

Arabic Legal AI through Knowledge Graphs: A Systematic Review, Ontology Design, and Jordanian Labor Law Case Study

Osama Alqasem¹ and Omar Qawasmeh^{1,*}

¹ Data Science Department, Princess Sumaya University for Technology, Amman, Jordan.

* Corresponding author: o.alqawasmeh@psut.edu.jo. Contributing author: osa20258070@std.psut.edu.jo.

Received: 24 May 2026 **Accepted:** 16 July 2026 **Published:** 6 October 2026

DOI: 10.15849/ijasca.128 **License:** Open access under CC BY 4.0. Authors retain copyright.

Publisher: Al-Zaytoonah University of Jordan (ZUJ), Amman, Jordan.

Cite this article: O. Alqasem and O. Qawasmeh, "Arabic legal AI through knowledge graphs: a systematic review, ontology design, and Jordanian labor law case study," *Int. J. Advance Soft Compu. Appl.*, vol. 18, no. 3, pp. 1–17, 2026, doi: <https://doi.org/10.15849/ijasca.128>

ABSTRACT

In this paper, the growing trend of legal documents and the need for efficient methods for processing and representing legal knowledge have been investigated. Even though there is significant progress in legal natural language processing, there is still little research related to Arabic legal documents due to the complexity of the Arabic language, specific legal terminology, and lack of available resources. Two main contributions are offered by the paper. First, the comprehensive survey of Arabic and multilingual legal document processing is provided with the analysis of existing methods for Legal AI, legal ontologies, and knowledge graphs. Second, the Arabic legal ontology and building a prototype knowledge graph based on Jordanian Labor Law are presented. The proposed approach provides a foundation for further applications of Arabic Legal AI.

Keywords: Natural Language Processing; Semantic Web; Knowledge Graph; Ontology; Legal.

1 Introduction

Legal texts belong to one of the most complex kinds of writing across the world [3][44]. A great number of legal texts are produced each day, much of which stays in an unstructured format [3]. Understanding and analyzing the text become complicated in light of the difficult nature of the legal language, ambiguities in professional vocabulary, and the need for precise thinking in law [44]. Such issues become even more complicated when applied in practice since errors in the process can have very serious consequences both legally and socially [32]. Given recent advancements in artificial intelligence (AI) and natural language processing (NLP), there has been much effort devoted by researchers in the construction of intelligent legal systems which could help lawyers perform such tasks as document categorization, legal information searching, legal argument analysis, summarization, contract analysis, and legal Q&A [3]. While early research mainly relied on machine learning and manual feature design, the recent efforts have turned to deep learning and representation learning [8].

On the other hand, Large Language Models (LLMs) have been shown to significantly improve their understanding of legal language [10][21] thanks to pre-training and fine-tuning on instructions. They are capable of modeling long contexts, generating summaries, and making inferences based on legal text [21]. However, LLMs-powered legal systems might be faced with the problems of factual reliability, domain specificity, explainability, and multilingual robustness, which are especially important in a specialized setting [5][10]. Simultaneously, Knowledge graphs (KGs) are a cornerstone representation of legal knowledge containing entities (people, court, etc.), relations (cites, amends, violates, etc.), and logical rules [20][34]. With their explainability and support for symbolic reasoning, KGs supplement data-driven models [20]. Thus, the combination of LLMs with KGs can be a good approach towards developing multilingual AI for the legal field as it combines NLP with legal knowledge and reasoning [33][34]. While advances have been made within NLP for law concerning English and few other

languages, there is relatively little exploration in relation to the Arabic legal language [2][19]. There are several challenges when building intelligent and interpretable Arabic legal multilingual systems. These include linguistic richness, complex morphology, and lack of annotated corpora [19]. Hence, there is a need to look into solutions to build intelligent Arabic legal multilingual systems that go beyond monolingual solutions [2]. In this study, the main contributions will be the following:

- A review of literature in the field of Arabic legal document processing, legal ontology, and knowledge graph.
- The development of a domain-specific Arabic legal ontology for Jordanian labor law, accompanied by a sample prototype KG to demonstrate its applicability

In this systematic review, we will focus only on KG-based approaches covering legal knowledge representations, legal ontologies, and graph-based legal reasoning. This also includes approaches combining language models (LM) and KG. List 1 describe an example of part of turtle representation of the proposed legal ontology.

List 1 Turtle representation of the proposed legal ontology

```
@prefix ex: <http://example.org/labor#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
ex:Employee rdfs:label "Employee"@en, "Muwazzaf"@ar, "Employe"@fr .
ex:Employer rdfs:label "Employer"@en, "Sahib_Amal"@ar, "Employeur"@fr .
ex:Article56 rdfs:label "Article 56"@en, "Al-Maddah_56"@ar, "Article 56"@fr .
ex:worksFor rdfs:label "works for"@en, "ya3mal_lada"@ar, "travaille_pour"@fr .
ex:Ahmad ex:worksFor ex:Factory .
```

The rest of the paper is organized as follows: In Section 2, the literature review section is given. Section 2.1 will present the Criteria for Included Studies. In Section 2.2, the systematic review on KG-related literature will be presented. In Section 3, the proposed ontology is introduced. In Section 4, discussions will follow. Finally, Section 5 concludes the paper with future directions while Section 6 the Acknowledgments and Section 7 the References.

2 Literature review

Document processing in the legal domain has moved from simple data storage and keyword-based search processes to using artificial intelligence for legal reasoning. While traditional systems do well at document processing, they cannot completely understand legal meanings such as responsibilities, referencing, timeliness, and legal interpretation, especially when dealing with multilingual and Arabic legal documents [3]. Therefore, recent research has focused on AI approaches to legal document management, particularly NLP, LLMs, and Knowledge Graphs. NLP supports key legal tasks such as document classification, information retrieval, named entity recognition, citation extraction, summarization, question answering, and legal judgment prediction [3]. LLMs, in addition, expand these capabilities by enabling more advanced applications, including legal and contract analysis, clause extraction, and assisted legal drafting; however, they face challenges related to hallucination, factual reliability, ambiguity, and interpretability [22]. KGs offer a complementary advantage by presenting legal entities and their relationships in a structured and understandable way, enabling legal reasoning, semantic retrieval, and multilingual or cross-jurisdictional knowledge organization [39]. Research papers related to KG addressed interpretability and the gap in structure through clear and explicit legal representation using entities, relationships, and ontological schemas. General surveys in this field provided a general understanding and methodological foundations for refinement, evaluation,

and maintenance [34]. While legal-domain KG studies demonstrate how structured legal metadata and semantic links between entities and legislation can be constructed and applied to applications such as answering legal questions [27][42]. There are also other techniques that deal with the same problem from different perspectives, such as the paper introduced by [1] that emphasize the need to integrate semantic understanding with systematic access to knowledge, since answering legal questions often relies on references, citations, and formal legal concepts [36] systematically mapped legal research related to NLP from 2015 to 2022, covering tasks such as classification, summarization, information extraction, and question answering. The study highlighted the need for curated datasets, ontologies, and organized legal resources, which supports the role of Knowledge Graphs in explainable legal AI. An overview of recent literature reviews, as summarized in Table 1. However, in the scope of our paper, we will focus on our contribution based on the KG related domain.

