DOI 10.52209/1609-1825_2025_4_188 UDC 69:004

Ontology Modeling of Normative Documents: Analysis of Approaches and Prospects of Automation

[1]**KABZHAN Zarina,** *Master's Degree, Head of Department, Kabzhan90@gmail.com,*
[1]**SHAKHNOVICH Alexandr,** *Cand. of Tech. Sc., Managing Director for Development and Digitalization, gipocentr@gmail.com,*
[1]***SHOGELOVA Nazym,** *Master's Degree, Specialist, nazymshogelova@gmail.com,*
[1]**GLYZNO Yevgeniya,** *Master's Degree, Senior Employee, e.glyzno@kazniisa.kz,*
[2]**GORSHKOV Sergey,** *Technical Director, serge@datavera.kz,*
[1]JSC «Kazakh Research and Design Institute of Construction and Architecture», 3 Microdistrict, 44a, Almaty, Kazakhstan,
[2]LLC «Datavera», Radostovets Street, 323a, Almaty, Kazakhstan,
**corresponding author.*

Abstract. *The use of ontological modeling for the formalization and automation of regulatory document processing in the construction industry is explored. The main goal of the study is to develop a methodological and technological framework for creating an automated regulatory control system. The methodologies for ontology development are reviewed, including upper-level ontologies (e.g., Basic Formal Ontology) and domain-specific ontologies (Industry Foundation Classes). Approaches for identifying duplications and inconsistencies in regulatory documents using knowledge organization models, such as SKOS, are described. Special attention is given to the automation of regulatory statement analysis through the creation of "semantic profiles". The application of natural language processing (NLP) technologies and large language models (LLMs) is proposed to enhance analysis efficiency. The need for expert involvement at the current stage of technological development to verify the obtained results is emphasized.*

Keywords: *ontological modeling, regulatory documents, construction industry, NLP, automation, semantic inconsistencies, machine learning.*

Introduction

This study examines ontological modeling applied in the context of regulatory documents in the construction industry. Ontologies, which originally emerged in philosophy as a tool for formalizing existing entities and their relationships within various theories, have been actively used in several applied fields since the late 20th century, including business process modeling, knowledge management, and data integration [1]. These approaches have found their place in areas such as software engineering, biomedicine, and enterprise management, where they help formalize knowledge, improve data exchange processes, and automate information processing [2].

The relevance of this study lies in the need to address the issue of automating the processing of regulatory documents in the construction industry, where ensuring strict compliance with regulatory norms plays a key role in improving the quality and safety of construction projects. Regulatory documents contain numerous complex and diverse terms, which creates challenges in their interpretation and compliance control. Ontological modeling can provide the structuring of these documents and automate their analysis processes, thereby significantly facilitating knowledge management and information exchange between different systems and users.

With the advancement of Natural Language Processing (NLP) methods and large language models (LLMs), there is a growing opportunity to significantly enhance the automation of regulatory document analysis. The integration of ontologies into the regulatory data processing not only promotes the unification of concepts but also simplifies compliance control in construction projects, making this area of research highly relevant.

The article by Neuhaus analyzes various

definitions of ontology, including Gruber's classic definition as "an explicit specification of a conceptualization". The authors note that many definitions violate the clarity rule for definitions. An alternative definition is proposed, which takes into account the variability of ontologies, the choice of vocabulary, and the role of annotations, as well as discussions on collaborative ontology development and the relationship between ontological conceptualism and realism [3].

Otero-Cerdeira presented a literature review on ontology matching over the past decade, analyzing more than 1,600 articles. The study classifies publications to identify trends and the distribution of scientific activity, helping new researchers navigate the field and determine promising research directions [4].

The article by Park presents the DocOnto methodology for building ontologies based on engineering documents, following a "bottom-up" approach. It consists of three stages: creating a terminology ontology, integrating it with semantic networks, and optimizing it for practical application. First-order logic and semantic networks are used for ontology representation, along with semantic mapping methods [5].

The article by Zhong and co-authors presents a review of the application of ontologies in the construction industry based on an analysis of Scopus publications (2007-2017). Scientometric methods, including co-authorship and cluster analysis, identified four key themes: domain ontologies, Industry Foundation Classes, automated compliance checking, and BIM. The study revealed the evolution of keywords from project management to modeling and compliance control [6].

