2018

"Peri Nomou" System: Automated Codification and Interrelation of Legal Elements Based on Text Mining

Zoi Lachana
University of Aegean, zoi@aegean.gr

Michalis-Avgerinos Loutsaris
University of Aegean, mloutsaris@aegean.gr

Charalampos Alexopoulos
University of Aegean, alexop@aegean.gr

Yannis Charalabidis
University of Aegean, Greece, yannisx@aegean.gr

Recommended Citation

"PERI NOMOU" SYSTEM: AUTOMATED CODIFICATION AND INTERRELATION OF LEGAL ELEMENTS BASED ON TEXT MINING

Research full-length paper

Track New Directions for Digital Governance: Towards Government 3.0

Zoi Lachana, University of the Aegean, GR, zoi@aegean.gr

Michalis-Avgerinos Loutsaris, University of the Aegean, GR, mloutsaris@aegean.gr

Charalampos Alexopoulos, University of the Aegean, GR, alexop@aegean.gr

Yannis Charalabidis, University of the Aegean, GR, yannisx@aegean.gr

Abstract

One of the most promising developments comes with the use of innovative technologies and thus with the availability of novel services. The combination of text mining with legal elements may contribute to the development of many innovative legal information systems. Moreover, in the case of public administrations and governments, the distribution, availability, and access towards legal information are essential and urgent. On the other hand, legal data and law texts are a potential open Government data category in order for innovation to be achieved, regarding the development of new, better, and more cost-effective services for citizens. Those data need to be available 24/7 basis and compliant towards a standard. Yet, there exist some severe issues at the moment regarding this access. This, in turn, makes the use of automated crawling and analysis more than difficult. This paper describes the "Peri Nomou" (about law) system: an innovative legal information system for Greek laws utilising text mining techniques to indexing legal documents, identifying correlations and dividing legal documents into their articles. The first version of the system has been evaluated by legal experts and the second version is developed based on the previous evaluation and presented in this paper. The results from the evaluation indicate the significance of the "Peri Nomou" system for the legal experts and allow us to promote the Peri Nomou system to other user groups, such as business, public administration.

Keywords: Parsing Legal Texts, Legal Text Mining, Laws / Connecting Graphs, Legal Elements Interconnections, Automated Codification, Automated Interrelation.

1 Introduction

Information and Communication Technology (ICT) being triggered by the vision of different concepts, such as digital transformation, smart cities, data-intensive policy making etc., provide entirely new opportunities for public sector and societies in order to solve major problems and issues of everyday life. The new domain of Government 3.0 which combines established ICTs with the use of disruptive ICTs (Pereira, et al., 2018), such as artificial intelligence, is trying to develop new innovative solutions targeting the creation of public value and quality of life for all citizens. The developed solutions will empower and engage the citizens of a state towards the achievement of this target (Nakamura, Nobuoka, & Shimazu, 2008; Bui, Nguyen, & Ho, Towards a Conceptual Search for Vietnamese Legal Text., 2014) mentioned the importance of the ability to structure and extract information automatically from legal texts. This ability will support experts to establish complete and consistent laws and will assist societies to understand laws. For this reason, governments promote the development of legal information systems, from which legal documents indexed, as modernization of public sector (Casanovas, Palmirani, Peroni,

van Engers, & Vitali, 2016). Furthermore, users' expectations have increased because of the ongoing development of new information systems (Bhardwaj & Madhusudhan, 2016).

