Towards Building Legal Knowledge Maps with Natural Language Processing Toolkits

Hung Q. Ngo*
hung.ngo@tudublin.ie
Technological University Dublin
School of Management
Dublin, Ireland

Hien D. Nguyen*
hiennd@uit.edu.vn
University of Information Technology
Ho Chi Minh City, Vietnam
Vietnam National Univeristy
Ho Chi Minh City, Vietnam

Nhien-An Le-Khac
an.lekhac@ucd.ie
University College Dublin
School of Computer Science
Dublin, Ireland

ABSTRACT

Today, the legal document system is increasingly strict and affects activities in many different fields. The construction of knowledge maps that involve one or a group of legal documents is an effective approach to represent actual knowledge domains. A legal knowledge graph constructed from laws and legal documents can enable a number of applications, such as question answering, document similarity, and search. In this paper, we describe the process of building a system of knowledge maps for the Vietnamese legal system from the source of more than 300,000 legal documents that span all fields of social life. The study also proposes an integrated ontology to represent the knowledge of a proposed legal text. This model integrates the ontology of relational knowledge and the graph of key phrases in the form of a concept graph. It can express the semantics of the content of the legal document. In addition, the study also describes the process of building and exploiting natural language processing tools to build a VLegalKMaps system based on the natural language processing toolkit (SpaCy). Based on this integrated model, the method of improving the self-attention network by linguistic-oriented semantic analysis is studied to recover wisdom on legal documents. Finally, the study proposes to apply it to build an intelligent support system for querying knowledge about Vietnam's Land Law. It can help users query some meanings of terms in land law and some administrative procedures related to land. We also highlight open challenges in the realization of a knowledge graph in a technical legal system that enables this approach at scale.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

LegalTech, Knowledge Map, Legal Ontology, Knowledge Representation, Legal Intelligent System

SoICT 2022, December 2-3, 2022, HaLong, Vietnam

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Hung Q. Ngo, Hien D. Nguyen, and Nhien-An Le-Khac. 2022. Towards Building Legal Knowledge Maps with Natural Language Processing Toolkits. In *Proceedings of the 11th International Symposium on Information and Communication Technology (SoICT 2022).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Knowledge graphs (KG) represent real-world facts as structured data. A knowledge graph includes a set of triples (subject, relation, object), where the subject and object entities represent as nodes, and their relations are labeled edges in the graph.

The CELLAR repository was introduced to provide semantic indexing, advanced search, and data retrieval for multilingual resources of the Publications Office of the European Union [22]. This system is based on the Common Metadata Model (CDM) as a core ontology-based approach, inherited from the Functional Requirements for Bibliographic Records (FRBR) model and described by RDF(S)/OWL technologies. It contains 24 official languages spoken in the 28 member states of the European Union. In general, this system is a multilingual document management system based on the ontology-based approach and RDF(S)/OWL technologies to retrieve an RDF/XML representation of work-level metadata. Another project from European is Lynx, which built a legal knowledge graph (LKG) for smart compliance services to assist companies in Europe with compliance needs [9, 10]. Lynx approach contains three main phases: data acquisition, data integration, and data exploitation to integrate and link heterogeneous compliance data sources including legislation, case law, regulations, standards, and private contracts [9].

The authors in [7] built interlinking Legal Data to handle e European Law Identifier (ELI) and the European Case Law Identifier (ECLI) as web identifiers and vocabularies that can be used to describe metadata related to legal documents. Moreover, we argue that the ELI and ECLI standards themselves and their associated metadata-guidelines serve as an excellent basis for a trans-national European legal knowledge graph (KG). Possible use case applications of such a knowledge graph are widespread ([4]), including: (1) supporting comparative analyzes of court decisions and different legal interpretations of legislations; (2) enabling the analyzes of the evolution of legislation and jurisdiction; (3) interlinking legal knowledge with other data (such as online discussions, news, etc.).