The need to develop specialized ontologies for representing the content of regulatory documents is evident, at least because almost all upper-level and domain ontologies are descriptive, meaning they focus on describing existing instances of objects and processes in the world. In contrast, legal ontologies are prescriptive, describing the proper state in which certain types of objects should exist in specific situations.

Research Goal and Objectives

Ontological modeling has gained the attention of the scientific community as a tool for the formalization and automation of regulatory document processing, which is important for both governmental and business processes. The goal of this article is to study approaches to ontology development and assess their potential for automating the control of regulatory compliance. To achieve this goal, the following objectives were set:

- Conduct a review of existing ontology development methodologies applicable to regulatory documents.

- Analyze the potential use of ontologies for automating compliance control with regulatory document requirements.

- Explore modern Natural Language Processing (NLP) methods, including large language models (LLMs), to assess their applicability for working with regulatory information.

- Identify prospects for the further development of ontological modeling and NLP technologies to address automation tasks in governmental and business processes.

Development of Automation Tools for Regulatory Document Content Processing

In the 2000s-2010s, approaches to developing legal ontologies for analyzing regulatory document texts were actively explored. The collection "Approaches to Legal Ontologies" [7] discusses methods and techniques in this field. The chapter "A Complex-System Approach" [8] describes a method for building ontologies that includes defining key terms, establishing their hierarchy, and analyzing semantic relationships between them. This approach is based on the empirical analysis of term usage without formal logical text analysis, assuming a consistent meaning of terms throughout the document.

The chapter "The Multi-Layered Legal Information Perspective" [9] addresses the challenges of ontological formalization of regulatory documents, including semantic drift during translations, legislative evolution, and differences in term usage at national and supranational levels. The example of EU legislation implementation illustrates the complexities that arise even within documents issued by the same authority. The authors emphasize the need to create multiple ontologies for different scenarios and ensure their interoperability.

The chapter "Legal Ontologies: The Linguistic Perspective" [10] presents a classification of elements for the semantic description of regulatory statements: metadata (authority, number, date), pragmatic classification (obligations, prohibitions, permissions), and syntactic structures with logical operators ("if", "and", "otherwise"). The role of deontic modality ("may", "must") and the dependence of concepts on the context of legal acts are emphasized. The authors highlight the challenges of automating meaning extraction due to differences in term interpretation across various documents.

Large Language Models and NLU

Ontological modeling and machine learning (ML) represent different approaches to automation: ontologies model logical reasoning, while ML models empirical reasoning. Ontologies provide explainable outcomes but re-

quire significant development effort, whereas ML results are difficult to interpret due to their probabilistic nature. Large language models (LLMs), such as BERT and GPT, have improved text processing through contextual word representations. Despite their success in logical tasks, LLMs require substantial computational resources. A promising direction is their integration with knowledge graphs and logical reasoning to automate regulatory compliance control.

The authors of Logic-LM [11] proposed an approach where LLMs convert textual descriptions of situations into sets of logical instructions for decision-making automation. LLMs generate ad hoc ontologies with predicates and classes to describe problem domains. The use of such models simplifies task automation but requires high computational resources. Available models, such as LLaMA, enable the community to use them in projects. Promising directions include using embeddings to analyze the semantic similarity of terms and QA models to extract structured information from regulatory texts.

Deontic Logics and Rule Representation in Ontologies

Legal norms express judgments about permissions, prohibitions, and obligations, for which deontic logics – a subset of modal logics – are used. A key element is the concept of subjunctive betterness, which evaluates the preference of possible worlds with different configurations of objects and processes.

M. Vacura proposes a deontic logic model where the central element is a deontic concept based on a normative principle from a legal source [12]. The concepts are interrelated: the obligation of one subject may be the right of another. The model describes the behavior of agents in situations with specific conditions ("states of affairs"). Physical objects, such as documents, can embody rights and obligations, for example, tickets or prescriptions.

Legal norms can be formalized using rules, for example, through SHACL (W3C) constraints, which allow for the representation and validation of ontological models with reasoner support. A more practical approach is implemented in LegalRuleML (OASIS), which enables the integration of legal norms into RDF/OWL ontological models for efficient legal data processing. LegalRuleML describes the structure of norms and their interconnections. L. Robaldo proposed using SHACL to structure norms based on deontic logics, including "reified input/output logic", which is suitable for the automatic control of regulatory compliance by extracting data from natural language texts.