The legislation of every state is very important to be available online in order for every part of the society to have free and unhindered access (Fulton, 2011). Legal texts are a kind of documents with specific characteristics, which is different from other daily-used documents due to their length and complexity (Nguyen, Bach, & Shimazu, 2011; Nguyen, Bach, & Shimazu, 2011). The complexity of the legislation is identified in the comprehending the legal terminology, in other words, legal language, in combination with the existence of correlations between laws as a reference. These references may modify one or even more different laws and, in many cases, we have to continue with an in-depth search of previous years' laws to detect which components are in force (different components from different laws may consist a valid regulatory). In Greece, National Printing Office (NPO) is the only official responsible organization for printing and publishing legal documents whenever a new law is voted by the Parliament. NPO offers to citizens legal documents as pdf file through a modern user-friendly website. There are different types of legislative documents such as laws, presidential decrees and ministerial decisions and every type contains a specific form including specific structural elements (components). In our research, we examine laws and their components which are chapters, articles, paragraphs, subparagraphs etc. Generally, the legislation network is a multi-relational network that accommodates the hierarchy between the sources of law and can represent relationships of various categories between legal documents (Koniaris, Anagnostopoulos, & Vassiliou, Network Analysis in the Legal Domain: A complex model for European Union legal sources, 2017). The kind of this reference is called dynamic reference. There are also static references which affects only a specific law (Winkels, 2015). The complexity of Greek legislative documents is intense as the usage of dynamic references is more often since static references are being used only in specific types of legislative texts such as presidential decrees and ministerial decisions. Furthermore, modifications to laws or parts of a law are being accomplished by the publication of a new one including the modification type (is substituted by etc.), the unique number of legal document and a description of applied changes (Koniaris, Anagnostopoulos, & Vassiliou, Network Analysis in the Legal Domain: A complex model for European Union legal sources, 2017).

This study develops an innovative legal information system by utilising the power of text mining towards the solution of critical problems faced by professionals trying to comply with the relevant legislation. Our contribution is to develop a useful, free and innovative law system for the Greek society with the aim of facilitating the citizens with the potential to enhance existing law systems' functionalities since the ability to retrieve automated interrelations of Greek legal elements has not been resolved. Our goals in our research were first to ensure that there is a capability of creating a system as mentioned above and secondly the offered functionalities to be as useful as possible. The rest of this paper is organized as follows. The next section describes the followed methodology. Section 3 presents existing systems and methods for automated law analysis. Then, the main elements of our proposed system are presented in Section "Peri Nomou: A System for analyzing and interrelating legal document in Greece", describing the various capabilities of the ICT platform that has been developed to enable this approach. An application scenario of the model on Greek Laws is presented in Section "Application", followed by its results. Finally, Section 6 concludes the paper by raising issues for further research.

2 Methodology

The methodological approach for the development of the "Peri Nomou" (about law) system was based on the content analysis (Krippendorff, 2012; Charalabidis, Alexopoulos, & Loukis, A taxonomy of open government data research areas and topics, 2016).In particular, it consisted of the following five steps as shown in Figure 1.

As a first step, Greek laws by the Government Gazette have been identified and acquired. All the document details have been analysed and mapped including the exact method of correlations in laws along with modification types and their unique number as it is presented in Table 1. Greek laws are publicly

available as open data to the wider society and they are offered as pdf documents in Greek language. Greek laws before 1982 use to have a different form of writing (a conservative form of the Modern Greek language) than modern Greek in Greek alphabet so we decided to exclude this form from our research and for the years between 1983 until 2017, we examined two legal documents per each year. We 've recovered three different types of law documents, which we have categorized by their published year (<2000, 2000-2005, 2005<) and we have written down more than twenty-four modification types (is abolished, is substituted, followed by etc.). This categorization was made since laws before the year 2000 are scanned images, while the other two have different encoding. The outcome of this step was a set of terms, which were used for constructing the first version of Peri Nomou system in step 3.
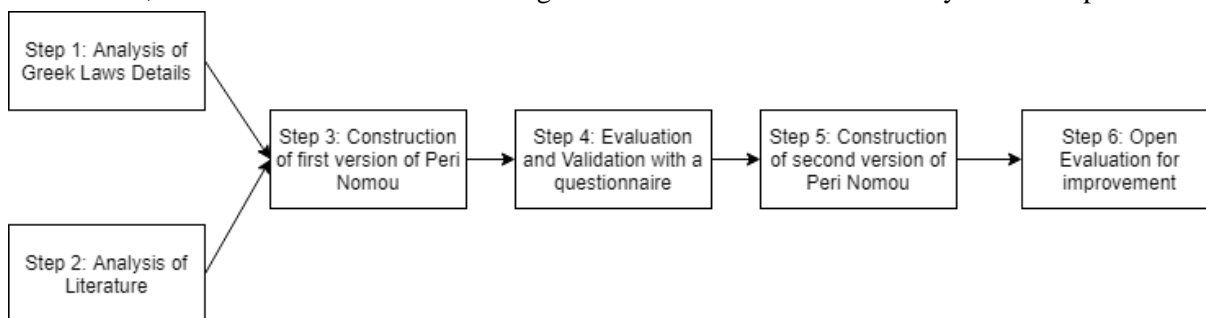

Figure 1: Steps of PeriNomou development method

In step 2, previous implementations proposing the use of text mining in legal elements have been identified and analysed. Obstacles and barriers coming from the development and utilisation of such a system have been also identified through the analysis of worldwide solutions and information systems presented in section 3. Google Scholar and Scopus libraries were used to search these keywords: "legal text mining", "legal elements text mining", "legal information system" and "legal element correlations". The outcome of this step was a set of possible techniques of using text mining in laws which were also used for the construction of the first version of Peri Nomou system in step 3.

