Knowledge Graphs in the Age of Language Models and Neuro-Symbolic AI
A. Salatino et al. (Eds.)

doi:10.3233/SSW240004

Leveraging Semantic Model and LLM for Bootstrapping a Legal Entity Extraction: An Industrial Use Case

Julien BRETON [a,b,1], Mokhtar Boumedyen BILLAMI [b], Max CHEVALIER [a] and Cassia TROJAHN [a]

[a] *Informatics Research Institute of Toulouse (IRIT), Toulouse, France*
[b] *Berger-Levrault, Toulouse, France*

ORCiD ID: Julien Breton https://orcid.org/0000-0001-8285-0399, Mokhtar Boumedyen Billami https://orcid.org/0000-0003-4428-4298, Max Chevalier https://orcid.org/0000-0001-5402-6255, Cassia Trojahn https://orcid.org/0000-0003-2840-005X

Abstract. Compliance with legal documents related to industrial maintenance is the company's obligation to oversee, maintain, and repair its equipments. As legal documents endlessly evolve, companies are in favour of automatically processing these texts to facilitate the analysis and compliance. The automatic process involves first, in this pipeline, the extraction of legal entities. However, state-of-the-art, like BERT approaches, have so far required a large amount of data to be effective. Creating this training dataset however is a time-consuming task requiring input from domain experts. In this paper, we bootstrap the legal entity extraction by levering Large Language Models and a semantic model in order to reduce the involvement of the domain experts. We develop the industrial perspective by detailing the technical implementation choices. Consequently, we present our roadmap for an end-to-end pipeline designed expressly for the extraction of legal rules while limiting the involvement of experts.

Keywords. Legal Entity Extraction, Semantic Model, Large Language Models

1. Introduction

The legal industry is characterized by an extensive volume of evolving documents, such as contracts, legislation, court rulings, and regulatory filings. These documents are dense, complex, and rich in specialized language, making their analysis and processing both time-consuming and prone to human error. Automatic processing of such documents is essential for several applications, including legal research, compliance monitoring, contract analysis, and case preparation. It not only accelerates their analysis but also improves compliance, accuracy, and accessibility of legal information. Moreover, as laws and regulations frequently change, automated systems ensure that analyses are up-to-date with the latest legal standards. As an example, in December 2021, the CNIL (French Na-

[1] Corresponding Author: Julien Breton, julien.breton@irit.fr

tional Commission for Computing and Liberties)[2] has applied a record penalty against Google of 150 million euros for non-compliance with the law. This example demonstrates that companies are required to comply with the law and, if they fail to do so, they will face severe penalties. As highlighted in [1], in France, there are “more than 10,500 laws, 120,000 decrees, 7,400 treaties, 17,000 community’s texts, tens of thousands of pages in 62 different codes. Some are constantly being modified: 6 modifications per working day for the 2006 Tax Code”.

Legal entity extraction is a backbone task in the automation process, involving the identification and categorization of entities. By automatically extracting entities, legal professionals can quickly locate relevant information, understand the relationships between different legal entities, and extract insights from large volumes of text. Different approaches for legal entity extraction have been proposed in the literature, including rule-based systems, Bi-directional Long Short-Term Memory (Bi-LSTM) networks [2], and BERT (Bidirectional Encoder Representations from Transformers) models [3]. Each of these methods has its own set of advantages and application contexts. Rule-based systems, due to predefined rules and patterns identified by experts, are highly interpretable but often less flexible, while BERT models leveraging transformer architectures [4], excel at automatically understand the context. However, a common challenge across all these techniques is their dependency on large amounts of annotated data required for training. Annotated datasets are crucial for training these models to the specific tasks they need to perform, but the creation of such datasets involves manually labelling text with the correct annotations. A process that is both time-consuming and resource-intensive. This dependency on extensive annotated data limits the scalability of deploying these models, especially in domains where annotated data is scarce or expensive to produce.

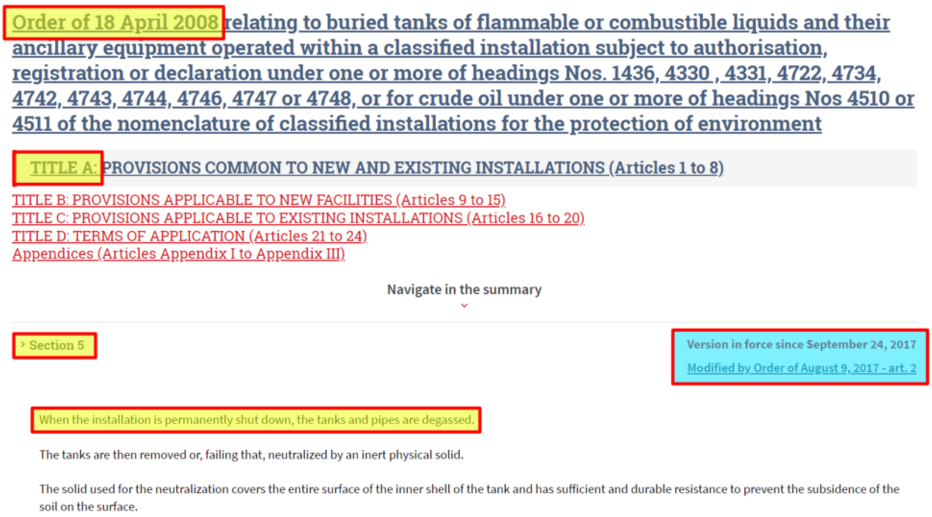
This paper bootstraps the legal entity extraction by leveraging a Large Language Model (LLM) and a semantic model in order to reduce the involvement of the domain experts. Our objective is to encapsulate expert knowledge within a semantic model and leverage the capabilities of the foundational model GPT-4 to extract legal entities based on this semantic framework. We develop the industrial perspective by detailing the technical implementation choices. Consequently, we present our roadmap for an end-to-end pipeline designed expressly for the extraction of legal rules while limiting the involvement of experts.

The rest of this paper is organized as follows. Section 2 introduces a motivated example. Section 3 discusses the main related works. Section 4 presents the semantic model followed by the pipeline in Section 5. Section 6 concludes the paper and discusses directions for future work.

