The Greek Legal Knowledge Graph: A Progress Report

Iosif Angelidis[1], Ilias Chalkidis[2], Christos Papaloukas[1], and Manolis Koubarakis[1]

[1] Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens
{iosang, koubarak, christospap}@di.uoa.gr
[2] Athens University of Economics and Business
ihalk@aueb.gr

Abstract. We present the Greek legal knowledge graph which has been under construction in our group since 2014. The graph has been constructed by extracting various kinds of important knowledge (e.g., which ministers signed for a law) from legal documents published in PDF by the Greek National Printing Office, as well as representing the documents themselves, covering the period 1990 to 2018. The Greek legal knowledge graph has been used to power our platform Nomothesia which aims to make Greek legal documents available on the Web as linked data so that they can be interlinked with other Greek open data, and be made available for the agile development of legal services for legal professionals and citizens.

Keywords: linked data, legal knowledge graph, named entity recognition and linking, dataset generation, entity reference representation, deep learning

1 Introduction

Legislation applies to every aspect of people's living and evolves continuously building a huge network of interlinked legal documents. Therefore, it is important for a government to offer services that make legislation easily accessible to the citizens aiming at informing them, enabling them to defend their rights, or to use legislation as part of their job. It is equally important to have law professionals (lawyers, judges, etc.) access legislation in ways that allow them to do their job easily (e.g., they might need to be able to see the evolution of a law over time).

In Greece, the previous statements hold true even more so. Greek bureaucracy is very complex, involving many governmental instruments and organizations, meaning a legal expert would need many hours and thorough research to properly process a case, its specifics and all parties involved, in order to reach a correct and valid conclusion. To make matters worse, documents are provided mostly in PDF format by the National Printing House of Greece, which makes things even more difficult, as the documents are not machine-readable and, therefore, impossible to use any sort of word search with sufficient accuracy. Moreover, in the age of the Web, it is important to enable software developers to develop applications for citizens and law professionals easily, by connecting the available laws with other kinds of government or private sector information, an aspect in which Greece needs to evolve soon in order to follow suit with the latest progress in Europe and internationally.

There are already many countries in Europe and elsewhere that have computerized the legislative process by developing platforms for archiving legislation documents and offering on-line access to them using standards such as Akoma Ntoso (aka LegalDocML) which is an OASIS standard, the European standard CEN-MetaLex, the European Legislation Identifier, the European Case Law Identifier etc. There are also private companies (e.g., ROSS Intelligence, LexisNexis, RAVEL, LexMachina etc.) that specialize on providing digital services for law, case law, compliance, contracts, etc.

Additionally, due to significant progress in research on Artificial Intelligence technologies and, more specifically, deep learning for NLP, state-of-the-art technologies evolve and change often, becoming more and more accurate in their tasks. Since it has already been shown that they can automate a great amount of everyday tasks and save hundreds of work hours and resources, it only makes sense to endeavor to incorporate them into existing workflows and pipelines.

Drawing inspiration from other successful efforts in Europe, we aim at disrupting the way Greek legislation is made available. As a result, we designed Nomothesia[3], a web platform that aims to provide advanced search capabilities for everyone (especially legal experts and professionals), but also enabling developers to utilize RESTful services to consume resources. A prototype version was presented in [9], while an updated version was discussed in [3]. Last, we provided a detailed documentation [2] of the incorporation of an RNN network (BILSTM-BILSTM-LR) with the aim of extracting named entities from Greek legal text, interlinking them with other sources and generating a subset of a knowledge graph.

In this paper, we present the knowledge graph we have, which results from all the above efforts. We built and expanded upon our efforts each time, until we finally obtained a rich knowledge graph, capable of becoming a significant point of reference for Greek legislation information representation. Speaking in numbers, the graph consists of legal documents of the National Printing House of Greece for the years 1990-2018 (about 8.000 Government Gazettes), 2.9 million triples regarding strictly Greek legislation, and 5 million triples with all data included (12.000 decisions and 195.000 referencs to entities). The current Greek legal knowledge graph is freely available online[4] to any interested party. The capabilities of the graph have so far been demonstrated to the following interested parties: the Greek Parliament, the National Printing Office, the Ministry

[3] http://legislation.di.uoa.gr/

[4] The dataset can be found at http://legislation.di.uoa.gr/data, under a non-commercial license. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/.

of Justice and the Ministry of Administrative Reform. The graph is curated by the members of our Knowledge Representation, Reasoning and Analytics[5] group at the National and Kapodistrian University of Athens under the lead of the authors of this paper. The main contributions of the graph (and, consequently, of this paper) are the following:

- The knowledge graph represents a network of legal texts, each one broken into increasingly detailed parts and encoded following the ELI specification.
- European legislation is represented as well as linked with Greek legislation with *transposed_by* links.
- Legal text contains textual references, which are *relevant_for* named entities.
- A full timeline (more specifically, the signature and publication dates, as well as the date it was set into force) of each document is represented, along with its modifications.
- Named entities are interlinked with third party datasets and incorporate all third-party information into them, further enriching their value.
- Signers and ministries that are involved in each document are represented, while information about ministers related to a piece of legislation includes a wealth of information due to interlinking them with the Greek DBpedia politicians dataset.
- The Greek Administrative Geography dataset[6], which contains all administrative units of Greece after the most recent administrative Kallikratis law, is incorporated.
- Geographical landmarks which are not officially recognized anywhere such as farms, forests and islets are represented for the first time in a consistent format, as well as in a digital format.
- The aforementioned landmarks carry *belongs_to* links towards corresponding administrative units of Greece, so that their potential is increased even further.