Table 1 Comparison of selected survey and review papers in the legal domain by paradigm type and primary focus.

Paper	Year	Paradigm Type	Primary Focus
Paulheim [34]	2017	KG	KG refinement and evaluation methods.
Sovrano et al. [42]	2020	KG + NLP	Legal knowledge extraction for KG-based QA.
Muninggar and Krisnadhi [27]	2021	KG	Semantic KG construction for legal documents.
Quevedo et al. [36]	2023	NLP	Evolution and trends in legal NLP research.
Abdallah et al. [1]	2023	KG + NLP	Legal QA using NLP and KGs.

2.1 Eligibility Criteria for Included Studies

This review examines current academic literature related to the topics of this research article. This covers journal papers, conference papers, and selected preprints concerning legal knowledge representation and legal AI. The used search query was as follows: “ (“Ontology” OR “Knowledge Graph” OR “KG” OR “Terminology”) AND (“Law” OR “Legal”) ”. The literature search was performed using several electronic databases¹. These databases were chosen because they offer comprehensive literature that is related to the topics under consideration. Peer-reviewed journal and conference papers have been preferred wherever possible. Also, a select few preprints from arXiv have been taken into account only in cases where they provided substantial advancements. The following inclusion and exclusion criteria were adopted:

1. Inclusion criteria: Legal-domain studies on ontologies, KGs, terminological resources, AI, reasoning, question answering, AI integration, and Arabic/multilingual/cross-lingual legal systems.
2. Exclusion criteria: Non-legal AI/NLP studies, editorials, commentaries, letters, tutorials, non-technical papers, studies with weak methodology, unavailable full text, or overlapping/low-contribution preprints

For each reviewed article, key information was analyzed, including the research objective, representation method, ontology or knowledge graph use, AI techniques, evaluation approach, and main findings. The papers were then grouped into major themes, enabling the identification of research trends and methodological differences.

¹ E.g.: Google Scholar, IEEE Xplore, ACM Digital Library, ScienceDirect, Informatica, and arXiv

2.2 Systematic review

Recent research in KGs in the legal field reveals a systematic evolution from basic ontology to graph construction, semantic retrieval, legal aid and integration with modern artificial intelligence systems. This development is based on studies that focus on formalizing legal knowledge through ontology construction and conceptual modeling, particularly in Arabic legal contexts where organized semantic resources are still scarce. In this direction Bouaziz Mezghanni and Gargouri [6] introduces the semiautomatic learning of an ontology from Arabic legal documents to build Arabic Panel ontology. The method uses a hybrid approach for concepts identification based on a linguistic phase that apply morphological and syntactical grammars followed by a filtering process relying on a machine learning algorithm. The whole approach exploits the peculiar properties of documents during the learning process. The learning algorithm is based on semantic, content and structural features. The approach combines linguistic tools like Nooj [41] to extract the term candidates, and FCA [15] and its extension RCA [17] to extract hierarchical and transversal relations. The authors aim to develop a mechanism for reformulating contextual, ontology-based queries to retrieve legal information.

A complementary effort by Mezghanni and Gargouri [25] that they introduce CrimAr, a hand-designed ontology of Arab criminal law intended to be the gold standard for evaluating semi-automatic ontology learning methods in the Arab legal field, where high-quality Ontologies are rare. The ontology was constructed using a middle-out strategy. Researchers began by selecting a corpus from the Tunisian Code of Criminal Procedure. Subsequent to the preprocessing of the text according to the Arabic language and linguistic analysis through NooJ [41], the resulting concepts were validated by legal experts and structured hierarchically. For structuring the higher hierarchy, CrimAr reuses top-level concepts of LRI-Core [7], which is an initial common sense legal ontology, and then translated into Arabic. Afterward, it performs corpus-based concept refinement downward from top to bottom. The identification of the relations was done based on extraction using patterns and the distinction between descriptive and normative law. The performance of the tool is compared with learned Arabic legal ontology (ALO) [6] using hierarchical similarity score (reported similarity ≈ 0.8) and precision/ recall (≈ 0.89 and ≈ 0.75 respectively). The scores reveal that the resulting ontology has enough consistency to serve as a reference standard. Based on the findings of previous works, we can conclude that in order to query the legal knowledge base and apply it in practical cases, it needs to be formalized into ontologies that reflect legal relations.

Based on this conceptual framework, another line of work focuses on building legal knowledge graphs directly from legal documents, which leads to the transformation of lengthy judicial texts into machine-readable graphic representations. Authors in [40] introduced JudgmentGraph, an ontology-based methodology for constructing legal knowledge graphs (LKGs) from Indian Supreme Court judgments. The goal is to transform lengthy and unstructured legal opinions into triples for scholarly analysis and legal reasoning. The process starts with the identification of named entities in law, that is, the identification of main participants and references, like parties, judges, court, legislation or clauses, and precedents. Entities are then matched to a legal ontology that defines legal relation types and constraints on their scope, which guarantees that only legitimate legal relations are created. Potential relations are created and filtered via disambiguation based on cues from text, such as distance/proximity scoring of mentions across sentences and paragraphs, trigger patterns, and rules that remove mutually exclusive relations. The resulting graph is saved as RDF and can be displayed in the form of graphs. For the assessment of the model, gold standards were created by manually annotating relations in fifteen decisions of the Supreme Court; the quality turned out to be very high (precision 81.4%, recall

88.4%, F1 84.7%), and a qualitative case study illustrated that extracted relations could be interpreted well, pointing out common mistakes, such as the false connection of legal provisions to other text elements.