2. Motivated Example

Figure 1 presents an example of a document from the French governmental website Légifrance, translated using Google Translate. This document specifically addresses the regulations concerning companies that manage buried tanks, with an emphasis on the requisite checks and inspections. As demonstrated in Figure 1, "Section 5" of the document includes cross-reference; it mentions a modification by the *Order of August 9, 2017 - art. 2* and directs readers to relevant provisions. These cross-references between docu-

[2] https://www.cnil.fr/



Order of 18 April 2008 relating to buried tanks of flammable or combustible liquids and their ancillary equipment operated within a classified installation subject to authorisation, registration or declaration under one or more of headings Nos. 1436, 4330 , 4331, 4722, 4734, 4742, 4743, 4744, 4746, 4747 or 4748, or for crude oil under one or more of headings Nos 4510 or 4511 of the nomenclature of classified installations for the protection of environment

TITLE A: PROVISIONS COMMON TO NEW AND EXISTING INSTALLATIONS (Articles 1 to 8)

TITLE B: PROVISIONS APPLICABLE TO NEW FACILITIES (Articles 9 to 15)
TITLE C: PROVISIONS APPLICABLE TO EXISTING INSTALLATIONS (Articles 16 to 20)
TITLE D: TERMS OF APPLICATION (Articles 21 to 24)
Appendices (Articles Appendix I to Appendix III)

Navigate in the summary

› Section 5

Version in force since September 24, 2017
Modified by Order of August 9, 2017 - art. 2

When the installation is permanently shut down, the tanks and pipes are degassed.

The tanks are then removed or, failing that, neutralized by an inert physical solid.

The solid used for the neutralization covers the entire surface of the inner shell of the tank and has sufficient and durable resistance to prevent the subsidence of the soil on the surface.

Figure 1. Example of an ‘Order’ from Légifrance translated in English.

ments are abundant and highlight the complexity of legal regulations, illustrating that the law is a continuously evolving. Consequently, in our work, we place significant emphasis on managing versions of the law, and in Section 4.2, we describe the corresponding solutions.

The perspective of automating the extraction and synthesis of relevant information from such documents is significant. In this specific case, the extracted information could enhance understanding of the regulations outlined in "Section 5". For instance, when a company employs a Computerized Maintenance Management System (CMMS) to service a tank, the system can identify the prerequisites for the shut-down procedure. This ability to automatically process legal rules enables the CMMS to recommend relevant actions, such as the obligation to degas the tank, like those found in the order of March 18, 2008, "Section 5".

Breaking down the legal rules extraction task, it can be viewed as comprising several subtasks:

1. Extraction of legal entities: This involves identifying and retrieving relevant instances of legal concepts from complex and structured legal documents.
2. Extraction of semantic relations: Linking legal entities together to generate a formalize model
3. Structuring of Information: Post-extraction, this step involves organizing the data into a formal format that is computational ready.

The forthcoming sections of this paper will delve deeper into the first task, offering a detailed methodology. We aim to provide clarity on how the legal entities extraction can be efficiently implemented, thereby reducing the workload for legal and technical experts tasked with compliance. To illustrate our method and the efficacy, we will later revisit and explain the highlighted sentence in Figure 1, applying our proposed techniques to demonstrate practical applications and outcomes.

3. Related Work

3.1. Information extraction

Information extraction is a broad field with multiple contributions from the scientific community. Recently, there has been a significant trend toward employing neural networks for labelling and classification tasks to extract relevant elements. For example, David B. et al. [5] developed a system to anonymize named entities in German financial documents. This system identifies entities such as first and last names, postal and email addresses, and locations. The study evaluated various architectures, including Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM) [6], and Conditional Random Fields (CRF) [7].

Transformers technology, introduced by Vaswani et al. [4], has surpassed many existing models. This architecture has proven effective for extracting named entities [8] and structuring texts into knowledge graphs [9]. The legal community has also adopted this architecture for Named Entity Recognition (NER), as demonstrated in Italian judgments [10] and Brazilian legal texts by Wang et al. [11]. These implementations achieved impressive results by combining BERT, Bi-LSTM, and CRF. However, all these methods require a substantial amount of data to train these architectures or time from the experts, a challenge we address in this article.

As described in the survey by Solihin et al. [12], the legal community has primarily focused on the task of named entity recognition (NER), with research and datasets dedicated to legal entity linking remaining scarce. Although these two tasks might appear similar at first glance, but they differ significantly. Named entities, such as persons or organizations, are typically represented as single entities and are usually short, comprising a few words. In contrast, legal concepts, detailed in section 4.2, can include broader concepts, named entities, or even complete sentences. Therefore, there is reasonable doubt that approaches effective for named entities [13,14,15,16] will be equally effective for legal entity extraction.

However, notable studies such as those by Sleimi et al. [17] have created their own datasets of legal entities and performed extraction using a rule-based approach. Other work, such as Dragoni et al. [18], prefer the name "terms extraction" when referring to the extraction process of legal rules components. Additionally, recent research by Castano et al. [19] explores a similar process by extracting both concepts and terms from European legal documents, which are then integrated and maintained in a knowledge management system.

A significant part of the work proposes an information extraction based on resources containing, in a structured way, the concepts of a domain as well as the relations between them. These resources are semantic models like, for example, ontologies or knowledge graphs. The semantic data model is a method for organizing data to represent it within a defined logical framework. It is a conceptual model that incorporates semantic information, thereby imparting essential meaning to the data and elucidating the relationships among them. The creation and the use of a semantic model has been the subject of different proposals in the literature. Munira A. et al. [20] proposed an Ontology-Based Information Extraction (OBIE) system with the objective of extracting, from textual documents, the land suitability for residential use. In the domain of industrial maintenance, [21] developed a system that relies on a semantic model and that allows managing the maintenance assets in industry. In the following section, we introduce the semantic models in the domain of the legal maintenance. We will, at least, present the two main domains related to the legal maintenance: the law and the industrial maintenance.

3.2. Semantic models

3.2.1. Semantic models related to the law.

One of the early works in semantic modelling related to the law [22] created the “Frame” model, which aims to structure legal rules. Many of its concepts will be found in the further work. Van E. et al. [23] detail two of these models dedicated to the representation of legislation: FOLaw and LRI-Core. LRI-Core has been used as a high-level ontology for the modelling of German administrative laws. LKIF [24] is an open source ontology alternative to LRI-Core which can be applied on multi-domain representation of the law in order to facilitate its reuse. This ontology contains for example the notion of “Right” which characterizes the permission, obligation or prohibition to perform an act according to the law. LegalRuleML [25], an XML standard for the legal domain, has been inspirited by LKIF to represent the legal knowledge and legal reasoning.