Step 3 includes the development and testing of the Peri Nomou system that is capable of retrieving and visualising laws' correlations in an automated way.

Step 4: A questionnaire was designed for the evaluation and validation of the first version of the "Peri Nomou" system aiming at the assessment of its main functionalities as well as the determination of additional functionality in order to meet user requirements. The questionnaire (is analysed in more details in section 5) is being sent to a number of professional users to use the prototype system.

Step 5: The second version of "Peri Nomou" system was developed based on the beta testing and the feedback collected from the questionnaire. This version of the system is described in section 4.

Step 6: Finally, the evaluation questionnaire will remain open for any citizen willing to improve the prototype.

3 The state of the art in legal text mining

One of the most challenging, but also most potential developments, comes with the web of data (Auer et al., 2007) and the inherent mass of freely-available information, i.e., open data (Zeleti, Ojo, & Curry, 2016). Legal data and law texts is a category of open government data and thus they possess the potential to unlock innovation in governments and businesses, regarding the development of new, better, and more cost-effective services for citizens (Zuiderwijk & Janssen, 2014). In case of public administrations and governments, the distribution, availability and access towards legal information is crucial. Yet, there exist some severe issues at the moment regarding this access. One of them is found in form of available APIs, which are not always up and running on a 24/7 basis, paired with slow systems and often non-

compliant data towards standard or even self-issued schemata. This in turn makes the use of automated crawling and analysis more than difficult (Charalabidis, et al., 2018).

Text mining, also known as text data mining, intelligent text analysis (Gupta & Lehal, 2009; Erhardt, Schneider, & Blaschke, 2006) or knowledge discovery from textual (structured) databases (Tan, 1999; Feldman & Dagan, 1995; Delen & Crossland, 2008), has been defined as "the discovery by computer of new, previously unknown, information by automatically extracting information from different written resources" (Hearst, 2003). Generally, refers to the process of extracting interesting and non-trivial patterns or knowledge from unstructured text documents (Tan, 1999; Gupta & Lehal, 2009). Legal text mining analyses legal texts in order to extract useful legal information such as an overview of text's content (Merkl & Schweighofer, 1997). Legal text documents are being unstructured stocked except in cases that online legal databases provide an easy access to citizens, businesses etc. Furthermore, as pointed out by (Hearst, 2003) legal text documents are stored using natural language, so text mining can be suitably used for efficient analysis of such documents.

This section presents the current state of the art in legal text mining by identifying and analysing the current projects and initiatives of the domain. During the last decades, a variety of applications have been carried out by governments and organizations as a solution to the complexity of legislation (Moens, 2001). These efforts have been stepped up in order for a legislative automated codification and by extension automated (?) legal text retrieval to be achieved. Artificial Intelligence techniques for information retrieval constitutes an important component in legal science (Harris, 1979; Moens, 2001; Stranieri & Zeleznikow, 2005).

The Openlaws.eu is a European project (Lampoltshammer, Sageder, & Heistracher, 2015; Winkels, 2015) and a next-generation compliance and legal information system providing linking legal information. Compliance in this project means the adherence to legal regulations, regulatory standards and the fulfillment of other standards and requirements, which a company voluntarily agree to. At this moment, this system integrates legal databases from Austria, Germany, and the EU and makes it possible to search multiple databases at once (Wass, 2017). In other words, Openlaws.eu is a network of legislation, case law, legal literature and legal expert which automatically collects data from different sources, specifically EUR-Lex, Legal Information of the Republic of Austria and all of the German federal laws which provided from the German Federal Ministry of Justice and Consumer Protection. Openlaws.eu currently offers some interesting tools which help users to increase their productivity, such as Highlighting, Personal Folders, Automatic Searches, Working in groups etc. Users can use highlighting to emphasize what's actually relevant from unstructured text, personal folders to collect laws that interest them and automatic search to constantly monitor specific terms and topics. One more tool that is interesting for users' experience is the working in groups. This tool allows users to share folders and comments with the general public or with specific users.

