Ontologies as a Set to Describe Legal Information

Anatolii Getman[1][0000-0002-1987-2760], Volodymyr Karasiuk[1][0000-0001-9092-2137]
and Yevhen Hetman[2][0000-0002-1801-7252]

[1] Yaroslav Mudryi National Law University, Kharkiv, Ukraine
[2] National Academy of Law Science of Ukraine, Kharkiv, Ukraine
karasiuk@yahoo.com

Abstract. The article discusses the features of legal knowledge ontology creation. It is determined that ontology is the most appropriate way to describe legal knowledge. The particular qualities of legal information and the features of the language of a right were investigated. A review of legal knowledge ontologies that are used in various branches of law was made. The properties of legal information and the requirements for regulatory documentation in Ukraine were described. The formalization of the structure of the ontology database was presented, taking into account the required attributes of the concepts. The methodology of the work with the knowledge base was proposed to use the independent work of many users. The legal knowledge ontology at the law university was filled by all users of the software package, but experts checked the quality of this content. Crowdsourcing was considered as the main technique of the ontology filling process. Several branches of the ontology of legal knowledge were filled. The results of the experimental operation of this ontology by university students were analyzed.

Keywords: artificial intelligence, legal information, ontology, knowledge representation, crowdsourcing

1 Introduction

The legal systems of continental Europe, including Ukraine, use regulatory law. That is, they pay more attention to judicial interpretation in assessing legal facts and actions than to judicial precedents. Therefore, the regulatory framework (legislation) first of all is to be easy of access and simply ordered. For many practical areas of application of law information should also be available on the daily activities of legal structures, that is, by-laws and regulations (on a second-priority basis). A modern way of structuring legal knowledge is to describe them using ontologies. Based on legal ontologies and inference rules, automated decision-making methods in the field of law can be implemented. However, to create legal ontologies, it is necessary to conduct a legal analysis by the efforts of legal experts and knowledge specialists. Features of the application of legal information requires the concept of legislative defini-

tions and attributes of regulatory documents in which they are introduced and used. One more serious problem is the organization of the process of filling the ontology with relevant content. In this process, enormous amounts of legislative (and regulatory) information as well as the presence of collisions and confusion of terminology are complicating conditions. With regard to legal information for application software, it is to be noted the high variability of its content. Continuous changes in the regulatory framework are difficult to track and model in the ontology. It is impossible to rely on only the efforts of legal experts to fill the ontology, due to its volume. In this case, you can count on the use of crowdsourcing and advanced rules for working with the content of the ontology.

The ideas and approaches considered in this article were implemented in the process of creating an ontology, which is applied in the educational process at the university. The present article is a further development of the ideas contained in [1-4].

2 Particular Qualities of Legal Information

Legal information has a great importance for the functioning of modern society. Due to its significance, it has significant features and otherness. Characteristics of legal information and the legal language in Ukraine in general are given in detail in reviews [5, 6].

With regard to legal terminology, it is special because of its versatility, since any social activity is closely linked to law. A clear definition and unambiguous interpretation of legal terms is an extremely important condition for becoming constitutional state, contributes to the improvement of legislative activity, the effective implementation of legislative acts, creates the conditions for their ordering and systematization of legislation. In order to carry out its functions, the term must satisfy the requirements that are scientifically substantiated and legally binding. Most experts consider systematic, accurate, unambiguous, stylistic neutrality (lack of emotionally expressive marking) and motivation to be inalienable attributes of the legal term.

Regulatory and legal terminology, moreover, are characterized by multi-component quasi-terms - long-term descriptive turnovers of different structures that give a fairly accurate definition of the subjects and objects of specific relationships and their interaction, thereby reflecting the essence of legal relations in each case, which is regulated in this case regulatory act.

There are certain rules for using regulatory terminology when creating legal documents [7]. And the volume of by-laws (instructions, regulations, certificates, and other) is generally difficult to assess.