The LKIF ontology has been constructed via a supervised approach in order to manually build a semantic model [26]. While most works have considered the construction of the models in a manual manner, several alternative approaches have considered the construction using automatic approaches on large corpus [27].

Beyond structuring knowledge semantically, other works focus on mathematical formalizations using, e.g., deontic logic rules. Propagated in the scientific community by [28], this formalization of philosophical concepts relies on symbolic reasoning with notions of modality (prohibition, permission, obligation).

3.2.2. Semantic models for industrial maintenance.

Semantic models dedicated to legal maintenance have also been addressed in scientific research. In 2004, Rasovska I. et al. [29] proposed a system composed of a conceptual model allowing to make decisions related to the industrial maintenance. A few years later, the IMAMO (Industrial MAintenance Management Ontology) [21] has been published. IMAMO is an ontological model with the objective of standardizing information exchanges related to maintenance. It aims at ensuring semantic interoperability while generating data to be used as a decision-making support. Many other works reflect this structure life cycle for industrial equipment [30,31,32,33,34]. We find in these models essential concepts for industrial maintenance. For example, the notion of ‘Maintainable activity’, ‘Maintainable Item’ or ‘Qualified Person’ that group different concepts specific to maintenance which are not mentioned in the legal models. Indeed, these concepts are extracted form an open source model [35] produced by the IOF group (Industrial Ontologies Foundry) which tries to "*create a set of reference and open ontologies covering the whole industry domain*"[3].

[3] https://www.industrialontologies.org/our-mission/

4. Formalizing legal maintenance with a semantic model

4.1. Why using a semantic model?

In the domain of legal formalization, the adoption of different methodologies reflects the diverse requirements and challenges faced by the scientific community. The first major approach centres around formal logic, leveraging systems like deontic logic or the PROLEG language [36]. These systems are structured to facilitate rule-based reasoning, enabling automated processes such as compliance verification. This capability is crucial in legal environments where consistency and respect to explicit regulations are paramount. However, the inherent complexity and specialized nature of formal logic systems make them less accessible to those without a background in this area, potentially limiting their usability in interdisciplinary applications.

On the other hand, the second approach employs semantic models, including ontologies and knowledge graphs, to encapsulate legal knowledge. These models strive to represent legal concepts and relationships in a way that is both semantically rich and intuitively understandable. This approach enhances transparency and eases communication of legal rules, making them more accessible to non-specialists. For instance, legal practitioners without technical expertise can interact with and contribute to the model, facilitating broader collaboration. However, while semantic models excel in knowledge representation, they often lack the native capability to perform complex reasoning tasks directly. This limitation necessitates additional computational mechanisms or integration with other technologies to enable practical reasoning processes.

In this paper, we opted for a semantic model because our approach requires validation by maintenance legal experts who are not familiar with formal logic. Our objective was to design a pipeline that enables legal experts to comprehend and modify the extracted information. Although the comprehensibility of the model may limit its computational and reasoning capabilities. The motivation to produce a comprehensible model was also driven by the need to provide a model that accurately represents legal maintenance rules. In the next section, we will revisit the model, discussing the purpose of this semantic model and the reasons behind our decision to develop SEMLEG v2.

4.2. SEMLEG for legal maintenance

As detailed in the previous section, we developed a semantic model to formalize legal maintenance rules. Our inaugural version[4], aimed to address a gap in the existing state of the art. At that time, there were semantic models designed specifically for industrial maintenance and others for legal contexts, but none that encompassed both aspects. In the forthcoming sections, we will detail the existing features of the model and the new improvements, illustrating these enhancements with examples that highlight previously unaddressed issues. Figure 2 provides an overview of the updated semantic model. In the following section, we will revisit the core features that have remained unchanged between the two versions.

[4] SEMLEG v1

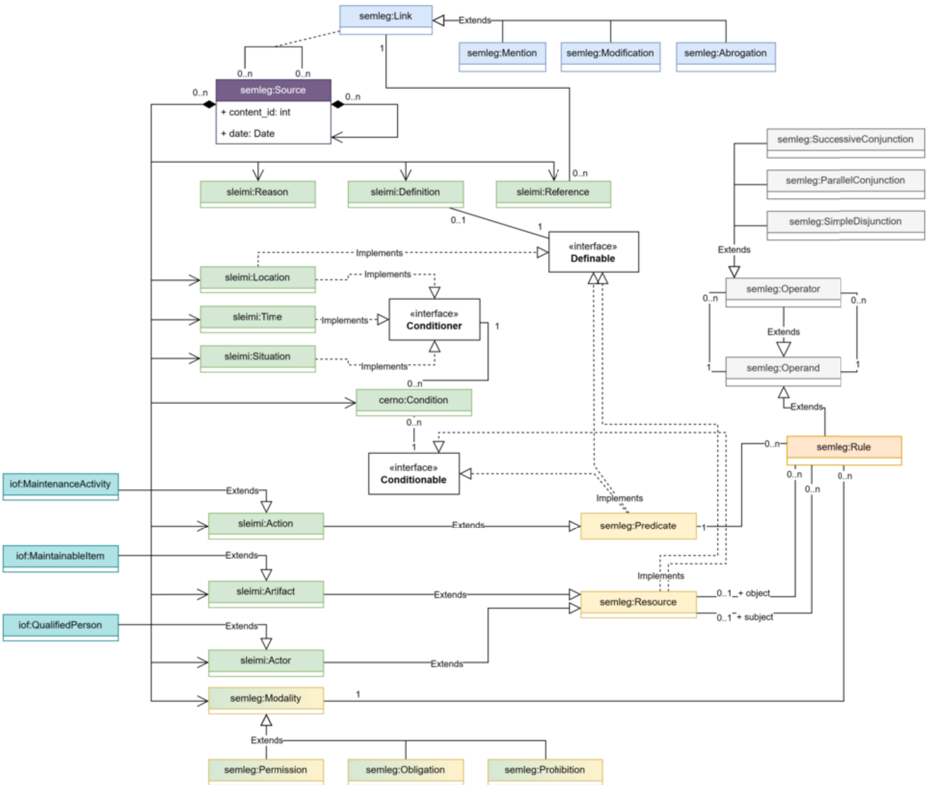


Figure 2. Semantic Model designed to formalize legal maintenance rules.

