Procesamiento del Lenguaje Natural, Revista nº 69, septiembre de 2022, pp. 105-116 recibido 31-03-2022 revisado 12-05-2022 aceptado 30-05-2022

Extraction and Semantic Representation of Domain-Specific Relations in Spanish Labour Law

Extracción y representación de relaciones específicas de dominio en la legislación laboral española

Artem Revenko,[1] **Patricia Martín-Chozas**[2]

[1]Semantic Web Company, Vienna, Austria
[2]Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain
artem.revenko@semantic-web.com, patricia.martin@upm.es

Abstract: Despite the freedom of information and the development of various open data repositories, the access to legal information to general audience remains hindered due to the difficulty of understanding and interpreting it. In this paper we aim at employing modern language models to extract the most important information from legal documents and structure this information in a knowledge graph. This knowledge graph can later be used to retrieve information and answer legal question. To evaluate the performance of different models we formalize the task as event extraction and manually annotate 133 instances. We evaluate two models: GRIT and Text2Event. The latter model achieves a better score of ≈ 0.8 F_1 score for identifying legal classes and 0.5 F_1 score for identifying roles in legal relations. We demonstrate how the produced legal knowledge graph could be exploited with 2 example use cases. Finally, we annotate the whole Workers' Statute using the fine-tuned Text2Event model and publish the results in an open repository.
Keywords: Information Extraction. Knowledge Graphs. Semantic Web. Legal Domain.

Resumen: A pesar de la actual libertad de información y del desarrollo de diferentes repositorios de datos abiertos, el acceso a la información jurídica al público general sigue suponiendo un problema debido a la dificultad de comprensión e interpretación de dicha información. En este artículo, nuestro objetivo es emplear modelos de lenguaje punteros para extraer información relevante de documentos jurídicos; así como estructurar esta información en un grafo de conocimiento, con el objetivo de que este grafo pueda utilizarse más adelante para recuperar información y responder preguntas sobre el dominio jurídico. Para evaluar el rendimiento de los diferentes modelos, hemos formalizado este proceso como una tarea como extracción de eventos, y hemos anotado manualmente 133 instancias. Evaluamos dos modelos: GRIT y Text2Event. El último modelo consigue mejores resultados, de ≈ 0.8 F_1 para identificar clases jurídicas y de 0.5 F_1 para identificar roles en relaciones jurídicas. Asimismo, ejemplificamos cómo el grafo producido podría explotarse con diferentes casos de uso. Finalmente, hemos anotado todo el Estatuto de los Trabajadores con el modelo Text2Event y publicado los resultados en un repositorio abierto.
Palabras clave: Extracción de Información. Grafos de Conocimiento. Web Semántica. Dominio Jurídico.

1 Introduction

Due to its specific nature, the legal domain has always been a complex area for non legal users. The challenges include *finding* the correct document for a purpose and *interpreting* the document. With the recent rise of the data sharing and open data technologies in the last decade, legal knowledge is more accessible than ever. Well-known legal practitioners have already exposed their legal data in open and machine readable formats, developing platforms such as the European Data Portal[1], a platform funded by the Eu-

[1]`https://data.europa.eu/`

ISSN 1135-5948. DOI 10.26342/2022-69-9

ropean Union and managed by the Publications Office that gathers legal data from different subdomains such as justice, legal system and government. At a national level, one of the most important sources of legal data in Spain is the Official State Gazette[2], which is constantly being updated and accessed by lawyers in Spain. It contains documents in the labour law domain, such as state collective agreements and the Spanish Workers Statute. This availability of legal data efficiently addresses the task for finding the correct documents and accessing them. Yet, the interpretation of legal data and, therefore, exploitation of the legal results by general public remains an important challenge (Robaldo et al., 2019). We are driven by the idea of tackling this challenge and enabling more human-friendly interfaces for accessing this kind of information. We choose the labour law sub-domain for our experimentation as this domain is relevant for everyday use by general audience and the tasks of enabling natural language search, information retrieval, question answering over the labour law are highly demanded. To solve these kind of tasks and ease the access to legal information from general audience, we aim at structuring legal data into a knowledge graph.

Modern (multilingual) language models have celebrated many successes on various NLP tasks (named entity recognition, relation extraction, paraphrase detection, question answering, etc.). We employ these models to tackle our challenge as well. We noticed that most models are trained with general corpora and, therefore, fail to identify the peculiarities of domain specific texts. The lack of domain specific annotated corpora and domain specific language resources published in machine readable formats hinders the fine-tuning of the models.

Consequently, the purpose of this paper is twofold: on the one hand, we test different language models on a domain specific corpus to extract relations amongst terms and, on the other hand, we provide annotated data in the labour law domain and structure the results in Semantic Web formats so they can be reused in the future by upcoming researchers.

This work (code, input and output data) is openly available and published in a GitHub repository[3].

