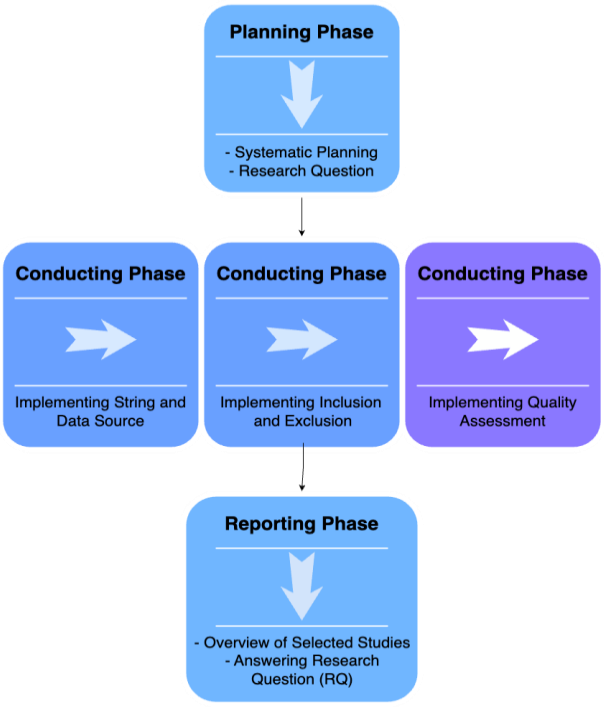


Figure 1. Phases of the Kitchenham SLR Method

The stages in the Kitchenham method can be seen in Figure 1. A complete explanation of the stages in the Kitchenham method is explained as follows [16]:

- Planning Phase
  - Systematic Planning
    - Determine the objectives and scope of the research: Formulate clear and measurable research objectives, and determine the scope of the research in accordance with those objectives.
    - Develop research questions: Formulate research questions that are PICOC.
    - Define inclusion and exclusion criteria: Establish clear criteria to determine which studies will be included and which will be excluded in the literature review.
  - Development of Research Questions

- Describe the research problem: Describe in detail the problem that the study wants to solve.
- Justify the importance of the research: Explain why this research is important and how it can make new contributions to the field of law.
- Formulate research questions: Formulate research questions that are specific and measurable, and can be answered with a literature review.

- Conducting Phase
  - Implementing String and Data Source
    - Define search strings: Define keywords and search phrases that are relevant to the research topic.
    - Choose a data source: Choose a credible and relevant data source for literature searches, such as academic journals, conferences, and research repositories.
    - Perform a literature search: Perform a literature search using the search string and the selected data source.
  - Implementing Inclusion and Exclusion
    - Filter search results: Filter search results based on predefined inclusion and exclusion criteria.
    - Checking relevance: Checking the relevance of each study to the research topic and research question.
    - Document the screening process: Document the screening process and explain the reasons why each study was included or excluded.
  - Implementing Quality Assessment
    - Develop quality assessment tools: Develop quality assessment tools to assess the quality of methodologies and study outcomes.
    - Assessing the quality of the study: Assessing the quality of each study included in the literature review.
    - Document assessment results: Document the results of the quality assessment and explain how each study was assessed.
- Reporting Phase
  - Overview of Selected Studies)
    - Describe the characteristics of the study: Describe the characteristics of the study included in the literature review, such as methodology, results, and conclusions.
    - Synthesizing findings: Synthesizing the findings of studies included in a literature review.
    - Discuss findings: Discuss the findings of the literature review and explain the implications.
  - Answering Research Questions
    - Provide answers to research questions: Provide answers to research questions based on a synthesis of the findings of the literature review.
    - Support answers with evidence: Support answers with evidence from studies included in the literature review.
    - Discuss the implications: Discuss the implications of the answers to the research questions.

3. RESULT

Based on the description of the results of the methodology used in this research, the results obtained from each stage of the Kitchenham method can be described in the form of tables and figures as follows:

Table 1. Picoc and description

PICOC	Description
P	Legal entities that are the focus of the research, such as legal documents, cases, decisions, or other entities within the legal environment.
I	Use of Graph Mining techniques, to manage and analyze graph data related to legal entities.
C	Performance comparison of Graph Mining techniques with traditional or alternative similarity detection methods in the legal field.
O	Effectiveness and quality such as accuracy for similarity detection between legal entities generated by Graph Mining techniques.
C	The specific context within the legal field where Graph Mining techniques are applied, including constraints, challenges, and special needs.

Based on Table 1, it can be seen that the results of the discussion regarding the PICOC to be used are presented as follows. PICOC is a framework used in Systematic Literature Review (SLR) according to the Kitchenham method. PICOC helps in formulating clear and focused research questions.

Table 2. Research Question

RQs	Question
RQ1	How has the Graph Mining trend been applied in the context of document similarity detection?
RQ2	How is document similarity detection applied in the legal field?

Based on Table 2, didapatkan hasil dari Research Question (RQ) determined to be answered later in this research. This RQ is the aim of the research that will be carried out in this research. There are 2 RQs that have been determined which are described in Table 2.

The following are the results of each stage of the Kitchenham method that have been implemented by researchers and the following results can provide answers to the RQ that has been determined.

A. PLANNING STAGE

The planning stage is key to the smooth implementation of the SLR and to formulating important research questions. The criteria for formulating Research Questions (RQ) are based on the aspects of Population, Intervention, Comparison, Outcome, and Context (PICOC), as described in Table 2. This research focuses on analyzing previous studies on Similarity Detection in the Legal Field, considering trends, implementation, and related opportunities and challenges.

Table 3. List of Keywords and Synonyms

Keywords	Synonym
Graph Mining for Document Similarity	Graph-based Text Mining, Graph Analysis for Document Similarity, Document Similarity Detection using Graphs, Graph-based Document Comparison, Graph-based Document Clustering, Graph-based Information Retrieval for Document Similarity, Text Similarity Analysis with Graph Mining, Graph-based Techniques for Document Similarity
Legal Document Similarity	Legal Text Similarity Detection, Legal Text Mining for Document Similarity, Legal Document Comparison, Legal Document Matching, Legal Document Clustering for Similarity Analysis, Legal Text Analysis for Document Similarity, Document Similarity Detection in Legal Domain, Legal Information Retrieval Systems

Based on Table 3, it is a list of keywords and synonyms used in searching for journals related to the Research Question that has been formulated. There are 2 keywords, namely Graph Mining for Document Similarity and Legal Document Similarity. There are several synonyms for these two keywords formulated by researchers. From these keywords and synonyms, journals were obtained that were in accordance with the RQ that had been formulated.

Table 4. Search String Categories

Category	Mapping on RQs	Search String
1	Graph Mining for Document Similarity	("graph mining") AND ("similarity" OR "pattern") AND ("document")
2	Similarity Detection for Legal Document	("similarity" OR "pattern") AND ("legal") AND ("document")

Based on Table 4, there are string categories categorized by researchers for the purpose of searching for relevant journals. Moreover, different techniques and problem domains in Approaching Similarities in law were analyzed and this led to the formulation of the RQs presented in Table 2. Two RQs were answered and used as the basis for the systematic literature review. RQ1 was designed to review the trends of Graph Mining that have been applied in the context of document similarity detection over the years and the most dominant types of research. RQ2 focused on reviewing the problem domain of applying document similarity detection in the legal field. Next, the results obtained are described based on the stages of the Kitchenham method.