Another area of study is focused on semantic legal retrieval and querying, where the task is not just limited to storing legal knowledge but also involves searchability, interpretation, and responding according to user needs. One example is the legal ontology and knowledge-graph based knowledge graph approach for legal information retrieval presented by Pham et al. [35]. This method combines a legal ontology layer made up of legal concepts, relations, and inference rules with a knowledge graph layer obtained from legal texts. The intent of this is to improve semantic retrieval and produce better answers regarding legality than simple keyword techniques. The model includes a pipeline for legal knowledge extraction, optimization of graph relations or triples, as well as a query graph matching technique that converts user query into graphs to retrieve the required legal knowledge in an efficient manner. Evaluation on Vietnamese road traffic law documents showed that the model performed excellently and achieved a success rate of approximately 82.6%. When compared with other NLP models (TF-IDF [26], BM25 [37], TIWS, TPS [23]), the method performs much better. Moreover, this paper also includes benchmarking of the proposed model against several other LLMs, like ChatGPT-40 [32], and Gemini 2.5 [16]. The authors state that LLMs do not offer as precise legal answers and that using ontology and knowledge graph provides more detailed answers grounded in legal articles.

A closely related effort done by Nguyen et al. [29] that they presented the LegalOnto (Legal Relational Model), a representation that integrates ontology and graphs to capture the semantics of norm legal documents, going beyond traditional keyword-based or article-based methods. The core idea is to combine the relational ontology layer- concepts, relationships between concepts, and inference rules -with the conceptual graph of key legal phrases. This technique allows for the interpretation of queries by the users based on relations and expanding them through rules. This representation is used by the authors to develop a new querying technique for the legal Vietnamese language queries including classification of user queries such as concept definition, procedure, etc., extraction and expansion of the key phrase by relations, and finally searching relevant articles and concepts from a structured legal document database. This technique is tested on the Intelligent Support System for Vietnamese Land Law 2013 where a subset of queries is used to train the system and assign objectives for each query. Experiments have shown that the correct legal document can be retrieved with more than 60% accuracy which validates that the key contribution is in organizing and relating knowledge semantically.

KGs have been used for legal aid in the form of a practical application, using the legal knowledge to optimize legal operations. According to Wei [43], a KG-based process can be used to generate better legal documents compared to other methods, showing better results regarding structural standards. It involves extracting the entities and relationships from legal literature to formulate case KGs that include entity recognition, relation extraction, and disambiguation. The relations extracted through KGs are then used to find related facts, thus making the process of document formulation more coherent. Multiple experiments were conducted in different fields of law, such as criminal, civil, administrative, and divorce. The comparison was made between the knowledge-schema and rule-based methods, resulting in better performance and more effective document generation using the former. In addition, the research showed that dependency on KGs prevents mistakes and makes the logic of the documents more coherent. Hence, it is important to create a proper model of legal knowledge and improve the connection with other parts, such as NLP and information retrieval systems. Moreover, KGs standardize references to all generated blocks.

A related applied perspective appears from Eddine Meftah et al. [11] where IALAS (Intelligent Arabic Legal Assistant System) is an ontology-based search method and consultation system designed to save the time and effort spent with traditional methods of searching Arabic laws and regulations. To address the scarcity of Arabic legal knowledge bases, the system combines Arabic NLP with OWL ontology to enable more concept-level semantic retrieval rather than keyword-based search. The researchers designed the ontology according to Ontology Development 101 [30] and implemented it using Protégé software [28], whereby they modeled basic classes such as JudicialLaw, Article, LegalRule, Authority, Sector, Keyword, and OfficialJournal, along with object/data properties that capture legal lifecycle relationships (e.g., publication, modification, cancellation, referencing) and article metadata. During runtime, the user enters a text query through the user interface (implemented using Django [14]). The system then performs Arabic preprocessing/analysis using CAMEL Tools [31], and issues SPARQL queries over the ontology via OwlReady2, with SQLite to store ontology links/auxiliary resources. The paper provides a well-defined and functional ontology of Arab law for organizational research, recognizing that such an endeavor is only a first attempt. The need for developing a broader Arab law knowledge model becomes apparent in order to increase both efficiency and usability.

The combination of KGs with sophisticated artificial intelligence structures such as LLMs and multi-agent architectures is a current advancement in the subject matter. An example of such an approach can be found in Cui et al. [9], where the authors present ChatLaw, a Chinese legal system that addresses the problem of LLMs' hallucination and improves their performance by adding three elements, namely: a knowledge graph-enhanced database, MoE-based legal LLMs, and multi-agents imitating law firm processes. First, the authors collect extensive legal data using several sources including question-answer pairs and legal advice. They then organized the knowledge using specialized knowledge graphs. For modeling purposes, Chatlaw selects a top-k of experts for each token (MoE Transformer) and assigns different experts to each different legal subfield. This reduces computational costs and increases the model's ability to recognize specialized terminology. For publication purposes, they implement a four-role Standardized Operating Procedure (SOP): Legal Assistant, Legal Researcher with RAG [24], Senior Lawyer, and Legal Editor. The Assistant fills in missing information on cases by answering questions posed by the knowledge base, while the Researcher retrieves up-to-date statuses/cases using retrieval-augmented generation (RAG). The rest of the agents create guidelines and prepare consulting documents based on safety firewall checks. In practice, the proposed model outperformed various other models, such as GPT-4 [32] using LawBench [13], by 7.73%, and the Chinese Unified Legal Professional Qualification Exam, by 11 points. Additionally, it received better evaluations in terms of aspects relevant to actual consulting, such as comprehensiveness, validity, guidance, and reliability.

Current studies expand the area of application of knowledge-based legal intelligence to include regulatory compliance and machine-verifiable legal analysis. Falcarin et al. [12] give a clear illustration of such studies in terms of their research-preview approach to automating compliance with regulations in view of NLP and knowledge graphs. In particular, such an innovative approach is driven by the rapid increase in digital EU regulations (AI Act, NIS2, Cyber Resilience Act, and DSA) together with high costs and slow progress related to manual audit processes conducted by legal experts. A key idea includes a compliance architecture for auditors where regulations and organizational information are translated into knowledge graphs and, thus, compared with each other. This will allow identifying the differences between them and gaps and possible breaches; in turn, this will make it easier to draw machine-verifiable conclusions concerning obligations, capabilities, and prohibitions. Keyword-based or article-based methods for testing their approach, the authors used a text-to-knowledge conversion tool SPRING [4] and

knowledge-graph analysis technique Neo4j [38]. Moreover, a knowledge-base generation from legal texts using LLMs, namely, ChatGPT-4o [32], to extract triples in SPO format was considered. It turned out that knowledge representations based on SPO triplets resulted in fewer nodes and better suited text structuring and compliance verification. Further plans for research involve systematic evaluation using actual regulations and SRS, compliance checking by domain experts, and GraphRAG [18] and inclusion-based similarity usage for providing more actionable compliance reports.

