Legal Knowledge and Information Systems
G. Sileno et al. (Eds.)

doi:10.3233/FAIA230983

Event Extraction and Semantic Representation from Spanish Workers' Statute Using Large Language Models

Gabriela ARGÜELLES TERRÓN [a], Patricia MARTÍN CHOZAS [a] and Víctor RODRÍGUEZ DONCEL [a]

[a] *Universidad Politécnica de Madrid*

Abstract. This work uses Large Language Models to process an important piece of Spanish legislation: the Workers' Statute. The proposed method extracts the relevant events in its articles using a GPT-3.5 model and represents the entities involved in the events and the relationships between them as RDF triples. The experiments carried out to select a high-performance strategy include both zero- and few-shot learning tests. Finally, this work proposes a strategy to uplift the extracted legal relations into a legal knowledge graph.

Keywords. Spanish Workers' Statute, Large Language Models, Knowledge Graph, Legal Domain, Event Extraction

1. Introduction

The legal domain is a complex and dynamic field that involves interpreting and applying laws and regulations. Legal data (court cases, legislations, contracts, etc.) is becoming a valuable source to push forward intelligent legal tools [1]. This work proposes an approach using event extraction and semantic graph modeling to bring these systems closer to the public. The event extraction task is being tackled in the state-of-the-art using deep learning models. However, it presents numerous challenges, especially for Spanish texts, including ambiguity, polysemy, and domain-specific terminology [2].

The recent development of Large Language Models (LLMs) [3, 4] has proven to be an excellent approach to mitigate these problems and an important tool to deal with limited data [5, 6] through natural language instructions, called prompts.

This research aims to improve the performance of the event extraction task within the legal domain and to link the information into a semantic graph representation. The data to be used will be the Spanish Workers' Statute, given its importance for legislators and the general public, and the availability of an annotated corpus of 133 sentences from the Statute gathered by Revenko and Martín-Chozas [7]. The low amount of tagged data will be tackled using the GPT-3.5 model, as it has been proven the high performance on Natural Language Processing (NLP) tasks like event extraction [3, 5]. Finally, the extracted events will be represented in a knowledge graph.

2. Related Work

This work defines an event as a textual region likely to compact relevant legal information encapsulated by the articles on the law. The most common event structure is formed by an event mention, an event trigger, an event argument, and an argument role. The argument role is the relationship between an argument and the event it participates in. The basic argument roles are *subject*, *object*, and *complement*. To classify the legal relations, many works [7, 8] use the Hohfeldian classes *Right*, *Duty*, *No-Right*, and *Privilege* [9].

State-of-the-art event extraction relies nowadays on deep learning models like graph neural networks (GAN) [10] and attention mechanisms [11]. However, these models rely on huge amounts of labeled data to improve the model's performance and are mainly used for English corpora. This research uses only 133 annotated sentences from the Spanish Workers' Statute [7], which are insufficient to achieve high-performance models.

To tackle this issue, common approaches use data augmentation techniques [5], transfer learning [7], and active learning [12]. In recent years, language models (LM) have also been used for this and other NLP tasks [6, 3]. In 2021, the work [13] presented an LM exclusively trained with legal corpora in the Spanish language called RoBERTalex, but its use is limited only to fill mask tasks.

OpenAI[1] defines the term in-context learning (ICL) to refer to the ability of an LLM (evolved from the LMs) to adapt or recognize the desired task rapidly at inference time using the learned patterns and skills in training time [3]. If the task definition contains examples of the desired output, it is called few-shot learning; otherwise, it is called zero-shot learning. Using *few-shot learning* [4] drastically reduces the number of task-specific training examples needed to adapt the model to a particular application.

Regarding the semantic representation of the legal relationships, several works [2, 7] were focused on the definition of strategies to structure legal information. The creation of a structure that allows a machine to quickly understand information and relations is the main goal of the Semantic Web [14].