The authors in [7] built interlinking Legal Data to handle e European Law Identifier (ELI) and the European Case Law Identifier

(ECLI) as web identifiers and vocabularies that can be used to describe metadata related to legal documents. Moreover, we argue that the ELI and ECLI standards themselves and their associated metadata-guidelines serve as an excellent basis for a trans-national European legal knowledge graph (KG). Possible use case applications of such a knowledge graph are widespread ([4]), including: (1) supporting comparative analyzes of court decisions and different legal interpretations of legislations; (2) enabling the analyzes of the evolution of legislation and jurisdiction; (3) interlinking legal knowledge with other data (such as online discussions, news, etc.). The authors in [7] built interlinking Legal Data to handle e European Law Identifier (ELI) and the European Case Law Identifier (ECLI) as web identifiers and vocabularies that can be used to describe metadata related to legal documents. Moreover, we argue that the ELI and ECLI standards themselves and their associated metadata-guidelines serve as an excellent basis for a trans-national European legal knowledge graph (KG). Possible use case applications of such a knowledge graph are widespread ([4]), including: (1) supporting comparative analyzes of court decisions and different legal interpretations of legislations; (2) enabling the analyzes of the evolution of legislation and jurisdiction; (3) interlinking legal knowledge with other data (such as online discussions, news, etc.).

[18] describe Knowledge of the Law in the Big Data Age. [3] used BERT and SpaCy for information extraction tasks on tourism domain.

[16] proposed Legal-Onto ontology for representing knowledge of a legal document.

[24] proposed knowledge graph construction approach for the legal domain.

[5] built Knowledge graph-based legal search for german court cases.

[21] builds Legal knowledge extraction for knowledge graph based question-answering.

[19] The Web Is Your Oyster-Knowledge-Intensive NLP against a Very Large Web Corpus.

[6] Similar Cases Recommendation using Legal Knowledge Graphs.

[23] proposed IFlyLegal, A Chinese Legal System for Consultation, Law Searching, and Document Analysis.

[20] proposed Lynx for A knowledge-based AI service platform for content processing, enrichment and analysis for the legal domain.

Thus, a support system for intellectual retrieval on law knowledge is very necessary for people.

2 VLEGALKMAPS - LEGAL KNOWLEDGE MAPS

2.1 Modeling

In this paper, we propose a knowledge map model to represent knowledge in legal documents. This proposed model is based on the brief previous models, including Rela-model [13, 16] and the OAK model [12]. In fact, OAK model is proposed to handle mined knowledge as knowledge maps from data mining tasks in digital agriculture [12].

The model for representing the legal document, called VLegalKMaps, consists of an ontology as Rela-model combining knowledge maps from legal documents. This model has the structure as follows:

$$\mathbb{K} = (\mathbb{C}, \mathbb{R}, \mathbb{RULES}) + (Key, Rel, weight)$$

In which:

- $(\mathbb{C}, \mathbb{R}, \mathbb{RULES})$ is a structure of the Rela-model [13, 16].
- *(Key, Rel, weight)* is a conceptual graph representing the relations between key phrases of legal documents. In which *Key* is a set of key phrases in the law document, *Rel* is the set of arcs, and *weight* is a map from Key to a binary vector.

The VLegalKMaps modeling is described in Figure 1.

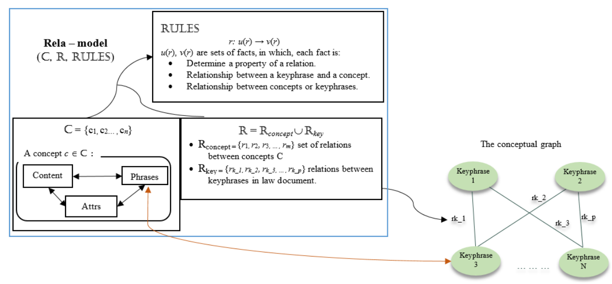

Figure 1: Example Legal KG model.

2.2 Ontology as Rela-model (C, R, Rules)

The Rela-model $(\mathbb{C}, \mathbb{R}, \mathbb{RULES})$ is an ontology to represent the knowledge of legal documents. In which, $\mathbb{C}$ is a set of concepts, each concept in $\mathbb{C}$ has an internal structure to organize its law information; $\mathbb{R}$ is a set of relations, those relations are between concepts, key phrases, and a database storing; and $\mathbb{RULES}$ is a set of inference rules of the knowledge domain.