Bui T. D et al. (2014) presented a rule-based approach in order for a search engine, which is deployed as an Information Retrieval system, based on Vietnamese legal documents, to be created. The mentioned approach includes two phases and aims to retrieve most relevant legal documents. The first one is to recognize the logical structure of legal elements and the second to build a legal domain ontology. A similar approach is mentioned by Bui T. D. & Ho Q. B. (2014) to solve semantic problems faced in the recognition logical parts in legal elements. In both approaches authors are using GATR framework and a set of manual rules.

Nomothesia (meaning legislation in Greek language) is a web platform tailored to the particularities of Greek legislation. It is based on European standards and best practices of recent years using semantic web and including a development of an Owl ontology called Nomothesia ontology (Chalkidis, Nikolaou, Soursos, & Koubarakis). The Nomothesia ontology is used for modeling the structural information of any Greek legal document including their accompanied metadata (such as title and publication date). Nomothesia aims to represent any legislative modification made in any legal document to capture the evolution of its content, as well as to formulate advanced queries over the content of Greek legislation.

Among all, as authors mentioned Nomothesia is able to interlink legislation with other linked data sources (e.g. the administrative geography of Greece) to enrich queries.

OpenLawsGR (http://www.openlaws.gr/) is a semi- automated system for the consolidation of Greek legislative texts and it was revealed from Garofalakis J. et al. (2016). Thie system is capable of consolidating historical revisions of legal documents by using natural language processing techniques (specifically regular expressions and pattern matching). As the authors mentioned it is characterized as a semiautomated system, because of the manual process that is required to correct potential errors, due to the original writing process of a law document (e.g. Syntax errors). Furthermore, considering the use of scanned images in law documents before 2004 (the authors referred to the necessity of an OCR tool as we also mention and analyze in section 4), the system is capable of analyzing only legal documents from 2004 onwards, using the UltraCl@rity tool (https://yperdiavgeia.gr/) for the automated acquisition of all legal texts. The authors mentioned that they came across with the Greek language's complexity in their effort including non-Latin alphabets, wrong punctuation, wrong numbering of structural elements e.tc.

Furthermore, in Greece there are three legislative database services, Nomos (https://lawdb.intrasoft-net.com/), Isokratis (http://www.dsanet.gr) and Nomotelia (http://www.nomotelia.gr/). By providing the entire Greek legislation to their users as pdf or word files along with several functionalities such as several categorizations for legal documents, multiple types of search, including correlations either as simple comments or as a pdf or word file of the correlated legal element, they are being used nationally, both by the private and by the public sector. Nevertheless, these databases are only commercial, and both Greek legislation encoding and all these, mentioned above, functionalities are performed manually, thus the Greek language is not an obstacle.

Solon (Solon was an Athenian legislator and the founder of democracy in Athens) consists a platform (Koniaris, Papastefanatos, Meimaris, & Alexiou, 2017) for moderating legal documents by using advanced text mining, modelling and managing techniques. The authors noted that Solon platform is able to extract semantic representations of legal sources, interlink and enrich them by using advanced classification features. Moreover, Solon allows users to connect or even to explore legal resources based on their personal needs, by providing advanced search results utilizing both legal resources' structure and specific features. The power of system's architecture is that its different components can also be exposed to the platform as REST service (RESTful HTTP API).

4 Peri Nomou: System Architecture

"Peri Nomou" is a System for analyzing and interrelating legal documents in Greece. The proposed solution is a Legal Information System providing the capability for identifying correlations and indexing legal documents based on their vocabulary. Peri Nomou system is available on the internet (http://www.perinomou.com) in Greek language and can be used by professional users, such as lawyers, with the aim of accelerating the workload. In this section, we present the architecture of Peri Nomou solution which consists of three layers (Figure 2). The main processes of our system are carried out in the second (service) layer and they are divided into six subprocesses.