3 LegalKG: A Legal Ontology-Driven KG

LegalKG² is introduced here, which represents the proposed framework for representing the knowledge obtained through the analysis of Jordanian laws. The framework utilizes ontologies as its foundation and focuses mainly on representing the knowledge related to Jordanian labor laws. The LegalKG framework represents key concepts related to laws including involved entities, rights, duties, procedures, and semantic connections between them. These goals have been achieved using knowledge extraction techniques to convert knowledge expressed in unstructured texts into structured forms suitable for computational applications.

The creation of LegalKG adheres to a seven-stage method discussed in [30], which includes ontology design. This method consists of the following stages: (1) definition of the domain and scope; (2) examination of relevant ontologies; (3) listing of relevant terms; (4) class structure construction; (5) property and relation definitions; (6) constraint identification; and (7) knowledge base establishment. In this work, this methodology was adapted to develop an ontology for Jordanian labor law. The development process was also supported and consulted by a legal expert to ensure that the modeled concepts, relations, and constraints accurately reflect the legal terminology, interpretation, and practical meaning of the domain. As a result, the ontology forms the conceptual layer of LegalKG, while the instantiated legal entities and relations constitute its knowledge graph layer.

Before delving into the details of the LegalKG structure, we will introduce the fundamental concepts required to understand ontology-driven knowledge graph construction. This includes the role of ontologies in representing knowledge, the difference between fundamental concepts such as TBox and ABox layers, and how these terms work together to produce a final product that presents knowledge in an organized manner.

3.1 Ontologies and knowledge graphs in the legal domain

Ontologies define the fundamental terms, concepts, and relationships within a specific field. In the legal field, ontologies provide a common conceptual framework for representing entities such as laws, articles, courts, legal cases, and institutions, and also present the semantic relationships between them. This helps ensure the accurate interpretation of legal terms and supports the automation of legal knowledge processing.

Knowledge graphs augment ontologies in that knowledge pertaining to a specific field can be captured by means of a series of nodes and edges in a graph structure. In the case of legal knowledge graphs, the nodes can refer to laws, articles, cases, or organizations, whereas the edges can represent relations such as citation, issued by, or applied to. The graph architecture is especially well-suited to capturing knowledge in the legal realm since information and connections within this realm are highly intertwined. Knowledge graphs are better suited than CSV data or relational tables for capturing semantic connections between data due to their triple format. Ontology-based knowledge can be divided into two separate but equally important parts:

² <https://github.com/osamaqsm/A-Legal-Knowledge-Graph-for-Arabic->

the Terminological Box (TBox) and the Assertional Box (ABox). While the TBox identifies and categorizes types of objects and relationships in the domain of interest, the ABox defines examples of each type found in legal documents. TBox, therefore, reflects the conceptual scheme of the legal field by defining categories such as Law, Article, Court, and Legal Case as well as relationships among them like has Article, issued by, and refers to. Thus, the concept of a legal domain is reflected in the TBox. On the other hand, ABox is a list of examples of each category mentioned above taken from actual legal texts. Hence, one particular law, article, or decision is described in the ABox through relationships specified in TBox.

3.2 Our proposal: legal Jordanian ontology

Description

Next, we describe an ontology created for the Jordanian Labor Law, which is the conceptual layer of our proposed LegalKG. Specifically, the ontology covers the main legal concepts, actors, structures, and semantic relationships found in the labor law, including the employment relation, its contractual nature, rights and obligations of the two sides involved, and legal violations resulting from legal duty breach. The ontology was created after consulting with a legal expert who contributed towards ensuring the correct identification of main legal entities and their relations from a semantic point of view.

The presented ontology is a legal-semantic ontology in contrast to document-based ontologies, as it is not restricted to modeling the structure of laws but includes also legal knowledge captured in terms of labor parties, contract types, labor parties' rights and obligations, and labor offenses. Therefore, the ontology serves best as a tool for knowledge organization, legal querying, legal reasoning, and then knowledge graph construction from Jordanian legal documents.

At the top level, the ontology consists of several overlapping groups, namely the structure of legal documents, legal authorities, employment parties, types of contracts and agreements, rights, obligations, and legal violations. Fig. 1 shows a high-level presentation of the categories proposed through ontology. And Fig. 2 shows a fragment of the proposed ontology indicating the main classes, some of their subclasses, and certain semantic relationships.

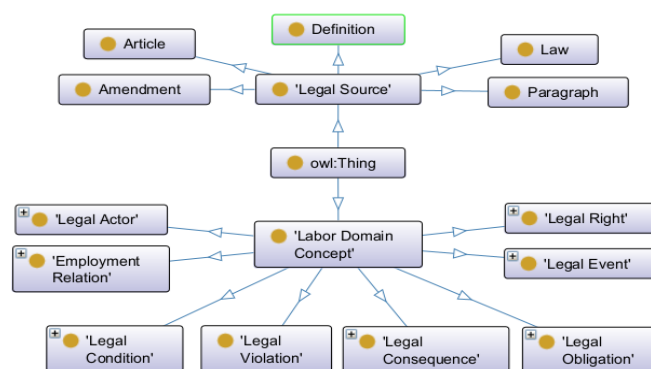


Fig. 1 High-level presentation of the categories proposed through ontology

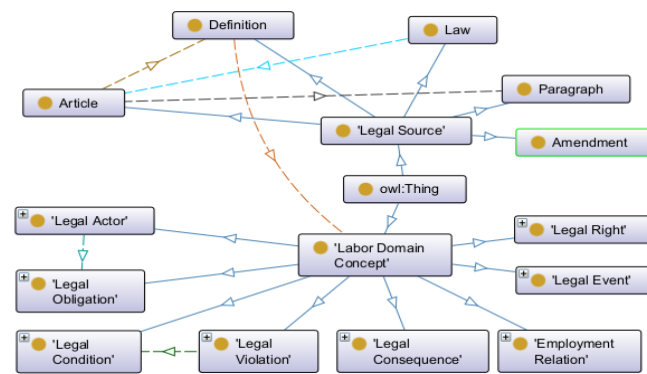


Fig. 2 Partial view of the proposed ontology for the Jordanian labor law domain, showing the main classes, selected subclasses, and representative semantic relations among key legal concepts.