To represent legal data, LegalRuleML [15] is the most promising specification, providing an RDFS meta-model for the deontic and defeasible logic operators applied in the legal domain. Another alternative is the Provision Model [8], which allows the representation of deontic relations and Hohfeld concepts, enabling the representation of different provision types, like Duty, Right, Power, and Liability, and related attributes, like Bearer and Counterpart.

3. Experimental Contributions with LLMs

To process the Spanish Workers' Statutes, several experiments with different prompt settings, like the number of examples and the example selection strategy, are performed. This step aims to select the best performance prompt for the event extraction from the Statute. Towards this goal, three main experiments were conducted: zero-, 3-, and 5-shot learning with the configurations shown in Table 1.

The base prompt was developed in an iterative process known as prompt engineering. It was passed to the GPT-3.5 model to develop a zero-shot experiment, which was

[1] https://openai.com/

Table 1. Settings of each performed experiment.

In-context learning	Examples in the prompt	Gold standard size	Selection Strategy
zero-shot learning	0	133	No selection
3-shot learning	3	130	Random
5-shot learning	5	128	Example selection

evaluated using the complete annotated dataset. Based on these results, three random examples from the annotated data were added to the prompt to demonstrate each step.

The results of the previous two experiments suggested that more examples chosen intelligently could improve the performance. In this regard, an example selection strategy, based on the Sequential Forward Selection (SFS), was defined. It starts with an empty set S and adds one example on every step from the full set of possible examples S^c ($|S^c| = N$ in the first stage). The aim is to get to $|S| = k$ examples. In every iteration i, the algorithm uses as examples the set S and adds one example e_i from the set S^c. Then, the prompt enclosing the $|S|+1$ demonstrations is applied to the set $S^c/\{e_i\}$, and the results are evaluated using metric E. The example e^* added to S is the one that maximizes the model's performance ($\max_i E(S \cup \{e_i\})$). This process is repeated until $|S| = k$. A run of a full set of iterations that leads to adding an example to S is called a *stage*. Given that this process is computationally expensive, the paper proposes reducing it by selecting a subset of S^c of size M in each stage. To perform a 5-shot learning experiment, this selection strategy was executed using the values $k = 5$, $M = 5$, $N = 133$, and strict F1 metric to measure performance. The overall metrics obtained for each of these methods according to the four types of matching are shown in Table 2.

Table 2. Summary of the overall precision, recall, and F1 scores from each tested in-context learning approach.

Experiment	Precision	Recall	F1 Score
Exact			
Zero-shot Learning	0.33	0.31	0.32
3-shot Learning	0.59	0.55	0.57
5-shot Learning	**0.69**	**0.56**	**0.62**
Partial			
Zero-shot Learning	0.38	0.36	0.37
3-shot Learning	0.61	**0.58**	0.60
5-shot Learning	**0.72**	**0.58**	**0.65**
Strict			
Zero-shot Learning	0.13	0.14	0.13
3-shot Learning	**0.58**	**0.58**	**0.58**
5-shot Learning	0.57	0.55	0.56
Type			
Zero-shot Learning	0.20	0.18	0.19
3-shot Learning	**0.55**	**0.55**	**0.55**
5-shot Learning	0.52	0.52	0.52

Although the zero-shot learning was necessary to establish a baseline from which to build and improve the prompts, none of its results are comparable with the other two approaches. The zero-shot learning always stays below the 0.5 threshold.

The other two approaches have similar results. The 5-shot learning is superior in the exact and partial match metric, which means it is more capable of recognizing the boundaries of the roles. The precision of this approach outperforms the 3-shot learning by approximately 0.1 points. On the other hand, the 3-shot learning model is better at classifying the entities, achieving the best results on all metrics with a strict and type match. However, compared with the 5-shot learning, the difference is not remarkable, ranging from 0.02 to 0.03. In all cases, the event object's recognition and classification are improved from 0.0 to the range of 0.50 to 0.65 with respect to the research [7].

The selection between the two approaches relies on the trade-off between entity recognition and role assignment, and the argument and trigger classification. Given that the difference in performances in the second case is insignificant, the best approach to extract the events from the Spanish Workers' Statute is the 5-shot learning approach.