Definition 3.1: Each concept $c \in \mathbb{C}$ consists of five elements, including *Name, Content, InnerRela, Phrases,* and *Attrs*, as belows:

$$c = (Name, Content, InnerRela, Phrases, Attrs)$$

where:

- *Name*: The name of the concept in the law.
- *Content*: the meaning of the corresponding concept.
- *InnerRela*: Articles in the document related to the corresponding concept.
- *Phrases*: Set of key phrases related to concepts in each article.
- *Attrs*: Set of components (or other concepts) builds the corresponding concept (if necessary).

Definition 3.2: Set $\mathbb{R}$ consists of two kinds of relations between concepts and keyphrases:

$$\mathbb{R} = \mathbb{R}_{concept} \cup \mathbb{R}_{key}$$

where:

- $\mathbb{R}_{concept}$ is a set of relations between concepts in $\mathbb{C}$. Those relations are "has-a", "is-a", "a-part-of", and other relations.
$$\mathbb{R}_{concept} \subseteq \mathbb{C} \times \mathbb{C}$$
- $\mathbb{R}_{key}$ is a set of relations between keyphrases in the law document. This set also contains relations between keyphrases and a concept to determine the meaning of that concept.
$$\mathbb{R}_{key} \subseteq Key \times Key \text{ and}$$
$$\mathbb{R}_{key} \subseteq Key \times \mathbb{C}.Phrases$$

$\mathbb{RULES}$-set includes deductive rules representing the constraint and inferring the relation between keyphrases and concepts. These rules help to deduce direct or indirect relationships to determine semantic similarity among keyphrases and concepts.

Definition 3.3:

a) The facts in VLgegalKMaps model are classified into three kinds:
   1. Determine a property of a relation.
      $[< r > is < property >]$ with $r \in \mathbb{R}$ is a relation.
   2. Determine a relationship between concepts or keyphrases.
      $[< d_1 >< r >< d_2 >]$ with $d_1, d_2 \in \mathbb{C}$ or $d_1, d_2 \in Key$.
   3. Determine the relationship between a keyphrase and a concept.
      $[< term >< r >< c.Phrases >]$ with $term \in Key$ and $c \in \mathbb{C}$.

b) Given a rule $r \in \mathbb{RULES}$. Then, the rule r has the form:

$$r : u(r) \longrightarrow v(r)$$

with $u(r)$ (and $v(r)$) is the set of hypothesis facts (and goal facts).

2.3 Conceptual graph *(Key, Rel, weight)*

The structure of Rela-model $(\mathbb{C}, \mathbb{R}, \mathbb{RULES})$ organizes the knowledge of a law document. However, in practice, when retrieving the content of law, there are some main key phrases in the query sentence that have been connected to knowledge through their semantics. In this study, the semantics of key phrases are organized by a conceptual graph.

Definition 3.4: The structure of the graph that represents the relations between key phrases in a legal document is a tube:

$$(Key, Rel, weight)$$

where:

- $Key = \{k | k$ is a key phrase of the legal document$\}$.
- $Rel = \{e = (k_1, k_2) \in Key \times Key |\ k_1$ and k_2 are key phrases appearing in the same article of the law document$\}$.
- $weight : Key \longrightarrow \mathbf{R} \times \mathbf{R}$ is a map to compute the similarly binary vector for each key phrase in Key. ($\mathbf{R}$ is the set of real numbers).

The measure for similarly key phrases is computed by the tube $(tf(v,d), idf(v,d))$, where $tf(v,d)$ is the term frequency representing the frequency of a key phrase v in a document d, and $idf(v,d)$ is the inverse articles frequency representing the specificity of the key phrase v in the document d. The formulas of $(tf(v,d), idf(v,d))$ are calculated in [16].

2.4 Hierarchy of Legislation

In Vietnam, under the law on the promulgation of legal documents in 1996 (amending and supplementing in 2002), the system of legal documents has three levels [14] as follows:

- Major legislations passed by the National Assembly is classified in the following order:
  – The Constitution, the highest law of the state;
  – Codes, Laws, and Decree-Laws (Ordinances) issued by the National Assembly;
  – Ordinances and Resolutions by the National Assembly's Standing Committee.
- Legal instruments issued by the President of the State and the Minister of the Ministry to implement the main laws. These sublaws are, in turn, classified in the following order.
  – Orders and Decisions signed by the State President;
  – Government's documents: Directives and Executive Orders signed by the Prime Minister;
  – Decisions, Directives, and Circulars signed by the Ministers or the Heads of government agencies.
- Local ordinances are issued by local People's Councils in order to implement the country's laws;
  – Resolutions of People's Councils;
  – Decisions and directives of the People's Committee.
  – International Treaties

In this system, major legislations are the most important because they affect the whole country, while local ordinances might affect locally at each province/city where it is issued.