Proposed classes

The proposed ontology is based on a two-level semantics approach to organizing Jordanian labor laws. Specifically, there are two levels involved – the legal source layer and the labor-domain concept layer. The former allows modeling the textual structure of the law and the latter provides for modeling the principal concepts of the legal system. Within the legal source layer LegalSource is used as the main superclass for all legal textual units. The following legal classes are included here: Law, Article, Paragraph, Definition, and Amendment. Law is the name of the actual Jordanian labor law. Article and Paragraph serve as its components that structure the law. Definition class contains legal definitions that can be found in Jordanian labor law, while Amendment models legal modifications that take place. All classes mentioned above are linked via hasArticle, hasParagraph, hasDefinition, and amendedBy relations, correspondingly. The second level of the ontology is presented by LaborDomainConcept class. It includes all of the principal legal concepts of labor-law domain, such as LegalActor, EmploymentRelation, LegalRight, LegalObligation, LegalViolation, LegalCondition, LegalConsequence, and LegalEvent. These classes are required in order to represent semantic meaning behind the text of the document in question. Actors of labor laws are represented in this ontology by LegalActor class that includes subclasses like Worker, Employer, Establishment, and NaturalPerson. Employment relations are represented by EmploymentRelation class that primarily covers EmploymentContract with such sub-classes as FixedTermContract, IndefiniteTermContract, WrittenContract, and OralContract. Relations between actors include isPartyTo, hasEmployer, and hasWorker relations. Workers' legal rights and their respective violations will be covered by LegalRight and LegalViolation classes, correspondingly. In particular, LegalRight class will cover such subclasses as WageRight, LeaveRight, SickLeaveRight, WeeklyRestRight, SafeWorkEnvironmentRight, LitigationRight, SocialSecurityRelatedRight, and EndOfServiceBenefitRight. On the other hand, LegalObligation class represents PayWages, ProvideSafeWorkEnvironment, AssignAgreedWork, PerformWork, PreserveConfidentiality, and FollowLawfulInstructions legal obligations. Legal violations are represented by WageDelay, ArbitraryDismissal, NonProvisionOfAgreedWork, AssignmentOfDifferentWork, DisclosureOfEmployerSecrets, and WorkNegligence subclasses. Legal events, conditions, and consequences are also needed to be added to the ontology for covering occurrence of violations and resulting legal consequences. Finally, it should be noted that some relationships connecting the two layers are also needed (defines, regulates, regulatedBy, and supportedByArticle). Table 2 illustrates main ontology classes.

Table 2 High-level ontology class groups

Class Group	Description
-------------	-------------

LegalSource	Symbolizes the architectural design of legal texts and resources, consisting of laws, articles, paragraphs, definitions, and amendments.
LaborDomainConcept	Presents the conceptual level of labor law, which comprises actors, relations, rights, duties, violations, conditions, consequences, and legally relevant events.
LegalActor	Reflects the key participants of labor law, such as employees, employers, enterprises, and individuals.
EmploymentRelation	Shows labor-related legal relations, which consist of labor agreements and their basic types (such as fixed-term and indefinite-term, written and oral agreements).
LegalRight	Covers rights recognized in labor law, such as wage rights, leave rights, social security rights, weekly rest, safe work environment, litigation, and end-of-service benefits.
LegalObligation	Describes duties imposed on parties, such as paying wages, providing safe working conditions, assigning agreed work, performing work, confidentiality, and following lawful instructions.
LegalViolation	Models breaches of labor law obligations or rights, including wage delay, arbitrary dismissal, non-provision of agreed work, and work negligence.
LegalCondition	Represents situations affecting interpretation or consequence within the law, such as discharge without justifiable cause, delays of more than seven days, contracts, and the status of social security protection.
LegalConsequence	Captures legal outcomes and remedies, including compensation, legal claims, termination of contract, employer liability, disciplinary action, and end-of-service benefits.
LegalEvent	Represents legally relevant events in the employment lifecycle, such as contract formation, wage payment, dismissal, leave requests, medical examination, and work injury.

Relations - Object Properties & Data Properties

In order to relate these classes and represent the meaning of Jordanian labor law, the ontology makes use of a set of object properties and datatype properties. These properties link the legal source layer with the labor-domain concept layer and allow the ontology to represent both the structure of the legal text and the legal meaning associated with it.

At the legal source level, the ontology uses structural object properties to model the organization of the law. The property `hasArticle` links a law to its articles, while `hasParagraph` connects articles to their paragraphs. In addition, `hasDefinition` links articles to the legal definitions they contain, and `amendedBy` connects a law to its amendments. These relations preserve the internal structure of the legal document and support article-level and paragraph-level referencing.

The ontology also includes bridge relations that connect the textual structure of the law with labor-domain concepts. The property `defines` links a legal definition to the concept it defines, while `regulates` connects an article to the labor-law concept it regulates. Its related inverse direction is represented through `regulatedBy`, which links a legal concept back to the article that regulates it. Similarly, `supportedByArticle` connects legal concepts such as rights, obligations, violations, conditions, and consequences to the articles that provide their legal basis.

Labor relations are represented through properties that connect employment contracts and legal actors. The property `isPartyTo` links a legal actor to an employment relation or contract, while `hasEmployer` and `hasWorker` connect an employment contract to the employer and worker involved in it. These relations allow the ontology to represent the basic structure of the employment relationship.

Rights, obligations, and violations are connected through semantic relations. The property `violatesRight` links a legal violation to the right affected by that violation, while

breachesObligation connects a violation to the obligation that has been breached. These characteristics are necessary because they define the relation between illegal behavior and specific legal rights and obligations. The ontology also includes datatype properties to store textual and descriptive values. For example, articleNumber records the number of a legal article, while bodyText stores the textual content of legal sources such as articles or paragraphs. In addition, Arabic and English labels are provided through annotation properties, allowing the ontology to support bilingual representation of Jordanian labor-law concepts. Table 3 shows main ontology relation groups and examples. As illustrated in Fig. 3 the ontology represents labor-law concepts, their hierarchy, semantic relations, and links to relevant legal articles and paragraphs.