The rest of the paper is organized as follows. Section 2 mentions related work and other approaches throughout Europe. Section 3 discusses the structure of the ontology the Greek legal knowledge graph is be based on. Section 4 documents the pipeline and procedures followed in order to generate the Greek legal knowledge graph. Section 5 presents SPARQL querying capabilities with RESTful services. Section 6 discusses the novelties of the new version of the knowledge graph overall and over its previous iteration. Section 7 mentions the applications and overall attention our knowledge graph has received thus far. The paper concludes with Section 8, where we summarize our work and discuss future work.

2 Related work

In Greece, so far, there has been a limited degree of computerization of the legislative process; even the discovery of legislation related to a specific topic

[5] http://kr.di.uoa.gr/
[6] http://www.linkedopendata.gr/dataset/greek-administrative-geography

can be a hard task. The legislative work of the Greek government has been published since 1907 in the form of a gazette by the National Printing Office [7]. Legislation is published on a daily basis in that gazette and it is distributed only in a PDF file format. An initial effort trying to remedy this is the Diavgei@ portal[8]. It is a Greek program introduced in 2010, enforcing transparency over government and public administration, by requiring that government and public administration upload their decisions on the Web. However, the files submitted are not restricted to a single type format, or to have a specific structure.

As already mentioned, there are many European Union countries which have had various degrees of success in computerizing the legislative process. A few examples include the MetaLex document server[9] [16], the official portal of the UK[10], a portal offering Finnish legislation as linked data[14] and a central content and metadata repository of EU[11] [13]. All of the above endeavors are based on semantic web technologies using web standards like XML, RDF, SPARQL. Such examples include XML schemata like Akoma Ntoso[12] [6], (CEN) MetaLex [7,8,21], and the European Legislation Identifier (ELI) [11], a new common framework that has to be adopted by the national legal publishing systems in order to unify and link national legislation with European legislation.. ELI, like Akomo Ntoso and MetaLex, is not a one-size-fit-all model but it has to be extended to capture the particularities of national legislation systems.

Both Akoma Ntoso and MetaLex have been keystones for the adoption of relevant practices in the legal domain. Akoma Ntoso has been the vocabulary of choice for the development of the XML schemata LegalDocML [20] and LegalRuleML [5]. LegalRuleML has been applied to numerous works and studies already, such as [24] which is a pilot study on the annotation of the contents of Scottish legal instruments, using key LegalRuleML elements as annotation. MetaLex was extended in the context of EU Project ESTRELLA[13], which developed the Legal Knowledge Interchange Format (LKIF) [17]. In the same spirit, the European Case Law Identifier (ECLI), a sister endeavor of ELI, was introduced recently for modeling case laws [19]. Other efforts, such as [12] propose an approach to represent the legal data (legal norms and court decisions) of Austria and show how this data can be used to build a legal knowledge graph.

Another recent work on legal knowledge graphs is Lynx [18], an attempt to create a European ecosystem of cloud services to better manage compliance, based on a legal knowledge graph modeled using semantic web technologies. Its objective is to integrate and link heterogeneous compliance data sources. Since our work conforms to European standards regarding legislative representation, it is highly compatible with the work of Lynx. Also, it is important to provide ways

[7] http://www.et.gr/
[8] https://diavgeia.gov.gr/
[9] http://doc.metalex.eu/
[10] http://www.legislation.gov.uk/
[11] Publicly available by the EUR-lex service http://eur-lex.europa.eu.
[12] http://www.akomantoso.org/
[13] http://www.estrellaproject.org/

to make legal data more easily accessible. One such example is the work of [10], which outlines an approach for question answering over regulatory documents with neural networks.

Legal text contains a wealth of information but also needs to be checked in order to ensure the information extracted is up to date. To that end, works such as [4] employ machine learning techniques for automatic detection or ranking of meaningful legal changes. However, it is also important to ensure ontologies and legal thesauri that represent the data can also be properly aligned, so that potential interlinking between datasets is easier and less error-prone. This can be addressed by either forcing everyone to conform to the same ontologies and thesauri when representing specific information, or finding ways to achieve alignments of ontologies and thesauri. [22] discusses the production of two gold-standard alignment datasets between the European Union thesaurus EuroVoc and two other notable resources adopted in legal environments: the thesaurus of the Italian Senate TESEO and the IATE European terminological resource.

Finally, regarding information extraction, which needs to be automated as much as possible, [1] introduces the BO-ECLI Parser, an open framework for the extraction of legal references from case-law issued by judicial authorities of European member States. The problem of automatic legal links extraction from texts is tackled for multiple languages and jurisdictions by providing a common stack which is customizable through pluggable extensions in order to cover the linguistic diversity and specific peculiarities of national legal citation practices.