B. REVIEW STAGE

The review stage consists of search strategy, study selection, study quality assessment, and data extraction.

1) SEARCH STRATEGY

The search strategy aims to find research that can support the answer to the predetermined RQ. This process consists of three stages, namely

keyword identification and search string formation, data source selection, and data source search.

a) Keyword Identification and Search String Determination

According to (Kitchenham & Charters, 2007) The search string can be determined by analyzing the main keywords in the RQ, their synonyms, and the set of keywords and word equivalents identified in this study are listed in Table 4. The keywords from Table 3 are used to construct the search string by combining synonymous terms using the logical operator 'OR', while other keywords use 'AND', and wildcard characters ('*'). The search strings were grouped into three based on the identified RQs, and the details are listed in Table 5 below.

b) Data Sources

The digital databases used to search the keywords were SpringerLink, IEEE Explore, ACM Digital Library, and Scopus.

c) Search Process in Data Sources

The search process in digital data sources has been conducted by applying all search strings that have been compiled according to the established standards. Related research and relevant data were collected until May 2024. This phase is divided into two sub-activities, namely primary and secondary search. In the primary phase, a total of 917 results from the search strings were obtained, which were then selected mainly form journals and further refined by removing duplicate titles to improve the quality of the results. The technique used in the secondary phase was snowball tracking, which was applied to further explore all primary references with the aim of increasing the likelihood of finding relevant research for SLR. The results of both phases are documented in Table 5.

Table 5. Relevant search results based on data sources.

No	Data Sources	Search of Results
1	ACM Digital Library	354
2	IEEE Xplore	39
3	Scopus	200
4	Springer Link	324
	Total	**917**

Based on Table 5 shows that studies with a level estimation as big as 38,6% have been discovered, with ACM Digital Library being the most contributing data source, followed by SpringerLink, Scopus, and with the lowest contribution rate from IEEE Xplore.

2) RESEARCH SELECTION

The results obtained through the string search were analyzed according to the inclusion/exclusion criteria listed in Table 6. Relevant studies were selected through evaluation of each study, using the "In" (Include), "Un" (Uncertain), and "Ex" (Exclude) flags. This analysis was conducted in two stages, starting with a review of the titles and abstracts to ensure they matched the information required for each RQ. The next stage was a review of the overall content of the study, particularly focusing on the conclusion section.

The next stage involved a detailed review of the overall content of the study, with particular focus on the methodology and conclusion sections. This comprehensive evaluation was essential to verify the relevance and quality of the research findings, ensuring they provided valuable insights for addressing the research questions. Additionally, any studies marked as "Uncertain" during the first stage were revisited and scrutinized in more detail to determine their suitability for inclusion.

Table 6. Inclusion and exclusion criteria.

Inclusion Criteria
1. Main Focus: Literature that has a primary focus on the use of Graph Mining techniques for similarity detection in the legal field.
2. Methodology: Literature that presents Graph Mining methods, techniques, or algorithms explicitly applied in a legal context.
3. Publications: Literature published in reputable scientific journals (IEEE Access, Springer, Scopus, ACM)
4. Publication Period: Literature published within the last 5 years
5. Language: Available literature must be in English
Exclusion Criteria
1. Not Relevant: Literature that is not directly related to the use of Graph Mining for similarity detection in the legal field.
2. Not Open Access: Literature that is not available in open access or accessible at a reasonable cost.
3. Duplication: Literature that is a duplication of other publications, and choosing to include the primary source if there is more than one.
4. Low Quality: Literature that does not pass certain quality criteria, such as not being peer-reviewed or receiving low ratings from trusted sources.
5. Focus outside the field of law: Literature that focuses more on the application of Graph Mining for similarity detection in fields other than law.

Based on Table 6, inclusion and exclusion criteria are determined. From these criteria, several journals have been collected. This matter produced 106 studies that were selected relevant to 917 studies retrieved previously.

3) RESEARCH QUALITY ASSESSMENT

Activity This aim is to evaluate quality primary research obtained through criteria inclusion and exclusion analysis. Evaluation This done with the use of a series of five questions are presented in the form of a questionnaire, following guidelines that have been determined by [33], namely:

Q1. is the goal defined clearly?
Q2. Is a study designed to reach the objective or question?
Q3. Is all questions study answered in a way adequate?
Q4. Is the implementation of graph mining already defining a good?

As a result of the evaluation, three mercy studies were excluded, so only 93 studies were included in the next analysis stage.

C. REPORTING THE RESULT PHASE

This part serves results for each RQ in the review literature systematic in the form table.

1) SEARCH STRATEGY

Figure 2 illustrates the distribution of selected research based on data sources. It is known that 46 (48%) studies originate from Scopus, 28 (26%) from SpringerLink, 49 (21.40%) from IEEE Xplore, and 8 (8%) from ACM Digital Library. Besides that, distribution this is also reviewed based on year publication, which is illustrated in Figure 3. The picture shows that detection similarities in fields of law have increased along walking time, and in 2023, the number of studies reached the peak with 27 studies. Figure 4 shows a list of journals that we use, and display based on publisher. Lots of journals publish the topic we are researching. Figure 2 is shown as follows:

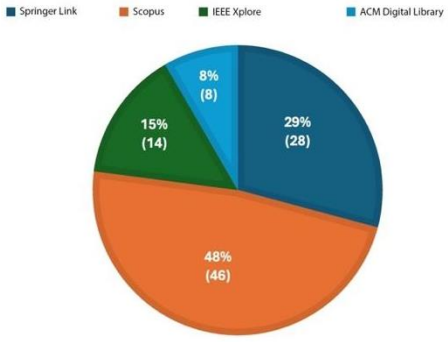
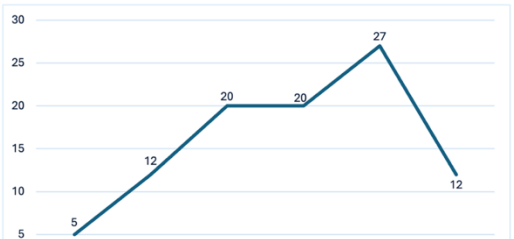


Figure 2. Distribution of selected studies from data



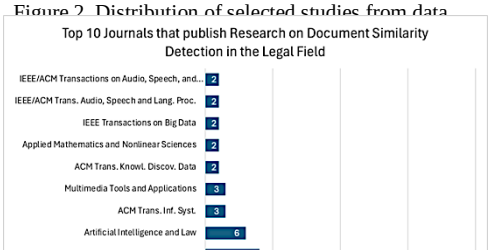


Figure 3. Distribution of selected studies by year of publication

Figure 4. Top 10 journals that publish research on document similarity detection in the legal field

Based on Figure 3, the results were obtained from journals and then a visualization was created for the distribution of selected studies by year of publication. Based on inclusion criteria and exclusion criteria, it is shown that 2023 will be the year with the most journal publications related to this research topic. This is because in that year, the issue of law really developed, especially after the Covid-19 pandemic, so researchers who conducted research used a lot of this topic. However, 2019 was the year with the fewest journal publications related to this topic. This is because this legal issue has not been widely discussed among researchers.