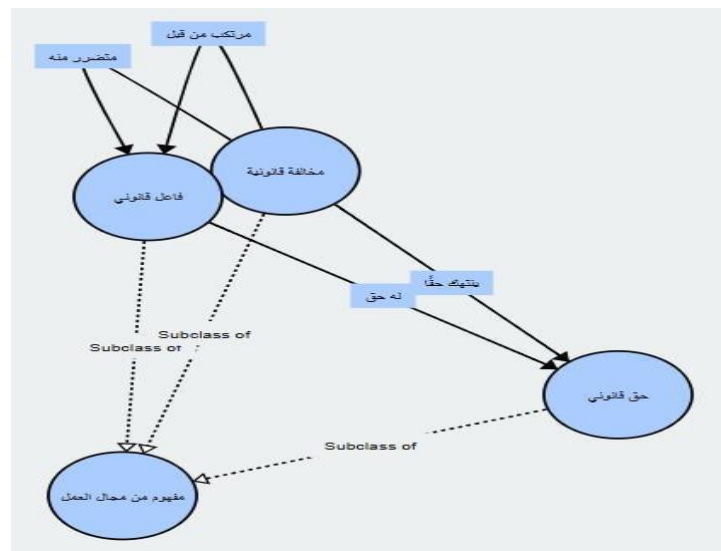


Fig.3 Illustrative Arabic example from the proposed ontology, showing selected class hierarchy relations (subClassOf) and semantic relations among core labor-law concepts.

Table 3 Main ontology relation groups and examples

Relation Group	Description	Examples
Legal text structure	Organizes the internal structure of legal sources, linking laws, articles, paragraphs, definitions, and amendments.	hasArticle, hasParagraph, hasDefinition
Semantic grounding	Connects legal provisions to labor-law concepts, identifying which articles define, regulate, or support them.	defines, regulates, regulatedBy
Employment relations	Represents contractual links between workers, employers, and employment contracts, including contract type and form.	isPartyTo, hasEmployer, hasWorker
Rights and obligations	Models the rights held by legal actors and the obligations imposed on them.	hasRight, hasObligation
Violations and parties	Represents violations, the rights or obligations affected, and the actors involved.	breachesObligation, violatesRight, committedBy
Conditions	Captures conditions that determine legal applicability or interpretation.	hasCondition, appliesTo
Consequences	Links violations or actors to legal outcomes and remedies.	resultsIn, entitlesTo
Events and context	Represents legally relevant events and the employment setting in which they occur.	basedOnEvent, occursIn-ContextOf

To illustrate how the proposed ontology works, a simple example is presented in List 2 to show how abstract concepts are instantiated and connected.

List 2 Turtle example of the proposed legal ontology.

```
@prefix : <http://example.org/jordan-labor-law#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

:Worker a owl:Class ; rdfs:subClassOf :Party ; rdfs:label "Worker"@en .

:Employer a owl:Class ; rdfs:subClassOf :Party ; rdfs:label "Employer"@en .

:Agreement a owl:Class ; rdfs:label "Agreement"@en .

:hasAgreement a owl:ObjectProperty ; rdfs:domain :Party ; rdfs:range :Agreement ;
  rdfs:label "hasagreement"@en
```

KG Prototype Construction and SPARQL-Based Validation

An example prototype of the KG was instantiated based on selected concepts and relationships from the Jordanian labor-law ontology, in addition to some extra operational elements. These additional elements include entities related to labor inspection, labor law violations, administrative measures, and monetary sanctions. In order to show the practical validity of the proposed KG prototype, the validation experiment was done via SPARQL query. Instead of simple structural inspection of the graph, the focus was on retrieving the semantically connected legal information with multi-hop relations. To demonstrate this process, one example competency question was formulated as follows: “What legal consequences, monetary penalties, and supporting legal articles are associated with a labor-law violation and its corresponding legal condition?” The above competency question has been translated to SPARQL query List 3. The SPARQL query starts from a labor-law violation instance and retrieves its related legal condition, consequences, monetary sanctions, and supporting article, as shown in Table 4. The results demonstrate that the proposed KG prototype supports multi-hop semantic retrieval and explainable access to connected legal knowledge.

List 3 SPARQL validation query of the proposed legal KG

```
SELECT ?violationLabel ?conditionLabel ?consequenceLabel ?minJOD ?maxJOD ?articleNo ?articleTitle
WHERE {
  :labor_inspection_violation rdfs:label ?violationLabel ;
    :hasCondition ?condition ;
    :resultsIn ?consequence ;
    :supportedByArticle ?article .

  ?condition rdfs:label ?conditionLabel .
  ?consequence rdfs:label ?consequenceLabel .
  ?article :articleNumber ?articleNo ;
    :articleTitle ?articleTitle .
  OPTIONAL {
    ?consequence :minAmountJOD ?minJOD ;
      :maxAmountJOD ?maxJOD .
  } }
ORDER BY ?consequenceLabel
```

Table 4 Semantic results of the SPARQL query

Violation	Condition	Consequences	Article
مخالفة أحكام التفتيش	عدم إزالة المخالفة خلال سبعة أيام من الإنذار الخطي	إغلاق المؤسسة لحين إزالة المخالفة أو صدور قرار المحكمة	١. مهام مفتش – Article 9 – العمل
		إلزام بإزالة المخالفة	٢.
		غرامة بين 50 و500 دينار	٣.

4 Discussion

Ontologies and knowledge graphs have become important methods for presenting, organizing, and reasoning about legal knowledge. Unlike traditional methods of managing legal documents, they provide structured and clear representations of legal terms, entities, and relationships. This makes them suitable for the legal field, where legal meaning depends on the relationships between entities, laws, clauses, parties involved, rights, duties, violations, and procedures. This literature study shows that the contribution of ontologies to the construction of knowledge graphs is the specification of legal concepts and relations in schemas. Knowledge graphs represent an expansion of the ontology schema approach by adding legal instances in terms of laws, actors, rights, duties, case law, events. Ontologies and knowledge graphs can be used in a variety of applications requiring semantic structuring of data, their retrieval, traceability, and reasoning. It is noted that there are problems associated with the interoperability and standardization of ontologies, especially in the area of Arabic law, where the reuse of ontologies, knowledge bases, and linguistic multilingual resources is not well developed. The main findings of this research suggest that ontologies and knowledge graphs have a strong potential to enable the presentation of legal knowledge in an interpretable and machine-readable format. Specifically, ontologies enable specifying key terms, entities, and relations in the domain, whereas knowledge graphs build on top of this representation by establishing connections between the concepts specified in the ontology and real world instances, such as laws, articles, actors, rights, duties, and violations. The resulting combination of ontology and knowledge graphs enables achieving such benefits as semantic consistency, interpretability of searches, traceability, and reasoning beyond mere keyword-based document searches. The developed Arabic legal ontology can also be leveraged to support different semantic layers in legal applications.

5 Conclusion

This paper examines the current research efforts related to legal documents processing, paying special attention to legal ontologies and knowledge graphs that serve as legal knowledge representation. The literature review reveals that legal documents should not be limited to storing and searching for keywords; rather, they require a comprehensive examination because legal meaning is established based on structured interrelationships between legal norms, articles, rights, duties, entities, violations, and legal actions. This implies that ontology- and KG-based approaches serve as a basis for knowledge organization and structuring.