LegalRuleML (OASIS) is an extension of RuleML for representing legal norms with support for integration into RDF/OWL ontologies. The specification describes the structure of norms and their interconnections: The References section contains links to regulatory documents, the Context section provides rule metadata, including conditions of application, and the Statements block forms logical constructs with conditions (if) and conclusions (then).

Upper-Level Ontologies and Domain Ontologies

The BFO ontology, including IAO for informational objects, is widely used for modeling processes and objects. An alternative is DOLCE, which relies on natural language categories, such as "Proposition", but its complex structure complicates modeling. For describing term hierarchies, SKOS is recommended, as it allows grouping concepts by meaning for analyzing semantic differences.

Among domain ontologies, the IFC library is important for describing building structures, but it does not cover aspects such as ergonomics and the natural environment. SKOS can enhance the formalization of norms by linking them to IFC elements. The inclusion of common knowledge ontologies helps describe spatial relations, but creating a complete model of all descriptions is not necessary, which reduces the workload.

Ontologies for Representing Regulatory Information

Among the existing core ontologies for representing regulatory information is the Core Legal Ontology, based on the upper-level ontology Dolce Ultra Lite. This ontology provides classifications of agent roles, types of regulatory norms, and sets a semantic framework for concepts and their definitions, as well as various types of statements (constraints, descriptions, etc.). However, despite its detailed description of specific types of statements, this ontology is not universal for a broad range of norms, making it somewhat eclectic and not always suitable for extensive applications. The use of this model in Natural Language Understanding (NLU) systems would require significant effort to determine the suitability of its elements for specific legal texts.

Another example is the LKIF Core Ontology of Basic Legal Concepts, which includes modules for describing processes, roles, actions (including legal actions), as well as norms and rules. This model also provides classifications of statements, modalities, and agent roles, making it highly detailed for the analysis of legal documents. However, its high level of detail complicates the automation of semantic analysis and is more suited for analysts creating models for specific regulatory acts. The main purpose of LKIF is to be used in decision

support systems based on logical reasoning, although the formalization of rules requires a significant amount of manual work.

The rule syntax of LKIF, based on Horn clauses, allows the formalization of legal statements, but the labor-intensive nature of this process makes it unsuitable for efficient automation. The ontologies under consideration are outdated and focus on the roles of agents, which limits their applicability to regulatory documents in the construction industry, where the emphasis is on requirements for physical elements.

Despite the absence of a universal ontology for the construction industry, elements of the listed models can be adapted to create a specialized ontological model.

Results and Discussion

Automated control of legal norms involves two tasks: compliance control and meta-processing of norms (classification, structuring, search, change management). Compliance control requires a detailed domain model and the formalization of norms as rules. For meta-processing, a meta-description of norms and their contexts without deep detailing is sufficient.

A key tool for such tasks is the use of ontological specifications, such as RDF/OWL and SHACL, recommended by the W3C consortium. However, these tools remain labor-intensive, requiring significant resources for the manual formalization of norms and situations. Despite significant advancements in Natural Language Understanding (NLU), systems that existed before the 2020s were unable to fully automate the process.

The emergence of large language models (LLMs) has opened new prospects for automation; however, their reasoning capabilities remain limited. At the current stage of LLM development, these models are not always able to generate correct answers to complex queries, as they can "hallucinate" or make errors when processing contradictory data. A logically complete model of norms and situations using LLMs remains unattainable.