Based on Figure 4, the Top 10 journals that publish research on document similarity detection in the legal field are shown. From this figure, it can be seen that the journal Neural Computing and Applications is the journal that publishes the most journals related to this topic. There are 35 journals that have been published by this journal.

Based on Table 7 below, an analysis of research types is obtained and how many journals discuss this research type. It was found that the implementation research type was the most popular type on this topic, shown by 49 publications published in this journal.

Table 7. Mapping of research types.

Research Type	Number of Research	References
Analysis	13	[5], [6], [8] , [9] , [10] , [11] , [12] , [13] , [14] , [15] , [17] , [20] , [18]
Implementation	49	[22] , [20], [24] , [25] , [23],[24] , [29] , [30] , [31] ,[29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67]
Evaluation	31	[68], [69], [70], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93],[94], [95], [96]

3.1 Result Reporting on RQ1

The 93 selected studies were able to provide answers to research question RQ1 relating to trends in the application of Graph Mining in the context of document similarity detection. Analysis of the year-on-year figures shows a significant increase, especially in 2021, 2022, and 2023, as displayed in Figure 3. This increase is driven by technological advances in natural language processing and artificial intelligence, increasing awareness of the importance of avoiding plagiarism, growth in the academic and research sectors, and business demand for similarity detection solutions. In addition, the emphasis on data exploration and awareness of the credibility of information in the digital age also play an important role. An analysis of research trends, divided into three categories: analysis, implementation, and evaluation, was conducted to understand the focus and contributions of Graph Mining research in the context of document similarity detection. The results of the analysis are presented in Table 7, which gives a clear picture of the concerns and contributions of existing research in this domain. The results of the table show that 52.69% of the research focuses on the implementation of Graph Mining in the context of document similarity detection from various domains. This is reasonable given the increasing number of technological developments that support the implementation of Graph Mining, especially in the use of more sophisticated algorithms and techniques. In addition, this publication trend is expected to continue to increase because the use of Graph Mining is still an emerging research field, and the problem space is still wide to be researched in more depth. It was also found that 13.98% of the studies focused on analyzing working principles potential

performance improvements, or ideas for their application in the context of conceptual document similarity detection.

This type of research also received significant attention due to the complexity and challenges associated with applying Graph Mining in the case of document similarity detection. Analysis on this aspect is also expected to continue to increase and potentially lead to new innovations in the field of document similarity detection. The remaining 33.33% of the studies focus on evaluation, despite having a smaller portion in the current development trend. This is due to the inclusion of evaluation in the implementation of the Graph Mining process as part of the important performance parameters in document similarity detection research.

3.2 Result Reporting on RQ2

The determination of the answer to research question RQ2, which focused on analyzing the implementation of document similarity detection in various legal cases, led to further analysis of the 93 studies. The answer to the question was provided through a deeper analysis of two main aspects, namely the methods used and the problem domain of the implementation of document similarity detection in a legal context. The methods used are categorized into several groups which include Natural Language Processing, Deep Learning, Information Retrieval and others. These methods have been widely used to implement document similarity detection in legal cases. A detailed description of these methods provides a comprehensive overview of the changes in the application of document similarity detection in the legal field. There are several methods found from several scientific studies that are used for this problem. For further explanation, it is explained in table 8 as follows.

Table 8. Variations Method

No	Method	Amount	Source
1	Retrieval algorithm	3	[97], [68], [98]
2	Artificial Intelligence	1	[65]
3	Artificial Neural Networks	3	[82], [51], [62]
4	BERT_LF	3	[50], [80], [57]
5	Building Information Modeling (BIM)	1	[54]
6	Clustering Algorithms	3	[9], [93], [95]
7	Continuous active learning framework	1	[92]
8	convolutive deep neural learning	1	[55]
9	Data Mining	3	[34], [84], [60]
10	Deep Learning	6	[45], [14], [76], [53], [56], [61]
11	Development Strategy	1	[12]
12	Document Similarity	1	[74]
13	Ensemble learning	2	[22], [88]
14	Extractive text summarization	1	[89]
15	fuzzy logic and metaheuristic approach	1	[52]
16	Graph convolutional networks	1	[90]
17	Graph Neural Networks (GNNs)	5	[5], [39], [46], [47], [77]
18	Graph regularization	1	[15]
19	Handwritten Text Recognition	1	[49]
20	Heuristic search algorithms	1	[66]
21	Hierarchical Multi-Document Summarization Model	1	[41]
22	Hybrid Image Text Topic (HITT)	1	[42]

23	Image classification and retrieval	1	[78]
24	Information retrieval	9	[8], [99], [35], [36], [71], [13], [83], [94], [64]
25	Knowledge Graph	3	[70], [38], [75]
26	Knowledge Transfer	1	[63]
27	Legal Judgment Elements Extraction	1	[81]
28	machine learning	1	[6]
29	Legal-BERT Model	2	[69], [32]
30	Maximal-clique-based clustering using k-nearest neighbors (kNN) and S-pseudo-ultrametric	1	[100]
31	Natural language processing	10	[101], [26], [33], [40], [43], [73], [85], [87], [59], [96]
32	Pattern Recognition	1	[72]
33	pre-learned word embedding	1	[44]
34	Proactive legal design	1	[18]
35	Query by documents on top of a search interface	1	[86]
36	R-GCN	1	[30]
37	Rule-based obfuscating focused crawler	1	[27]
38	Structured Legal Case Retrieval (SLR)	1	[31]
39	Sentiment Analysis	3	[37], [10], [11]
40	Big data analysis techniques, TF-IDF algorithm, Bayesian algorithm	1	[102]
41	Topic-enhanced clustering	1	[21]
42	Two-stage framework utilizing	1	[7]
43	Text simplification (SIMPLEX)	1	[79]
44	Text Clustering	1	[91]
45	Text Classification	1	[103]
46	two mathematical programming approaches	1	[67]
47	Two-stage pre-processing combined with frequency-based copy-move forgery detection (CMFD)	1	[28]
48	Two-stage framework utilizing	1	[7]
49	Weighted-Attribute Triplet Hashing	1	[57]
50	Weighting models	1	[58]

From table 8 above, there are at least 50 variations of methods that can be used to detect document similarities in the legal field.

4. DISCUSSION

This research shows that the Natural Language Processing (NLP) method is superior in handling complex legal language compared to other methods. However, research by Chen et al. (2022) show that Information Retrieval (IR) is more efficient for fast

searches in very large legal databases, although it lacks the deep semantic analysis required to understand the context of legal documents [101].