2.5 Legal Ontology Design

According to the description of categories and the structure of legislative documents in Section 2.4, the deepest and most complex structure of one legal document has six levels, including document, section, part, chapter, article, clause, and point (Figure 2). Based on the structure and content of legal documents, we propose an ontology to represent knowledge in legal documents. The diagram illuminates the relationships between LegalItem and related objects (such as organizations, people, and locations) (Figure 3).

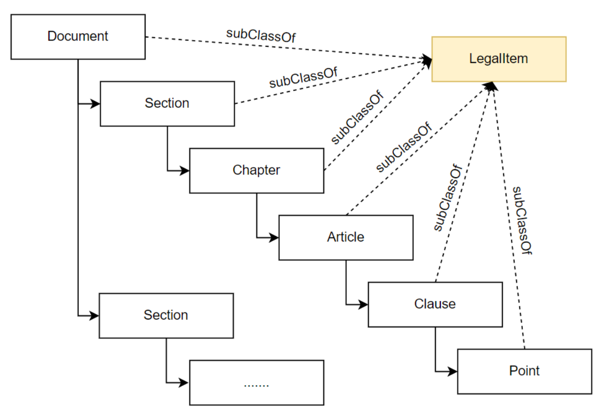

Figure 2: Structure of a General Legal Document/Law

This proposed diagram supports connecting legal documents, issued organizations, signed people, affected locations or organizations, and related entities (as shown in Figure 3).

The VLegalO ontology was implemented by Protege, which is a well-known tool for developing and maintaining ontologies. In Protege, object properties are used to create and modify relationships where the source of a relationship is called the domain class, and the destination of the relationship is called the range class.

Furthermore, several classes have been referred to concepts in the Document Components Ontology (DoCO[1]) [4]. DoCO provides

[1]http://purl.org/spar/doco

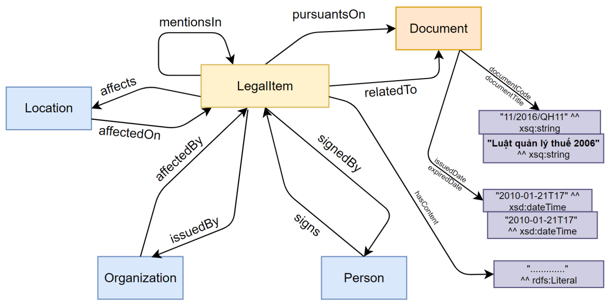

Figure 3: Legal Ontology for VLegalKMaps

Table 1: Raw legal documents for LegalKG

Document Type	*Count*
Code, Law	494
Decree	5,081
(Joint) Circulars	21,086
Ordinates/Orders	244
National Technical Regulation	1,550
Introduction	1,550
Decisions	172,346
Documentary	119,221
Total	350,000

a general framework to represent document components, such as chapter, abstract, section, paragraph, sentence, figure, table, glossary, and bibliography.

2.6 Linked Data Construction

The knowledge maps representing each legal document contain two parts, 1) one part for structured items and relations of the documents; and 2) the other part for conceptual graphs of legal knowledge in the documents. According to the theory of conceptual graph and design of the ontology (Section 2.2 and Section 2.5), each legal document will be extracted and transformed into a triple sets (Key, Rel, weight). In which,
{Key} is a set of key phrases in the document,
{Rel} is a set of pre-defined relations between key phrase pairs, and
{weight} is a set of
Both {Key} and {Rel} are instances of generated knowledge maps and they, they are linked to concepts in the pre-defined ontology (Section 2.5). All maps extracted from raw data are transformed into RDF triples[2], which is a well-known format for linked data and the foundation of graph databases.