4.2.1. A quick look at SEMLEG v1

The development of the semantic model was structured into two primary phases: (1) the extraction of three representative industrial maintenance decrees from Légifrance, covering different domains: pressured equipment[5], lifting machines[6], and buried tanks[7]; (2) the identification of recurring components within the decrees and the collaborative construction of the semantic model by four experts, each specializing in a distinct field: (a) industrial legal maintenance, (b) ontological knowledge modelling, (c) generalized information systems, and (d) automated natural language processing.

The legislative hierarchy is modelled through a series of nested groups, which facilitate the structuring of legal information. For instance, in Figure 1, "Section 5" is nested within Title A. The various components such as chapters, sections, and articles are modelled using the concept **Source**, and their hierarchical relationships are defined by the composition relation. To interconnect these sources, as exemplified by the blue box in Figure 1, we introduced the concept **Link**. This allows one source to abrogate, mention, or modify another legal source. This feature allows SEMLEG to track history and evolu-

[5] https://www.legifrance.gouv.fr/loda/id/JORFTEXT000036128632/2022-06-01/
[6] https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000439029/2022-07-01/
[7] https://www.legifrance.gouv.fr/loda/id/JORFTEXT000018820571/2022-06-01/

tion of legal rules. For example, in Figure 3, the order from 2017, article 2 modified the "section 5" from 2008 changing underlying rules.

In the model, a rule is characterized by four components: a **Subject**, a **Predicate**, an **Object**, and a **Modality**. This framework is heavily inspired by the semantic triple, making it well-suited for the formalization of legal rules. We have introduced **Modality** to capture the granularity of the rule, which can be categorized as a **Permission**, an **Obligation**, or a **Prohibition**. This addition enhances the model's ability to accurately represent the nuances of legal mandates. Some **Modality**, like the one in our example, can be implicit due to the lack of verbs.

Similar to mathematical operations, we have introduced **Operator** and **Operand**, which facilitate reasoning about the sequence and interrelationship between rules. These concepts fill in the gaps in the semantic model, particularly in terms of reasoning in relation to formal logic. To date, we have identified three distinct operators: Simple Disjunction, Parallel Conjunction, and Successive Conjunction.

- **simpleDisjunction** is represented as: "I do A <u>or</u> B," indicating a choice between two actions.
- **parallelConjunction** is represented as: "I do A <u>and at the same time</u> B," implying that two actions occur simultaneously.
- **successiveConjunction** is represented as: "I do A <u>and only after</u> B," suggesting that one action follows the other in sequence.

These operators enable the structuring of rule chains (the operands) within legislative texts, providing a clear framework for understanding the logic and flow of legal stipulations.

4.2.2. New features introduced in SEMLEG v2

Since this first version, we improved SEMLEG by applying modification based on uncovered cases from new legal documents. The introduction of new features will be explained using an example in Figure 3 extracted from the sentence in Figure 1: «When the installation is permanently shut down, the tanks and pipes are degassed. ».

The first major enhancement in SEMLEG v2 concerns the legal concepts it employs. We opted to align our model with the work of Sleimi et al. [17], who compiled various semantic models dedicated to legal domains and proposed a unified model. This unified model incorporates well-known frameworks such as LegalRuleML [25], LKIF [24], and Cerno [37]. Therefore, we introduced new legal concepts, which are highlighted in green in the Figure 2. The Table 1 gives a summary about all the different concepts from the semantic model in Figure 2. For instance, prior to this update, both **Artifact** and **Actors** were subsumed under the category **Resource**. Now, our model distinguishes these elements more precisely, enhancing the legal representation. Additionally, utilizing a composition relation with **Source** allow storing all extracted entities from our pipeline even if they are not yet linked to a rule. In our example, «Installation » is categorized as **Maintenable Item** but never involved in a rule.

Since the introduction of the concepts **Artifact**, **Action**, and **Actor** into SEMLEG v2, we have achieved precise alignment with the IOF maintenance ontology. Consequently, we have defined **Maintenance Activity** as a subclass of **Action**, **Maintainable Item** as a subclass of **Artifact**, and **Qualified Person** as a subclass of **Actor**. The alignment with external resources was a crucial aspect of the initial version of our model

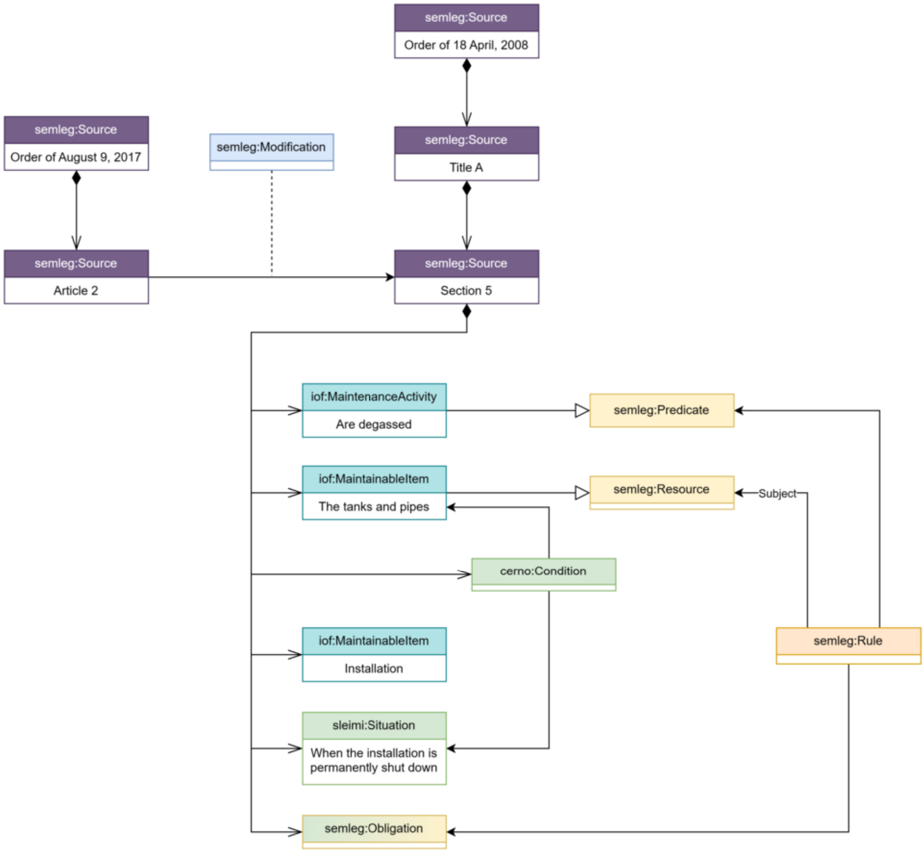


Figure 3. Example of SEMLEG v2 applied on the sentence : «When the installation is permanently shut down, the tanks and pipes are degassed. »