3 Formulation of the Problem

The main purpose of this research are: explore legal information to spot features that are fundamental to the formation of the electronic structure of legal knowledge; study the problem area of knowledge engineering to build legal information ontology, and then provide an ontology model and technology to work with it. Features of legal information imposes their requirements on the structure of the ontology and the relationship between concepts. Attributes of concepts are very important for practical activities in the field of law. The task is also to investigate the crowdsourcing method for ontology filling-in.

4 The Use of Ontologies to Represent the Legal Information

Intelligent knowledge-based systems have various models of data processing and display. The most common in practice are taxonomies, semantic networks, and ontologies.

Taxonomy. Taxonomies display hierarchy of concepts associated subordination relation, wherein each subsequent level taxonomy includes previous subtypes. Formally taxonomy can generally be described: $T = (C, H)$, where C is the set of terms (concepts), H - hierarchy relationships between terms (concepts).

Semantic network. Network models represent a directed graph wherein vertices - are concepts and arcs - the relationship there between: $N = (C, R)$, where C is the set of terms (concepts), R - set of relationships between concepts. This is the closest view to what natural language information appears. Network model at the same time does not give an idea of the domain hierarchy; in addition, the creation and modification of the subject area model is difficult; moreover, for the processing of network models a special apparatus for formal inference and planning is necessary [8].

Ontology. Ontologies are a variety of frame models that are used to describe resources in WEB applications, corporate databases, document processing applications, and the like [9-11]. Ontologies have a large number of element types and relationships between them. Ontologies contain such elements: concepts (notions), classes, relations, interpretation functions. Concepts are objects - the vertices of the ontology graph that have their own semantic representation; classes combine many of the same type of concepts. Relationships and interpretation functions define the relationship between terms and their properties.

Ontology should introduce a minimum basic set of concepts, but sufficient for modeling purposes, and given the description of complex situations, determine the ratio between these concepts, their characteristics, effectively transmit certain value concepts necessary to describe situations and interpretation functions.

A more relevant definition of ontology is given in the IEEE Std 1872 ™ -2015 standard: ontology is the basis that defines the main concepts, their properties, rela-

tionships between concepts and domain rules. Taxonomies provide an ordered set of vocabulary and a single type of relationship between the terms, and an ontology provides more relationships, restrictions, and rules. Ontologies provide relevant knowledge about the subject area explicit in a computer-interpreted format, which allows software (SW) to reason about this knowledge to output new information. Ontologies are an excellent tool to reduce the ambiguity in transferring knowledge between groups of people, information systems, knowledge bases, control systems and other objects that are built on the same abstraction [12].

With regard to law, ontologies here allow us to emphasize [13, 14] the relationship of legal norms with each other; the relevance of official legal conclusions; take into account the degree of priority of legal norms; the identity of the diverse branches of law and more. Ontological engineering in law therefore has its own characteristics: 1) a large number of generally accepted concepts with their own specific application; 2) the difference in the structures of various branches of law; 3) the presence of a general theoretical legal level between the ontology of the upper level and the ontology of the subject area; 4) a large number of theoretical assumptions and abstract constructions, which depend on the specifics of legal views. Therefore, it should be considered that the most effective approach is founded on the learning of the terminology of law and the construction of ontologies on large text blocks.

Today, reliable methods for creating ontologies have not been created, and there is no methodology for a comprehensive assessment of ontologies. Although general approaches imply the idea of an assessment [15]: epistemological identity (clarity, intuition, relevance, completeness), executive authenticity (sequence, computability), the possibility of multiple use (task and method, subject area). These criteria are anticipated to be developed for use as a standard for describing all ontologies.

In Ukraine, work on the creation of legal ontologies is also underway [16 - 18].

5 Overview of Legal Information Ontologies

Currently, a large number of software packages have been developed that implement information systems in law based on ontologies. Consider the most interesting.

PrOnto [19], the ontology is designed to describe information about products and their groups, and is currently effectively used to describe the GDPR (General data protection regulation). This ontology considers the GDPR as a starting point, however it is meant to be extended to the concepts and relative relations of other legal frameworks.

LRI-Core - ontology, which covers all branches of law. Book [20] illustrates the use of LRI-Core for legal ontologies in the Netherlands.