3 IMPLEMENTATION

3.1 Materials

As mentioned in Section 2.4, there are many different types of legal documents in three groups; however, the most important documents are major legislative documents, including laws, decrees, and circulars (as shown in Table 1). All raw documents are crawled from existing public legal websites. They are pre-processed The set of raw documents is crawled from many different existing resources, and each document has a code, title, issued date, and issued organization to identify. All documents are pre-processed to create structured documents, then transformed into the first version of knowledge maps. Approximately 27,000 of total 350,000 documents are focused to extract conceptual graphs of legal knowledge in each document, then transformed into the second version of knowledge maps.

3.2 Building VLegalKMaps

The VLegalKMaps is built from raw legal documents based on the VLegalO ontology framework and NLP toolkits. There are four modules in the system as shown in Figure 4. Firstly, the data collection module is used to crawl data and collect data from the Internet. The document is parsed into a structured document, and then each element of the document is extracted key phrases and entities to create legal knowledge maps, as described in Figure 3. Finally, all elements of the structured document are converted into RDF triples as linked data.

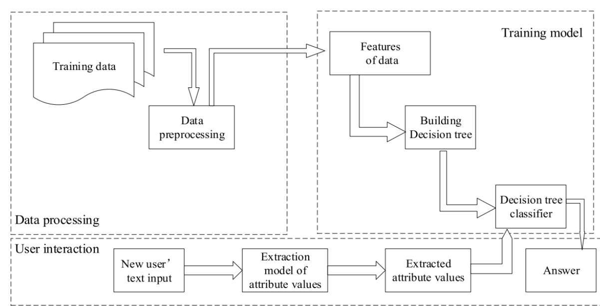

Figure 4: Process of building VLegalKMaps

3.3 Vietnamese NLP toolkits

Underthesea[3] is Vietnamese NLP Toolkit, which includes most common tasks in NLP, including sentence segmentation, word Segmentation, POS Tagging, Chunking, Dependency Parsing, Named Entity Recognition, Text Classification, and Sentiment Analysis.

SpaCy for Vietnamese. (NLTK, OpenNLP, StanfordNLP, SpaCy, Gate. **SpaCy** is used as an end-to-end practical tool to implement NLP applications using the Python ecosystem [1].

NLP Tasks for Building VLegalKMaps. Based on the requirements of VLegalKMaps, NLP toolkits need to support following tasks:

- *NP Chunking*, for key phrase extraction.
- *Named Entity Recognition*, for identifying location, organization, person, date, and document entities.
- *Dependency Parsing*, for extract relations in documents.

Basically, SpaCy is the most suitable toolkit for all the above tasks because it includes all well-known foundation NLP tasks, however, it only well supports English. For Vietnamese, it only

[2]https://www.w3.org/TR/n-triples/

[3]https://underthesea.readthedocs.io/

has the Vietnamese tokenizer[4] and needs to retrain the model for dependency parsing, noun phrase chunking, and named entity recognition. For dependency parsing, this study uses Vietnamese Tree Bank (VTB[5]) from VLSP project [15] for training. This dataset is available as a universal dependency format, therefore, it is used to train the dependency parsing model for Vietnamese SpaCy. For the recognition of Vietnamese named entities, we used the Vietnamese part of EVNECorpus [11] as the dataset, fine-tuned the SpaCy model under SpaCy's guidelines[6], then used the trained model for tagging.

4 VALIDATIONS

4.1 Experiment Setup

Basically, each domain creates a group of legal documents and they are linked to real entities in actual life. Table **??** presents the raw legal documents for LegalKG and Figure 5 shows a part of Legal KG model.

Listing 1: Triples of Law on Road Traffic

```
VLegalKMaps:LawDoc0349
    rdf:type owl:NamedIndividual ,
             VLegalO:Law ;
    rdfs:label "Law on Road Traffic,
                Law No. 23/2008/QH12." .
    VLegalO:hasKeyphrase "Regressor 004" .
            VLegalKMaps:Article_004 ;
    VLegalO:relatedTo VLegalKMaps:LawDoc0349;
VLegalKMaps:LawDoc0349 rdf:type owl:
    NamedIndividual ,
                       VLegalO:NTR ;

VLegalO:Keyphrase
    VLegalO:hasAlgorithm
            VLegalO:Algorithm_DTR ,
            VLegalO:Algorithm_LR  ,
            VLegalO:Algorithm_RF ,
            VLegalO:Algorithm_GBRT;
    VLegalO:hasCondition
            VLegalKMaps:SoilPH_004 ;
    VLegalO:predicts
            VLegalKMaps:SoilPH_004x ;
    VLegalO:hasDataset
            VLegalKMaps:Dataset_CONSUS_001 ;
VLegalKMaps:SoilPH_004
    rdf:type owl:NamedIndividual ,
            VLegalO:SoilPH ,
    VLegalO:hasTransformation
            VLegalO:Transformation_SoilPH_Max ,
            VLegalO:Transformation_SoilPH_Min ,
            VLegalO:Transformation_SoilPH_Avg ,
[...]
```