and remains vital for standardizing across computational systems. The incorporation of the IOF ontology will facilitate future integrations with existing systems that utilize this framework.

We also developed a new system for handling conditions within SEMLEG v2, which includes two interfaces: **Conditioner** and **Conditionable**. In this revised structure, **Predicate** and **Resources** are classified as **Conditionable**, meaning they can be conditioned by a **Conditioner**. **Conditioner** includes entities like **Location**, **Time**, and **Situation**. This modification allows for a more nuanced and flexible representation of conditions, aligning resources and predicates with specific contextual elements.

A similar process has been developed, enabling the definition of **Resource**, **Predicate**, and **Location**. Indeed, certain legal documents include definitions to clarify complex or specialized vocabulary. To date, definitions have been identified for these three concepts only. However, the modularity of the definable interface facilitates easy adaptation to new use cases.

Following our pipeline, after the formalization by domain experts and legal entities, we aimed to extract information with minimal involvement from experts using LLM.

The subsequent section explores the capabilities of LLM in zero-shot and few-shot configurations.

5. Bootstrap legal entity extraction by leveraging LLMs

5.1. Methodology and strategies

5.1.1. The dataset

In our study, the LLM processed input data consisting of raw sentences extracted from the Luxembourg Traffic Law. Utilizing various techniques, the LLMs are capable of identifying predefined legal entities. To provide an in-depth explanation of this process, we will begin with the dataset and describe the eight semantic concepts. The dataset we use in our work has been introduced in 2018 by Sleimi et al. [38], extracted from the Luxembourg traffic law. This study accomplishes the extraction of legal entities by employing rule-based methodologies. The authors report that their rule-based system is capable of achieving a precision score of 0.874 and a recall score of 0.855. However, achieving this level of performance necessitates significant time investment from experts in annotation process and rule based pattern creation. Furthermore, in a subsequent paper [17], they addressed the issue of limited generalizability by applying their approach to various legal codes, including the Code of Commerce, the Penal Code, the Code for Healthcare, the Labour Code, and the Code for the Environment. This application resulted in a significant reduction in precision, which declined by around 14.8%, indicating a significant decrease from the original domain to new domains.

In their paper [38], experts annotated 200 French selected statements from the Luxembourg Traffic Law and identified 1339 phrases. They focus on 14 legal concepts and publish the dataset[8]. However, in our paper, we'll work on a subpart of these concepts by using only 8 of them : Action, Actor, Object (Artifact), Condition, Location, Modality, Reference and Time. This selection was guided by our observation that certain concepts were either underrepresented in the dataset or did not align with our interpretation of their definitions.

A distinctive feature of this dataset and the associated tasks is the occurrence of entity overlap. Parts of a sentence may be annotated with two different concepts. For instance, in the Figure 3, the segment «installation »could be annotated as a Maintainable Item and as a Situation.

5.1.2. Prompt Engineering Pipeline

Prompt engineering is a subfield within artificial intelligence and human-computer interaction (HCI) focused on the design, analysis, and optimization of prompts, structured inputs or queries, to elicit desired responses or behaviours from AI systems, particularly those based on LLMs. This domain encompasses a range of activities, including the formulation of prompts to guide AI in legal entity extraction, thereby enhancing its performance [39,40].

Unlike conventional methods, prompt engineering does not modify the underlying architecture or weights of an AI model. Instead, it leverages the pre-existing knowledge

[8] https://sites.google.com/view/metax-re2018/

Table 1. Concepts and definitions in SEMLEG v2

Concept	Definition	Example
Source	Allows, like Russian dolls, to encapsulate a set of legal concepts or other sources.	Article L. 512-5
Link	Semantic link between two sources	
Mention	Mention of an other source	
Modification	Modification of an other source	
Abrogation	Abrogation of an other source	
Reason	The rational for an action	To improve the birth rate
Definition	Legal provision defining the meaning of concepts	A functional test of a lifting device is ...
Reference	Textual mention of another legal source	as defined in the Article L. 512-5
Location	A place where an action can be performed	In the park
Time	Moment, duration or occurrence of an action	Every two weeks
Situation	Description of something that has happened or can happen	When the installation is shut down
Condition	A constraint stating the properties that must be met	When the door is open
Action	The process of doing something	Playing rugby
Artifact	Material or immaterial object involved in an action	Basket ball
Actor	Entity that has the capability to act	The president
Maintenance Activity	Action of maintaining something	are degassed
Maintainable Item	Object that can be maintained	Lift
Qualified Person	Person who is qualified	The technician
Operand	Represents the elements on which a logical operation will operate.	
Operator	Represents the logical operation operator.	
Successive Conjunction	Represents performing one action after another.	I do A **and only after** B
Parallel Conjunction	Represents the performance of an action at the same time as another.	I do A **and at the same time** B
Simple Disjunction	Represents the completion of one action or another.	I do A **or** B
Predicate	Predicate in the meaning of the semantic triple	are degassed
Resource	Resource in the meaning of the semantic triple	Lifting devices
Modality	Represents the enforcement constraint of a rule	
Permission	The possibility to perform	Can
Obligation	Mandatory action to perform	Must
Prohibition	Forbidding an action to happen	Must not
Rule	Describe the rules and allows aggregating the necessary entities.	The periodic general verification of lifting devices must be done every twelve months.

and capabilities encoded within the model during its initial training phase, which typically involves exposure to vast datasets spanning diverse domains. By crafting prompts, the input queries given to the model can guide the AI to generate outputs that align more closely with specific user intentions or requirements.