CLO – (Core Legal Ontology) SLO is an instrumental framework for building legal ontologies for any branches of [21].

LKIF - is, first of all, several libraries of related concepts and their properties. LKIF is a universal means for the legal domain [22].

LTS - Legal Taxonomy Syllabus - is an ontology-based legal curriculum [23].

Eurovoc thesaurus - is a multilingual thesaurus that forms the basis of domain names for the European Union terminology database. Eurovoc is available in 23 official languages.

Eunomos - is a software for managing legal documents and terminology. Eunomos is focused on tracking changes in regulatory documents. This software supports multiple languages and focuses on many areas of application in various branches of law.

LegalRuleML - a package that uses Legal XML to represent a model of legal knowledge and uses the rules to build a legal ontology. It includes NLP tools. Very interesting project.

NRV - following development of LegalRuleML, which contains a set of basic primitives for identification and classification.

ODRL - a structured language focused on legal support of open licenses in publishing digital data on the network - photos, software, news and other data.

LDR - a modification of the language ODRL, which is focused on the use of model LOD, including the multilingual data.

CC - language for describing the use of Creative Commons copyright licenses. RDF concept used.

L4LOD - ontology for one problem area - for a universal description of licenses in a computer network for data exchange.

ELI - ontology for a standardized description of European law through XML. This ontology can be used online on the internet.

LOTED2 - legal ontology, which provides information on tendering in Europe. It supports the legal grounding for purchases.

PPROC - an ontology in which a semantic description is used to support public procurement procedures. Ontology prevents misconduct in procurement.

And other ontologies [24, 25].

6   Features of Legal Knowledge Ontology

The mathematical description of the legal ontology that we are developing is presented in detail in [3].

Features of legal information that were considered earlier should be taken into account when developing an ontology scheme:

- the presence of a synonymous description of some legal terms (nodes of ontology);

- the presence of restrictions on the duration of wording of a legal term;

- the presence of a mandatory relationship between definitions (ontology nodes) and strict wording (legislative definitions) in regulatory documents.

-and more others.

That is, the database implements the ontology, includes the following parts:

- concept (legal term) and connections;

- connections of groups of synonyms of legal terms;
- texts - definitions of legal terms in accordance with regulatory documents;
- lexical mapping of legal terms and relationships;
- indices the use of legal terms and relationships in the source document.

More about each part.

• concept and relations: are recorded with unique identifiers, with a name string for output in the graphical interface.

• texts - definitions of legal terms: presented in the form of sentences, with fixation of belonging to the section and the text, that is, among them the following entities can be distinguished:

- text - in accordance with the structure of the normative document, this is a sequence of paragraphs, type, title, list of service attributes of the document;

- paragraph - description of the structure of the part of the document, which includes the title of the paragraph, the sequence of blocks and separate sentences of the text;

- sentence - a structural unit of the source text in the form of a string and the type of sentence (indicating membership in the source regulatory document or its paragraph).

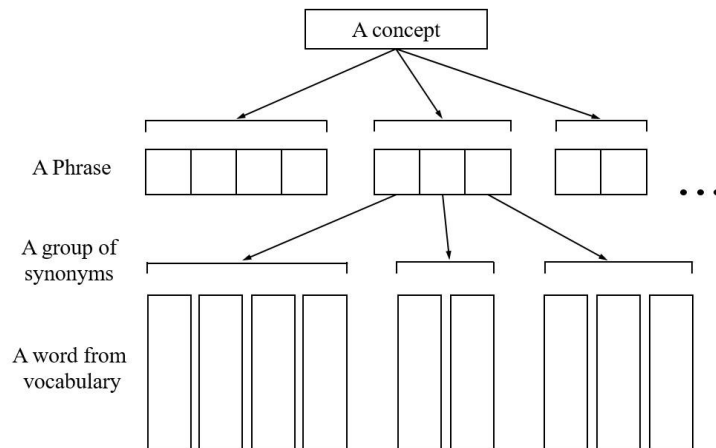Fig. 1 schematically shows the hierarchy between the legal term (concept), sentences, groups of synonyms.