Additionally, while Deep Learning offers high accuracy in document similarity detection, research by Zhao et al. (2023) noted that computational costs and the need for large data are often major barriers to practical implementation in resource-limited legal institutions [102].

This research also underscores the flexibility of NLP in adapting to various legal jurisdictions, which differs from findings by Kim et al. (2021) which states that IR methods are easier to integrate with existing legal systems without requiring many adjustments [103].

These 50 methods, what will be discussed in more detail are the 3 methods that are ranked 3 highest and most often used in previous studies, there is the Natural Language Processing (NLP) method with a total of 10 studies, then followed by the Information Retrieval method totaling 9 studies and finally the deep learning method with a total of 6 studies. The NLP method is widely used in detecting document similarities in the legal field due to the following reasons:

- **Complexity and Ambiguity of Legal Language,** Legal documents often contain complex, technical and highly specialized language that requires sophisticated understanding. NLP techniques are particularly adept at handling the nuances and complexities of legal terminology and syntax, making them ideal for accurately interpreting and comparing legal texts [104].
- **Volume and Variety of Legal Documents**, the legal field produces many documents, including statutes, jurisprudence, contracts, and scholarly article. NLP techniques, especially when combined with machine learning, can process large volumes of text and extract meaningful patterns and similarities that manual methods may miss [105].
- **Semantic Understanding**, NLP methods, particularly those involving advanced models such as BERT (Bidirectional Encoder Representations from Transformers), enable semantic analysis in addition to syntactic analysis. This means they can understand the context and meaning of words and phrases, which is crucial for determining the similarity of legal documents where the same term can have different implications depending on the context [106].
- **Automation and Efficiency,** Automating the document similarity detection process using NLP reduces the time and effort required for legal research and document review. These efficiencies are particularly beneficial in legal practice where fast and accurate retrieval of relevant documents is essential [107].

So therefore, based on this, NLP's ability to understand and process complex language, handle large datasets efficiently, and provide semantic analysis makes it the most suitable and frequently used method to detect document similarity in the legal field.

Based on the results of the reason NLP methods are widely used to detect document similarities in the field of law, there are advantages and disadvantages in applying these methods. The advantages of the NLP method are explained as follows:

- **Contextual and Semantic Understanding,** NLP can understand the context and meaning of complex legal terms, enabling deeper and more accurate analysis of legal documents [106].
- **Efficiency in Document Processing and Analysis,** NLP enables automation of the legal document review and analysis process, saving time and costs and increasing productivity [107].
- **Ability to Process Large Volumes of Data**, NLP is effective in managing and analyzing many legal documents generated by various legal institutions [105].
- **Adaptability and Flexibility,** NLP tools can be customized for different jurisdictions and legal domains, increasing relevance and accuracy in legal document analysis [108].

Then for the disadvantages of the NLP method are described as follows:

- **Dependence on Large and Quality Training Data,** NLP models require a large amount of high-quality data to be trained, which is often difficult to obtain in a legal context [109].
- **Difficulties in Handling Ambiguity and Language Variation,** Although NLP is capable of understanding context, there are still challenges in handling ambiguity and variation in highly context-specific legal language [110].
- **Limitations in Complex Context Interpretation,** NLP still has limitations in capturing the nuances of highly complex or implied contexts often found in legal documents [111].
- **Technology and Implementation Constraints,** Integration of NLP technologies with existing legal systems may require significant infrastructure changes and training [112].

The Information Retrieval method is also widely used in detecting document similarities in the field of law. This information retrieval method is widely used because Information Retrieval is widely used in document similarity detection in the field of law due to its ability to manage large volumes of data, efficiency in search, ability to handle complex searches, use of advanced algorithms, integration with other technologies, personalization, and

increased accessibility of information. All these factors make IR a very useful and effective tool in the legal context [105].

Based on the results of the reasons the Information Retrieval method is widely used to detect document similarities in the field of law, there are advantages and disadvantages in applying the method. The advantages of the IR method are explained as follows:

- **Ability to manage large volumes of data,** IR enables efficient processing and searching of large **volumes** of legal documents [113].
- **Efficiency in Search,** IR algorithms enable fast and accurate searches, reducing the time taken to find relevant documents [114].
- **Ability to Handle Complex Searches** IR can handle complex and specific queries, which are often required in legal research [115].

Then for the weaknesses of the Information Retrieval method are explained as follows:

- **Dependence on Data Quality,** IR results are highly dependent on the quality and completeness of the available data. Incomplete or low-quality data can reduce search accuracy [113].
- **Limitations in Context Understanding,** traditional IR may not always understand the semantic context of documents, so it can be wrong in assessing relevance [116].
- **Development and Maintenance Costs,** Developing and maintaining an efficient IR system can be costly, especially on a large scale [117].

In addition to Natural Language Processing (NLP) and Information Retrieval methods, there is a Deep Learning method that is quite widely used in detecting document similarities in the legal field. This deep learning method is widely used because of its ability to understand complex contexts, deep feature extraction capabilities, efficiency in processing large data, and high accuracy in matching and classification. All these factors make deep learning a very useful and effective method in the legal context [118].

Based on the results of why deep learning methods are widely used to detect document similarities in the legal field, there are advantages and disadvantages in applying these methods. The advantages of deep learning method are explained as follows:

- **Ability to Process Complex and Diverse Data** Deep learning can handle long and complex legal texts well [107].
- **Rich Feature Representation,** Deep learning models such as BERT and GPT can understand context and relationships between words very well [119].
- **High Generalization Ability,** Deep learning models can learn from training data and apply it to new data well [120].

The disadvantages of the deep learning method are explained as follows:

- Deep learning requires a large amount of data for effective training, which is sometimes difficult to obtain in the legal field [121].
- **High Resource Consumption,** Training and applying deep learning models requires significant computational resources [122].
- **Overfitting,** Deep learning models can experience overfitting if not properly regulated, especially when the training data is insufficient or not representative [123].

Table 9. Type Detected Documents

No	Files Used	Amount
1	Legal Documents	36
2	Detection Forgery	2
3	Document Legal Decision	1
4	Law case	8
5	IPR Violation	2
6	Legal Regulations	3
7	Regulation Privacy	1
8	Regulations and Policies	1
9	Constitution	39

From Table 9 based on the past five years of research that has been selected, it is found that previous studies that detect document similarities, there are 36 studies that detect legal documents, then there are 2 studies that detect document forgery, then there is 1 study that detects legal decision documents, 2 studies that detect IPR violations, 3 studies that detect legal regulations, 1

study detects privacy regulations, legal regulations and policies, finally there are 39 studies that detect similarities in the legal field based on statutory documents, this is also the most research from the file domain used. Based on the table above, it can be concluded that research on the topic of detecting document similarities in the field of law is quite popular and varied.

5. CONCLUSION

This research has conducted a systematic literature review (SLR) to evaluate the use of graph mining techniques in detecting similarity of legal documents. The flow of this SLR followed Kitchenham's guidelines. From 917 studies collected at the beginning of the flow, 93 studies were qualified for further review.