4.2 Case Study

Based on the knowledge base on Road Traffic Law[7] collected from [2, 8] and the architecture proposed in Section 4.2, an intelligent searching system on Vietnam's road traffic law is designed. This section presents some test results of the system through some kinds of inputted queries.

[4]https://github.com/trungtv/vi_spacy
[5]https://github.com/UniversalDependencies/UD_Vietnamese-VTB
[6]https://spacy.io/usage/training
[7]https://vanbanphapluat.co/law-no-23-2008-qh12-of-november-13-2008-on-road-traffic

Table 2: Results of testing queries on Vietnamese Traffic Law

Kinds	Quantities	Correct	Rate
Queries about concepts / definitions	41	34	83%
Queries about penalties and fines	55	42	76%
Total	**96**	**76**	**79%**

Example 5.1: The inputted query q_1 = *"What is motorcycle?"*

The system will extract keywords from the query q_1: *"What is"*, *"motorcycle"*. From this, it returns the following results:

"Motorcycle means a motor vehicle that has two or three wheels with a cylinder capacity of 50 cm3 or higher, maximum speed over 50 km/h and net weight not exceeding 400 kg."

In this query, the word *"what is"* is used to classify the query as a kind of declaration of the meaning of a concept. The keyword *"motorcycle"* helps to find the concept.

The result is retrieved from Article 3, Clause 3.31 of the National Technical Regulation on Traffic Signs and Signals [17].

Example 5.2: The inputted query q_2 = *"The fines of operator of motorbike does not wear helmet"*.

The keywords of the query q_2 are *"fines"*, *"not wear"*, *helmet"*, *"operator of motorbike"*. The word *"fines"* used to classify the query as penalties and fines for offenses. The word *"operator of the motorbike"* consists of *"motorbike"* which is similar to the word *"motorcycle"*. The word *"helmet"* is a part of the concept *"motorcycle helmet"*. Therefore, the concepts of query q_2 are *"operator of motorcycle"* and *"motorcycle helmet"*. The relational keyword is *"not wear"*. With the concepts and the relation, the rules were used to match them and the result was found. The result is returned:

"Through article 6, Decree 100/2019/ND-CP [8]: Penalties imposed on operators of mopeds and motorcycles (including electric motorcycles) and the like that violate the rules of road traffic.

2. A fine ranging from VND 200,000 to VND 300,000 shall be imposed upon a vehicle operator who commits any of the following violations:

i) The operator or the passenger on the vehicle does not wear a motorcycle helmet or does not wear it properly;"

The designed system can perform some common searches on the traffic law. It is effective with finding usual penalties and fines in the road. This system was tested on 95 queries about road traffic codes. The results are shown in Table 2:

5 CONCLUSIONS AND FUTURE WORKS

In conclusion, with the proposed legal knowledge map model, the knowledge management system based on the VLegalKMaps architecture is a potential architecture to expand it to other knowledge types and knowledge domains. As a result, knowledge items are imported into the knowledge management system for retrieval. Moreover, the knowledge management system also supports a knowledge browser application as a further way to access knowledge for both data scientists and agronomists with different levels of interpretations of pre-model and post-model stages of AI applications.