4.1 Application Layer

The application layer is the web interface between users and Peri Nomou system in order to gain access to explore and interact with legal information. Users can use any browser supporting Java Servlets either for a successful legal information search to be achieved or for uploading legal documents to support system's community. The client-side web interface is implemented using the PHP scripting language, HTML, CSS, Javascript, jQuery and is designed to be user-friendly and fast, allowing users to view, search data and upload files. Furthermore, the features that can promote Peri Nomou users' interaction

is the three different types of search. The main search capability is that users can ask Peri Nomou for any legal document by typing its unique number and thus they can understand the content because of the generated law tags and, also, they can identify any correlation which is included. Furthermore, users can ask for laws associated with any keyword or for a divided law into its articles.
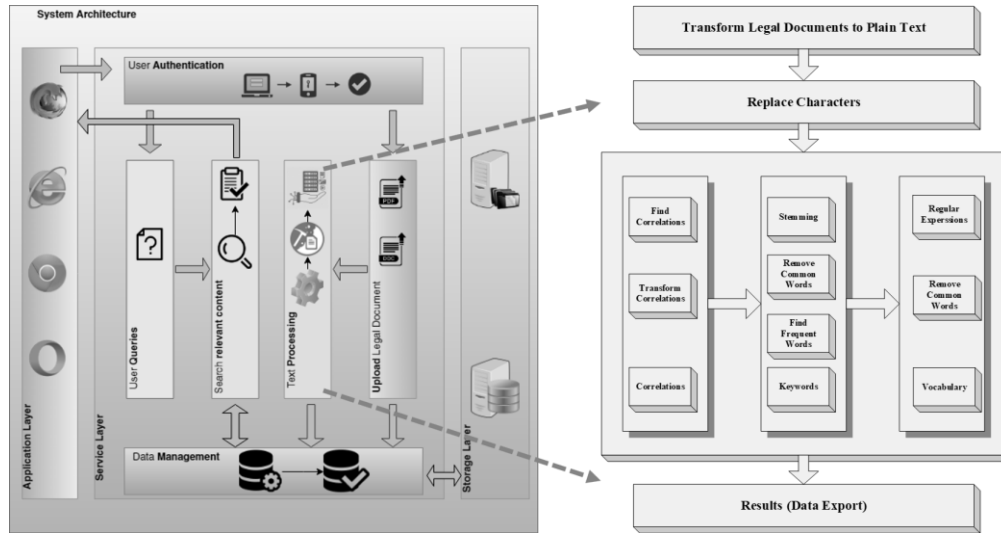

Figure 2: PeriNomou System Architecture

4.2 Service Layer

Service layer has the following subprocesses and the main one is the text processing in which any legal document acquires a structured form.

(a) User Authentication ensures that our system is secured against unauthorized use and attacks by the usage of the SHA-256 algorithm during the validation process. The permissions for any user in User mode and Administration mode are regulated by the administration panel. After registration users can log in to the system and access the system's content.

(b) Upload Legal Document requires processing any legal document that is missing from the database. In this sub-process, users have to upload the legal document of their interest by downloading it, due to the lack of an official stream for automated acquisition of a legal document, such as RSS feed, API etc. As we mentioned above, the only free electronic source to retrieve any law document is the NPO. After uploading a legal document to our system, all file's metadata are, automatically, imported to the database and the "core" file placed in a directory on the file server. Furthermore, legal documents, according to their publication date, are being processed in the corresponding Java Servlet, as we describe below.

(c) Text Processing is responsible for the codification and the identification of correlations of a legal document and, also this process is separated into four main steps:

  i. Transformation of Legal Document to Plain Text: NPO provides legal documents as pdf files so, during this process, we faced the following specificities (Figure 3):
      - The text is divided into two columns
      - Documents include images which are not related to the content
      - Many types of dashes are being used to "cut" a word that continues in the next row.