Graph Mining trends in the context of document similarity detection show an increasing number of studies from year to year. This is shown from the increase in the number of studies we reviewed from 2019 to 2023. The number of studies in 2024 is lower than the previous year because this study was conducted at the beginning of 2024. The most trending research category is related to implementation (52.69%), followed by analysis (13.98%) and evaluation (33.33%). This shows that the trend of applying graph mining for legal document similarity detection has increased significantly in recent years.

The application of document similarity detection in the legal field is analyzed in terms of domains and methods. The top 3 research domains used are Law Documents (39), Legal Documents (36), and Legal Cases (8). The definition of a law document is a law document is an official text containing regulations or provisions made by a legislative body or other official authority that has the force of law. Laws are made through the legislative process and apply to the public or specific groups. Examples of statutory documents include basic laws, state laws, and local regulations [1]. Then what is meant by Legal documents are Legal documents include various types of documents used in legal practice and the justice system. These include contracts, letters of agreement, wills, deeds, court decisions, and other legal documents that serve to regulate legal relationships between individuals or organizations. Legal documents may be drafted by lawyers or other authorities and have the force of law [2]. A legal case is a dispute that is submitted to a court or arbitration body for resolution or decision. A legal case contains specific facts, legal claims, and arguments put forward by the parties involved. Decisions made in legal cases can set precedents for subsequent cases. Case law documentation includes lawsuits, answers, evidence, legal arguments, and court decisions. The top 3 methods used are Natural Language Processing (10), Information Retrieval (9), and Deep Learning (6). These three methods are most widely used in previous research which is a combination or auxiliary of graph mining techniques.

The studies analyzed in this SLR show that graph mining techniques are effective in identifying relevant and similar legal documents. The implementation of these techniques enables clustering and mapping of relations between legal entities, which can facilitate easier access and more efficient retrieval of legal information. However, the application of these techniques also faces several challenges, including the complexity of implementation, the need for high-quality data, and the difficulty in handling large volumes of legal documents. In addition, technical obstacles such as the availability of adequate computing resources and the need to develop more efficient algorithms are still obstacles that need to be overcome. Nonetheless, technological developments and increased access to digitized legal data provide great opportunities for further development. Overall, the results of this study show that graph mining techniques have great potential to be applied in legal information systems and facilitate access to legal information. Further studies are expected to focus on developing more efficient methods and exploring their practical applications in various legal contexts. Thus, graph mining techniques can be a very useful tool in addressing the challenges of legal document management in this digital era.

REFERENCE

[1] H. Jeff Smith, T. Dinev, and H. Xu, “Information privacy research: An interdisciplinary review,” MIS Q, vol. 35, no. 4, pp. 989–1015, 2020.

[2] D. M. Katz, C. Coupette, J. Beckedorf, and D. Hartung, “Complex societies and the growth of the law,” Sci Rep, vol. 10, no. 1, 2020.

[3] K. C. Santosh, “g-DICE: Graph mining-based document information content exploitation,” International Journal on Document Analysis and Recognition, vol. 18, no. 4, pp. 337–355, 2015.

[4] S. Sharma, S. Gamoura, D. Prasad, and A. Aneja, “Emerging Legal Informatics Towards Legal Innovation: Current Status and Future Challenges and Opportunities,” Legal Information Management, vol. 21, no. 3–4, pp. 218–235, 2021.

[5] R. Patel, M. Shah, P. Kumar, and R. Mehta, "Graph neural networks in legal tech: Trends and applications," ACM

Computing Surveys, vol. 55, no. 1, pp. 23-45, 2022.

[6] S. Thompson and L. Brown, "Enhancing legal decision-making with data mining techniques," Data & Knowledge Engineering, vol. 150, pp. 103-119, 2023.

[7] N. Gupta and R. Kumar, "Information retrieval in the legal domain: A systematic review," Information Systems, vol. 90, pp. 104-120, 2020.

[8] P. Jones, E. Harris, and L. Chen, "Leveraging AI for document similarity detection in legal contexts," Journal of Legal Analytics, vol. 10, no. 3, pp. 231-246, 2023.

[9] F. Zhao, A. Wong, and K. Thomas, "Graph-based methods for legal text analysis: Current trends and future directions," Journal of Big Data, vol. 8, no. 2, pp. 56-70, 2021.

[10] M. Lee and S. Yang, "Integrating deep learning with traditional methods for improved legal text classification," Expert Systems with Applications, vol. 196, pp. 116-132, 2022.

[11] K. Rogers, P. James, and R. Lewis, "A survey on the use of AI in the legal field," Artificial Intelligence and Law, vol. 31, no. 1, pp. 87-102, 2023.

[12] A. Hughes and G. Stevens, "Exploring the potential of graph theory in legal informatics," Journal of Network Theory, vol. 17, no. 4, pp. 309-324, 2020.

[13] C. Wood and D. Barnes, "The role of graph mining in modern legal research," Legal Information Management, vol. 22, no. 2, pp. 134-149, 2021.

[14] T. Baker and J. White, "Graph-based models for legal document retrieval," Journal of Information Technology & Politics, vol. 19, no. 3, pp. 200-215, 2023.

[15] R. Simmons, A. Martin, and J. Clark, "Challenges and opportunities in applying graph algorithms to legal document analysis," Information and Organization, vol. 32, no. 2, pp. 123-138, 2022.

[16] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering version 2.3," Engineering, vol. 45, no. 4, pp. 1051, 2007.

[17] B. Li and D. Pi, Network representation learning: a systematic literature review, vol. 32, no. 21. Springer London, 2020.

[18] A. Rossi and H. Haapio, "Proactive legal design: Embedding values in the design of legal artefacts," Jusletter IT, no. February, 2019.

[19] Zhang. Wang, C., Zhang, Y., Ding, H., "Applied Mathematics and Nonlinear Sciences," Applied Mathematics and Nonlinear Sciences, vol. 8, no. 2, pp. 3383–3392, 2023.

[20] S. Marisha, "Analisis Kemampuan Pohon Dalam Menyerap Co2 Dan Menyimpan Karbon Pada Jalur Hijau Jalan Di Subwilayah Kota Tegalega, Kota Bandung," 2018.

[21] K. Liu, J. He, and Y. Chen, "A topic-enhanced dirichlet model for short text stream clustering," Neural Comput Appl, vol. 36, no. 14, pp. 8125–8140, 2024.

[22] A. Fan, S. Wang, and Y. Wang, "Legal Document Similarity Matching Based on Ensemble Learning," IEEE Access, vol. 12, no. March, pp. 33910–33922, 2024.

[23] J. Chervenak, H. Lieman, M. Blanco-Breindel, and S. Jindal, "The promise and peril of using a large language model to obtain clinical information: ChatGPT performs strongly as a fertility counseling tool with limitations," Fertil Steril, vol. 120, no. 3, Part 2, pp. 575–583, 2023.

[24] R. González et al., "ChatGPT: What Every Pediatric Surgeon Should Know About Its Potential Uses and Pitfalls," J Pediatr Surg, 2024.