Our future work will focus on two areas: (a) Building the explanation interface (such as the service to interact with several

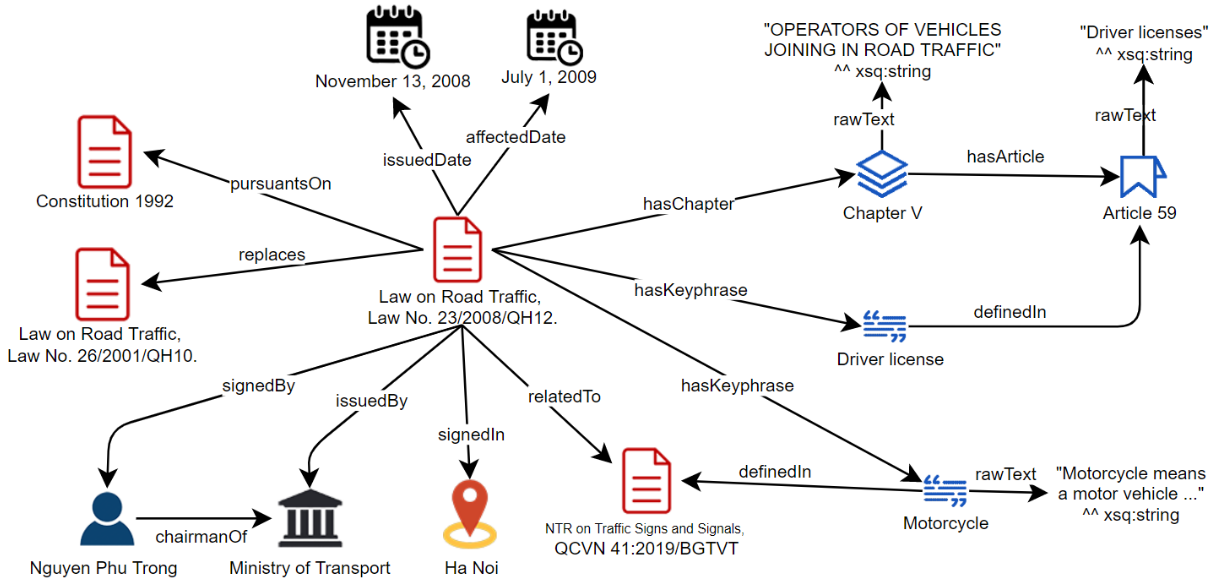

Figure 5: Example Legal KG model.

groups of users, taking user questions as input and retrieving existing knowledge in this system (extracted from AI systems); (b) Expanding the proposed model to compromise several algorithms to predict tasks based on explainable approaches.

Moreover, representations can be supported by different legal services that would support several user groups, by taking user's questions as input, retrieving existing knowledge in the OAK model (extracted from AI systems), and then developing a reasoning module to decide the most suitable explanation.