[2] `https://www.boe.es/`

[3] `https://github.com/pmchozas/term_relex/`

1.1 A look at the Semantic Web

More than 20 years ago, the *World Wide Web Consortium (W3C)* promoted the publication of data in structured, machine-readable and interlinked formats, in which the meaning of data can be interpreted by machines to achieve more complex and effective data understanding. This initiative is known as the *Semantic Web* or the *Web of Data* (Berners-Lee, Hendler, and Lassila, 2001).

The most common format for publishing data on the Semantic Web is the *Resource Description Framework (RDF)*. RDF is at the core of the Linked Open Data paradigm for publishing information, based on the *Linked Data Principles* (Berners-Lee, 2006). According to these principles, resources need to be identified by a *Uniform Resource Identifier (URI)* (a unique identifier that follows the HTTP standard web protocols), and that resources need to contain pointers to other resources.

The inner structure of Linked Data is determined by the *ontology* (also known as *model* or *vocabulary*) that defines how to represent the concepts of a certain domain (Chandrasekaran, Josephson, and Benjamins, 1999). These ontologies are composed of classes, relations, rules and restrictions.

In this paper, we make use of RDF to structure the extracted knowledge following the Linked Data Principles, publishing the results in a machine readable and linked dataset, also called *knowledge graph*.

1.2 Motivation

We have already approached the task of creating legal knowledge graph in one of the pilots (Spanish labour law pilot) of the [Lynx][4] project, an Innovation Action funded by the European Union's Horizon 2020, that was active from 2017 until 2021. Although the project already ended, this pilot served as an inspiration to work deeply on the extraction of knowledge over labour law documentation, since several issues were spotted during its development:

1. The labour law texts are highly domain specific. This fact hinders the reuse of already pre-trained language models that are usually trained with texts from the general domain.

[4] `https://lynx-project.eu/`

2. Annotated corpora in this domain are scarce; even after annotating a part of the Statute, the size of the resulting annotated corpus is not sufficient for a language model to obtain good results.

Therefore, within this paper we tackle those issues trying to untangle labour law data, easing the understanding of duties and rights related to the labour domain to anyone willing to know them. For instance, if we suppose that a worker is in trouble with a company for taking a leave from work, we may want to answer questions such as:

Q1) *In which situations can a worker claim for a right?*

Q2) *When must a worker come back to work after a leave from work?*

The rest of the paper is divided as follows: Section 2 explains the statement of the task; Section 3 describes the corpus and the annotation process; Section 4 identifies the language models applied in the experiment; Section 5 reports on the results and evaluation; Section 6 covers the conversion of the results to Semantic Web formats; Section 7 shows examples on how to exploit the resulting knowledge graph; Section 8 gathers related work on event extraction and legal ontologies and, finally, Section 9 collects conclusions and future work.

2 Task Statement

As described in previous work by the authors (Martín-Chozas and Revenko, 2021), we analysed the nature of conceptual relations in legal texts and noticed that labour law texts are full of Hohfeld *deontic relations*, part of the *Hohfeldian fundamental relations* (Hohfeld, 1913), that are divided into two sets of relations:

- *Deontic relations*, that are those that modify ordinary actions: Right, Duty, No-Right and Priviledge.
- *Potestative relations*, that are those that modify deontic relations: Power, Liability, Disability and Immunity.

In this work, we focus on the extraction of deontic relations (see Figure 1), since they are the basis of the fundamental relations and the most common within the Spanish labour law.

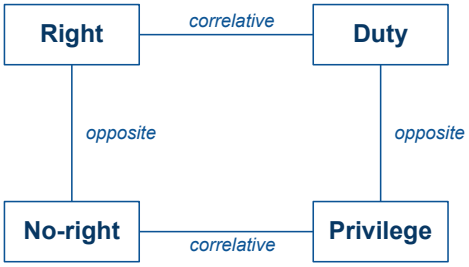

Figure 1: Hohfeld's Deontic Relations.

Hohfeld classes provide information about the particular type of legal relation. However, it is difficult to use this information alone. In fact, more complex use cases, such as question answering over legal texts or the merge of different legal documents into a single knowledge graph, require extracting additional information about the participants of these relations.

We identify the following roles of the participants: subjects of relations, objects of relations and complements of relations. These roles correspond to the classes identified by the Provision Model (see Section 6):

- The *subject* is the agent of the action, who performs the action.
- The *object* is the patient of the action, who receives the action.
- The *complement* is the item which is handled in the relation.

Consequently, the model should be capable of (1) classifying a string from a legal text into one of Hohfeld relation classes (if there is one); (2) identifying the roles of participants of the extracted relation. We formalise the task as *sentence-level event extraction*[5]. Event extraction is an essential task for natural language understanding, aiming to transform the text into structured event records (Doddington et al., 2004; Ahn, 2006). These event records can further be transformed into a knowledge graph, see Section 6.

