Research on Knowledge Graph Construction Technology Based on Intellectual Property Legal Documents

Linrui Han
The Institute for Data Law, China University of Political Science and Law, Beijing, China

Abstract: Amidst the dual thrust of artificial intelligence development and the construction of legal informatization, knowledge graphs are increasingly applied in areas such as judicial judgment assistance, legal retrieval, and question answering. Addressing the heterogeneity of intellectual property rules and the complexity of case semantics, this paper aims to propose a construction method for a legal knowledge graph in the field of intellectual property. This method establishes an ontological framework covering 84 concepts, 89 relationships, and 105 attributes. It then employs the CasRel and BiLSTM-CRF models to extract entities and relationships from 3,713 intellectual property judicial cases, integrating these with the conceptual-level ontology using the TransE knowledge alignment model. Finally, a visualization platform is built on the Neo4j database to display and manage the knowledge graph, and a knowledge reasoning model is developed using the TransD algorithm to facilitate intelligent question answering. The knowledge graph constructed in this study provides a critical knowledge resource for legal research, case analysis, and decision support.

Keywords: Legal Knowledge Graph; Intellectual Property; Legal Documents; Knowledge Reasoning; Intelligent System

1. Overview

1.1 Research Background

In the context of a globalized economy, intellectual property has become a crucial indicator of a nation's competitiveness. Legal protection, serving as the cornerstone for safeguarding intellectual property, faces unprecedented challenges in case handling and law enforcement. Particularly, the differentiation in types of intellectual property, such as copyrights, trademarks, and patents, presents unique challenges in infringement determination and liability attribution, adding complexity to legal practice [1].

With the rapid development of information technology, knowledge graph technology offers a new solution for the legal protection field of intellectual property with its unique data organization form and powerful knowledge representation capability. Knowledge graphs effectively organize and manage the complex knowledge structures in legal documents, supporting sophisticated knowledge queries and reasoning, thereby enhancing the efficiency and accuracy of legal document processing.

However, despite significant achievements of knowledge graphs in other fields, their application in the realm of intellectual property legal documents is still in its infancy [2]. Building a knowledge graph specifically for intellectual property legal documents faces multiple challenges: on one hand, the graph needs to deal with the complex rules and concepts within the field of intellectual property; on the other hand, it must capture and express the rich semantic information of actions, events, and more within case contents. This study is dedicated to constructing a comprehensive knowledge graph, by defining entity types, relationship types, and attribute types, among others, to hierarchically and structurally represent the knowledge in the field of intellectual property [3]. Furthermore, this knowledge graph overcomes the aforementioned challenges by efficiently and accurately organizing and utilizing the knowledge within intellectual property legal documents, providing robust support for intellectual property protection, case analysis, and legal research. This research aims to offer a new tool for professionals in the field of intellectual property, to better understand and apply legal knowledge, further promoting the development and application of legal technology.

1.2 State of the Research

With the continuous advancement of artificial intelligence technology, the application of knowledge graphs in the judicial domain has showcased its unique potential and value. As a bridge connecting complex data points, knowledge graphs not only enhance the efficiency and accuracy in case analysis, legal consulting, and judgment prediction but also provide new perspectives and methodologies for judicial practice.

Recent studies have demonstrated that knowledge graphs can recommend similar cases to legal professionals by deeply analyzing the relationships and similarities between cases, thereby optimizing case handling processes and improving decision-making efficiency and precision [4]. Furthermore, the ability of knowledge graphs to infer legal rules and logical relationships supports legal interpretation and case analysis, enhancing the quality of legal decisions.

Legal provisions retrieval and analysis represent another important application of knowledge graphs. By linking legal articles and case information, they facilitate rapid location and interpretation of legal texts, greatly aiding legal professionals in finding and applying legal bases during case handling [5]. In the construction of evidence chains, knowledge graphs, through structured representation and relationship modeling, clearly display the logic and support relationships between pieces of evidence, providing powerful tools for evidence analysis[6].

Particularly in handling intellectual property cases, the application of knowledge graphs has extended to building legal intelligent question-answering systems, offering quick and accurate legal consulting services through natural language interaction, highlighting their importance in enhancing the efficiency of legal information acquisition and processing [7].

In summary, the application of knowledge graphs in the judicial domain not only enriches the toolbox for legal research and practice but also opens new avenues for improving the overall efficiency of the judicial system. The technology for constructing knowledge graphs based on intellectual property legal documents proposed in this paper aims to further enhance the application capabilities of knowledge graphs in legal document processing, case analysis, and intelligent question answering, with the hope of bringing more accurate and efficient solutions to the judicial field.

1.3 Innovations in Research

There are mainly two approaches to constructing knowledge graphs: top-down and bottom-up, with the top-down approach being suitable for specific industries that have a fixed knowledge system or can define fixed pattern data [8]. Given that legal documents have a standardized format and components, this study opts for a top-down approach for construction. The construction process, as shown in Figure 1, consists of four parts: data collection and preprocessing, entity and relationship extraction, knowledge alignment, and knowledge graph expansion.

The innovations of the knowledge graph constructed in this study are twofold. Firstly, in terms of the design of the conceptual layer of the knowledge graph, this study conducts an in-depth analysis of the diversity and heterogeneity of intellectual property protection. Compared with existing research, this study not only focuses on technological and model improvements but also performs meticulous design at the legal level, constructing specific ontologies for different types of rights, and revealing the core elements and characteristics of intellectual property cases. Additionally, this study proposes an innovative approach to case ontology construction, distinguishing whether the conclusion of intellectual property infringement in cases is established or not, thereby structurally showcasing the dual nature of cases, significantly enhancing the accuracy and coverage of the knowledge graph.