3 Representing our knowledge graph using semantic web technologies

Before we describe the pipeline necessary to produce the RDF knowledge graph, it is important to discuss the structure of the ontology it will be based on. Our primary concern is offering a dataset which conforms to existing ontologies and their standards so that interlinking of linked data and augmenting represented information is easier to handle. To that end, since our original work, we opted to adopt the ELI ontology (see Figure 1), since it is a proposed European standard for the proper representation of legal texts.

Focused on our initial ontology, we re-engineered it[14] to become even more consistent with ELI. Further on, we extended it to handle more efficiently the intricacies of Greek legal text as well as for the additional representation of entity types for entity recognition and linking tasks. Using our extended ontology, we manage to model the structural information of legal documents, their metadata (e.g., dates, title, gazette etc.), their evolution through time in response to modifications and also all the entities (e.g., persons, landmarks etc.) extracted from legal documents through our Named Entity Recognizer and Linker component.

All legal documents (*eli:LegalResource*) are organized in fragments (*eli:Legal ResourceSubdivision*). A legal document has successive versions (*eli:LegalExpres*

[14] The full ontology we utilize can be found at http://legislation.di.uoa.gr/data/ontology.

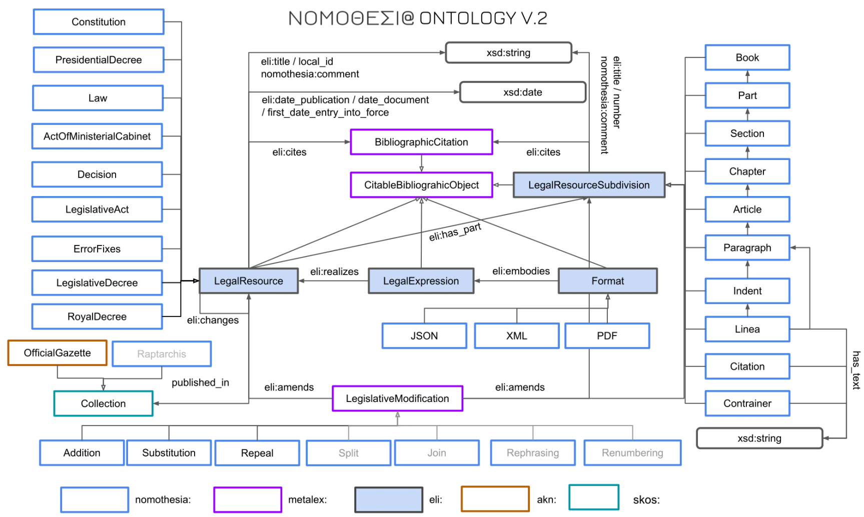

Fig. 1. ELI's (expanded) ontology

sion) that modify (*eli:realizes*) the original, which are differentiated by the previous ones based on subsequent modifications (*eli:LegislativeModification*) that are referred to in subsequent legal documents and affect (*eli:amends*) specific fragments. The latest version changes (*eli:changes*) the previously enacted legal documents. Legal documents also include references (*eli:BibliographicCitation*), which cite (*eli:cites*) other documents, or their fragments (*eli:CitableBibliographicObject*).

Greek legal documents are organized in specific types of fragments (e.g., Article, Paragraph, Indent). Indents (i.e., cases, subcases) and passages are the bottom elements of the main body's hierarchy that define text (*has_ text*). Legislative modifications are also classified in three predefined types: *Insertion*, *Repeal* and *Substitution*. Given the above additions, we are able to model and store legislation natively only in RDF format. Using the RDF encoding of a legal document, any additional format (e.g., PDF, JSON) can be produced dynamically, without the need to include special information.

Moreover, essential metadata to identify the legal documents are described using both the ELI vocabulary and our own extensions (which are further augmented over our initial prototype): type of document (e.g., Law, Presidential Decree), title (*eli:title*), serial number (*eli:id_ local*), the Official Gazette issue (*eli:OfficialGazette*) in which a document has been published (*eli:published_ in*) and especially the legal document's related dates. We provide a thorough timeline for each document as there are three important dates: the date of publication (*eli:date_ publication*), the date it was signed (*eli:date_ signature*) and the day it is set in force (*eli:date_ inforce*). When dealing with the temporal dimension of a law, it is very important to capture all three dates to do proper reasoning and inferencing. This is even more evident when tackling the codification of a law.

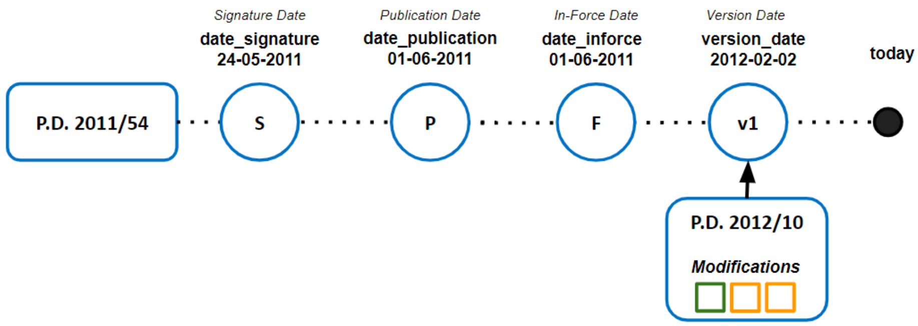

Fig. 2. Timeline representation

Finally, we enhance the knowledge regarding a legal document by capturing additional information (not present in our original work) a few secondary metadata such as signers (*Signer*) of each legal document that can be either a minister (*Minister*) or a ministry (*Organization*) and for the first time we utilize even more European standards regarding legislation. Towards this, we employ tags with thematic categories (*tag*) by utilizing Eurovoc for each document and also, in cases where European legislation is transposed by Greek legislation, we link the ELI URIs of directives that are transposed (*eli:transposes*) by national legal documents.