[5]In preliminary experiments we also considered an alternative formalisation as a relation extraction task (Hendrickx et al., 2019). However, in that case, we need to extract the entities in a separate step and then use relation extraction to identify relation between those entities. Moreover, the roles that we want to extract are better described as roles within a sentence rather than being in a relation with a particular entity. In sight of these difficulties, we refrained from using relation extraction as the task formulation.

We illustrate the event extraction task in Figure 2 with an example from Spanish labour law. From the sentence, we extract an event record of type "Right" corresponding to the Hohfeld deontic class "Right", together with the roles of the different participants.

Generally, event extraction allows the definition of different sets of roles for each event. For our use case we do not exploit this flexibility of task formulation as we define the same set of roles for all event types.

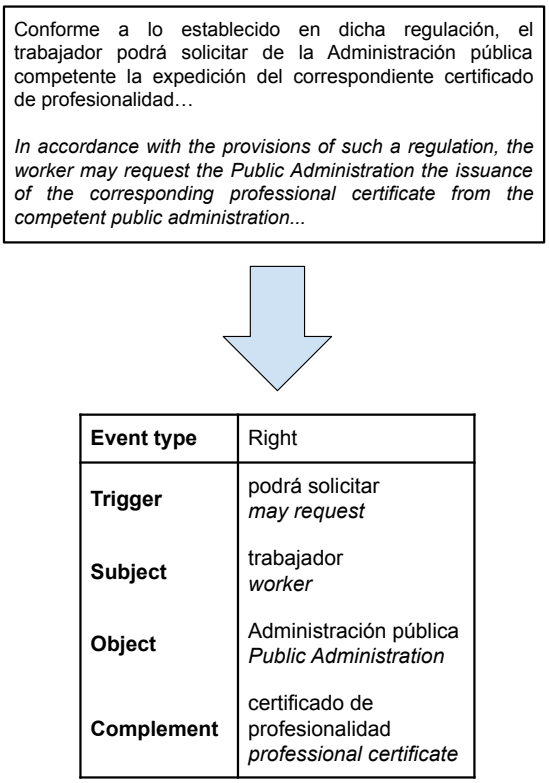

Figure 2: Annotated sample of an event event from a sentence from Spanish labour law, with approximate translations.

3 Training Data

As mentioned before, this experiment is based on the Spanish Workers' Statute, that is published in the Official State Gazette. The text is the main legislative labour law document in Spain, therefore, it clearly corresponds to our aim. The Spanish Workers' Statute is a representative example of a domain-specific legal corpus, therefore, any obtained results could be extrapolated to other legal sub-domains.

The Statute is divided into three main sections named as "titles". The first title covers individual labour relations; the second title covers the rights of collective representation and workers' assemblies inside companies, and the third title covers collective bargaining and collective agreements. In total, the three sections gather 92 articles, containing approximately 50.000 tokens. With the current state of analysis we estimate the density of relations in the Spanish labour law to be 3.65 relations per article.

Regarding the manual annotation of this document, it started by identifying its most relevant terms. To speed up this process, we made use of an open source terminology extraction tool that extracts the most frequent terms in the statute. For more information about the evaluation of the tool's performance, we refer the reader to its research paper (Oliver and Vàzquez, 2015).From those most frequent terms, we identified those that could hold the roles of subject and object of a Hohfeld relation. This is, *legal agents*, such as worker or employer, and *legal entities*, such as company or worker union.

Having the document automatically annotated with these entities, we focused on discovering Hohfeld relations amongst them and extracted the corresponding text excerpts. Optionally, these excerpts could also include a complement, usually an object that takes part in the relation. Not only positive samples were annotated, but also negative samples: text excerpts with legal entities that do not present any relation at all amongst them. Corpus and annotated data statistics are shown in Table 1. Figure 2 shows an example of a positive annotation. Regarding the negative ones, we have identified 2 types: 1) Annotations with entities but no relations and 2) Annotations with neither entities nor relations. Examples of these types are shown below:

1. Mediante los convenios colectivos, y en su ámbito correspondiente, los trabajadores(e1) y empresarios(e2) regulan las condiciones de trabajo y de productividad. (*By means of the collective agreements, in their corresponding field, workers(e1) and employers(e2) regulate working and productivity conditions.*)

2. Igualdad de remuneración por razón de sexo. (*Equaly pay based on sex*)

Type of Element	Total number
Sentences in the corpus	1568
Tokens in the corpus	54849
Annotated samples	133
Positive samples	97
Negative samples	36
Legal agents	127
Legal entities	86
Subjects	90
Objects	69
Complements	100

Table 1: Statistics of the corpus and the annotated data.