The pre-prompt, available in our GitLab repository [9], is structured as follows: initially, a role is assigned to the LLM: "NLP expert", along with the task: "extracting entities from sentences". Indeed, the guidance provided in the OpenAI documentation[10] suggests that clearly defining the role and capabilities of a model, can significantly enhance the quality and relevance of the outputs generated. The second part entails a description of the eight chosen concepts from SEMLEG v2 and definitions created by domain experts. This instance represents the only necessary introduction of external knowledge, necessitating the involvement of an expert. Lastly, details regarding the output are provided, specifying that it should be in JSON format without further explanation and adhering to a non-rephrase constraint.

To construct the final prompt, one simply concatenates the pre-prompt with the sentence. This composite sentence is then provided to the LLM for processing. As noted in the dataset description, our sentences are derived from legal documents from Luxembourg, specifically focusing on traffic law. Here is an example of an added sentence, after the pre-prompt, from which to extract entities: "When one or more major or critical defects or non-conformities are found on a Luxembourg-registered vehicle, the roadworthiness inspector may decide that the vehicle must undergo a full roadworthiness inspection within a given timeframe". This methodology is called "zero-shot" because it does not provide any examples to the LLM. Conversely, in the "one-shot" approach, an example along with the desired output is included at the end of the pre-prompt, offering a direct illustration of the task to be performed by the model. This article evaluates both of these options.

5.2. *Evaluation and Results*

5.2.1. *Evaluation methodology*

In the assessment of information extraction models' accuracy, a range of metrics is routinely utilized to evaluate their precision and reliability. Recall, precision, F1 score, and F2 score play a crucial role in delineating a model's performance. These measures are especially valuable in contexts where achieving an equilibrium between false positives and false negatives is critical for the intended task.

In this study, we add another metric to address certain limitations inherent in traditional evaluation. A primary concern arises from discrepancies in boundary annotation between experts and LLMs. Consider the sentence: "The very old blue car has to pass the technical inspection". Here, an expert may annotate "very old blue car" as an artifact, whereas LLMs might only identify "blue car" as the relevant segment. Under conventional evaluation methodologies, the LLM's annotation would be dismissed as a False Positive. However, we contend that such annotation could still represent a valid response (i.e. True Positive).

[9] https://gitlab.irit.fr/ala/legal-concepts-extraction
[10] https://platform.openai.com/docs/guides/prompt-engineering

To accommodate this perspective, we have refined the True Positive category into two distinct subcategories: a) the perfect match, the expert's annotation is equal to the LLM's annotation; b) the partial match, which acknowledges instances where the LLM correctly identifies the overarching concept (in this instance, artifact) and its annotation is a subset of, or is encompassed by, the expert's annotation. Table 2 presents examples to summarize our evaluation methodology.

Expert annotation	LLM annotation	Result
very old blue car	very old blue car	Perfect Match (True Positive)
very old blue car	blue car	Partial Match (True Positive) (NLD = 0.529)
blue car	very old blue car	Partial Match (True Positive) (NLD = 0.529)
blue car	red bus	False Positive

Table 2. Examples for the evaluation methodology including : Perfect Match, Partial Match (with the Normalized Levenshtein Distance: NLD) and False Positive.

To assess the "partial march" category, we will introduce an additional metric: the normalized Levenshtein distance. The Levenshtein distance, or edit distance, serves as a metric to gauge the similarity between two strings, by calculating the minimal number of single-character alterations needed to transform one string into the other. The goal is to determine the difference between the LLM's annotation and the expert's annotation. We will normalize this metric in order to compare it across all the partial match annotations.

5.2.2. Overall Performance

Table 3 presents a comprehensive overview of the model GPT-4, detailing its configurations and results. Notably, GPT-4, configured with a one-shot prompt style, demonstrated superior performance, achieving an F1 score of approximately 0.690.

	Strategy		Result			
Model	Prompt Style	Fine-tuning	Precision	Recall	F1	F2
Rule-based [38]			0.972	0.958	0.965	0.961
GPT-4	Zero-Shot	No	0.645	0.684	0.664	0.676
	One-Shot	No	**0.677**	**0.704**	**0.690**	**0.698**

Table 3. Precision, Recall, F1 and F2 results for legal entity extraction using GPT-4. It achieves optimal performance across all evaluated metrics with the following configuration: one-shot prompt style and no fine-tuning.

Within the evaluation dataset, which comprises 973 legal entities, GPT-4 accurately extracted 270 entities with an exact match. Furthermore, approximately 400 additional entities were recognized as "partial matches" of the annotations provided by experts.

As described in the previous section, we introduced the normalized Levenshtein metric in order to compare the edit distance between the expert and LLMs' annotation. Figure 4 demonstrates a gradual increase in the F1 score as the allowance for partial matches in LLMs is varied. Specifically, in the absence of permitted partial matching,

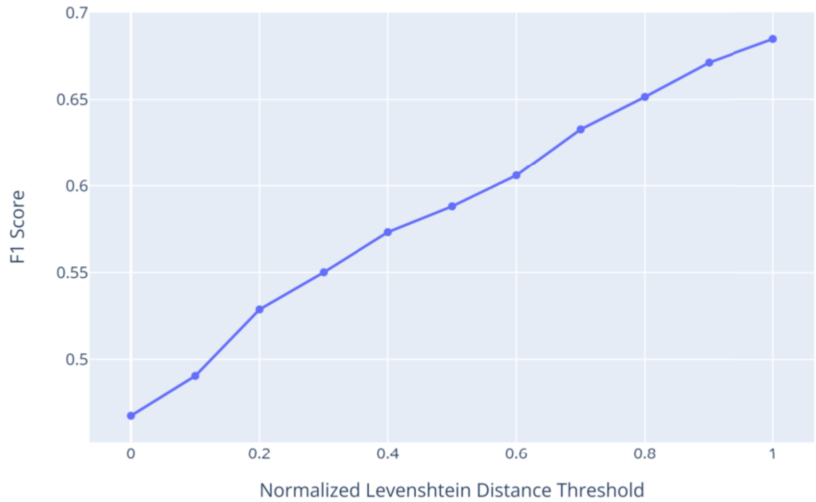


Figure 4. Impact of Normalized Levenshtein Distance Threshold (NLDT) on F1 score. An NLDT of O allows no variation between the annotations made by experts and the partial matches from LLMs. By permitting a sentence variation of approximately 60%, GPT-4 can achieve an F1 score of around 0.6.

the performance of GPT-4 initiates at a level below 0.5. However, with an allowance for up to 60% variation in partial matches, GPT-4 achieves an F1 score of approximately 0.6. For reference, rule-based system reach an average Jaccard index about 0.46 [17], which gives a very similar performance. Figure 4 complements the insights provided by Table 3, highlighting that model performance is not binary and is significantly influenced by the choice of threshold applied to variations in the Levenshtein distance. As the threshold approaches 1, there is an increased risk of incorporating errors, underscoring the delicate balance required in setting this parameter to optimize performance without compromising accuracy.