4. Event Extraction

Using the 5-shot learning approach selected in Section 3, the Statute's text can be processed to extract the events and their attributes. It is important to notice that this phase is analogous to the prediction phase in the classic machine learning methodologies, so the results summarized here correspond to the insights extracted from this prediction output.

The prompt was designed to receive sentences, not a full raw text. Because of this, a data processing step was performed before the event extraction task. The aim was to transform the Spanish Workers' Statute's raw form into atomic sentences to be passed to the model. An atomic sentence has meaning by itself and does not contain more than one idea, generally separated by a period. After a division process of all the Statute's parts and a cleaning phase (both automatic and manually), a dataset containing 1235 refined sentences is obtained. This dataset is available in the Zenodo platform with DOI 10.5281/zenodo.8143596 (*https://doi.org/10.5281/zenodo.8143596*).

The normalized text from the Spanish Workers' Statute is the input to the 5-shot learning approach selected in Section 3. From this extraction, the composition of the Statute can be studied. The 56% of the legal relationships within the Statute are duties, while the rest are divided between rights, no-rights, and privileges. Only 232 objects were identified for all sentences, while 1141 phrases contained a subject. Additionally, more than half of the entities recognized per role at a sentence level were duplicated. Also, 105 entities between subjects, objects, triggers, and complements are repeated in different roles.

5. Semantic Representation

The representation strategy followed is based on the Provision Model explained in Section 2. The elements `prv:Right`, `prv:Duty`, `prv:Prohibition` and `prv:Permission` correspond to the Hohfeld classes used to classify the events: Right, Duty, NoRight and Privilege, respectively. The attributes Bearer, Counterpart, and Object correspond to the extracted argument roles, i.e. subject, object and complement.

The provision attributes need to be represented using other ontologies. The subject and object entities are defined with the `sem:Actor` class from the Simple Event Model (SEM) ontology. In contrast, the type of the subject and object will be represented using the `sem:actorType` property. For this, three new resources of type `sem:ActorType` will be defined: `LegalAgent`, `LegalEntity`, and `Legal Concept`. The complement and the event trigger are instances of `skos:Concept` and `schema:Action`, respectively. The labeling will be done in all cases using the schema.org resources like it was proposed in the research [7].

The final result of mapping the extracted events with this strategy, along with the intermediary datasets created, is publicly available through Zenodo platform with DOI 10.5281/zenodo.8147616 (*https://doi.org/10.5281/zenodo.8147616*).

The built RDF has many potential uses in legal systems. Although implementing such systems is outside the scope of this research, a simple example of its potentialities can be shown by answering the question *As an employer, what are the ways that I have to pay my workers?* using the semantic graph built and a SPARQL[2] (SPARQL Protocol and RDF Query Language) query. This question relates two entities: "empresario" (*employer*) and "salario" (*salary*). Additionally, it defines a right relation where the *employer* is the subject and the *salary* is the object. Figure 1 shows the SPARQL query executed into the graph and the corresponding output, processed for more readability.

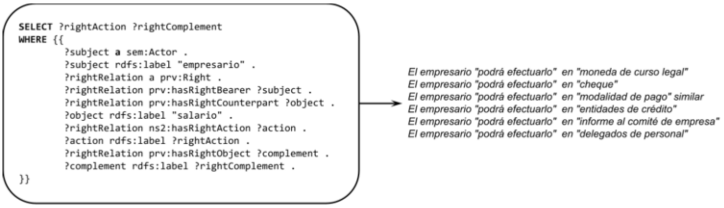

Figure 1. SPARQL query and the corresponding output to answer the question made in Section 1.

6. Conclusions and Future Work

The Workers' Statute represents the reference legal text in labor relations in Spain. The presented research focused on extracting events from the Statute text and representing them in a semantic graph. For this, the 5-shot learning approach was selected after thoroughly evaluating different prompts. There was a significant improvement in extracting the object role, which achieved a 0.61 F1 score, a 100% improvement compared with the research [7]. After this selection, the GPT-3.5 model was applied to a normalized version of the entire Statute, and the extracted information was represented in a semantic graph.