4 Models

For this work we focused on joint multi-task deep learning classification models as they achieve state of the art results on common event extraction datasets, see also Section 8. As these common datasets are in English, for the final choice of the model it was important for us that the model code is publicly available, so that we can reuse the model and that the base model is either multilingual or can be changed to multilingual. Finally, we proceed with two models: GRIT (Du, Rush, and Cardie, 2021) and Text2Event (Lu et al., 2021). For both models

GRIT The model GRIT is a generative role-filler transformer model, i.e. it is capable of identifying the (predefined) roles of entities in text. In order to apply GRIT to event extraction task we found it necessary to declare the trigger as a separate role and extend the model to also classify the class (event type) of the input text. The extended model is a joint generative event extraction model. The base model of GRIT is BERT (Devlin et al., 2018), we used the model with BETO (Cañete et al., 2020) that is a BERT model trained on a big Spanish corpus.

Text2Event The event extraction model is Text2Event. This model relies on the description or names of roles, more precisely the names of roles are input to the language model together with the legal text. Therefore, we translated the names of the roles and also the names of the Hohfeld classes to Spanish. As recommended by the authors of the original paper, we have pre-trained the model on ACE dataset, and only then fine-tuned the model on our dataset[6]. The base model of Text2Event is T5 (Raffel et al., 2020) pre-trained on English corpora, we used Text2Event with multilingual T5 (mT5) (Xue et al., 2020).

Roles The final set of roles for both models consists of trigger, subject, object, and complement roles as described in Section 2.

5 Results and Evaluation

For the evaluation of the performance of our model we will use well established metrics such as precision (P), recall (R) and F_1 score. Let the *gold standard* be the correct manually annotated data. Let the *true positives (TP)* be all the correctly predicted relations; *false positives (FP)* – incorrectly predicted relations; *false negatives (FN)* – those cases when a relation is not predicted, though it does exist in the gold standard; *true negatives (TN)* – the relation is not predicted and it does not exist in the gold standard. Then $P = \frac{TP}{TP+FP}$, $R = \frac{TP}{TP+FN}$ and $F_1 = 2 * \frac{P*R}{P+R}$. These measures are well established and widely used for evaluation of different classification models, see also Section 8. It should be noted that we use strict scores, i.e. only exactly correct matches are counted as true predictions.

For computing the results we used the 126 samples from the training set in the following split: 116 samples used for training and 10 samples used for development set. The test set consists of 20 samples, all the results are reported for the test set. This setup is applied to both models. The training was performed with default parameters as set by the authors of the original models, except for the extension of GRIT and the change of the base models as described in Section 4.

The comparison of the models is in Table 2. The two columns present the F_1 scores for the task of classifying all roles including "trigger" role and classifying the Hohfeld class, respectively. The Text2Event model outperforms the GRIT model by a significant margin. Manual checks of the results confirm this finding. A possible explanation of the difference in performance could by the ability of the Text2Event model to include the

[6]Though the roles and the language in the ACE dataset are quite different from our use case, we use ACE in the pre-training to enable the model to learn the constrained generative language as suggested by the authors of the original Text2Event paper.

model name	F_1 roles	F_1 Hohfeld class
GRIT	0.26	0.65
Text2Event	0.47	0.82

Table 2: Comparison of F_1 scores achieved by GRIT and Text2Event models. "F_1 roles" is the F_1 score of extracting all the different roles, including the trigger. "F_1 Hohfeld class" is the F_1 of the legal relation classification.

data	F_1	p	r
all roles	0.47	0.45	0.50
↳ trigger	0.82	0.75	0.90
↳ subject	0.57	0.50	0.67
↳ object	0.00	0.00	0.00
↳ complement	0.45	0.42	0.50
Hohfeld class	0.82	0.75	0.90

Table 3: Detailed scores of Text2Event model.

names of roles and classes, i.e. "derecho", "sujeto", etc., as input to the model. This ability allows for pre-training on the large ACE dataset and efficient knowledge transfer for the few-shot learning with our labour law dataset even despite the different types of events, different domain and different language (English) of the pre-training dataset.

More detailed results for the better performing Text2Event model are presented in Table 3. The model can classify and extract triggers with good confidence, however, the model experiences difficulties with other roles, in particular with identifying objects. Manual investigation of the final results reveals additional problems in identifying the *complements*. Moreover, model rarely predicts other classes than "Right" and "Duty". Nevertheless, the Hohfeld classes, subjects, objects, and triggers are predicted with F_1 scores in the range of 0.5-0.8, which we consider to be reasonably good.

6 Triplification

The second part of our work is focused on the publication of the results obtained following Semantic Web formats. Normally, the results of this type of experiments are usually published in unstructured formats, such as txt, or semi-structured, such as csv.

In this case, we consider that it is highly important to publish these data in structured, open and machine-readable formats, to support their reuse and update. This is possible thanks to the data models of the Semantic Web. In this specific case, we have combined linguistic data representation models with legal information representation models, that are described in Section 6.1. Consequently, we have transformed our results following those vocabularies, as explained in Section 6.2, and the resulting dataset is presented in Section 6.3.