Secondly, in terms of entity and relationship extraction and knowledge reasoning, a variety of advanced machine learning algorithms are employed, including but not limited to Word2Vec, BERT, CasRel, BiLSTM-CRF, TransD, TransE, etc. This semi-supervised extraction method, which combines multiple algorithms, ensures text extraction efficiency while maintaining the precision of the extraction results, guaranteeing usability and reliability when processing large-scale data.

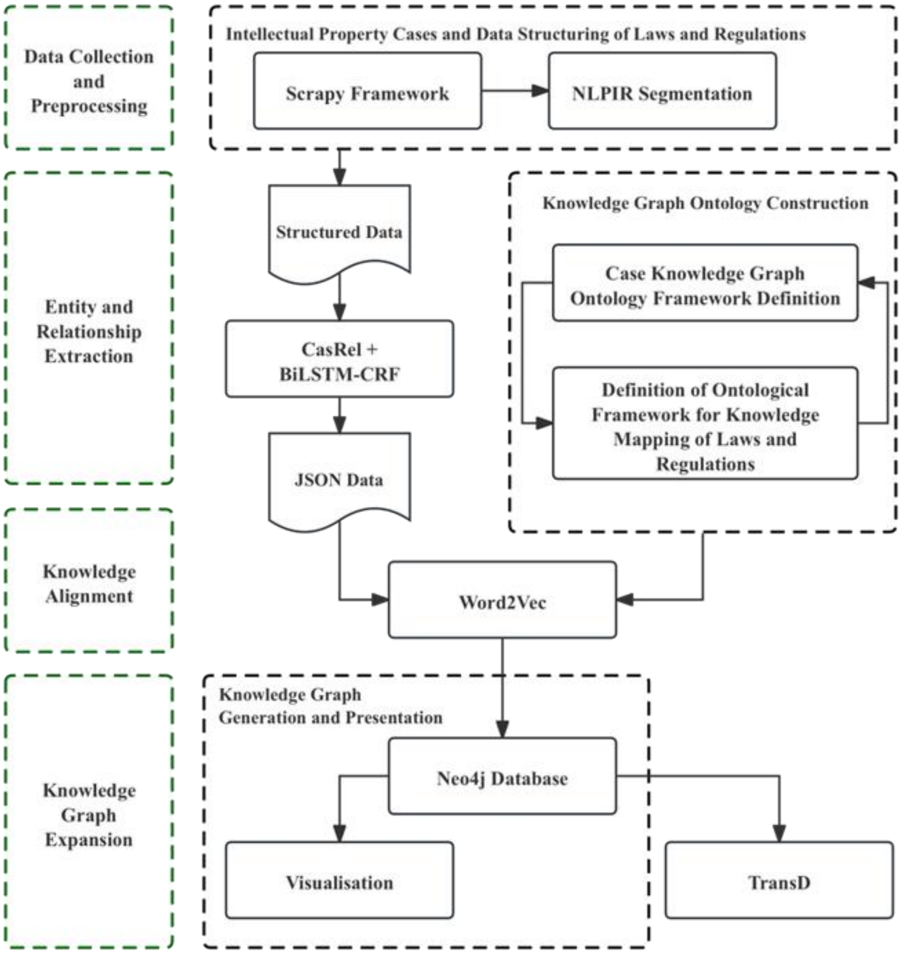

Figure 1. Framework for Constructing a Knowledge Graph Based on Intellectual Property Legal Documents

2. Data Collection and Preprocessing

2.1 Data Collection

In the data collection phase for constructing the intellectual property protection knowledge graph, the Scrapy framework is used to develop web crawlers to gather data related to intellectual property laws and cases. Scrapy, a powerful Python open-source framework, is specialized in efficiently executing web scraping and data extraction [9]. The dataset covers a variety of information related to intellectual property laws, including fields such as the name of the legal regulation, issuing body, publication date, effective date, category, and access link, ensuring the completeness and accuracy of legal regulation information.

2.2 Data Preprocessing

Based on the collected data, the NLPIR toolkit is configured to read the data on intellectual property laws and cases gathered, while removing specific punctuation marks, identifying and preserving proprietary terms specific to the intellectual property field. The NLPIR toolkit is a natural language processing tool [10], which can be used for tasks such as text segmentation, part-of-speech tagging, keyword extraction, and proprietary term identification. The data preprocessing process is shown in Figure 2.

The NLPIR-provided segmentation function is utilized for batch segmentation processing on numerous documents pertaining to intellectual property laws and cases. Based on the return value of the segmentation function, segmentation results can be obtained, namely a list of segmented words or tuples corresponding to words and their parts of speech. Subsequently, the results are saved into variables for further processing and analysis in subsequent steps.