The above description covers just the legal text representation and additions, modifications and tweaks we did on the original ELI ontology. However, since we also provide named entities as linked data (a novel feature over our original prototype), we have enriched the ontology even further. More specifically, we focus on extracting 6 entity types, when present:

- **Person** Any formal name of a person mentioned in the text (e.g., Greek government members, public administration officials, etc.).
- **Organization** Any reference to a public or private organization, such as: international organizations (e.g., European Union, United Nations, etc.), Greek public organizations (e.g., Social Insurance Institution) or private ones (e.g., companies, NGOs, etc.).
- **Geopolitical Entity** Any reference to a geopolitical entity (e.g., country, city, Greek administrative unit, etc.).
- **Geographical Landmark** References to geographical entities such as local districts, roads, farms, beaches, which are mainly included in pieces of legislation related to topographical procedures and urban planning.
- **Legislation Reference** Any reference to Greek or European legislation (e.g., Presidential Decrees, Laws, Decisions, EU Regulations and Directives, etc.).
- **Public Document Reference** Any reference to documents or decisions that have been published by a public institution (organization) that are not considered a primary source of legislation (e.g., local decisions, announcements, memorandums, directives).

Additionally, Greek geographical landmarks are a major asset for our legal recognizer since they are related to planning and architectural interests. However, there is no such public dataset to interlink between the references and the actual entities. We proceed in generating a new dataset by applying linguistic heuristics in order to form the entities and classify their type in 4 different abstract categories (classes):

– **Local District** such as *villages* and *small local communities.*
– **Area** sub-classified into *agricultural*, *forest*, *coastal* and *marine areas.*
– **Road** sub-classified into *highway*, *local*, *bypass roads.*
– **Point of Interest** such as a *farm*, an *islet*, or a *peninsula* which are commonly referred to urban planning legislation.

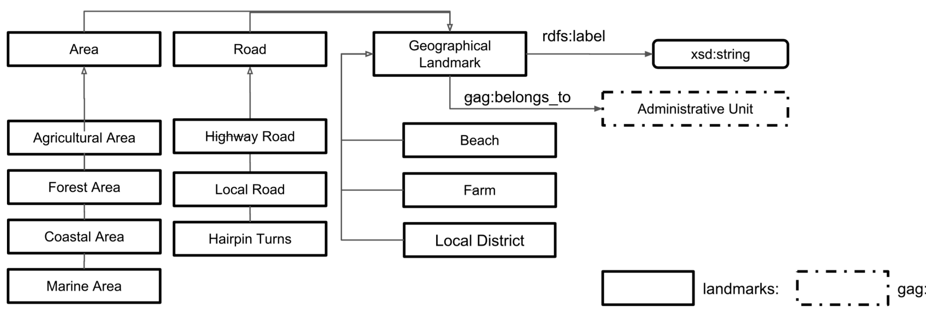

Fig. 3. Ontology types for geographical landmarks

Further on, we interlink this new dataset of geographical landmarks with the Greek administrative units (as described in the Greek Administrative Geography dataset) when there is a connection between them (*belongs_to*), indicated in terms of text (e.g., "Beach Kavouri at Municipality of Varis-Voulas-Vouliagmenis"). The process of extracting this information is described in detail in Section 4.2 below.

When we have normalized the labels, after post-processing and extracting entities from the text, we need to represent the entities within the text following the RDF specification. Legal text may contain (*has_reference*) a *Reference* to an entity (e.g., a law passage referring to a specific law that it modifies). This reference is realized in an interval of characters. In other words, it *begins* and *ends* on specific sequential characters inside the text of the subdivision of a law. This *Reference* most likely refers to (or, in another sense, is *relevant_for*) an Entity, which is probably described in open public datasets. Therefore, a *LegalResourceSubdivision* contains references to persons, administrative units and legal resources (e.g., laws, decisions etc.).

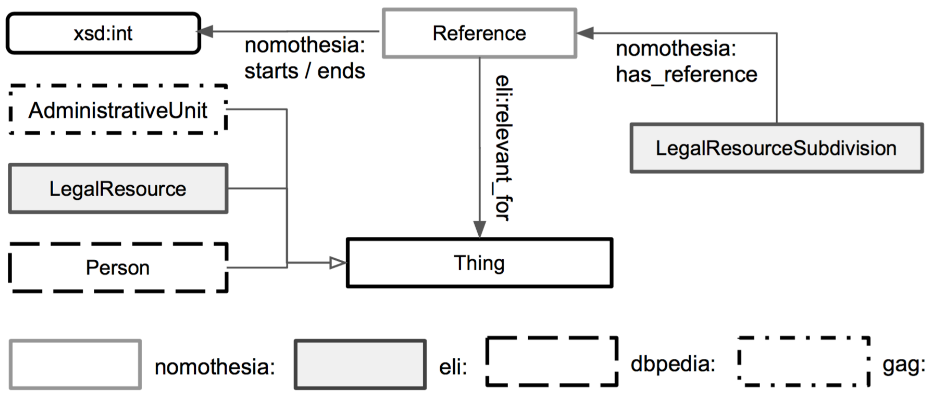

Fig. 4. Textual references vocabulary

4 Building the Greek legal knowledge graph

In this section, we document the entire pipeline process we follow in order to produce the current version of the Greek legal knowledge graph. Since it consists of many and complex components (see Figure 5), we need to properly introduce each one individually.