6.1 Ontology selection

As mentioned in the Motivation (Section 1.2), the results of this experiment are to be transformed into a labour law knowledge graph, with the aim of generating a rich resource of concepts and relations that can be applied to other NLP tasks in the future. Therefore, to represent these relations, we chose the Provision Model mentioned in Section 8 since, although LegalRuleML also allows the representation of deontic operators, the Provision Model contains classes corresponding to the potestative relations in case we would like to include them in the future.

On the other hand, to represent the linguistic information apply SKOS[7] and labelling properties from RDF Schema[8].

In this case, since the envisioned output is a rich terminological resource with many different kind of data, we need both vocabularies to represent the complexity of the information contained. Additionally, we combine them with other ontologies such as the Schema data models[9], to add extra information to the relation, as explained in the following section.

6.2 Schema

The heuristics behind the semantic representation of this dataset are as follows:

- Every time we find a Hohfeld relation, we create a new Hohfeld class with the Provision Model, depending on the nature of the relation. Therefore, this would be `prv:Right`, `prv:Duty`, `prv:Prohibition` or `prv:Permission`.
- Following the nomenclature of the Provision Model, every class needs to have a *Bearer*, a *Counterpart* and, optionally, an *Object*. These elements correspond

[7] `https://www.w3.org/TR/skos-reference/`
[8] `https://www.w3.org/TR/rdf-schema/`
[9] `https://schema.org/`

to the *Subject*, *Object* and *Complement* from our annotations, respectively. These items are represented with the class `skos:Concept`, and they are linked to the Hohfeld class with different properties: `prv:hasRightBearer`, `prv:hasRightCounterpart` and `prv:hasRightObject`.

- The `skos:Concepts` are thought to be URIs, to assign unambiguous identifiers to each Hohfeld element. To represent their labels, we use the `rdfs:label` property.
- Additionally, we also include the relation trigger in this data model, that is represented with the class schema:Action, and linked to the Hohfeld class with the properties `prv:hasRightAction`, `prv:hasDutyAction`, `prv:hasProhibitionAction` or `prv:hasPermissionAction`.

6.3 Dataset in RDF

First, we have split the Statute into individual sentences, that were used one by one as input to our fine-tuned Text2Event model. Then we recorded the results. For each sentence that is classified into one of Hohfeld classes we created a subgraph as described in Section 6.2. The statistics of the resulting dataset are presented in Table 4.

Type of Element	Total number
Hohfeld classes	791
Right	578
Duty	213
Subjects	659
Objects	31
Complements	312

Table 4: Statistics of the resulting dataset.

7 Exploitation

In this section, we translate the questions in natural language formulated at the end of Section 1.2 into SPARQL queries[10], to exemplify how to navigate through the generated graph.

First, in Listing 1, we collect all the prefixes that are needed to formulate the rest of the queries. These prefixes identify the vocabularies that have been used to structure the data: RDF and RDF Schema, Provision Model, Schema and SKOS.

Now, Listing 2 formalises Question 1, as stated in Section 1.2: *In which situations can a worker claim for a right?*. Therefore, we look for something (?s) that is a right (prv:Right), which has a right bearer (prv:hasRightBearer), whose ID we do not know (?bearer), and that has right action (prv:hasRightAction), whose ID we do not know either (?action). However, we do know their labels (rdfs:label), which are *trabajador* in Spanish (@es) for the bearer, and *podrá reclamar* in Spanish (@es) for the action. The result of this query are three URIs which are the identifiers of the right instances that satisfy these criteria: http://www.testuri.com/test_hohfeld#1032; http://www.testuri.com/test_hohfeld#762; http://www.testuri.com/test_hohfeld#764.

On the other hand, Listing 3 formalises the Question 2: *When must a worker come back to work after a leave from work?*. In this case, we look for something (?s) that is a duty (prv:Duty), with a duty bearer (prv:hasDutyBearer) and a duty action, which is the trigger (prv:hasDutyAction). Here, the label of the duty bearer is *trabajador*, and the label of the duty action, which is of the type schema:Action, is *deberá reincorporarse*. The result of this query is a URI which is the identifier of the right instance that satisfies this criteria: http://www.testuri.com/test_hohfeld#1051

Additionally, since we already have obtained the ID of a given instance, we could make a simple query to retrieve the textual excerpt (with the property skos:note) from which the relation has been extracted, as exposed in Listing 4. The result of this query is the following excerpt: *En los supuestos de suspensión por ejercicio de cargo público representativo o funciones sindicales de ámbito provincial o superior, el trabajador deberá reincorporarse en el plazo máximo de treinta días naturales a partir de la cesación en el cargo o función.*

8 Related Work

Event Extraction In this work we focus on sentence-level event extraction and consider document-level extraction for future work. Event extraction task has recently received widespread attention (Liu, Min, and Huang, 2021). Most work in event extraction

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

Model name	Trigger F_1	Roles F_1
OneIE	72.8	54.8
Text2Event	71.8	54.4

Table 5: State of the art results on ACE dataset.

has focused on the ACE sentence level event task (Walker et al., 2006), which requires the detection of an event trigger and extraction of its arguments from within a single sentence. This dataset consists of 599 documents and includes 8 event types and 6000 individual events. Further important dataset is MUC-4 (muc, 1992) with 1700 documents, 400 tokens per document on average. Note that these datasets are significantly larger than the one we consider in this paper. Most existing event extraction dataset are available in English language, no event extraction dataset in Spanish is known to us.