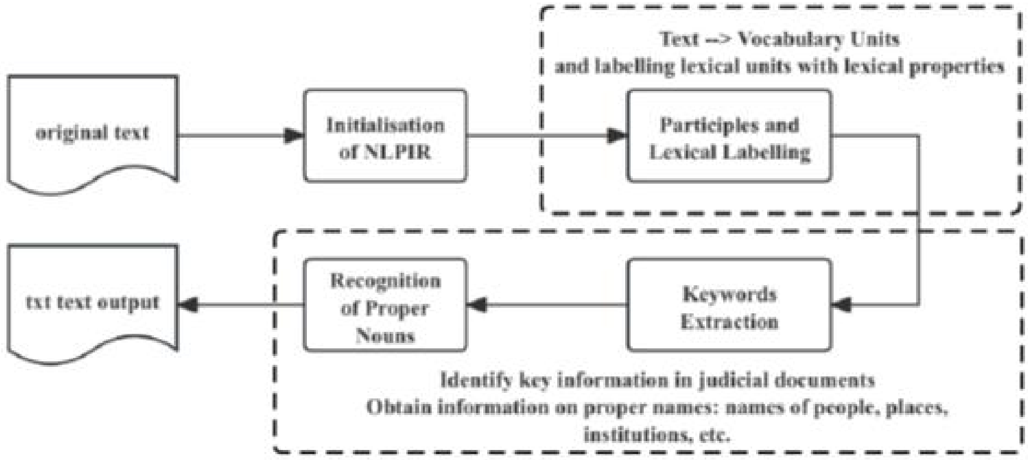

Figure 2. NLPIR Data Preprocessing Flowchart

The segmentation results reveal key nouns, verbs, and other vocabulary in the cases, and can accurately divide the structure of sentences, helping research capture the core elements of the cases. For example, in an intellectual property dispute case, data preprocessing can yield relevant nouns such as “patent,” “infringement,” “litigation,” etc. These terms accurately reflect the basic situation of the case. Furthermore, the recognition function of the NLPIR toolkit can also extract proprietary nouns in cases, such as company names and product names, which are significant for understanding the background and involved parties of the cases. Through the segmentation results, key information about the enterprises, products, technologies, etc., involved in the cases can be quickly located.

3. Knowledge Graph Ontology Construction

3.1 Construction Approach for Ontology

The construction of knowledge graphs in vertical domains, especially in the field of intellectual property law, faces unique challenges. A significant challenge during the ontology construction process is how to efficiently and accurately reflect the heterogeneity and complexity of intellectual property protection. Although some studies attempt to construct ontologies through semi-automated methods with data support, expert manual intervention is ultimately required to ensure the accuracy and practicality of the ontology [11]. Considering that the intellectual property law knowledge graph is still in its early stages of development, and the requirement for precision in conceptual terminology is extremely high [12], this study opts for manual construction of the ontology by experts. This includes high-level abstraction of concepts, knowledge points, and rules within the field of intellectual property, followed by the definition of additional relationships and attributes based on this foundation.

The purpose of ontology construction is to provide precise and effective knowledge support for professionals within the industry, such as lawyers, judges, and legal personnel. By observing and analyzing the fixed format and standardized terminology of legal judgments, this study finds that they contain rich information about case details and the thought processes of judicial personnel, all of which are important references for ontology construction. Therefore, when constructing the knowledge graph ontology, this study pays special attention to the visualization and structured representation of entity relationships in the judgment documents, to facilitate a quick understanding of personal relationships, case information analysis, and mastery of the review process.

Moreover, during the construction process, this study also refers to mature knowledge graph application cases in other fields such as finance and medicine, as well as knowledge sources in the field of law, such as legal textbooks and trial guides, providing important guidance for the construction of the ontology. Especially in the application of the intellectual property law field, to ensure the accuracy of ontology concepts, this study also follows relevant technical standards, such as the “Basic Vocabulary of Intellectual Property Literature and Information (GB/T 21374)" and the "Standards for Information Extraction of Intellectual Property Infringement Judgment Documents (FYB/T 51018-2020)".

3.2 Ontology Structure and Definitions

This study, integrating existing ontologies in the field of intellectual property cases with the basic structure of judicial judgment documents, and guided by experts in the judicial field from universities and enterprises, has constructed an ontology for knowledge graphs based on intellectual property infringement judgment documents as shown in Figure 3