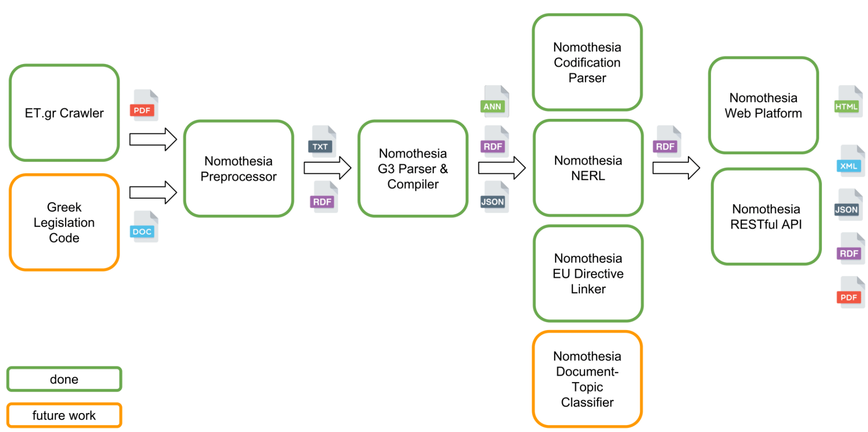

Fig. 5. Knowledge Graph generation pipeline

4.1 Constructing the core knowledge base

The first step towards composing the graph starts by building the core knowledge base. To begin with, we developed a web crawler that runs automatically at timed intervals with the aim of obtaining the latest official government gazettes from the National Printing House of Greece website, in order to process them and augment the knowledge graph.

However, since the initial stages of this work, the need for an advanced automated legislative system in Greece became apparent. The fact that the government gazettes are only published in PDF format, many times without compliance with the specified encoding, poses even greater challenges. Going a step further, taking into consideration the lack of a standard encoding for the codification of repeals and amendments, the recognition of such legislative modifications becomes even more sophisticated.

So, initially, we had to built an extensible and robust preprocessing component to overcome all those complications. This component aims to transform the double column PDF documents in one-column plain text, split the text in individual legislative documents (one document per law) according to specific rules, segmenting of laws into zones (introduction, main body, signatories) and finally analyze the beginning of each legislative document to extract its primary metadata (i.e., title, dates, issue number etc.). The output of this component is an annotated document with all the aforementioned data, directly pipelined to the next main subsystem, the Greek Government Gazette (G3) Parser.

At this step, the main legal corpus of our knowledge graph is being built. G3 Parser analyzes the previously annotated document by following a context-free grammar expressing the encoding of Greek legislation, as described in [9]. During this phase, G3 parser is able to parse the textual content to identify the main structural parts (i.e., Sections, Chapters, Articles etc.), organize it into fragments (e.g., articles and their paragraphs, cases and passages) and generate the corresponding RDF triples following the knowledge graph's ontology described in Section 3. It is important to emphasize the fact that the G3 Parser of this verion of the knowledge graph is a more refined version over the original parser we utilized in our original work.

Finally, the Codification Parser component performs an even deeper level of analysis focused on bottom-level fragments (i.e., passages), aiming to identify legislative amendments (i.e. addition, replacement) along with the modification type, the targeted element and the new element (text). Working on such a low level of text representation and with proper rules, we are able to detect a considerable amount of these modifications. However, as related works have also mentioned [8,15], the detection of these legislative modifications in text is a demanding task so we are experimenting with further state-of-the-art techniques (see Section 8). Again, this component is improved over the original version.

4.2 Named entity recognition and linking

The previous steps are necessary, not only for vizualization purposes and enabling us to extract very specific legal information, but also to accommodate named entity recognition, a task that will generate valuable information for the knowledge graph and which was not present in our previous work. Each passage is fed into a deep neural network (BiLSTM-BiLSTM-LR to be more specific), alongside the shape and Word2Vec representations of each token. The network predicts if a token (set of tokens, respectively) represents an entity. For our purposes, we extract persons, geopolitical entities, geographical landmarks, legal references, public documents and organizations.

After obtaining the output, a thorough post-processing must take place. Labels for entities must be normalized, abbreviations must be extended, irrelevant text should be removed and so on. That is a necessary step, since it allows the generation of unique entity nodes in the graph, but to do so, entities of different legal texts must be correlated, provided they have similar types and labels.

The next step involves interlinking of open public datasets with the data we produce by using Silk [23]. For our purposes, we interlink the entities we have discovered with the Greek DBpedia politicians and Greek administrative geography datasets. The former consists of a subset of Greek DBpedia with Greek politicians and some metadata about them (i.e., their depiction, place and year of birth, political parties they belong to and those parties affiliation to European political parties). The latter consists of geospatial information about the various administrative divisions of Greece, including their polygon geometries and was made publicly available by our group. The interlinking process uses substring (or Levenshtein, in some cases) similarity to discover *owl:sameAs* links on the knowledge graph.