The most prominent approaches to solving the task include

1. Decomposition into subtasks such as entity recognition and argument classification (Ma et al., 2020; Zhang et al., 2020);
2. Semantic grounding, i.e. mapping entities to external knwoledge sources (Zhang, Wang, and Roth, 2020; Huang et al., 2018);
3. Question-answering based approaches (Zhou et al., 2021; Liu et al., 2020);
4. Joint multi-task classification models (Lin et al., 2020; Paolini et al., 2021; Du, Rush, and Cardie, 2021; Lu et al., 2021).

Recent state of the art results are achieved by the joint multi-task deep learning models. GRIT (Du, Rush, and Cardie, 2021) achieves joint (for classifying all roles) F_1 score of 54.5 on MUC-4 dataset. The results on ACE dataset are collected in Table 5.

Legal Ontologies Regarding the representation of legal information in Semantic Web formats, we find many approaches depending on the type of data and on the purpose of the ontology. Likewise, the type of data depends on the legal subarea to which a resource belongs (labour law, civil law, etc.) and on the type of legal document (provisions, rules, licenses...). The purpose of the ontology is also varied. Some are merely intended to structure the information while others are designed to reason over data and infer knowledge.

Amongst the most used ontologies to structure legal documents we find the ELI ontology[11], the European Legislation Identifier. This vocabulary is widely used by the publishers of legal data in the European Union to represent metadata of legislative documents as Linked Data. To complement ELI, the Publications Office of the European Union also applies the Common Data Model (CDM) vocabulary and the Functional Requirements for Bibliographic Records (FRBR) to represent legal resources and their relationships.

Apart from those ontologies intended to represent legal documents, we also find well-known vocabularies to represent common general terms in the legal domain such as Akoma Ntoso, which was created as an XML standard and afterwards evolved to an ontology (Palmirani and Vitali, 2011), and LegalRuleML (Athan et al., 2015), that is able to represent the particularities of the legal normative rules with a rich, articulated, and meaningful markup language. Similarly, we find the Provision Model (Biagioli, 1996), to annotate rules and rule ammendments in normative provisions, that was subsequently extended in (Francesconi, 2016), to cover Hohfeld's relations (which are described in Section 2).

In this section, we have mentioned those ontologies that are directly related to our work. For more information about legal ontologies, we refer the reader to more comprehensive surveys such as (Valente, 2005) and (de Oliveira Rodrigues et al., 2019).

9 Conclusions and Future Work

In this paper we experimented with pretrained multilingual language models for extracting knowledge from a domain-specific labour law corpus in Spanish. We formalised the task as *sentence-level event extraction* and applied two models: GRIT and Text2Event. To train and evaluate the model, we annotated the Workers' Statute with 133 individual sentences containing Hohfeld roles and relations. The latter model (Text2Event) outperforms the former by a significant margin of around 0.2 of F_1 scores both on identifying the roles and classifying

[11] `https://op.europa.eu/es/web/eu-vocabularies/eli`

into Hohfeld classes. Text2Event obtains a satisfying results above 0.8 F_1 score for classifying Hohfeld classes. The model also efficiently extracts the triggers ($F_1 \approx 0.75$), but looses the quality for other roles ($F_1 \approx 0.5$ for all roles including trigger).

Furthermore, we split the Workers' Statute into individual sentences and apply the fine-tuned Text2Event model on this input. The results of this automatic information extraction tasks have been also triplified, following existing Semantic Web vocabularies, such as SKOS for concepts and the Provision Model for Hohfeld relations. The resulting labour law knowledge graph is publicly available for to be reused by the community.

Future Work In the short term, we plan to develop a post-processing script based on NLP rules to clean the results with the aim of improving precision. This idea is based on the observation that the current models sometimes interchange relation agents with the relation complement, which we think that could be avoided with a role labeling task over the results. Furthermore, we focus on extracting sentence-level roles and relations as most relations are expressed in a single sentence. However, we note that the number of relations spanning over multiple sentences is not negligible. Hence, in the future work we also plan to experiment with extracting relations beyond sentence level.

Regarding the representation in RDF, we plan to add more linguistic information to the graph, linking it to existing legal resources published in Semantic Web formats, such as EuroVoc[12]. We may also want to link our labour law graph with more general resources such as Wikidata[13], to extend the graph with information from a wider scope. To represent this additional linguistic data we plan to use Ontolex[14] to complement SKOS. This combination is widely applied to represent language resources in the Semantic Web: while SKOS is used for thesauri and concept schemes, Ontolex is intended to represent lexical information such as dictionaries.