Despite these notable achievements, the impressive performance of LLM is not without its drawback. While LLM possesses the capability to extract the entities, relying on them exclusively for extraction purposes could result in incomplete extractions. The rule-based strategy [17] recorded only 175 instances of "partial matches" whereas LLMs encountered 415 instances. This issue is inherently linked to the fundamental nature of Large Language Models, underscoring a significant open issue regarding the ability of LLMs to delineate information with precise boundaries. This highlights a critical area for future research and development in the field of natural language processing, specifically in enhancing the precision of LLMs in information extraction tasks. Furthermore, the scientific community has only reported results for rule-based extraction using this dataset. As future work, we intend to explore the capabilities of BERT models on this dataset and establish benchmark results.

6. Conclusion

In our paper, we delved into the utilization of semantic model and LLMs specifically for the task of legal entity extraction, aiming to reduce the involvement of the domain experts. The necessity for automated legal processing cannot be overstated. Historically, regulatory frameworks have been documented across extensive collections of texts, requiring businesses to dedicate considerable human resources to interpret, monitor, and ensure adherence to these legal requirements. This process is not only resource-intensive but also prone to human error, given the complexity and volume of legal documents. Moreover, the dynamic nature of legal regulations, with legislative bodies and political institutions regularly revising and updating laws, further complicates the landscape for compliance. Such frequent changes demand continuous monitoring and analysis to understand their implications for business operations, a task that is both cumbersome and costly for companies.

In section 4, we discussed the methodologies adopted for formalizing legal maintenance rules, focusing on the development of a semantic model. The first approach detailed is based on formal logic, such as deontic logic and the PROLEG language, while the second approach uses semantic models like ontologies and knowledge graphs. Semantic modelling was chosen to facilitate comprehension and interaction by legal experts unfamiliar with formal logic. Following that, we enhanced our first semantic model version by incorporating new legal concepts from a unified model proposed by Sleimi et al. [17], which incorporates frameworks like LegalRuleML [25], LKIF [24], and Cerno [37]. Additionally, we introduced a more nuanced handling of conditions.

In section 5, our paper not only explored the potential of GPT-4 in streamlining legal entity extraction, but also has found the strategy that offers the best results. Through our investigation, we aimed to highlight how LLMs can be used in information extraction task and harness their full power without excessive involvement of experts. State-of-the-art work yields excellent results, achieving an F1 score of 0.961. However, this performance is a trade-off considering the time-consuming tasks such as annotation and rule creation. In contrast, GPT-4, using a one-shot prompt, delivered inferior performance with an F1 score of approximately 0.690, but it required minimal expert input aside from concept definitions and a single example. A significant limitation has been identified when utilizing LLMs for legal information extraction. Although these systems successfully extract and classify entities, there has been an observed increase in partial matches. Partial matches occur when the LLM extracts only a subpart of the intended annotation, rather than the entire entity. Our research highlights a trade-off inherent in employing more generalized systems capable of zero-shot extraction. The flexibility and broad applicability of LLMs may come at the expense of increased partial matches and potentially reduced boundary precision in entity extraction tasks.

Finally, in future work, we aim to address the boundary issue by filtering and enhancing the quality of the extraction process. This can be achieved by employing simple rule-based methods. Specifically, we propose that concepts, such as artefact for example, could be restricted to noun phrases identified through syntactic parsing. This strategy would help eliminate unnecessary words, resulting in more precise extracted concepts. Additionally, various methods, including active learning or human feedback, can be explored to further refine these concepts.

References

[1] Sassier P, Lansoy D. Ubu loi. France: Arthème Fayard; 2008.

[2] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:150801991. 2015.

[3] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018.

[4] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. 2017;30.

[5] Biesner D, Ramamurthy R, Stenzel R, Lübbering M, Hillebrand LP, Ladi A, et al. Anonymization of German financial documents using neural network-based language models with contextual word representations. International Journal of Data Science and Analytics. 2022:151-61.

[6] Sherstinsky A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. ArXiv. 2018.

[7] Lafferty J, McCallum A, Pereira F, et al. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Icml. vol. 1. Williamstown, MA; 2001. p. 3.

[8] Yan H, Deng B, Li X, Qiu X. TENER: Adapting Transformer Encoder for Named Entity Recognition. ArXiv. 2019.

[9] Friedman SE, Magnusson IH, Sarathy V. From Unstructured Text to Causal Knowledge Graphs: A Transformer-Based Approach. ArXiv. 2022.

[10] Pozzi R, Rubini R, Bernasconi C, Palmonari M. Named Entity Recognition and Linking for Entity Extraction from Italian Civil Judgements. In: Basili R, Lembo D, Limongelli C, Orlandini A, editors. AIxIA 2023 – Advances in Artificial Intelligence. Springer Nature Switzerland;. p. 187-201.

[11] Wang Z, Wu Y, Lei P, Peng C. Named Entity Recognition Method of Brazilian Legal Text based on pre-training model. Journal of Physics: Conference Series;1550(3):032149. Available from: https://dx.doi.org/10.1088/1742-6596/1550/3/032149.

[12] Solihin F, Budi I, Aji RF, Makarim E. Advancement of information extraction use in legal documents. International Review of Law, Computers & Technology;35(3):322-51. Available from: https://doi.org/10.1080/13600869.2021.1964225.