Furthermore, it is important to explain the process of generating the geographical landmarks dataset, as it is non-trivial. As explained above, a category of entities extracted from legal text is geographical landmarks. These are unofficial locations which do not exist in any dataset, at least in a consistent or useful form. Areas, roads, farms, local districts and beaches are some of those categories (we have a strict set of categories of interest as shown in Figure 3), while we also make sure to introduce sub-types, when necessary. In order to render them useful, however, it is still necessary to pinpoint the administrative division each geographical landmark belongs to. Geographical entities encountered throughout text and are in a "neighbourhood" that also contains a geopolitical entity produce a *belongs_ to* link. Although this generates links, it is likely that a geographical landmark might have some of its true neighbors in one occurrence and some in another (e.g., a beach is referred to be in Attica, while in another text it is referred to be in Pireus). This is addressed by "merging" all *belongs_ to* links for each entity, across all its occurrences. As a result, a single geographical landmark entity is produced, as a representative.

4.3 EU directive linker

Meanwhile, as the interlinking of the other datasets takes place, another component addresses the challenge of linking our legislation to Europe's, a task we omitted in the original prototype. More specifically, we focus on European directives which have been transposed by Greek legislation. To that end, we have extracted a corpus of European documents from EUR-Lex (EUR-Lex provides them translated in Greek), as well as metadata, which we use to produce the EU dataset.

Utilizing the metadata, we can discover the Greek legal document that transposes the corresponding EU directive. In original format, the metadata cannot directly provide a properly encoded reference to Greek legislation. The reason for this is that it mostly contains phrases such as "Newspaper of the Government Issue A", "serial number 3978", "2011-06-16" and "Law", which are being processed with regular expressions in order to properly codify the legal link (for that example, it would be http://legislation.di.uoa.gr/eli/law/2011/3978). As a result, we generate the corresponding links in the graph. Also, we make sure to include transposition information in the knowledge graph. To the best of our knowledge, our work is the first to interlink EU directives with Greek laws and decisions which transpose the aforementioned directives.

5 Querying our endpoint using SPARQL

As explained above, developers can find RESTful resources particularly useful. All legal documents are represented in our knowledge graph with persistent URIs, based on the format `[leg]/eli/{type}/{year}/{id}` ([leg] is an abbreviation for `http://legislation.di.uoa.gr`), while named entities are represented based on the format `[leg]/entity/{type}/{id}`. For example, the first paragraph of article 2 of presidential decree 2013/10 is represented by the URI `[leg]/eli/pd/2013/10/article/2/paragraph/1`. In order to make graph search easier, we also provide a SPARQL endpoint[15], where a user or developer can pose queries against. A few useful sample queries are shown in Table 1. The prefixes defined here for convenience are:

- eli: http://data.europa.eu/eli/ontology#.
  Prefix describing the ELI ontology (and dataset).
- lego: http://legislation.di.uoa.gr/ontology#.
  Prefix describing the knowledge graph ontology (and dataset).
- gag: http://geo.linkedopendata.gr/gag/ontology/.
  Prefix describing the Greek Administrative Geography ontology (and dataset).
- dbpedia-owl: http://dbpedia.org/ontology/.
  Prefix describing the dbpedia ontology (and dataset).

6 Novelties & improvements

At this point, it is important to point out that our approach produces RDF data strictly. As a result, the entire knowledge graph we generate is an RDF graph, while it is also the *only* knowledge base we utilize; every piece of information, representation and ontology exists there. Taking after ELI, we extend it in order to reach a very low level of expressing legal text, by splitting each document into increasingly specific parts (books, parts, sections, chapters, articles, paragraphs,

[15] See http://legislation.di.uoa.gr/endpoint.

Q1: *Retrieve any legal act that refer to municipalities, which belong to the regional unit of Florina.*

Query	**Municipality**	**Act ID**
SELECT DISTINCT ?municipality_name (group_concat(?act;separator=", ") as ?act_ids) ?act eli:id_local ?act_id. ?act eli:has_part+ ?part. ?part lego:has_reference ?reference. ?reference eli:relevant_for ?gpe. ?gpe owl:sameAs ?municipality. ?municipality rdfs:label ?local_district_name. ?municipality a gag:Municipality. ?municipality lego:belongs_to ?reg_unit. ?reg_unit rdfs:label "REGIONAL UNIT OF FLORINA"@en. } GROUP BY ?municipality_name LIMIT 5	"DIMOS PRESPON" "DIMOS AMINTAIOU" "DIMOS FLORINAS"	"leg:pd/1998/310, ..." "leg:law/1997/2539, ..." "leg:law/2013/4109, ..."