Figure 3 shows an example of the semantic representation of a right relation amongst the subject "trabajador" (*worker*), the object "administración pública" (*public administration*) and the complement "*certificado de profesionalidad*" (professional certification), being these labels represented with the Ontolex vocabulary. In the example, terms are also linked with matches in EuroVoc and Wikidata.

[12] https://eur-lex.europa.eu/browse/eurovoc.html

[13] https://www.wikidata.org/

[14] https://www.w3.org/2016/05/ontolex/

Acknowledgements

This work has been supported by the European Union's Horizon 2020 research and innovation programme through the Prêt-à-LLOD[15] project, with grant agreement No. 825182, and by COST (European Cooperation in Science and Technology) through NexusLinguarum, the "European network for Web-centred linguistic data science" COST Action (CA18209)[16].

[15] https://pret-a-llod.eu/

[16] https://nexuslinguarum.eu/

References

1992. *Fourth Message Uunderstanding Conference (MUC-4): Proceedings of a Conference Held in McLean, Virginia, June 16-18, 1992*.

Ahn, D. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8, Sydney, Australia, July. Association for Computational Linguistics.

Athan, T., G. Governatori, M. Palmirani, A. Paschke, and A. Wyner. 2015. Legalruleml: Design principles and foundations. In *Reasoning Web International Summer School*, pages 151–188. Springer.

Berners-Lee, T. 2006. Design issues.

Berners-Lee, T., J. Hendler, and O. Lassila. 2001. The semantic web. *Scientific american*, 284(5):34–43.

Biagioli, C. 1996. Law making environment: model based system for the formulation, research and diagnosis of legislation. *Artificial Intelligence and Law*.

Cañete, J., G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez. 2020. Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*.

Chandrasekaran, B., J. R. Josephson, and V. R. Benjamins. 1999. What are ontologies, and why do we need them? *IEEE Intelligent Systems and their applications*.

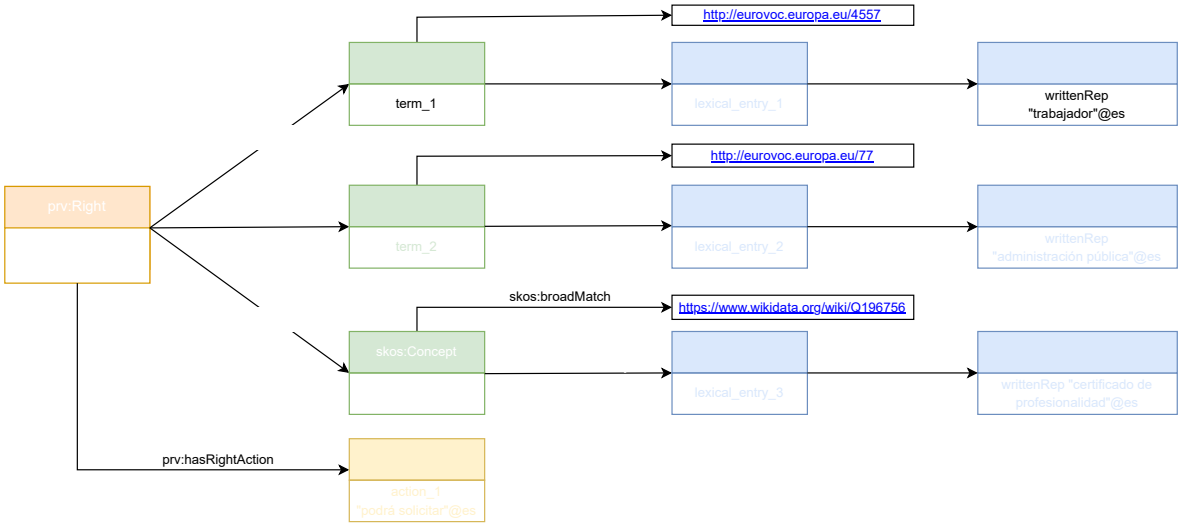

Figure 3: Example of a right modeled with SKOS and Ontolex, and linked with external knowledge bases.

de Oliveira Rodrigues, C. M., F. L. G. de Freitas, E. F. S. Barreiros, R. R. de Azevedo, and A. T. de Almeida Filho. 2019. Legal ontologies over time: A systematic mapping study. *Expert Systems with Applications*, 130:12–30.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv e-prints*, page arXiv:1810.04805, October.

Doddington, G., A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel, and R. Weischedel. 2004. The automatic content extraction (ACE) program – tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, May. European Language Resources Association (ELRA).

Du, X., A. Rush, and C. Cardie. 2021. GRIT: Generative role-filler transformers for document-level event entity extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 634–644, Online, April. Association for Computational Linguistics.

Francesconi, E. 2016. Semantic model for legal resources: Annotation and reasoning over normative provisions. *Semantic Web*, 7(3):255–265.