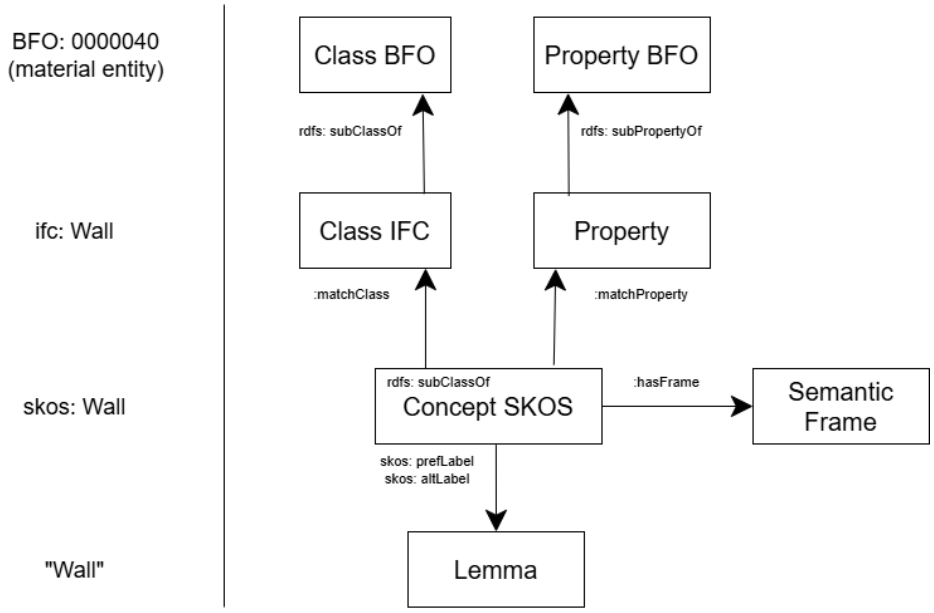
The most promising direction for future research is the combination of approaches. This involves creating a logically coherent model of norms or a meta-model using NLU and LLMs for the automatic generation of models based on regulatory texts.

All existing approaches to modeling regulatory information share several important aspects: the need for the classification of modalities of regulatory statements (such as necessity, possibility, prohibition), the classification of subjects of regulatory requirements, and a conceptual model for describing the situations to which the norms apply.

Conclusion

Based on the conducted review, the requirements for a regulatory document processing system in construction using NLU were formulated. The information support includes:

- An upper-level ontology (e.g., BFO) for general concepts.
- A domain ontology for the construction industry based on IFC.
- An ontology for structuring statements and elements of deontic logic.



Relationship of elements in the proposed ontologies

- A terminology ontology based on the SKOS model, taking into account contexts and correspondence to elements of the domain and upper-level ontologies.

The proposed approach involves dividing the text of regulatory documents into term definitions and norms, with the creation of a "semantic profile" for each norm (Figure). This "profile" should include a subject, a predicate with modality, an object, as well as additional elements (circumstances, definitions, etc.).

The formalization should be performed at the meta-level without fully detailing each situation where the norms apply. The ad hoc concepts created during the process, which are absent from the model, will be represented by the lemma of the corresponding word. These concepts may be unique to a specific set of documents and will help enhance the model during its development.

REFERENCES

1. Honderich, T. (Ed.), The Oxford Companion to Philosophy.: Oxford University Press. Retrieved 15 Oct. 2024, from DOI: https://www.oxfordreference.com/view/10.1093/acref/9780199264797.001.0001/acref-9780199264797.
2. Calero, C., Ruiz, F., & Piattini, M. (2010). Ontologies for Software Engineering and Software Technology., 339-339. https://doi.org/10.1007/3-540-34518-3.
3. Neuhaus, F. (2018). What is an ontology? arXiv. https://arxiv.org/abs/1810.09171v1
4. Otero-Cerdeira, L., Rodríguez-Martínez, F., & Gómez-Rodríguez, A. (2015). Ontology matching: A literature review. Expert Syst. Appl., 42, 949-971. https://doi.org/10.1016/J.ESWA.2014.08.032.
5. Park, J., Nam, J., Hu, Q., & Suh, H. (2008). Product Ontology Construction from Engineering Documents. 2008 International Conference on Smart Manufacturing Application, 305-310. https://doi.org/10.1109/ICSMA.2008.4505663.
6. Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H., & He, L. (2019). A scientometric analysis and critical review of construction related ontology research. Automation in Construction. https://doi.org/10.1016/J.AUTCON.2018.12.013.
7. Sartor, Giovanni, Casanovas, Pompeu, Biasiotti, Mariangela, Fernández-Barrera, Meritxell, Approaches to Legal Ontologies, Springer Science+Business Media B.V. 2011 https://doi.org/10.1007/978-94-007-0120-5
8. P. Mazzega, D. Bourcier, P. Bourgine, N. Nadah and R. Boulet, «A complex-system approach: legal knowledge, ontology, information and networks», In Approaches to Legal Ontologies. Theories, Domains, Methodologies. G. Sartor, P. Casanovas, M. Biasiotti and M. Fernández Barrera (Eds.), Springer, 2010, pp. 117-132. DOI:10.1007/978-94-007-0120-5_7
9. Boella, G., Rossi, P. (2011). The Multi-Layered Legal Information Perspective. In: Sartor, G., Casanovas, P., Biasiotti, M., Fernández-Barrera, M. (eds) Approaches to Legal Ontologies. Law, Governance and Technology Series, vol 1. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0120-5_8
10. Biasiotti, M.A., Tiscornia, D. (2011). Legal Ontologies: The Linguistic Perspective. In: Sartor, G., Casanovas, P., Biasiotti, M., Fernández-Barrera, M. (eds) Approaches to Legal Ontologies. Law, Governance and Technology Series, vol 1. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0120-5_9
11. Pan, L., Albalak, A., Wang, X., & Wang, W. Y. (2023). Logic-LM: Empowering large language models with symbolic solvers for faithful logical reasoning. https://arxiv.org/abs/2305.12295
12. Vacura, M. (2019). Towards New Ways of Modeling Deontic Concepts in Ontologies. JOWO. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://ceur-ws.org/Vol-2518/paper-CAOS6.pdf