Fig. 1. The structure of concepts and their relationships in the ontology database.

The problem of terminological confusion. The basic principle of legal technology requires that different terminology in one text designate different legal phenomena. Moreover, ideally, they try to extend this principle to the entire legal system as a whole, so that only one term answers each legal phenomenon within it. This problem is not confined to the particular case of terminological confusion in any particular

document - it is quite common, has long been known and has deep roots. One can find a series of many terminological inconsistencies and contradictions in the Ukrainian legislation. An analysis of the laws of Ukraine from the point of view of their terminological certainty gives reason to conclude that the meaning of the term used in one law is not always the same in content with the term contained in another law. However, this is due to the fact that in addition to general legal terminology, which combines the basic terms of the entire system of legislation, there is intersectoral and industry terminology. Tautological constructions, polysemy and the like lead to confusion in understanding the regulatory requirements of the laws of Ukraine [26].

In addition, conflicts between the provisions of regulatory documents of various branches of law are not uncommon. In the era of the widespread dissemination of electronic documents, digital databases and systems of related documents, the problem of adhering to terminological unity and its gradual implementation in current legislation and practice of its application is of fundamental importance. The terminological toolkit gives lawyers such an opportunity. However, on February 4, 2020 Law No. 469-IX was introduced to improve the electronic form of workflow in the Verkhovna Rada. Previously, it was assumed that publications on the Verkhovna Rada website are considered the official publication of laws (along with the publication in the Voice of Ukraine newspaper and in the Verkhovna Rada Vedomosti). Now the official is considered the publication of laws and other acts of the Verkhovna Rada in these media. That is, the electronic form of the document was made dependent on the printed form. And the issues of streamlining the conceptual and terminological apparatus of legal information remain relevant for modern Ukrainian jurisprudence. Naturally, when creating an ontological structure, these circumstances should be kept in mind.

7 How to Work with This Ontology

Crowdsourcing. Only the efforts of experts are not enough to build ontology of this scale. In addition, the legal knowledge system is very dynamic, «alive». One can recall Wikipedia, which was filled by everyone and a high level of reliability of the information contained in it. It means that the approach does works.

A number of projects for the creation and improvement of ontologies are known in which the crowdsourcing method was applied [27, 28].

We managed to attract a significant group of students to fill the ontology. As part of the training course, they analyzed the information of the subject area, identified concepts, and created (filled) the ontology. The work was carried out under the general guidance of a teacher, and after acquiring some experience, without assistance. The results turned out to be rather optimistic both in relation to the initial ontology, covering one subject area [4].

The completion of the first stage of creating an ontology of legal knowledge showed the following results. The quality of extracting legal concepts from legislative documents and textbooks was 90%. Students identified and contributed a total of

more than 6,000 legal terms. Experts noted some of the shortcomings identified in the ontology. Students were not able to correctly identify some concepts that have complex definitions consisting from several words. Conversely, several concepts were defined verbose, more complex than they actually were. Also, when filling in the ontology, some concepts were not placed in the branches where they should have been located according to the hierarchy of concepts. However, in general, the result is quite good.

8 Discussion

After creating the ontology, it is necessary to evaluate the characteristics and reliability of the obtained ontology. In accordance with [29], evaluation is required throughout the entire ontology life cycle. However, in analyzing the different methods for evaluating of an ontology it is to be noted that the proposed metric [30] difficult to apply ontologies in various problem areas that have serious differences of the structure and output functions.

Ontology can be actively used if it is complete and has a convenient interface. Users got used to modern information systems with a developed interface. Shortcomings in the interface repel potential users.

The experience of filling and training operation of an exemplary ontology has shown that the problem of updating (supplementing) the ontology with new concepts is acute. Users rarely refer to concepts that are needed infrequently, and these branches remain unfilled. That is, to achieve perfect coverage of the subject area is very difficult. Perhaps, when scaling the system, with the access of a huge number of users (all at once), this problem will be solved, as is now the case on Wikipedia — the classic crowdsourcing system.