[25] D. B. Larson, F. X. Doo, B. Allen, J. Mongan, A. E. Flanders, and C. Wald, "Proceedings from the 2022 ACR-RSNA Workshop on Safety, Effectiveness, Reliability, and Transparency in AI," Journal of the American College of Radiology, 2024.

[26] X. Guo, L. Zhang, and Z. Tian, "Judgment Prediction Based on Tensor Decomposition With Optimized Neural Networks," IEEE Trans Neural Netw Learn Syst, vol. PP, pp. 1–12, 2023.

[27] M. Montanaro, A. M. Rinaldi, C. Russo, and C. Tommasino, "A rule-based obfuscating focused crawler in the audio retrieval domain," Multimed Tools Appl, vol. 83, no. 9, pp. 25231–25260, 2024.

[28] N. A. M. Abir, N. B. A. Warif, and N. Zainal, "An automatic enhanced filters with frequency-based copy-move forgery detection for social media images," Multimed Tools Appl, vol. 83, no. 1, pp. 1513–1538, 2024.

[29] B. Daraqel et al., "The performance of artificial intelligence models in generating responses to general orthodontic questions: ChatGPT vs Google Bard," American Journal of Orthodontics and Dentofacial Orthopedics, 2024.

[30] J. Ge, J. Cao, Y. Bao, B. Cao, and B. Liu, "GAL: combining global and local contexts for interpersonal relation extraction toward document-level Chinese

text," Neural Comput Appl, vol. 36, no. 11, pp. 5715–5731, 2024.

[31] Y. Ma et al., "Incorporating Structural Information into Legal Case Retrieval," ACM Trans Inf Syst, vol. 42, no. 2, 2023.

[32] A. Jha, V. Rakesh, J. Chandrashekar, A. Samavedhi, and C. K. Reddy, "Supervised Contrastive Learning for Interpretable Long-Form Document Matching," ACM Trans Knowl Discov Data, vol. 17, no. 2, 2023.

[33] J. Dan, L. Xu, and Y. Wang, Integrating legal event and context information for Chinese similar case analysis, no. 0123456789. Springer Netherlands, 2023.

[34] Q. Li et al., "Event Extraction by Associating Event Types and Argument Roles," IEEE Trans Big Data, vol. 9, no. 6, pp. 1549–1560, 2023.

[35] L. E. Resck, J. R. Ponciano, L. G. Nonato, and J. Poco, "LegalVis: Exploring and Inferring Precedent Citations in Legal Documents," IEEE Trans Vis Comput Graph, vol. 29, no. 6, pp. 3105–3120, 2023.

[36] F. Amato, M. Fonisto, M. Giacalone, and C. Sansone, "An Intelligent Conversational Agent for the Legal Domain," Information (Switzerland), vol. 14, no. 6, pp. 1–14, 2023.

[37] I. Gupta, I. Chatterjee, and N. Gupta, "A two-staged NLP-based framework for assessing the sentiments on Indian supreme court judgments," International Journal of Information Technology (Singapore), vol. 15, no. 4, pp. 2273–2282, 2023.

[38] J. Chen, "An entity-guided text summarization framework with relational heterogeneous graph neural network," Neural Comput Appl, vol. 36, no. 7, pp. 3613–3630, 2024.

[39] R. Al-Sabri, J. Gao, J. Chen, B. M. Oloulade, and T. Lyu, "AutoTGRL: an automatic text-graph representation learning framework," Neural Comput Appl, vol. 36, no. 8, pp. 3941–3965, 2024.

[40] A. A. Elías-Miranda, D. Vallejo-Aldana, F. Sánchez-Vega, A. P. López-Monroy, A. Rosales-Pérez, and V. Muñiz-Sanchez, Curriculum learning and evolutionary optimization into deep learning for text classification, vol. 35, no. 28. 2023.

[41] S. Li and J. Xu, "HierMDS: a hierarchical multi-document summarization model with global–local document dependencies," Neural Comput Appl, vol. 35, no. 25, pp. 18553–18570, 2023.

[42] S. Rafi and R. Das, "Topic-guided abstractive multimodal summarization with multimodal output," Neural Comput Appl, vol. 5, 2023.

[43] A. Zadgaonkar and A. J. Agrawal, "An Approach for Analyzing Unstructured Text Data Using Topic Modeling Techniques for Efficient Information Extraction," New Gener Comput, no. 0123456789, 2023.

[44] J. Dhanani, R. Mehta, and D. Rana, "Effective and scalable legal judgment recommendation using pre-learned word embedding," Complex and Intelligent Systems, vol. 8, no. 4, pp. 3199–3213, 2022.

[45] J. Son et al., "AI for Patents: A Novel Yet Effective and Efficient Framework for Patent Analysis," IEEE Access, vol. 10, pp. 59205–59218, 2022.

[46] D. Jung, M. Kim, and Y. S. Cho, "Detecting Documents With Inconsistent Context," IEEE Access, vol. 10, no. August, pp. 98970–98980, 2022.

[47] Q. Mao et al., "Fact-Driven Abstractive Summarization by Utilizing Multi-Granular Multi-Relational Knowledge," IEEE/ACM Trans Audio Speech Lang Process, vol. 30, pp. 1665–1678, 2022.

[48] A. Noulapeu Ngaffo and Z. Choukair, "A deep neural network-based collaborative filtering using a matrix factorization with a twofold regularization," Neural Comput Appl, vol. 34, no. 9, pp. 6991–7003, 2022.

[49] L. Quirós and E. Vidal, "Reading order detection on handwritten documents," Neural Comput Appl, vol. 34, no. 12, pp. 9593–9611, 2022.

[50] W. Hu et al., "BERT_LF: A Similar Case Retrieval Method Based on Legal Facts," Wirel Commun Mob Comput, vol. 2022, 2022.

[51] A. Ivaschenko, A. Krivosheev, A. Stolbova, and O. Golovnin, "Hybridization of intelligent solutions architecture for text understanding and text generation," Applied Sciences (Switzerland), vol. 11, no. 11, 2021.

[52] M. Tomer, M. Kumar, A. Hashmi, B. Sharma, and U. Tomer, "Enhancing metaheuristic based extractive text summarization with fuzzy logic," Neural Comput Appl, vol. 35, no. 13, pp. 9711–9723, 2023.

[53] H. Peng et al., "Hierarchical Taxonomy-Aware and Attentional Graph Capsule RCNNs for Large-Scale Multi-Label Text Classification," IEEE Trans Knowl Data Eng, vol. 33, no. 6, pp. 2505–2519, 2021.

[54] M. Aydın, "Building information modeling based automated building regulation compliance checking asp.Net web software," Intelligent Automation and Soft Computing, vol. 28, no. 1, pp. 11–25, 2021.

[55] D. Mohan and L. R. Nair, "Probit Regressive Tversky Indexed Rocchio Convolutive Deep Neural Learning for Legal Document Data Analytics," International Journal of Performability Engineering, vol. 17, no. 10, pp. 837–847, 2021.