Нормативтік құжаттарды онтологиялық модельдеу: автоматтандыру тәсілдері мен перспективаларын талдау

[1]КАБЖАН Зарина Еркебулановна, *магистр, бөлім меңгерушісі, Kabzhan90@gmail.com,*
[1]ШАХНОВИЧ Александр Юльевич, *т.ғ.к., даму және цифрландыру жөніндегі басқарушы директор, gipocentr@gmail.com,*
[1]*ШОГЕЛОВА Назым Тулегеновна, *магистр, маман, nazymshogelova@gmail.com,*
[1]ГЛЫЗНО Евгения Дмитриевна, *магистр, аға қызметкер, e.glyzno@kazniisa.kz,*
[2]ГОРШКОВ Сергей, *техникалық директор, serge@datavera.kz,*
[1]«Қазақ құрылыс және сәулет ғылыми-зерттеу және жобалау институты» АҚ, 3 шағын ауданы, 44а, Алматы, Қазақстан,
[2]«Датавера» ЖШС, Радостовец көшесі, 323а, Алматы, Қазақстан,
**автор-корреспондент.*

Аңдатпа. *Құрылыс саласындағы нормативтік құжаттарды формализациялау және өңдеу процесін автоматтандыру үшін онтологиялық модельдеуді қолдану мәселесі зерттеледі. Жұмыстың негізгі мақсаты – нормативтік актілерді бақылауға арналған автоматтандырылған жүйені құру үшін әдістемелік және технологиялық база әзірлеу. Онтологияларды құру әдіснамалары, соның ішінде жоғары деңгейлі онтологиялар (мысалы, Basic Formal Ontology) және пәндік онтологиялар (Industry Foundation Classes) қарастырылады. SKOS сияқты білімді ұйымдастыру модельдерін қолдану арқылы нормативтік құжаттардағы қайталаулар мен қайшылықтарды анықтау тәсілдері сипатталады. Нормативтік тұжырымдарды «семантикалық портреттерін» құру арқылы талдау процесін автоматтандыруға ерекше назар аударылады. Табиғи тілді өңдеу технологиялары мен ірі тілдік модельдерді қолдану талдаудың тиімділігін арттыру үшін ұсынылады. Қазіргі технологияларды дамыту кезеңінде алынған нәтижелерді тексеру үшін сарапшылардың қатысуы қажеттілігі атап өтіледі.*

Кілт сөздер: *онтологиялық модельдеу, нормативтік құжаттар, құрылыс саласы, NLP, автоматтандыру, семантикалық қайшылықтар, машиналық оқыту.*

Онтологическое моделирование нормативных документов: анализ подходов и перспективы автоматизации

[1]КАБЖАН Зарина Еркебулановна, *магистр, зав. отделом, Kabzhan90@gmail.com,*
[1]ШАХНОВИЧ Александр Юльевич, *к.т.н., управляющий директор по развитию и цифровизации, gipocentr@gmail.com,*
[1]*ШОГЕЛОВА Назым Тулегеновна, *магистр, специалист, nazymshogelova@gmail.com,*
[1]ГЛЫЗНО Евгения Дмитриевна, *магистр, старший сотрудник, e.glyzno@kazniisa.kz,*
[2]ГОРШКОВ Сергей, *технический директор, serge@datavera.kz,*
[1]АО «Казахский научно-исследовательский и проектный институт строительства и архитектуры», 3 микрорайон, 44а, Алматы, Казахстан,
[2]ТОО «Датавера», ул. Радостовца, 323а, Алматы, Казахстан,
**автор-корреспондент.*