Based on the literature survey results, the main strengths of using ontologies and knowledge graphs include ensuring semantic consistency, facilitating explainable searching, improving the traceability of information, and performing legal inference. On the one hand, legal ontologies allow modeling legal concepts, while knowledge graphs combine ontologies with legal examples and relations. On the other hand, the current representation of Arabic legal knowledge is poor; namely, it lacks reusable ontologies, knowledge bases, benchmarks, and multilingual semantic resources.

As a result, the study develops an approach to representing Arabic legal knowledge in the form of ontology and knowledge graph. It includes identifying legal entities, concepts, and document

templates, along with mapping of laws to legal elements. While this work demonstrates the feasibility of constructing a knowledge graph prototype from selected provisions of the Jordanian Labor Law, future work will focus on expanding the graph to encompass the entire legislation. The resulting large-scale legal knowledge graph would include all articles, legal entities, rights, obligations, violations, sanctions, and procedural rules, thereby enabling richer semantic retrieval and more sophisticated reasoning capabilities. Furthermore, automated knowledge graph construction and population techniques, together with tighter integration with large language models, will be investigated to support explainable legal question answering and other Arabic Legal AI applications.

DECLARATIONS

Funding. This research was supported by the King Abdullah I School of Graduate Studies and Scientific Research at Princess Sumaya University for Technology under grant number (21) 3 - 2025/2026.

Conflicts of interest. The authors declare that they have no competing interests.

Ethics approval. Ethics approval was not required for this study because it did not involve human participants, human biological materials, or animal subjects.

Informed consent. Not applicable.

Data availability. The data used and/or analyzed during the current study are available from the corresponding author upon reasonable request. Certain datasets and materials cannot be made publicly available due to licensing, confidentiality, and institutional restrictions.

Author contributions. Both authors were involved in the conceptualization, methodology, investigation, validation, formal analysis, and writing of the manuscript.

Use of artificial intelligence. Artificial intelligence tools were used solely to improve the language, grammar, readability, and formatting of the manuscript. All scientific content, methodology, analysis, interpretation of results, and conclusions were conceived, verified, and approved by the authors, who take full responsibility for the content of this manuscript.

Acknowledgements. This research was supported by the King Abdullah I School of Graduate Studies and Scientific Research at Princess Sumaya University for Technology under grant number (21) 3 - 2025/2026.

REFERENCES

- [1] Abdallah A, Piryani B, Jatowt A (2023) Exploring the state of the art in legal qa systems. *Journal of Big Data* 10(1):127. <https://doi.org/10.1186/s40537-023-00802-8>
- [2] Al-qurishi M, Alqaseemi S, Souissi R (2022) AraLegal-BERT: A pretrained language model for Arabic legal text. In: Aletras N, Chalkidis I, Barrett L, et al (eds) *Proceedings of the Natural Legal Language Processing Workshop* 2022. <https://doi.org/10.18653/v1/2022.nllp-1.31>
- [3] Ariai F, Mackenzie J, Demartini G (2025) Natural language processing for the legal domain: A survey of tasks, datasets, models, and challenges. *ACM Comput Surv* 58(6). <https://doi.org/10.1145/3777009>
- [4] Bevilacqua M, Blloshmi R, Navigli R (2021) One spring to rule them both: Symmetric amr semantic parsing and generation without a complex pipeline. *Proceedings of the AAAI Conference on Artificial Intelligence* 35(14):12564–12573. <https://doi.org/10.1609/aaai.v35i14.17489>
- [5] Borji A (2023) A categorical archive of chatgpt failures. <https://arxiv.org/abs/2302.03494>
- [6] Bouaziz Mezghanni I, Gargouri F (2015) Towards an arabic legal ontology based on documents properties extraction. In: *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, pp 1–8, <https://doi.org/10.1109/AICCSA.2015.7507219>
- [7] Breuker J, Hoekstra R, et al (2004) Epistemology and ontology in core ontologies: Folaw and Iri-core, two core ontologies for law. In: *Proceedings of EKAW Workshop on Core ontologies*. CEUR

- [8] Chalkidis I, Fergadiotis M, Malakasiotis P, et al (2020) LEGAL-BERT: The muppets straight out of law school. In: Cohn T, He Y, Liu Y (eds) Findings of the Association for Computational Linguistics: EMNLP 2020. <https://doi.org/10.18653/v1/2020.findings-emnlp.261>
- [9] Cui J, Ning M, Li Z, et al (2023) Chatlaw: A multi-agent collaborative legal assistant with knowledge graph enhanced mixture-of-experts large language model. arXiv preprint arXiv:230616092
- [10] Dehghani F, Dehghani R, Naderzadeh Ardebili Y, et al (2025) Large language models in legal systems: A survey. Humanities and Social Sciences Communications 12(1):1977
- [11] Eddine Meftah MC, Soussa A, Herzallah A (2025) An intelligent arabic legal assistant system (ialas) based on ontology. Transportation Research Procedia 84:440– 447. <https://doi.org/10.1016/j.trpro.2025.03.094>
- [12] Falcarin P, Chowdhury P, Carbone E, et al (2025) Legal requirements compliance using nlp and knowledge graphs. In: 2025 IEEE 33rd International Requirements Engineering Conference Workshops (REW), pp 412–418, <https://doi.org/10.1109/REW66121.2025.00060>
- [13] Fei Z, Shen X, Zhu D, et al (2024) LawBench: Benchmarking legal knowledge of large language models. In: Al-Onaizan Y, Bansal M, Chen YN (eds) Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Miami, Florida, USA, pp 7933–7962, <https://doi.org/10.18653/v1/2024.emnlp-main.452>
- [14] Forcier J, Bissex P, Chun WJ (2008) Python web development with Django. Addison-Wesley Professional
- [15] Ganter B, Wille R (1999) Formal Concept Analysis - Mathematical Foundations. Springer, <https://doi.org/10.1007/978-3-642-59830-2>
- [16] Gemini Team, Anil R, Borgeaud S, et al (2023) Gemini: A Family of Highly Capable Multimodal Models. arXiv e-prints arXiv:2312.11805. <https://doi.org/10.48550/arXiv.2312.11805>, arXiv:2312.11805 [cs.CL]
- [17] Hacene MR, Huchard M, Napoli A, et al (2013) Relational concept analysis: mining concept lattices from multi-relational data. Ann Math Artif Intell 67(1):81–108. <https://doi.org/10.1007/S10472-012-9329-3>
- [18] Han H, Wang Y, Shomer H, et al (2024) Retrieval-augmented generation with graphs (graphrag). arXiv preprint arXiv:250100309
- [19] Hijazi F, AlHarbi S, AlHussein A, et al (2024) ArabLegalEval: A multitask benchmark for assessing Arabic legal knowledge in large language models. In: Habash N, Bouamor H, Eskander R, et al (eds) Proceedings of the Second Arabic Natural Language Processing Conference. Association for Computational Linguistics, Bangkok, Thailand, pp 225–249, <https://doi.org/10.18653/v1/2024.arabnlp-1.20>
- [20] Hogan A, Blomqvist E, Cochez M, et al (2021) Knowledge graphs. ACM Comput Surv 54(4). <https://doi.org/10.1145/3447772>
- [21] Katz DM, Bommarito MJ, Blackman J (2023) Gpt takes the bar exam. arXiv preprint arXiv:230308774 <https://arxiv.org/abs/2303.08774>
- [22] Lai J, Gan W, Wu J, et al (2024) Large language models in law: A survey. AI Open 5:181–196. <https://doi.org/10.1016/j.aiopen.2024.09.002>
- [23] Le HH, Nguyen CT, Ngo TP, et al (2023) Intelligent retrieval system on legal information. In: Nguyen NT, Boonsang S, Fujita H, et al (eds) Intelligent Information and Database Systems. Springer Nature Singapore, Singapore, pp 97–108