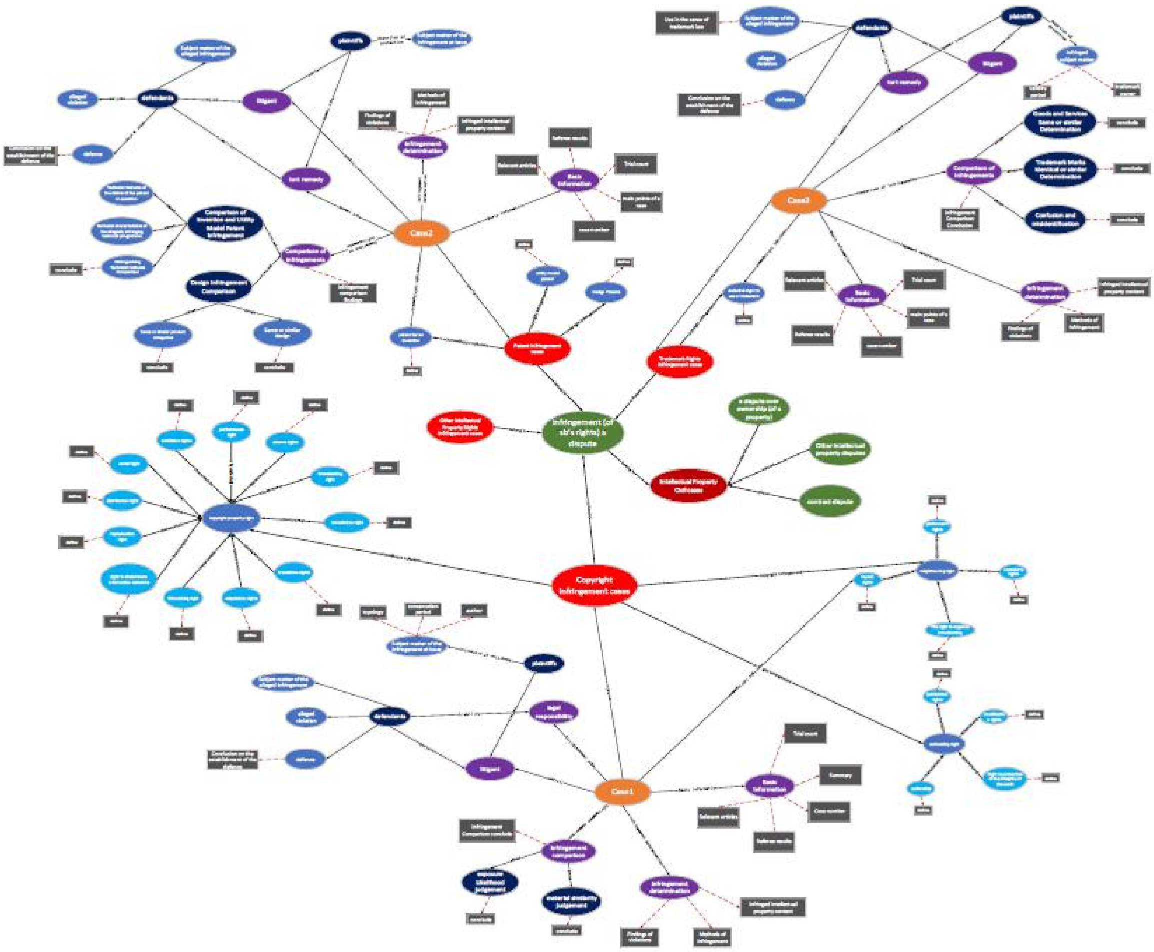

Figure 3. Ontology of Intellectual Property Infringement Cases

Given the variety of case types in the field of intellectual property civil cases, including infringement disputes, ownership disputes, contract disputes, etc., with infringement disputes being notably frequent and complex, this study limits the construction of the knowledge graph structure to cases classified under "intellectual property ownership and infringement disputes" as defined by the "Civil Case Cause Regulations," excluding ownership disputes [13]. Additionally, intellectual property can be subdivided into copyright, trademark rights, patent rights, and other types of rights, each with significant differences in legal rules and judicial approaches for infringement determination and liability recognition. Therefore, this study focuses on constructing separate ontologies for copyrights, patent rights, and trademark rights, initially assigning cases to the corresponding domain based on the specific rights involved in the case. The adjudication procedures for cases also vary, including first instance, second instance, retrial, and review, and the appellations for litigants differ across these procedures. Thus, for the ease of analysis and construction of the knowledge graph, the terms "plaintiff" and "defendant" in the ontology refer to those in the first-instance judgment, although the construction of the knowledge graph is not limited to first-instance civil judgment documents.

3.2.1 Definition of concepts

Concepts, also known as classes, are the foundation and core of ontology modeling. To form a concept system conducive to sharing and reuse, the definition of concepts must follow principles of objectivity and accuracy. For intellectual property infringement cases, the industry standard “Intellectual Property Infringement Case Decision Information Extraction Specification (FYB/T 51018-2020)” has already decomposed the main content of documents into eight major categories and primary fields. Additionally, to ensure usability, this study incorporates authoritative standards and guidelines applicable in practice, such as the “Trademark Infringement Judgment Standards” interpretation and application by the National Intellectual Property Administration and the “Copyright Infringement Case Trial Guide” by the Beijing Higher People’s Court, extracting 84 core fields as concepts from several judgments [14]. Given the separate construction of ontologies for different intellectual property types, the ontology structure can be divided into four parts: types of intellectual property civil cases, copyright infringement cases, patent infringement cases, and trademark infringement cases.

3.2.2 Definition of relationships

Relationships define the connections between concepts, which can be either within the same category or across different categories, with both the domain and range being concepts. Since cases are generally independent of each other, this study defines certain relationships between entities within specific cases, clarifying the hierarchical relationships between different case types. A total of 21 groups of relationships have been extracted.

3.2.3 Definition of attributes

Attributes complement and refine the information about an entity itself, further describing the characteristics of a concept from the attribute dimension and enriching the concept’s connotation. Based on the characteristics of intellectual property infringement cases, 27 attributes are defined. While the overall attributes related to copyright, patent, and trademark infringement cases are generally the same, there are exceptions that require separate definitions, such as attributes related to the “disputed infringing subject matter” differing significantly among copyright, patent, and trademark infringement cases.

The knowledge graph constructed in this study, reflecting the substantive content and reasoning of intellectual property infringement, covers 84 concepts, 89 relationships, and 105 attributes, and can be accessed at: https://gitcode.com/H66778899/KnowledgeGraph.

4. Knowledge Extraction

4.1 Entity Extraction

Entity extraction, a key task in the field of natural language processing, plays a crucial role in the construction of knowledge graphs. Legal regulations and documents in the intellectual property domain contain a vast amount of information on entities such as patents, trademarks, and copyrights, making the steps involved in entity extraction particularly complex. In this study, the combined use of Bidirectional Encoder Representations from Transformers (BERT) and the Conditional Random Field (CRF) model offers robust support for entity extraction.