Hendrickx, I., S. N. Kim, Z. Kozareva, P. Nakov, D. Ó. Séaghdha, S. Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz. 2019. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals. *arXiv e-prints*, page arXiv:1911.10422, November.

Hohfeld, W. N. 1913. Some fundamental legal conceptions as applied in judicial reasoning. *Yale Lj*, 23:16.

Huang, L., H. Ji, K. Cho, I. Dagan, S. Riedel, and C. Voss. 2018. Zero-shot transfer learning for event extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2160–2170, Melbourne, Australia, July. Association for Computational Linguistics.

Lin, Y., H. Ji, F. Huang, and L. Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7999–8009, Online, July. Association for Computational Linguistics.

Liu, J., Y. Chen, K. Liu, W. Bi, and X. Liu. 2020. Event extraction as machine reading comprehension. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1641–1651, Online, November. Association for Computational Linguistics.

Liu, J., L. Min, and X. Huang. 2021. An overview of event extraction and its applications. *arXiv e-prints*, page arXiv:2111.03212, November.

Lu, Y., H. Lin, J. Xu, X. Han, J. Tang, A. Li, L. Sun, M. Liao, and S. Chen. 2021. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction. In *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)*, pages 2795–2806, Online, August. Association for Computational Linguistics.

Ma, J., S. Wang, R. Anubhai, M. Ballesteros, and Y. Al-Onaizan. 2020. Resource-enhanced neural model for event argument extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3554–3559, Online, November. Association for Computational Linguistics.

Martín-Chozas, P. and A. Revenko. 2021. Thesaurus enhanced extraction of hohfeld's relations from spanish labour law. In *Proceedings of the 2nd International Workshop on Deep Learning meets Ontologies and Natural Language Processing (DeepOntoNLP 2021) co-located with 18th Extended Semantic Web Conference 2021*, volume 2918, pages 30–38. CEUR-WS.org.

Oliver, A. and M. Vàzquez. 2015. Tbxtools: a free, fast and flexible tool for automatic terminology extraction. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*.

Palmirani, M. and F. Vitali. 2011. Akomantoso for legal documents. In *Legislative XML for the semantic Web*. Springer, pages 75–100.

Paolini, G., B. Athiwaratkun, J. Krone, J. Ma, A. Achille, R. ANUBHAI, C. N. dos Santos, B. Xiang, and S. Soatto. 2021. Structured prediction as translation between augmented natural languages. In *International Conference on Learning Representations*.

Raffel, C., N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Robaldo, L., S. Villata, A. Wyner, and M. Grabmair. 2019. Introduction for artificial intelligence and law: special issue "natural language processing for legal texts".

Valente, A. 2005. Types and roles of legal ontologies. In *Law and the semantic web*. Springer, pages 65–76.

Walker, C., S. Strassel, J. Medero, and K. Maeda. 2006. Ace 2005 multilingual training corpus. *Linguistic Data Consortium, Philadelphia*, 57:45.

Xue, L., N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, and C. Raffel. 2020. mT5: A massively multilingual pre-trained text-to-text transformer. *arXiv e-prints*, page arXiv:2010.11934, October.

Zhang, H., H. Wang, and D. Roth. 2020. Unsupervised Label-aware Event Trigger and Argument Classification. *arXiv e-prints*, page arXiv:2012.15243, December.

Zhang, Z., X. Kong, Z. Liu, X. Ma, and E. Hovy. 2020. A two-step approach for implicit event argument detection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7479–7485, Online, July. Association for Computational Linguistics.

Zhou, Y., Y. Chen, J. Zhao, Y. Wu, J. Xu, and J. Li. 2021. What the role is vs. what plays the role: Semi-supervised event argument extraction via dual question answering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(16):14638–14646, May.

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX prv:<http://www.ittig.cnr.it/ontologies/def/ProvisionModel#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX schema:<https://schema.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

Listing 1: Prefixes used in the SPARQL queries over the resulting dataset.

```
SELECT ?s
WHERE {
    ?s rdf:type prv:Right;
        prv:hasRightBearer ?bearer ;
        prv:hasRightAction ?action .
    ?bearer  rdfs:label "trabajador"@es .
    ?action rdfs:label "podrá reclamar"@es.
}
```

Listing 2: Example of Question 1 translation into SPARQL.

```
SELECT ?s
WHERE {
    ?s rdf:type prv:Duty;
        prv:hasDutyBearer ?bearer ;
        prv:hasDutyAction ?action .
    ?bearer  rdfs:label "trabajador"@es .
    ?action rdf:type schema:Action ;
    rdfs:label "deberá reincorporarse"@es . }
```

Listing 3: Example of Question 2 translation into SPARQL.

```
SELECT *
WHERE {
    <http://www.testuri.com/test_hohfeld#1051> skos:note ?note .
    }
```

Listing 4: SPARQL query to retrieve the textual excerpt of a given duty.