[13] Zhou W, Zhang S, Gu Y, Chen M, Poon H. UniversalNER: Targeted Distillation from Large Language Models for Open Named Entity Recognition. arXiv;. Available from: http://arxiv.org/abs/2308.03279.

[14] Zaratiana U, Tomeh N, Holat P, Charnois T. GLiNER: Generalist Model for Named Entity Recognition using Bidirectional Transformer. arXiv;. Available from: http://arxiv.org/abs/2311.08526.

[15] Oliveira V, Nogueira G, Faleiros T, Marcacini R. Combining prompt-based language models and weak supervision for labeling named entity recognition on legal documents. Artificial Intelligence and Law. Available from: https://doi.org/10.1007/s10506-023-09388-1.

[16] Pakhale K. COMPREHENSIVE OVERVIEW OF NAMED ENTITY RECOGNITION: MODELS, DOMAIN-SPECIFIC APPLICATIONS AND CHALLENGES;.

[17] Sleimi A, Sannier N, Sabetzadeh M, Briand L, Ceci M, Dann J. An automated framework for the extraction of semantic legal metadata from legal texts. Empirical Software Engineering;26.

[18] Dragoni M, Villata S, Rizzi W, Governatori G. Combining NLP Approaches for Rule Extraction from Legal Documents. In: 1st Workshop on MIning and REasoning with Legal texts (MIREL 2016);. Available from: https://hal.archives-ouvertes.fr/hal-01572443.

[19] Castano S, Ferrara A, Furiosi E, Montanelli S, Picascia S, Riva D, et al. Enforcing legal information extraction through context-aware techniques: The ASKE approach. Computer Law & Security Review;52:105903. Available from: https://www.sciencedirect.com/science/article/pii/S0267364923001139.

[20] Al-Ageili M, Mouhoub M. An Ontology-Based Information Extraction System for Residential Land Use Suitability Analysis. ArXiv. 2022.

[21] Hedi Karray M, Chebel-Morello B, Zerhouni N. A Formal Ontology for Industrial Maintenance. In: Terminology & Ontology : Theories and applications, TOTh Conference 2011.; 2011. p. 1-20. Available from: https://hal.archives-ouvertes.fr/hal-00602182.

[22] Kralingen Rv. A Conceptual Frame-based Ontology for the Law. In: Proceedings of the First International Workshop on Legal Ontologies; 1997. p. 6-17.

[23] Van Engers T, Boer A, Breuker J, Valente A, Winkels R. In: Ontologies in the Legal Domain. Boston, MA; 2008. p. 233-61.

[24] Hoekstra R, Breuker J, Di Bello M, Boer A, et al. The LKIF Core Ontology of Basic Legal Concepts. LOAIT. 2007:43-63.

[25] Athan T, Boley H, Governatori G, Palmirani M, Paschke A, Wyner A. Oasis legalruleml. In: proceedings of the fourteenth international conference on artificial intelligence and law; 2013. p. 3-12.

[26] Kingston J, Schafer B, Vandenberghe W. Towards a Financial Fraud Ontology: A Legal Modelling Approach. Artificial Intelligence and Law. 2004 Dec;(4):419-46. Available from: http://link.springer.com/10.1007/s10506-005-4163-0.

[27] Saravanan M, Ravindran B, Raman S. Improving legal information retrieval using an ontological framework. Artificial Intelligence and Law. 2009 Jun:101-24. Available from: http://link.springer.com/10.1007/s10506-009-9075-y.

[28] von Wright GH. Deontic Logic. Mind. 1951:1-15. Available from: https://www.jstor.org/stable/2251395.

[29] Rasovska I, Chebel-Morello B, Zerhouni N. A conceptual model of maintenance process in unified modeling language. IFAC Proceedings. 2004:497-502.

[30] Hodkiewicz M, Klüwer JW, Woods C, Smoker T, Low E. An ontology for reasoning over engineering textual data stored in FMEA spreadsheet tables. Computers in Industry. 2021 Oct. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0166361521001032.

[31] Smith B, Ameri F, Cheong H, Kiritsis D, Sormaz DN, Will C, et al. A First-Order Logic Formalization of the Industrial Ontologies Foundry Signature Using Basic Formal Ontology. In: JOWO : Joint Ontology Workshops; 2019. .

[32] Otte JN, Kiritsi D, Ali MM, Yang R, Zhang B, Rudnicki R, et al. An ontological approach to representing the product life cycle. Applied Ontology. 2019 Apr:179-97. Available from: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/AO-190210.

[33] Matsokis A, Kiritsis D. An ontology-based approach for Product Lifecycle Management. Computers in industry. 2010:787-97.

[34] Koukias A, Nadoveza D, Kiritsis D. Semantic Data Model for Operation and Maintenance of the Engineering Asset. In: Advances in Production Management Systems. Competitive Manufacturing for Innovative Products and Services. Berlin, Heidelberg; 2013. p. 49-55.

[35] Woods C, Selway M, Bikaun T, Stumptner M, Hodkiewicz M. An ontology for maintenance activities and its application to data quality. Semantic Web. 2023;(Preprint):1-34.

[36] Satoh K, Asai K, Kogawa T, Kubota M, Nakamura M, Nishigai Y, et al. PROLEG: an implementation of the presupposed ultimate fact theory of Japanese civil code by PROLOG technology. In: JSAI international symposium on artificial intelligence. Springer; 2010. p. 153-64.

[37] Kiyavitskaya N, Zeni N, Breaux TD, Antón AI, Cordy JR, Mich L, et al. Automating the extraction of rights and obligations for regulatory compliance. In: Conceptual Modeling-ER 2008: 27th International Conference on Conceptual Modeling, Barcelona, Spain, October 20-24, 2008. Proceedings 27. Springer; 2008. p. 154-68.

[38] Sleimi A, Sannier N, Sabetzadeh M, Briand L, Dann J; IEEE. Automated extraction of semantic legal metadata using natural language processing. 2018 IEEE 26th International Requirements Engineering Conference (RE). 2018 October:124-35.

[39] Park D, An Gt, Kamyod C, Kim CG. A Study on Performance Improvement of Prompt Engineering for Generative AI with a Large Language Model. Journal of Web Engineering:1187-206. Available from: https://journals.riverpublishers.com/index.php/JWE/.

[40] Sahoo P, Singh A, Saha S, Jain V, Mondal S, Chadha A. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications;.