Organizing the process of filling the legal ontology, there is a proposal on the feasibility of forming an ontology at the time of adoption and the description of the normative act, due to a competent and proper analysis of correlation of concepts by experts in the divisions of the Verkhovna Rada are responsible for work with legal information.

In any case, ontologies are a better mechanism than using XML markup of legal documents. Although it requires large expenditures for maintenance and development.

The lack of state support for the development and filling of legal ontology dooms dozens of enthusiastic groups to proactive, unrelated (multidirectional) work in this area.

9   Conclusions

As a result of the study, analysis of the advantages and disadvantages of various knowledge representation systems, it is advisable to use an ontological representation to describe legal knowledge.

The full implementation of the task of developing an ontological description of legal information allows lawyers to provide a tool for access to relevant regulatory information in real time. And given the volume of the regulatory framework and its variability, this will have a great practical effect.

In the framework of this work, the structure of the ontology database is formed, the structure takes into account the attributes of concepts related to legislative definitions and the time frame of validity of regulatory documents.

The created ontology is filled with key concepts of one branch of law (which is considered as the core on the basis of which a full ontology will be developed). This ontology is used in the educational process for specially prepared tasks in the study of selected topics.

To fill in the ontology, the crowdsourcing method was used with the attracting of a sufficient number of law students. We hope that with the accumulation of experience with the created ontology, the quality of filling in the ontology will increase.

Based on the work performed, promising for future work, we consider the following:

- development of automatic ontology comparison tools, which will be an important tool for assessing the completeness, consistency of ontologies created by different users, as well as comparing ontologies of different branches of law (or ontologies of different countries);

- development of an interface in a natural language, to attract non-professional users of information systems to work.

References

1. Tatsyi, V., Getman, A., Ivanov, S., Karasiuk, V., Lugoviy, O., Sokolov O.: Semantic network of knowledge in science of law. In: Shokin Yu., Bychkov, I., Potaturkin O. (Eds.) Proceedings of the IASTED International Conference on Automation, Control, and Information Technology (ACIT 2010), 218-222. ACTA Press, Anaheim, USA (2010).
2. Karasiuk, V.: Ontological paradigm of process of content for education purposes. Bulletin of V. Karazin Kharkiv National University 19 (1015), 148-154 (2012).
3. Getman, A., Karasiuk, V., Hetman, Ye., Shynkarov, O.: Ontological Representation of Legal Information and an Idea of Crowdsourcing fot Its Filling. In: O. Chertov et al. (Eds.): ICDSIAI 2018, AISC, 836, 179-188, 2019. DOI: 10.1007/978-3-319-97885-7_18.
4. Getman, A., Karasiuk, V.: A crowdsourcing approach to building a legal ontology from text. Artificial Intelligence and Law. 22(3) 313-335 (2014). DOI: 10.1007/s10506-014-9159-1.