BERT, as a pre-trained deep bidirectional representation model, has acquired rich linguistic features through pre-training on a large corpus of text data [15]. In the task of entity extraction, BERT can understand and capture the complex semantic relationships and context within texts, which is particularly important for accurately identifying proprietary nouns and terms in legal texts. When employing BERT for entity recognition in this study, legal texts are first input into the BERT model, which then uses its powerful semantic analysis capability to predict the label for each word in the text, such as B-Entity (beginning of entity), I-Entity (inside entity), or O (non-entity) [16]. This deep learning-based method significantly improves the accuracy and efficiency of entity extraction compared to traditional models.

Despite the excellent performance of the BERT model in entity recognition, the dependency relationships between words still need to be considered in sequence labeling tasks. Therefore, this study further incorporates the CRF model on top of BERT to optimize the identification of entity boundaries and the coherence within entities. CRF is a statistical model designed for labeling sequence data, capable of considering the entire sentence's label sequence through global optimization, thereby enhancing labeling consistency and accuracy [17]. With CRF making the final sequence decision based on features extracted by BERT, this approach can more accurately define the start and end of entities, showing clear advantages, especially when dealing with legal texts of complex structures.

The combined use of BERT and CRF models in this study achieves high-precision performance in entity extraction tasks, providing a high-quality data foundation for the construction of knowledge graphs.

4.2 Relation Extraction

A key component of knowledge graphs is relational facts, most of which consist of two entities connected by a semantic relation. These facts are formed as (subject, relation, object) or (s, r, o), referred to as relational triples (or triples for short). Extracting relational triples from natural language texts is a critical step in constructing large-scale knowledge graphs. The general approach to relation extraction is $f(s,o) \rightarrow r$, which primarily involves identifying entities within the text, distinguishing them as subjects and objects, and then finding their corresponding relations [18]. However, in the domain of intellectual property knowledge graphs, relations should not be merely treated as discrete labels for pairs of entities. Instead, relations should be modeled, making them a function that maps the main entity to the target entity, to ensure the usability, extensibility, and rationality of relations. Therefore, this study posits that the method of relation extraction should be $f_r(s) \rightarrow o$. Under this decision-making pattern, this study selects the CasRel algorithm as the method for relation extraction in this knowledge graph. CasRel is an end-to-end cascading binary tagging framework. It consists of a BERT-based encoder module, a subject tagging module, and a relation-specific object tagging module [19]. Under the CasRel framework, triple extraction is a two-step process: first, identifying all possible subjects in a sentence; then, for each subject, applying a relation-specific tagger to simultaneously identify all possible relations and corresponding objects. This method's approach embodies the idea of.

CasRel is fundamentally a joint entity-relation extraction method based on parameter sharing, often referred to as a cascading pointer network. In fact, the core idea of CasRel, or the focus of the authors' improvement on existing models, lies in the design of the sublayers [20]. Because CasRel divides the task of relation extraction differently, the subtasks and the order of solving these subtasks also differ. Specifically: CasRel first identifies all possible subjects (head entities); then, given a category relationship, it identifies the objects (tail entities) related to the subject.

4.3 Knowledge Alignment

Knowledge alignment is a key step in the process of constructing a knowledge graph, especially when ontology construction and entity-relationship extraction are carried out asynchronously, and when entity-relationship extraction is done independently. As a result, the final output often contains different entities or relationships that express the same situation, manifesting as diverse expressions of knowledge. To address this issue, knowledge alignment needs to be performed before finalizing the construction of the knowledge graph.

Knowledge alignment typically employs Aggregation-based fusion methods and Translation-based fusion methods [21]. In this study, the TransE algorithm is used to perform Translation-based knowledge fusion operations. TransE maps triples (h, r, t) into the same space, aiming to achieve the goal where the head entity (h) points to the tail entity (t) through the relation (r).

Before mapping the triples, the Word2Vec algorithm is used for vectorizing the text. Word2Vec aims to embed words into a continuous vector space to capture semantic relationships between words, offering two main training algorithms: Continuous Bag of Words (CBOW) and Skip-gram [22]. Through these steps, the data processed for word vectorization is imported into the TransE model for knowledge alignment.

After importing the word vector data processed by Word2Vec into the TransE model, the knowledge alignment algorithm process based on the TransE model is shown in Figure 4. The process starts with embedding the knowledge graph through knowledge representation learning techniques, i.e., the embedding module; then, it maps the embedding spaces of different knowledge graphs into the same vector space, i.e., the interaction module; and finally, the entity alignment results are obtained based on the distance or similarity between entities in the vector space, i.e., the alignment module [23]. This process also introduces an iterative mechanism to continuously feedback the alignment results to optimize and improve the accuracy of the alignment.