Аннотация. *Исследуется использование онтологического моделирования для формализации и автоматизации обработки нормативных документов в строительной отрасли. Основной целью работы является разработка методологической и технологической базы для создания автоматизированной системы контроля нормативных актов. Рассматриваются методологии построения онтологий, включая верхнеуровневые онтологии (например, Basic Formal Ontology) и предметные онтологии (Industry Foundation Classes). Описываются подходы к выявлению дублирования и противоречий в нормативных документах с помощью моделей организации знаний, таких как SKOS. Особое внимание уделяется автоматизации анализа нормативных высказываний путем построения их «семантических портретов». Предлагается применение технологий обработки естественного языка и больших языковых моделей для повышения эффективности анализа. Подчеркивается необходимость участия экспертов на текущем этапе развития технологий для проверки полученных результатов.*

Ключевые слова: *онтологическое моделирование, нормативные документы, строительная отрасль, NLP, автоматизация, семантические противоречия, машинное обучение.*

REFERENCES

1. Honderich, T. (Ed.), The Oxford Companion to Philosophy.: Oxford University Press. Retrieved 15 Oct. 2024, from DOI: https://www.oxfordreference.com/view/10.1093/acref/9780199264797.001.0001/acref-9780199264797.
2. Calero, C., Ruiz, F., & Piattini, M. (2010). Ontologies for Software Engineering and Software Technology., 339-339. https://doi.org/10.1007/3-540-34518-3.
3. Neuhaus, F. (2018). What is an ontology? arXiv. https://arxiv.org/abs/1810.09171v1
4. Otero-Cerdeira, L., Rodríguez-Martínez, F., & Gómez-Rodríguez, A. (2015). Ontology matching: A literature review. Expert Syst. Appl., 42, 949-971. https://doi.org/10.1016/J.ESWA.2014.08.032.
5. Park, J., Nam, J., Hu, Q., & Suh, H. (2008). Product Ontology Construction from Engineering Documents. 2008 International Conference on Smart Manufacturing Application, 305-310. https://doi.org/10.1109/ICSMA.2008.4505663.
6. Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H., & He, L. (2019). A scientometric analysis and critical review of construction related ontology research. Automation in Construction. https://doi.org/10.1016/J.AUTCON.2018.12.013.
7. Sartor, Giovanni, Casanovas, Pompeu, Biasiotti, Mariangela, Fernández-Barrera, Meritxell, Approaches to Legal Ontologies, Springer Science+Business Media B.V. 2011 https://doi.org/10.1007/978-94-007-0120-5
8. P. Mazzega, D. Bourcier, P. Bourgine, N. Nadah and R. Boulet, «A complex-system approach: legal knowledge, ontology, information and networks», In Approaches to Legal Ontologies. Theories, Domains, Methodologies. G. Sartor, P. Casanovas, M. Biasiotti and M. Fernández Barrera (Eds.), Springer, 2010, pp. 117-132. DOI:10.1007/978-94-007-0120-5_7
9. Boella, G., Rossi, P. (2011). The Multi-Layered Legal Information Perspective. In: Sartor, G., Casanovas, P., Biasiotti, M., Fernández-Barrera, M. (eds) Approaches to Legal Ontologies. Law, Governance and Technology Series, vol 1. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0120-5_8
10. Biasiotti, M.A., Tiscornia, D. (2011). Legal Ontologies: The Linguistic Perspective. In: Sartor, G., Casanovas, P., Biasiotti, M., Fernández-Barrera, M. (eds) Approaches to Legal Ontologies. Law, Governance and Technology Series, vol 1. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0120-5_9
11. Pan, L., Albalak, A., Wang, X., & Wang, W. Y. (2023). Logic-LM: Empowering large language models with symbolic solvers for faithful logical reasoning. https://arxiv.org/abs/2305.12295
12. Vacura, M. (2019). Towards New Ways of Modeling Deontic Concepts in Ontologies. JOWO. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://ceur-ws.org/Vol-2518/paper-CAOS6.pdf