5. Law of Ukraine "On information" No. 2657-XII. In: Summaries of the Verkhovna Rada of Ukraine (VRU), 48 (1992).
6. Besedna, L. Terminology and definition of the concepts in regulatory acts. Information and Law. 2(5) 39-47 (2012).
7. Gladkivska, O.: Requirements for normative and legal terminology. Information and Law. 1(13) 55-62 (2015).
8. Borge-Holthoefer, J. and Arenas, A.: Semantic Networks: Structure and Dynamics. Entropy. 12 1264-1302 (2010). DOI: 10.3390/e12051264.
9. Gruber, T.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Int Journal of Human-Computer Studies. 43 (5-6) 907–928 (1995).
10. Jones, T.: Artificial Intelligence: A Systems Approach. Jones and Bartlett Publishers, Inc. (2008).
11. Doerr M., Kritsotaki A., Christophides V., Kotzinos D. Reference Ontology for Knowledge Creation Processes. In: Moen A., Morch A.I., Paavola S. (eds) Collaborative Knowledge Creation. Technology Enhanced Learning, vol 7. SensePublishers, Rotterdam. 31–52 (2012). DOI: 10.1007/978-94-6209-004-0_3.
12. The National Artificial Intelligence Research And Development Strategic Plan: 2019. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf, last accessed 2020/02/11.
13. Noy, N., McGuiness, D.: Ontology development 101. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 (2001). http://www.ksl.stanford.edu/people/dlm/papers /ontology-tutorial-noy-mcguinness.pdf, last accessed 2015/06/10.
14. Leone, V., Caro, L.D., Villata, S.: Legal Ontologies and How to Choose Them: the InvestigatiOnt Tool. In: International Semantic Web Conference (2018). http://ceur-ws.org/Vol-2180/paper-36.pdf, last accessed 2020/03/10.
15. Visser, P., Bench-Capon, T.: A Comparison of Four Ontologies for the Design of Legal Knowledge Systems. Artificial Intelligence and Law. 6, 27-57 (1998).
16. Khala, K.: Development of an OWL ontology for representing conceptual knowledge in the legal field on the example of court cases. Ukrainian Journal of Educational Studies and Information Technology, 4(2), 51-55 (2016).
17. Kosenko, S.: The main statements of ontology theory and its implementation in the system of legal knowledge. Information Technology and Security, 4, 2(7), 154-171 (2016).
18. Kiridon, A. (Ed.): The science of the ambush that theoretical and methodological principle of creating basic encyclopedias. Kiyv, Encyclopedic View (2015).
19. Palmirani M., Martoni M., Rossi, A., Bartolini, C., Robaldo L.: PrOnto: Privacy Ontology for Legal Reasonin. Springer International Publishing (2018) DOI: 10.1007/978-3-319-98349-3_11.
20. Chen, H., Brandt, L. et al. (Eds). Digital Government: E-Government Research, Case Studies, and Implementation. Springer US 233-261 (2008) DOI: 10.1007/978-0-387-71611-4.
21. Gangemi, A.: Design patterns for legal ontology construction. CEUR Workshop Proceedings. 321, 65-85 (2007).
22. Hoekstra, R., Breuker, J., Di Bello M., Boer, A.: The LKIF core jntology of basic legal concepts. In: Proceedings of the CEUR Conference, 321, 43-63 (2007).
23. Schmitz, P., Francesconi, E., Batouche, B., Landercy S. P., Touly, V.: Ontological models of legal contents and users' activities for EU e-Participation services. EGOVIS 2017. LNCS, 10441 99-114 (2017).

24. Peruginelli G. and Faro S. (Eds.): Knowledge of the Law in the Big Data Age. IOS Press. Netherlands 2019. DOI: 10.3233/FAIA317.
25. Jimenez, C., Alarcon, D., Royo, C.: Comparative Analysis of Legal Ontologies, a Literature Review (TFR). University of Catalonia (2019). https://pdfs.semanticscholar.org/a8cd/7aeb45b61d95ccd6ad71f1d585de1e93a798.pdf, last accessed 2020/02/11.
26. Klochko, M.: Legal Terminology: Concepts, Features. State Building and Local Self-Government 18 148-154 (2009).
27. Sivula, A., Kantola, J. Ontology focused crowdsourcing management. Procedia Manufacturing 3 632 – 638 (2015). DOI: 10.1016/j.promfg.2015.07.286 https://www.researchgate.net/ publication/283960344_Ontology_Focused_Crowdsourcing_Management, last accessed 2020/02/19.
28. Amini, R., Cheatham, M., Grzebala, P. and McCurdy, H. B.: Towards Best Practices for Crowdsourcing Ontology Alignment Benchmarks. In: 15th International Semantic Web Conference (ISWC) 2016 1-12 (2016). http://disi.unitn.it/~pavel/om2016/papers/om2016_Tpaper1.pdf, last accessed 2020/01/19
29. Zurek, T.: Conflicts in legal knowledge base. Foundations of computing and decision sciences, 37 (2) 129-145 (2012). DOI: 10.2478/v10209-011-0006-9.
30. Zhu, H. et al.: Quality Model and Metrics of Ontology for Semantic Descriptions of Web Services. Tsinghua science and technology, 22 3 254 – 272 (2017). https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7914198, last accessed 2020/02/29.