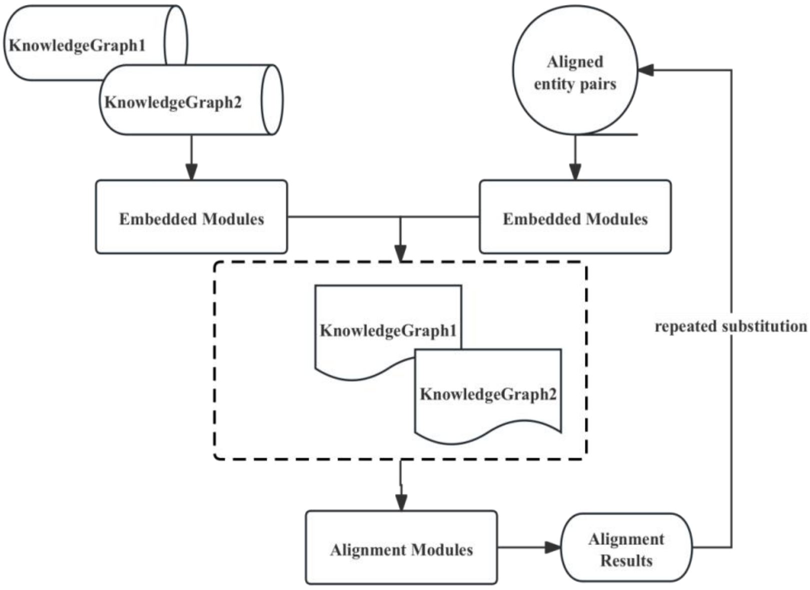

Figure 4. Knowledge Alignment Process Flowchart

5. Knowledge Graph Generation and Visualization

5.1 Knowledge Graph Generation

In this study, the automated construction of the knowledge graph is implemented through a combination of the Py2neo library and the Neo4j graph database. Py2neo is a powerful Python library specifically designed for creating, querying, and managing graph data in the Neo4j database. The use of this library makes the integration of processed knowledge datasets into the Neo4j database straightforward and efficient.

The construction process first involves structuring the data that has been preprocessed and aligned in knowledge, stored in JSON format. Then, using the interface provided by Py2neo, these JSON-formatted data are imported into the Neo4j graph database, creating a specific instance of the knowledge graph. In this process, entities are converted into nodes in the graph, while the relationships between entities are converted into edges connecting these nodes.

The choice of Neo4j database is based on its efficient graph data processing capabilities and excellent query performance. It can conveniently represent and store complex networks of entity relationships [24], providing strong support for deep querying and analysis of the knowledge graph. Furthermore, the flexibility and scalability of Neo4j ensure that the knowledge graph constructed in this study can accommodate a large volume of data and support the continuous updating and expansion of the knowledge graph.

After completing data import, complex query operations are executed using Neo4j's graph query language (Cypher) to verify the accuracy and completeness of the knowledge graph [25]. This step not only verifies the correctness of the data but also demonstrates the potential of the knowledge graph in providing legal decision support.

5.2 Visualization Display

To enhance users' understanding and interactive experience with the knowledge graph, this study utilized the advanced visualization tools of the Neo4j database to implement an intuitive user interface. This interface not only displays the entities in intellectual property cases and their relationships but also provides rich interactive features, allowing users to easily navigate and explore the data in depth.

In this visualization display, entities such as cases, legal provisions, and participants appear as graphical nodes, with their relationships represented by lines connecting these nodes. To improve the readability of the visualization, different types of nodes and relationships are assigned different colors and shapes, enabling users to quickly identify and distinguish various types of information.

Moreover, by implementing filtering and highlighting functionalities, users can highlight relevant nodes and connections based on specific conditions or keywords, effectively filtering out information of interest. This interactive exploration mechanism greatly enhances the usability and user experience of the knowledge graph, enabling legal professionals and researchers to deeply analyze the complex connections in intellectual property cases through an intuitive graphical interface as shown in Figure 5.