- [24] Lewis P, Perez E, Piktus A, et al (2020) Retrieval-augmented generation for knowledge-intensive nlp tasks. In: Larochelle H, Ranzato M, Hadsell R, et al (eds) *Advances in Neural Information Processing Systems*, vol 33. Curran Associates, Inc., pp 9459–9474.
- [25] Mezghanni IB, Gargouri F (2017) Crimar: A criminal arabic ontology for a benchmark based evaluation. In: Zanni-Merk C, Frydman CS, Toro C, et al (eds) *Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 21st International Conference KES-2017, Marseille, France, 6-8 September 2017, Procedia Computer Science*, vol 112. Elsevier, pp 653–662, <https://doi.org/10.1016/J.PROCS.2017.08.113>
- [26] Mishra A, Vishwakarma S (2015) Analysis of tf-idf model and its variant for document retrieval. In: 2015 International Conference on Computational Intelligence and Communication Networks (CICN), pp 772–776, <https://doi.org/10.1109/CICN.2015.157>
- [27] Muninggar NS, Krisnadhi AA (2023) Lexid: The metadata and semantic knowledge graph construction of indonesian legal document. *Jurnal Ilmu Komputer dan Informasi* 16(1):15–46. <https://doi.org/10.21609/jiki.v16i1.1096>
- [28] Musen MA (2015) The protégé project: a look back and a look forward. *AI Matters* 1(4):4–12. <https://doi.org/10.1145/2757001.2757003>
- [29] Nguyen TH, Nguyen HD, Pham VT, et al (2022) Legal-onto: An ontology-based model for representing the knowledge of a legal document. *Proceedings of the 17th International Conference on Evaluation of Novel Approaches to Software Engineering*. <https://doi.org/10.5220/0011066300003176>
- [30] Noy NF, McGuinness DL, et al (2001) *Ontology development 101: A guide to creating your first ontology*
- [31] Obeid O, Zalmout N, Khalifa S, et al (2020) CAMEL tools: An open source python toolkit for Arabic natural language processing. In: Calzolari N, Béchet F, Blache P, et al (eds) *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pp 7022–7032, <https://aclanthology.org/2020.lrec-1.868/>
- [32] OpenAI (2023) GPT-4 technical report. CoRR abs/2303.08774. <https://doi.org/10.48550/arXiv.2303.08774>
- [33] Pan S, Luo L, Wang Y, et al (2024) Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering* 36(7):3580–3599. <https://doi.org/10.1109/TKDE.2024.3352100>
- [34] Paulheim H (2016) Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8(3):489–508. <https://journals.sagepub.com/doi/pdf/10.3233/SW-160218>
- [35] Pham VT, Dung DV, Ngo HQ, et al (2025) Ontology-based knowledge graph approach for legal queries. <https://doi.org/10.15388/25-INFOR617>
- [36] Quevedo E, Cerny T, Rodriguez A, et al (2024) Legal natural language processing from 2015 to 2022: A comprehensive systematic mapping study of advances and applications. *IEEE Access* 12:145286–145317. <https://doi.org/10.1109/access.2023.3333946>
- [37] Robertson SE, Zaragoza H (2009) The probabilistic relevance framework: BM25 and beyond. *Found Trends Inf Retr* 3(4):333–389. <https://doi.org/10.1561/15000000019>
- [38] Robinson I, Webber J, Eifrem E (2015) Graph databases: new opportunities for connected data. "O'Reilly Media, Inc."
- [39] Rodríguez-Doncel V, Montiel-Ponsoda E (2020) Lynx: Towards a legal knowledge graph for multilingual europe. *Law Context: A Socio-Legal J* 37:175
- [40] SAM, Easwararkumar KS (2025) Judgmentgraph: Benchmarking legal knowledge graph construction from supreme court rulings. In: 2025 3rd International Conference on Self

Sustainable Artificial Intelligence Systems (ICSSAS),
<https://doi.org/10.1109/ICSSAS66150.2025.11081258>

- [41] Silberztein M (2005) NooJ: a linguistic annotation system for corpus processing. In: Byron D, Venkataraman A, Zhang D (eds) Proceedings of HLT/EMNLP 2005 Interactive Demonstrations. Association for Computational Linguistics, Vancouver, British Columbia, Canada, pp 10–11, <https://aclanthology.org/H05-2006/>
- [42] Sovrano F, Palmirani M, Vitali F, et al (2020) Legal knowledge extraction for knowledge graph based question-answering. *Frontiers in Artificial Intelligence and Applications* 334:143–153
- [43] Wei H (2024) Intelligent legal document generation system and method based on knowledge graph. In: Proceedings of the 2024 International Conference on Machine Intelligence and Digital Applications. Association for Computing Machinery, New York, NY, USA, MIDA '24, p 350–354. URL <https://doi.org/10.1145/3662739.3669909>
- [44] Zhong H, Xiao C, Tu C, et al (2020) How does NLP benefit legal system: A summary of legal artificial intelligence. In: Jurafsky D, Chai J, Schluter N, et al (eds) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp 5218–5230. URL <https://aclanthology.org/2020.acl-main.466/>