Q2: *Retrieve any legal acts that contain references to persons that have been born in Athens.*

Query	**Signer**	**Act ID**

SELECT DISTINCT ?person_name
(group_concat(?act_id;separator=", ") as ?act_ids)
WHERE {
?act eli:id_local ?act_id.
?act lego:published_by ?signer.
?signer lego:relevant_for ?person.
?person a lego:Person.
?person owl:sameAs ?per.
?per rdfs:label ?person_name.
?per dbpedia-owl:birthPlace <http://el.dbpedia.org/resource/Αθήνα>.
}
GROUP BY ?person_name
LIMIT 5
	"G. GENNIMATAS"
	"P. PIKRAMENOS"
	"G. PAPAKONSTANTINOY"
	"D. AVRAMOPOULOS"
		"leg:law/1990/1874, ..."
		"leg:la/1_14.06.2012, ..."
		"leg:la/1_31.12.2009, ..."
		"leg:dec/0546_M.6551_ΑΣ55, ..."

Q3: *Retrieve any legal act and its written code that is being referred to in Law 2014/4261.*

Query	**Act URI**	**Act Code**

SELECT DISTINCT ?ref_act ?ref_code
WHERE {
<http://legislation.di.uoa.gr/eli/law/2014/4261> eli:has_part+ ?part.
?part lego:has_reference ?ref.
?ref lego:relevant_for ?ref_act.
{?ref_act a lego:PresidentialDecree} UNION {?ref_act a lego:Law}.
?ref lego:has_original_label ?ref_label.
}
LIMIT 5
	leg:law/2017/3356
	leg:law/2007/3606
	leg:law/2008/3691
	leg:law/1920/2190
	leg:law/2006/3455
		"ν. 3556/2007"
		"ν. 3606/2007"
		"ν. 3691/2008"
		"κ.ν. 2190/1920"
		"ν. 3455/2006"

Table 1. Entity-based queries

indents, passages, citations). This enables us to properly handle and display the full codification of each legal document, including possible amendments. To extend the knowledge graph, we employ state-of-the-art deep learning techniques to extract entities, both textual and named. Finally, we make our data freely and publicly available to all, ensuring transparency to the best of our abilities and we present enterpreneurship potential due to offering our data in reusable format with a RESTful API, enabling developers to create innovative applications.

Comparing our work against the first prototype, which we thoroughly presented in [9], we had explicitly expressed our interest in:

- Re-engineering the existing parsing system with natural language components for document zoning and entity recognition.
- Interlinking of Greek legislation with European legislation and case laws following the European Case Law Identifier scheme.
- Extending our ontology to capture more complex legislative modifications.

As analyzed in previous sections, we are happy to report that all three goals have been successfully incorporated into our platform.

Regarding the amount of data our knowledge graph consists of, in our previous work we only captured legal documents from the National Printing House of Greece for the years 2006-2015. By designing an optimized web crawler and refining the whole pipelined building process, the range has now became 1990-2018 (about 8.000 Official Government Gazettes). However, the year range could be even wider, if documents before 1990 existed in a more exploitable format for our pipeline. To outline the importance of these improvements, the old RDF graph contained about 1.85 million triples, while the new knowledge graph contains 2.9 million triples regarding strictly Greek legislation, and 5 million triples with all data included (12.000 decisions and 195.000 referencs to entities).

7 Applications on the knowledge graph

The Greek knowledge graph we generate is a research prototype that has the potential to be utilized for enhanced information extraction. Having enriched the graph with named entitities, European legislation and complex links between entities and references, as well as the text structure itself, the applications on our data increase tenfold.

One major application is the Nomothesia platform which hosts the graph. It is of particular interest, since we turn much of the information of the graph into a visual format. The signers of a document, the relevant ministries, potential transpositions with European legislation, the timeline of a law, its fundamental parts, the entities extracted (which are enriched by our successful efforts into interlinking our sources with other third-party datasets), are all shown visually and provide a wealth of information, as a result. This is but one example of how the knowledge graph can utilized. We speculate that the incorporation of the dataset into the workflow of legal experts and professionals, as well as public organizations can greatly improve performance and reduce resources needed for information extraction tasks from legal documents.

Another major application is the project Choronomothesia in which we participate in collaboration with the company Geoapikonisis, which has very good expertise in landscape engineering. The goal of Choronomothesia is to extract geospatial information from legal documents, including maps, tables of coordinates and, if possible, even polygons and geometries that describe land. The task becomes much easier by employing our knowledge graph, since it already has a wealth of information available to exploit. In addition, the incorporation of the landmarks as extracted from legal text, assists in discovering potential links, with heuristics or otherwise, to more detailed geospatial information, as is the goal of the project.

8 Conclusion and future work

We described the process of generating the current version of the Greek legal knowledge graph, hosted on our web platform Nomothesia which is, to the best of our knowledge, the first portal functioning with an extended knowledge graph of Greek legislation. We compared our work to other successful efforts throughout Europe and we pointed out the old graphs's shortcomings, comparing it with the new version, outlining the optimizations and new features added. We strongly believe that the new version is vastly superior and we hope that it will be adopted by the public sector.

Although we are very enthusiastic to see our efforts in designing the knowledge graph and hosting it in Nomothesia come to fruition, we have planned a number of things that could improve it even further. We already consider the potential integration of a chatbot on the platform that will make it more user friendly (allow the public to ask questions via Facebook Messenger, Twitter, Skype), while we work towards implementing a document topic classifier for the legal documents we host using deep learning techniques. We plan to design a question answering system which will support SPARQL generation in natural language, in order to make complex features more appealing to the public. Modification types are a bit difficult to properly detect at all times, so that is another shortcoming we aim to overcome by using machine learning techniques. In order to increase the credibility of our knowledge graph, we plan to enable easy conversion and editing of legal documents by legal experts, so that everything the knowledge graph contains is verified. As explained above, we have already interlinked Greek legislation with European directives, but we aim to improve it even more, by finding new ideas of providing better connections between the two and adding links for treaties and agreements. Also, we eagerly wait for feedback in order to extend the services offered by our platform to improve the working conditions of employees in the legal domain (private and governmental).