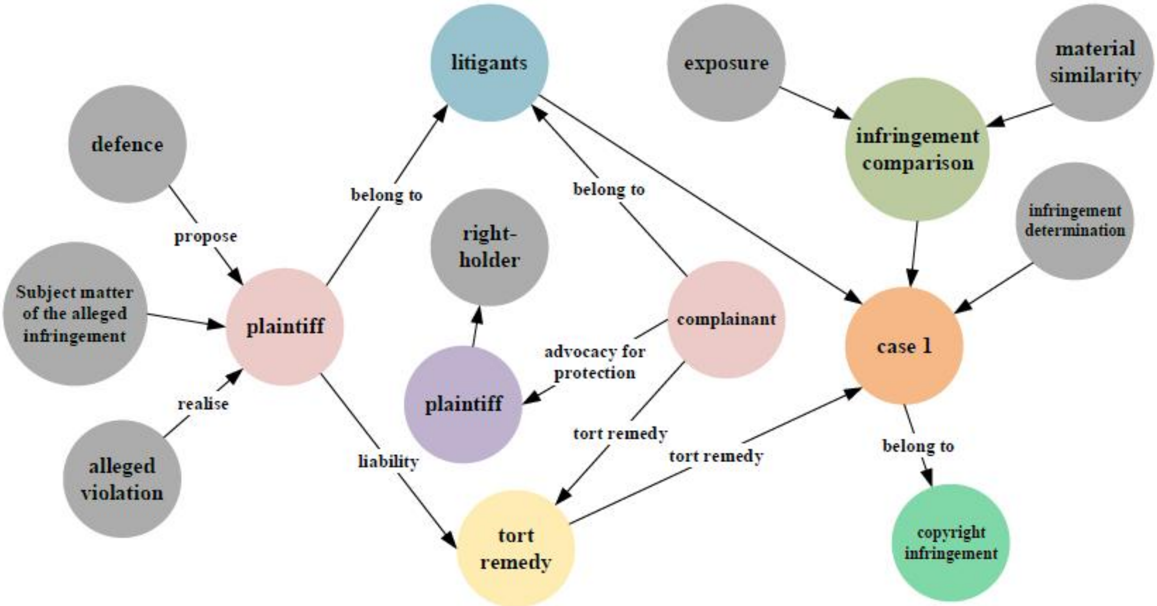

Figure 5. Example of Visualization Nodes for Individual Cases

6. Knowledge Graph Reasoning

Knowledge graph reasoning involves inferring new knowledge or identifying errors in the existing knowledge on a knowledge graph based on the facts it contains. This introduces two downstream tasks: knowledge graph completion and knowledge grapBased on the constructed ontology of the knowledge graph, this study finds that in the knowledge graph of the intellectual property domain, the corresponding relationships between entities are quite complex. There are 1 - 1 relationships, such as different case numbers corresponding to their individual parties; 1 - N relationships, such as copyright corresponding to various rights like attribution right, modification right, etc.; and N - N relationships, such as the mutual correspondence between infringement responsibility and infringement comparison with plaintiffs and defendants. Therefore, when constructing a knowledge reasoning model, the complexity of the intellectual property domain's graph itself must be considered. Currently, knowledge representation learning methods can be divided into two types: structure-based and semantics-based methods. Structure-based embedding representation methods include TransE, TransH, TransR&CTransR, TransD, etc., which learn the representation of entities and relations in the KG from the structure of triples; semantics-based embedding representation methods include NTN, SSP, DKRL, etc., which learn the representation of entities and relations in the knowledge graph from the perspective of textual semantics. After comprehensive analysis, this study chose the TransD model to complete the knowledge graph reasoning task.

The TransD model, proposed by Zhao Jun and Liu Kang from the NLPR Institute of Automation, Chinese Academy of Sciences, is an improvement over the TransE and TransR models [26]. Previous Trans series models established a unidirectional mapping between the entity vector space and the relationship vector space. Taking the TransE model as an example, in its implementation, a positive triple (h,r,t) is first selected, and then a negative triple (h',r',t') is sampled from Δ'. The scores $f_r(h,t)$ and $f_{r'}(h',t')$ for the positive and negative examples are calculated respectively. If $f_r(h,t)+\gamma-f_{r'}(h',t')>0$, the parameters h,r,t,h',r',t' are updated through gradient descent. Clearly, the TransE model encounters difficulties in handling complex relationship modeling (one-to-many, many-to-one, many-to-many relationships). This is because, for different relationships r, the representation of entity vectors is always the same.

In the TransD model, each named symbolic object (entity and relation) is represented by two vectors. The first vector captures the meaning of the entity (relation), and the second is used to construct the mapping matrix. For example, given a triple (h,r,t), its vectors are h,h_p,r,r_p,t,t_p, where the subscript p denotes the projection vector, $h,h_p,t,t_p \in R_n$ and $r,r_p \in R_m$. For each triple (h,r,t), this study sets two mapping matrices $M_{rh},M_{rt} \in R_{n*m}$ to project entities from the entity space into the relation space.

Therefore, the mapping matrices are determined jointly by the entity and the relation, allowing the two projection vectors to interact fully, as each element can satisfy entries from the other vector. The mapping relationship is illustrated in Figure 6.

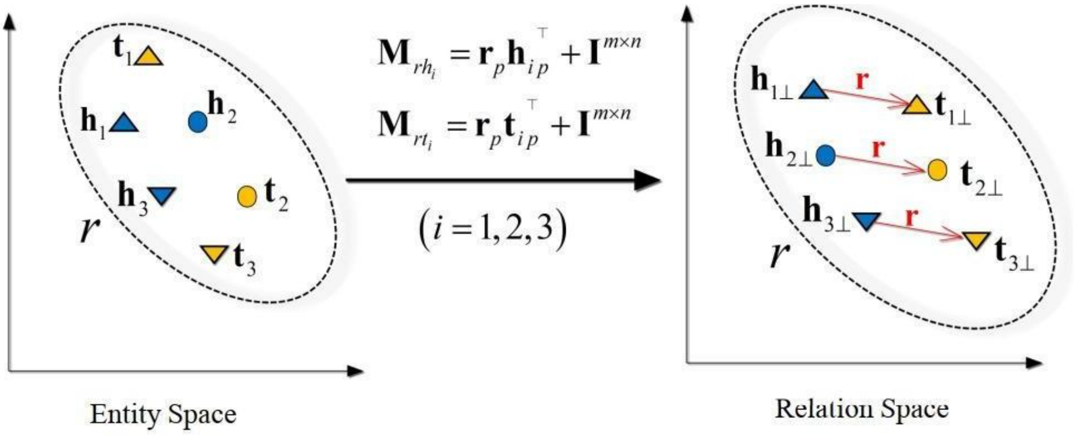

Figure 6. Mapping Relationship in the TransD Model

By constructing the mutual mapping relationship between the entity vector space and the relation vector space through TransD, this study enables the knowledge graph to perform knowledge completion for each triple. Thus, when facing retrieval requirements, it can intelligently search for answers to the questions posed by users, yielding appropriate results.

7. Conclusion

Through the efforts of this study, the complete process of knowledge graph construction, from data acquisition to usage, was accomplished. A total of 166 intellectual property domain documents were extracted, 3713 cases of original judicial data were obtained, cleaned, and structured, and a knowledge graph reflecting the substantive content of intellectual property infringement and the rationale for judgment was constructed, covering 84 concepts, 89 relationships, and 105 properties at the conceptual layer. Additionally, the extraction of entities and relationships was completed, and knowledge alignment was performed to eliminate the diversity of word expressions and remove deviant data, laying the foundation for the construction of the entity layer of the knowledge graph.