Despite these enhancements, this research exposed some limitations of the GPT-3.5 model for event extraction, which forced the application of a more intelligent example selection strategy. It is not robust enough because of its dependency on the quality of demonstrations and is highly susceptible to inconsistencies in the examples. In this sense, this research recommends a data quality check before using examples in the prompt.

As future work, this research also proposes to take advantage of the GPT-3.5 model chatting capabilities to develop reinforcement learning through human feedback (RLHF) approach to increase the model performance. On the other hand, the resulting knowledge graph might be linked to other resources to gain value. A comparative evaluation between QA over the semantic graph and QA directly over the LLM is suggested too. Finally, we propose to explore the neurosymbolic approach, using legal semantic resources to uplift existing LLMs models [3].

[2]https://www.w3.org/TR/rdf-sparql-query/

[3]Acknowledgements: Projects Knowledge Spaces (10.13039/501100011033), the EU H2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813497 (PROTECT) and Infraestructura para la Investigación de Espacios de Datos distribuidos en UPM (INESData).

References

1. Feng Y, Li C, and Ng V. Legal Judgment Prediction via Event Extraction with Constraints. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, 2022 May :648–64. DOI: 10.18653/v1/2022.acl-long.48. Available from: https://aclanthology.org/2022.acl-long.48
2. Martín Chozas P. Towards a Linked Open Data Cloud of language resources in the legal domain. PhD thesis. ETSI_Informatica, 2018
3. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, et al. Language models are few-shot learners. Advances in neural information processing systems 2020; 33:1877–901
4. Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, Barham P, Chung HW, Sutton C, Gehrmann S, et al. Palm: Scaling language modeling with pathways. arXiv preprint arXiv:2204.02311 2022
5. Pouran Ben Veyseh A, Lai V, Dernoncourt F, and Nguyen TH. Unleash GPT-2 Power for Event Detection. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, 2021 Aug :6271–82. DOI: 10.18653/v1/2021.acl-long.490. Available from: https://aclanthology.org/2021.acl-long.490
6. Wei X, Cui X, Cheng N, Wang X, Zhang X, Huang S, Xie P, Xu J, Chen Y, Zhang M, et al. Zero-shot information extraction via chatting with chatgpt. arXiv preprint arXiv:2302.10205 2023
7. Revenko A and Martín-Chozas P. Extraction and Semantic Representation of Domain-Specific Relations in Spanish Labour Law. Procesamiento del Lenguaje Natural 2022; 69:105–16
8. Francesconi E. Semantic model for legal resources: Annotation and reasoning over normative provisions. Semantic Web 2016; 7:255–65
9. Hohfeld WN. Fundamental legal conceptions as applied in judicial reasoning. The Yale Law Journal 1917; 26:710–70
10. Zhu H, Lin Y, Liu Z, Fu J, Chua Ts, and Sun M. Graph Neural Networks with Generated Parameters for Relation Extraction. 2019. arXiv: 1902.00756 [cs.CL]
11. Jettakul A, Wichadakul D, and Vateekul P. Relation extraction between bacteria and biotopes from biomedical texts with attention mechanisms and domain-specific contextual representations. BMC bioinformatics 2019; 20:1–17
12. Shen S, Li Z, and Qi G. Active Learning for Event Extraction with Memory-based Loss Prediction Model. arXiv preprint arXiv:2112.03073 2021
13. Gutiérrez-Fandiño A, Armengol-Estapé J, Gonzalez-Agirre A, and Villegas M. Spanish Legalese Language Model and Corpora. 2021. arXiv: 2110.12201 [cs.CL]
14. Berners-Lee T, Hendler J, and Lassila O. The semantic web. Scientific american 2001; 284:34–43
15. Tara A, Harold B, Governatori G, Palmirani M, Adrian P, Adam W, et al. LegalRuleML: from Metamodel to Use Cases-A Tutorial. RuleML2013-Theory, Practice, and Applications of Rules on the Web 2013 :1–6