Acknowledgements

This research has been cofinanced by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the RESEARCH CREATE INNOVATE call (project number: T1EDK-01000, project name: Choronomothesia).

References

1. Agnoloni, T., Bacci, L., Peruginelli, G., van Opijnen, M., van den Oever, J., Palmirani, M., Cervone, L., Bujor, O., Lecuona, A.A., García, A.B., Caro, L.D., Siragusa, G.: Linking European Case Law: BO-ECLI Parser, an Open Framework for the Automatic Extraction of Legal Links. In: JURIX (2017)
2. Angelidis, I., Chalkidis, I., Koubarakis, M.: Named Entity Recognition, Linking and Generation for Greek Legislation. In: JURIX (2018)

3. Angelidis, I., Chalkidis, I., Nikolaou, C., Soursos, P., Koubarakis, M.: Nomothesia: A Linked Data Platform for Greek Legislation. MIREL workshop (2018)
4. Asooja, K., Foghlú, O.Ó., Domhnaill, B.Ó., Marchin, G., McGrath, S.: Automatic Detection of Significant Updates in Regulatory Documents. In: JURIX (2017)
5. Athan, T., Governatori, G., Palmirani, M., Paschke, A., Wyner, A.Z.: LegalRuleML: Design Principles and Foundations. In: RW Summer School (2015)
6. Barabucci, G., Cervone, L., Palmirani, M., Peroni, S., Vitali, F.: Multi-layer Markup and Ontological Structures in Akoma Ntoso. In: AICOL Workshops (2009)
7. Boer, A., Hoekstra, R., Winkels, R., van Engers, T., Willaert, F.: ^{META}lex: Legislation in XML. In: JURIX: The Fifteenth Annual Conference. London (2002)
8. Boer, A., Winkels, R., van Engers, T., de Maat, E.: Time and versions in ^{META}lex XML. In: Proceeding of the Workshop on Legislative XML. Kobaek Strand (2004)
9. Chalkidis, I., Nikolaou, C., Soursos, P., Koubarakis, M.: Modeling and querying greek legislation using semantic web technologies. In: ESWC (2017)
10. Collarana, D., Heuss, T., Lehmann, J., Lytra, I., Maheshwari, G., Nedelchev, R., Schmidt, T., Trivedi, P.: A Question Answering System on Regulatory Documents. In: JURIX (2018)
11. ELI Task Force: ELI - A technical implementation guide (2015)
12. Filtz, E.: Building and Processing a Knowledge-Graph for Legal Data. In: ESWC (2017)
13. Francesconi, E., Küster, M.W., Gratz, P., Thelen, S.: The Ontology-Based Approach of the Publications Office of the EU for Document Accessibility and Open Data Services. In: EGOVIS (2015)
14. Frosterus, M., Tuominen, J., Wahlroos, M., Hyvönen, E.: The finnish law as a linked data service. In: ESWC (2013)
15. Hallo Carrasco, M., Martínez-González, M.M., De La Fuente Redondo, P.: Data models for version management of legislative documents. J. Inf. Sci. (2013)
16. Hoekstra, R.: The metalex document server - legal documents as versioned linked data. In: ISWC (2011)
17. Hoekstra, R., Breuker, J., Di Bello, M., Boer, A.: LKIF Core: Principled Ontology Development for the Legal Domain. In: LOSW (2009)
18. Montiel-Ponsoda, E., Rodríguez-Doncel, V., Gracia, J.: Building the legal knowledge graph for smart compliance services in multilingual europe. In: CEUR workshop proc. (2018)
19. Opijnen, M.V.: European Case Law Identifier: Indispensable Asset for Legal Information Retrieval. In: Biasiotti, M.A., Faro, S. (eds.) From Information to Knowledge. FAIA (2011)
20. Palmirani, M., Vitali, F.: Akoma-Ntoso for legal documents. In: Legislative XML for the Semantic Web (2011)
21. for Standardization (CEN), E.C.: CEN Workshop Agreement: Metalex (Open XML Interchange Format for Legal and Legislative Resources). Tech. rep. (2006)
22. Stellato, A., Turbati, A., Fiorelli, M., Lorenzetti, T., Schmitz, P., Francesconi, E., Hajlaoui, N., Batouche, B.: Towards the Assessment of Gold-Standard Alignments Between Legal Thesauri. In: JURIX (2018)
23. Volz, J., Bizer, C., Gaedke, M., Kobilarov, G.: Silk - A Link Discovery Framework for the Web of Data. In: LDOW Workshop on Linked Data on the Web, WWW2009 (2009)
24. Wyner, A.Z., Gough, F., Lévy, F., Lynch, M., Nazarenko, A.: On Annotation of the Textual Contents of Scottish Legal Instruments. In: JURIX (2017)