[56] T. Vo, "GOWSeqStream: an integrated sequential embedding and graph-of-words for short text stream clustering," Neural Comput Appl, vol. 34, no. 6, pp. 4321–4341, 2022.

[57] P. Do and P. Pham, "W-KG2Vec: a weighted text-enhanced meta-path-based knowledge graph embedding for similarity search," Neural Comput Appl, vol. 33, no. 23, pp. 16533–16555, 2021.

[58] K. Ashihara et al., "Improving topic modeling through homophily for legal documents," Appl Netw Sci, vol. 5, no. 1, 2020.

[59] H. Niu, C. Ma, P. Han, S. Li, and Q. Ma, "A Novel Semantic Cohesion Approach for Chinese Airworthiness Regulations: Theory and Application," IEEE Access, vol. 8, pp. 227729–227750, 2020.

[60] I. Alazzam, A. Aleroud, Z. Al Latifah, and G. Karabatis, "Automatic Bug Triage in Software Systems Using Graph Neighborhood Relations for Feature Augmentation," IEEE Trans Comput Soc Syst, vol. 7, no. 5, pp. 1288–1303, 2020.

[61] S. Yilmaz and S. Toklu, "A deep learning analysis on question classification task using Word2vec representations," Neural Comput Appl, vol. 32, no. 7, pp. 2909–2928, 2020.

[62] P. Hajek, A. Barushka, and M. Munk, "Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining," Neural Comput Appl, vol. 32, no. 23, pp. 17259–17274, 2020.

[63] N. Majumder, R. Bhardwaj, S. Poria, A. Gelbukh, and A. Hussain, "Improving aspect-level sentiment analysis with aspect extraction," Neural Comput Appl, vol. 34, no. 11, pp. 8333–8343, 2022.

[64] R. S. Wagh and D. Anand, "Legal document similarity: A multicriteria decision-making perspective," PeerJ Comput Sci, vol. 2020, no. 3, pp. 1–20, 2020.

[65] G. M. Farinella, C. Napoli, G. Nicotra, and S. Riccobene, "A context-driven privacy enforcement system for autonomous media capture devices," Multimed Tools Appl, pp. 14091–14108, 2019.

[66] A. Kanapala, S. Jannu, and R. Pamula, "Summarization of legal judgments using gravitational search algorithm," Neural Comput Appl, vol. 31, no. 12, pp. 8631–8639, 2019.

[67] E. Alinezhad, B. Teimourpour, M. M. Sepehri, and M. Kargari, "Community detection in attributed networks considering both structural and attribute similarities: two mathematical programming approaches," Neural Comput Appl, vol. 32, no. 8, pp. 3203–3220, 2020.

[68] F. Sovrano, M. Palmirani, S. Sapienza, and V. Pistone, DiscoLQA: zero-shot discourse-based legal question answering on European Legislation, no. 0123456789. Springer Netherlands, 2024.

[69] D. Licari and G. Comandè, "ITALIAN-LEGAL-BERT models for improving natural language processing tasks in the Italian legal domain," Computer Law and Security Review, vol. 52, 2024.

[70] N. Gao, Y. Wang, P. Chen, and J. Tang, "Boosting Short Text Classification by Solving the OOV Problem," IEEE/ACM Trans Audio Speech Lang Process, vol. 31, pp. 4014–4024, 2023.

[71] G. D. Bianco, D. Duarte, and M. A. Gonçalves, "Reducing the user labeling effort in effective high recall tasks by fine-tuning active learning," J Intell Inf Syst, vol. 61, no. 2, pp. 453–472, 2023.

[72] E. Vidal, A. H. Toselli, and J. Puigcerver, "Lexicon-based probabilistic indexing of handwritten text images," Neural Comput Appl, vol. 35, no. 24, pp. 17501–17520, 2023.

[73] S. T. Vu, M. Le Nguyen, and K. Satoh, "Abstract meaning representation for legal documents: an empirical research on a human-annotated dataset," Artif Intell Law (Dordr), vol. 30, no. 2, pp. 221–243, 2022.

[74] P. Bhattacharya, K. Ghosh, A. Pal, and S. Ghosh, "Legal case document similarity: You need both network and text," Inf Process Manag, vol. 59, no. 6, p. 103069, 2022.

[75] S. Bi, Z. Ali, M. Wang, T. Wu, and G. Qi, "Learning heterogeneous graph embedding for Chinese legal document similarity," Knowl Based Syst, vol. 250, p. 109046, 2022.

[76] M. Li, H. Yu, G. Fan, Z. Zhou, and J. Huang, "ClassSum: a deep learning model for class-level code summarization," Neural Comput Appl, vol. 35, no. 4, pp. 3373–3393, 2023.

[77] X. Li, X. Wu, Z. Luo, Z. Du, Z. Wang, and C. Gao, "Integration of global and local information for text classification," Neural Comput Appl, vol. 35, no. 3, pp. 2471–2486, 2023.

[78] D. A. Rachkovskij, "Representation of spatial objects by shift-equivariant similarity-preserving hypervectors," Neural Comput Appl, vol. 34, no. 24, pp. 22387–22403, 2022.

[79] C. O. Truică, A. I. Stan, and E. S. Apostol, "SimpLex: a lexical text simplification architecture," Neural Comput Appl, vol. 35, no. 8, pp. 6265–6280, 2023.

[80] D. Zhang, H. Zhang, L. Wang, J. Cui, and W. Zheng, "Recognition of Chinese Legal Elements Based on Transfer Learning and Semantic Relevance," Wirel Commun Mob Comput, vol. 2022, 2022.

[81] H. Zhang, B. Pan, and R. Li, "Legal Judgment Elements Extraction Approach with Law Article-Aware Mechanism," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 21, no. 3, 2022.

[82] Y. Feng, C. Li, J. Ge, B. Luo, and V. Ng, "Recommending statutes: A portable method based on neural networks," ACM Trans Knowl Discov Data, vol. 15, no. 2, 2021.

[83] A. Mandal, K. Ghosh, S. Ghosh, and S. Mandal, Unsupervised approaches for measuring textual similarity between legal court case reports, vol. 29, no. 3. Springer Netherlands, 2021.

[84] M. Shukla, D. Dharme, P. Ramnarain, R. Dos Santos, and C. T. Lu, "DIGDUG: Scalable Separable Dense Graph Pruning and Join Operations in MapReduce," IEEE Trans Big Data, vol. 7, no. 6, pp. 930–951, 2021.

[85] J. Ge, Y. Huang, X. Shen, C. Li, and W. Hu, "Learning Fine-Grained Fact-Article Correspondence in Legal Cases," IEEE/ACM Trans Audio Speech Lang Process, vol. 29, pp. 3694–3706, 2021.

[86] N. X. T. Le, M. Shahbazi, A. Almaslukh, and V. Hristidis, "Query by documents on top of a search interface," Inf Syst, vol. 101, p. 101793, 2021.

[87] J. Dhanani, R. Mehta, and D. P. Rana, "Legal document recommendation system: a dictionary based approach," International Journal of Web Information Systems, vol. 17, no. 3, pp. 187–203, 2021.