Despite the initial achievements in constructing a knowledge graph for intellectual property case studies, this research has room for improvement. There are further exploration and perfection needs in case type, data identification and extraction accuracy, application of natural language processing technology, expansion of legal tech application features, and deepening of algorithm research. Future research could focus on constructing more comprehensive and richer hierarchical knowledge graphs, improving model performance to enhance accuracy and application scope, exploring more legal tech application functionalities, and improving the algorithms for automatic alignment, updating, and completion of the knowledge graph expansion part, to further promote the application and development of knowledge graph technology in the field of intellectual property protection.

References

[1] Wang Guozhu. The Special Jurisprudence of Intellectual Property Infringement Liability. Legal Science (Journal of Northwest University of Political Science and Law), 2022, 40(04):114-125.

[2] Liu Huabing. Exploration of the Application Patterns of Knowledge Graphs in Big Data of Intellectual Property. Science Technology and Publishing, 2018(12):95-98.

[3] Xia Xiangyang, Xiao Wen. Knowledge Map Analysis of International Patent Information Research in the Field of Social Sciences. Journal of Zhejiang University (Humanities and Social Sciences Edition), 2019, 49(6): 64-77.

[4] Gao Xiang. Construction of Legal Knowledge Graph for Civil Judicial Application of Artificial Intelligence—Based on the Theory of Element-Fact Type Civil Judgments. Law and Social Development, 2018, 24(06):66-80.

[5] Li Zonghui. Theoretical Analysis and Practical Consideration of Artificial Intelligence-Assisted Legislation. Science and Society, 2022, 12(04):84-105.

[6] Wang Di. Data Modeling of the Evidence Standard System — Based on the Innovation and Application of Legal Knowledge Graphs. People's Procuratorate, 2020(23):33-36.

[7] Wang Chunlei, Wang Xiao, Liu Kai. A Review of Multimodal Knowledge Graph Representation Learning. Journal of Computer Applications, [Online] 1-19.

[8] Wang Haofen, Qi Guilin, Chen Huajun. Knowledge Graph: Methods, Practice and Application. Nanjing: Electronics Industry Publishing House, 2019: 425.

[9] Liu Duolin, Lü Miao. Simulation of Distributed Web Crawler Data Collection Algorithm under Scrapy Framework. Computer Simulation, 2023, 40(06):504-508.

[10] Zhou L, Zhang D. NLPIR: A Theoretical Framework for Applying Natural Language Processing to Information Retrieval. Journal of the American Society for Information Science and Technology, 2003, 54(2): 115-123.

[11] Wang Ning, Liu Wei, Lan Jian. Construction and Completion of Legal Knowledge Graphs Based on Court Judgment Documents. Journal of Zhengzhou University (Science Edition), 2021, 53(03):23-29.

[12] Yang Zexia: "A Review of Legal Knowledge Graph Construction Technology", published on the WeChat public account "Tsinghua University Institute for Intelligent Legal Systems", August 1, 2022.

[13] Lü Feiyi, Liu Jianxin. Determination of Case Cause and Jurisdiction in Intellectual Property Civil Cases. Intellectual Property, 2011(09):41-44.

[14] "Data Analysis Study on Common Intellectual Property Infringement Cases Based on Public Documents (September 2006 - September 2020)". China Applied Jurisprudence, 2021(06):158-188.

[15] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805, 2018.

[16] Wu Jun, Cheng Yao, Hao Han, et al. Research on Chinese Professional Term Extraction Based on BERT Embedding BiLSTM-CRF Model. Journal of Information Science, 2020, 39(04):409-418.

[17] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF Models for Sequence Tagging. arXiv preprint arXiv:1508.01991, 2015.

[18] Lample G, Ballesteros M, Subramanian S, et al. Neural Architectures for Named Entity Recognition. arXiv preprint arXiv:1603.01360, 2016.

[19] Sun Z, Hu W, Li C. Cross-lingual Entity Alignment via Joint Attribute-Preserving Embedding. In: The Semantic Web–ISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 21–25, 2017, Proceedings, Part I. Springer

International Publishing, 2017: 628-644.
[20] Wei Z, Su J, Wang Y, et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction. arXiv preprint arXiv:1909.03227, 2019.
[21] Cai L, Wang W Y. KBGAN: Adversarial Learning for Knowledge Graph Embeddings. arXiv preprint arXiv:1711.04071, 2017.
[22] Ling W, Dyer C, Black A W, et al. Two/too Simple Adaptations of Word2Vec for Syntax Problems. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015: 1299-1304.
[23] Saeedizade M J, Torabian N, Minaei-Bidgoli B. KGRefiner: Knowledge Graph Refinement for Improving Accuracy of Translational Link Prediction Methods. arXiv preprint arXiv:2106.14233, 2021.
[24] Miller J J. Graph Database Applications and Concepts with Neo4j. In: Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA. 2013, 2324(36): 141-147.
[25] Hölsch J, Grossniklaus M. An Algebra and Equivalences to Transform Graph Patterns in Neo4j. In: EDBT/ICDT 2016 Workshops: EDBT Workshop on Querying Graph Structured Data (GraphQ). 2016.
[26] Fusar - Poli P. TRANSD Recommendations: Improving Transdiagnostic Research in Psychiatry. World Psychiatry, 2019, 18(3): 361.