REFERENCES

[1] Duygu Altinok. 2021. *Mastering spaCy: An end-to-end practical guide to implementing NLP applications using the Python ecosystem.* Packt Publishing Ltd.
[2] Vietnam National Assembly. 2008. *Law on Road Traffic, Law No. 23/2008/QH12.*
[3] Chantana Chantrapornchai and Aphisit Tunsakul. 2021. Information extraction tasks based on BERT and SpaCy on tourism domain. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)* 15, 1 (2021), 108–122.
[4] Alexandru Constantin, Silvio Peroni, Steve Pettifer, David Shotton, and Fabio Vitali. 2016. The document components ontology (DoCO). *Semantic web* 7, 2 (2016), 167–181.
[5] Ademar Crotti Junior, Fabrizio Orlandi, Damien Graux, Murhaf Hossari, Declan O'Sullivan, Christian Hartz, and Christian Dirschl. 2020. Knowledge graph-based legal search over german court cases. In *European Semantic Web Conference.* Springer, 293–297.
[6] Jaspreet Singh Dhani, Ruchika Bhatt, Balaji Ganesan, Parikshet Sirohi, and Vasudha Bhatnagar. 2021. Similar Cases Recommendation using Legal Knowledge Graphs. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*
[7] Erwin Filtz, Sabrina Kirrane, and Axel Polleres. 2018. Interlinking Legal Data.. In *SEMANTiCS Posters&Demos.*
[8] Vietnam Government. 2019. *Decree on Administrative penalties for road traffic and rail transport offences, No. 100/2019/ND-CP.*
[9] Martin Kaltenboeck, Pascual Boil, Pieter Verhoeven, Christian Sageder, Elena Montiel-Ponsoda, and Pablo Calleja-Ibáñez. 2022. Using a Legal Knowledge Graph for Multilingual Compliance Services in Labor Law, Contract Management, and Geothermal Energy. In *Technologies and Applications for Big Data Value.* Springer, 253–271.
[10] Elena Montiel-Ponsoda and Víctor Rodríguez-Doncel. 2018. Lynx: Building the legal knowledge graph for smart compliance services in multilingual Europe. In *Proceedings of the 1st workshop on LREC (Language Resources and Technologies for the Legal Knowledge Graph) workshop.* 19–22.
[11] Quoc Hung Ngo, Dinh Dien, and Werner Winiwarter. 2014. Building English-Vietnamese named entity corpus with aligned bilingual news articles. In *Proceedings of the Fifth Workshop on South and Southeast Asian Natural Language Processing.* 85–93.
[12] Quoc Hung Ngo, Tahar Kechadi, and Nhien-An Le-Khac. 2020. OAK: Ontology-based Knowledge Map Model for Digital Agriculture. In *Future Data and Security Engineering: 7th International Conference (FDSE)*, Vol. LNCS 12466. Springer.
[13] Hien D Nguyen, Vuong T Pham, Trung T Le, and Dat H Tran. 2015. A Mathematical Approach for Representing Knowledge about Relations and Its Application. In *Proceedings of 7th International Conference on Knowledge and Systems Engineering (KSE 2015).* IEEE, 324–327.
[14] Phuong-Khanh Nguyen. 1999. How to Conduct Research in Vietnamese Law: Overview of the Legal System of the Socialist Republic of Vietnam. *International Journal of Legal Information* 27, 3 (1999), 307–331.
[15] Phuong Thai Nguyen, Xuan Luong Vu, Thi Minh Huyen Nguyen, Hong Phuong Le, et al. 2009. Building a large syntactically-annotated corpus of Vietnamese. In *The Third Linguistic Annotation Workshop-The LAW III.* 6p.
[16] Thinh H Nguyen, Hien D Nguyen, Vuong T Pham, Dung A Tran, and Ali Selamat. 2022. Legal-Onto: An Ontology-based Model for Representing the Knowledge of a Legal Document.. In *Proceedings of 17th Evaluation of Novel Approaches to Software Engineering (ENASE 2022), Online streaming.* 426–434.
[17] Vietnam Ministry of Transport. 2019. *National Technical Regulation on Traffic Signs and Signals. QCVN 41:2019/BGTVT.*
[18] Ginevra Peruginelli and Sebastiano Faro. 2019. *Knowledge of the Law in the Big Data Age.* Vol. 317. ios Press.
[19] Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Dmytro Okhonko, Samuel Broscheit, Gautier Izacard, Patrick Lewis, Barlas Oğuz, Edouard Grave, Wen-tau Yih, et al. 2021. The Web Is Your Oyster–Knowledge-Intensive NLP against a Very Large Web Corpus. *arXiv preprint arXiv:2112.09924* (2021).
[20] Julián Moreno Schneider, Georg Rehm, Elena Montiel-Ponsoda, Víctor Rodríguez-Doncel, Patricia Martín-Chozas, María Navas-Loro, Martin Kaltenböck, Artem Revenko, Sotirios Karampatakis, Christian Sageder, et al. 2022. Lynx: A knowledge-based AI service platform for content processing, enrichment and analysis for the legal domain. *Information Systems* 106 (2022), 101966.
[21] Francesco Sovrano, Monica Palmirani, and Fabio Vitali. 2020. Legal knowledge extraction for knowledge graph based question-answering. In *Legal Knowledge and Information Systems.* IOS Press, 143–153.
[22] Sebastian Thelen. 2015. The Ontology-Based Approach of the Publications Office of the EU for Document Accessibility and Open Data Services. In *Electronic Government and the Information Systems Perspective: 4th International Conference, EGOVIS 2015, Valencia, Spain, September 1-3, 2015, Proceedings,* Vol. 9265. Springer, 29.
[23] Ziyue Wang, Baoxin Wang, Xingyi Duan, Dayong Wu, Shijin Wang, Guoping Hu, and Ting Liu. 2019. IFlyLegal: A Chinese Legal System for Consultation, Law Searching, and Document Analysis. In *Proceedings of the 2019 Conference on*

Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations. 97–102.

[24] Haoze Yu, Haisheng Li, et al. 2021. A knowledge graph construction approach for legal domain. *Tehnički vjesnik* 28, 2 (2021), 357–362.