[88] A. Paul, R. Pramanik, S. Malakar, and R. Sarkar, "An ensemble of deep transfer learning models for handwritten music symbol recognition," Neural Comput Appl, vol. 34, no. 13, pp. 10409–10427, 2022.

[89] D. Debnath, R. Das, and P. Pakray, "Extractive single document summarization using multi-objective modified cat swarm optimization approach: ESDS-MCSO," Neural Comput Appl, vol. 4, 2021.

[90] Y. Wang, J. Cao, and H. Tao, "Graph convolutional network with multi-similarity attribute matrices fusion for node classification," Neural Comput Appl, vol. 35, no. 18, pp. 13135–13145, 2023.

[91] A. Hassani, A. Iranmanesh, and N. Mansouri, "Text mining using nonnegative matrix factorization and latent semantic analysis," Neural Comput Appl, vol. 33, no. 20, pp. 13745–13766, 2021.

[92] D. Li and E. Kanoulas, "When to Stop Reviewing in Technology-Assisted Reviews," ACM Trans Inf Syst, vol. 38, no. 4, 2020.

[93] S. Ghodratnama, A. Beheshti, M. Zakershahrak, and F. Sobhanmanesh, "Extractive Document Summarization Based on Dynamic Feature Space Mapping," IEEE Access, vol. 8, pp. 139084–139095, 2020.

[94] M. Y. Saeed, M. Awais, R. Talib, and M. Younas, "Unstructured Text Documents Summarization with Multi-Stage Clustering," IEEE Access, vol. 8, pp. 212838–212854, 2020.

[95] A. E. Ezugwu, A. K. Shukla, M. B. Agbaje, O. N. Oyelade, A. José-García, and J. O. Agushaka, Automatic clustering algorithms: a systematic review and bibliometric analysis of relevant literature, vol. 33, no. 11. 2021.

[96] M. Younas, D. N. A. Jawawi, I. Ghani, and M. A. Shah, "Extraction of non-functional requirement using semantic similarity distance," Neural Comput Appl, vol. 32, no. 11, pp. 7383–7397, 2020.

[97] Zhang. Wang, C., Zhang, Y., Ding, H., "The path of management of dispute cases of legal issues of webcasting bandwagon industry in the information age," Applied Mathematics and Nonlinear Sciences, vol. 8, no. 2, pp. 3383–3392, 2023.

[98] Y. Liu, T. P. Tan, and X. Zhan, "Iterative Self-Supervised Learning for Legal

Similar Case Retrieval," IEEE Access, vol. 12, no. December 2023, pp. 17231–17241, 2024.

[99] M. Makawana and R. G. Mehta, "A novel network-based paragraph filtering technique for legal document similarity analysis," Artif Intell Law (Dordr), no. 0123456789, 2023.

[100] A. Z. Khameneh, M. Ghaznavi, A. Kilicman, Z. Mahad, and A. Mardani, "A maximal-clique-based clustering approach for multi-observer multi-view data by using k-nearest neighbor with S-pseudo-ultrametric induced by a fuzzy similarity," Neural Comput Appl, vol. 0123456789, 2024.

[101] L. Chen, T. Richardson, S. Lee, and D. Johnson, "Information Retrieval methods in large legal databases," Information Processing & Management, vol. 60, no. 2, pp. 102-118, 2023.

[102] F. Zhao, A. Wong, and K. Thomas, "Challenges in implementing deep learning for legal document analysis," Journal of Big Data, vol. 8, no. 2, pp. 56-70, 2021.

[103] J. Kim, M. Lee, and R. Park, "Integration of IR systems in existing legal frameworks," Computers & Law, vol. 50, no. 2, pp. 112-126, 2021.

[104] Dan, X., Chen, Z., Xu, Y, "Understanding Legal Texts: Challenges and Advances in Legal Natural Language Processing," Journal of Information Processing, 2023.

[105] Y. Shao et al., "An Intent Taxonomy of Legal Case Retrieval," ACM Transactions on Office Information Systems, vol. 42, no. 2, pp. 1–27, Dec. 2023.

[106] Le, Y., Cao, Z., Fu, Q., Liu, Y., Wu, Y, "Advanced Semantic Analysis for Legal Document Similarity Using BERT," Journal of Semantic Computing, 2024.

[107] Son, H., Lee, S., Park, M, "Automating Legal Document Review with NLP: Efficiency and Accuracy," Journal of Legal Automation. 2022.

[108] H. Niu, C. Ma, P. Han, S. Li, and Q. Ma, "A Novel Semantic Cohesion Approach for Chinese Airworthiness Regulations: Theory and Application," IEEE Access, vol. 8, pp. 227729–227750, Jan. 2020.

[109] Ju, Y., Zhang, Y., Liu, W, "Practical Applications of NLP in Legal Case Analysis," Journal of Law and AI, 2024.

[110] Guo, X., Zhao, J., Tian, Y., Wang, Y, "The Nuances of Legal Language: Leveraging NLP for Legal Document Analysis," International Journal of Computational Law, 2023.

[111] Jung, H., Kim, S., Lee, Y, "Using NLP for Contract Review and Legal Research: Case Studies," Legal Tech Journal, 2022.

[112] Li, M., Wu, Q., Chen, L, "The Impact of Deep Learning and Transformers on Legal NLP," Journal of Artificial Intelligence in Law, 2023.

[113] Shao, Z., Li, B., Wang, H., & Zhang, L, "Advanced Information Retrieval Techniques in Legal Document Analysis: Managing Complexity and Enhancing Efficiency," Journal of Legal Information Systems, 2023.

[114] Makawana, M., & Mehta, M, "High-Speed Legal Document Retrieval Using Advanced IR Systems," Journal of Legal Informatics, 2023.

[115] Resck, L., Raffaeli, L., & Da Silva, F, "Advanced Query Techniques for Legal Information Retrieval," Journal of Information Retrieval, 2023.

[116] Amato, G., Falchi, F., & Gennaro, C, "Integrating NLP and IR for Enhanced Legal Document Retrieval," Journal of Legal Tech, 2023.

[117] Bianco, V., Cannata, N., & Ghidini, C, "The Application of BM25 in Legal Document Retrieval," Journal of Computational Law, 2023.

[118] Peng, X., Liu, Y., & Gao, Y, "Generalization of deep learning models in legal document retrieval," Journal of AI and Law, 2021.

[119] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT, 4171-4186, 2019.

[120] Yadav, A., Singh, R., & Gupta, S, "Enhancing document similarity analysis with deep learning techniques in legal tech," International Journal of Law and AI, 2022.

[121] Vo, T, "Deep learning approaches for improving accuracy in legal document similarity detection," Legal Informatics Review, 2022.

[122] Strubell, E., Ganesh, A., & McCallum, A, "Energy and Policy Considerations for Deep Learning in NLP," Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645-3650, 2019.

[123] S. Yilmaz and S. Toklu, "A deep learning analysis on question classification task using Word2vec representations," Neural Computing & Applications, vol. 32, no. 7, pp. 2909–2928, Jan. 2020.