- Different writing types of articles separation, such as Article 1, First Article.
- Files contain information that are not related to the content of the law, but they are general, such as information about National Printing Office.
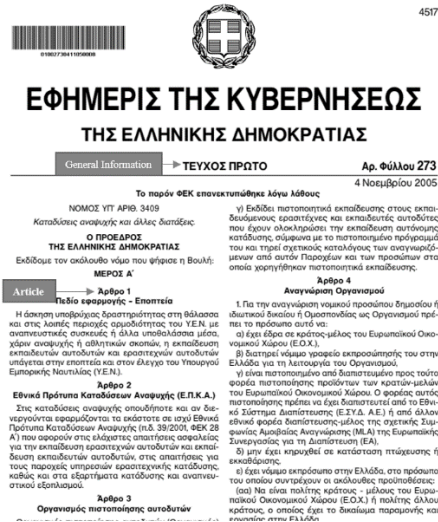

Figure 3: Government Gazette Example

We have managed to eliminate some of these challenges by using RapidMiner operator, which is responsible for reading documents in any file type.

ii. Replace Characters step was necessary to eliminate the problem of using English language characters that are similar to Greek, such as N, O, T etc. Nevertheless, we decide to adjust those characters with the corresponding Greek characters replacement because of the possibility to cause, in our next steps, a sequence of errors.

iii. In the next step, plain text is analyzed, using text mining techniques, to retrieve legal document correlations based on their number format (Table 1). After the completion of the above process, and by identifying and replacing any typing error (such as spaces), correlations are being adjusted into a correct form. Moreover, in this process, we generate tags for laws with the use of text tokenization, stemming and examination of the word frequency using TF-IDF algorithm. Furthermore, Peri Nomou divides any legal document into its articles by using Regular Expressions based on the different writing types, as we mentioned above. Lastly, we keep the vocabulary of the legal document by subtracting the most common words.

Legal Document Unique Number	Rapidminer Identification Formats
3409/2005	([0-9][0-9][0-9][0-9][/] [0-9][0-9][0-9][0-9])
2014/335 /EE	([0-9][0-9][0-9][0-9][/][0-9][0-9][0-9][/][a-zA-Z][a-zA-Z])
L512/3.9.05	([a-zA-Z][0-9][0-9][0-9][/][0-9][\.-][0-9][0-9][\.-][0-9][0-9])

Table 1: Examples of Legal Document Unique Numbers

iv. Results (Data Export) is the final sub-process in which we generate tables for correlations, tags vocabulary and articles.

(d) User Queries is the first step of the search process in PeriNomou system, and users can ask for legal information, according to search types. Any unique number of a legal document as well as any keyword relevant to user's case can be used.

(e) Search Relevant Content is the engine which initially receives the query message from the user and tries to retrieve the matching legal documents. The order of results depends on user query similarity.

(f) Data Management represents the exchange of data between storage layer and service layer during the import of user's register data, legal document files and the text processing results, to the database and fileserver. Furthermore, it is responsible for sending the user queries to the database and forwarding the replies to Search Relevant Content process.

4.3 Storage Layer

The storage layer handles the access to locally stored data objects of PeriNomou System. This component constituted by a MySQL database and a Linux file server according to store legal documents data. The MySQL database stores legal document file metadata in the main table as well as generates tables by text processing sub-process. These include tables of correlations, vocabulary tables and tags tables with the names of legal documents unique numbers. The Linux file server contains all document files. The service layer communicates with storage layer via connectors which are implemented in PHP scripting language for File Server and in Java for Text mining results.

Figure 4: Encoding Government Gazette Problems

As we mentioned above, Peri Nomou system uses three Java servlets according to legal document publication data because of issues relating to their structure. Legal documents which were published before 2001, need a "preparation step" in order to continue with the text processing step as these documents consist of scanned images. In this sub-process, Peri Nomou carries out optical character recognition (OCR) in document, by using Tesseract engine and then unifies exported texts in a new one document and replaces the original in the file server. Nevertheless, Tesseract engine needs training in order for improved recognition in poor-quality images to be achieved as replacement character issues may occur.

Greek Character	Character Issue	Character based on ISO843
ν	v	n
ό	ü	o
μ	ì	m
o	ï	o
I	É	I
υ	1	y
ρ	æ	r

Table 2: Font encoding character issues

Furthermore, Legal Documents which were published between 2001 and 2005 need additional amendments during the Transformation of Legal Document to Plain Text process because of some font encoding issues (Figure 4). Our system replaces each document character based on a matching table (sample on Table 2), and then our plain text proceeds to the text.

5 Application Scenario

In this Section, the operation of Peri Nomou is presented through a case of analyzing a Greek Law, and so all system features are disclosed. An overview of the application process is provided in Figure 6 including the respective interfaces of the platform, while the detailed steps and data are described in the following paragraphs.

In the first step, the user has to complete the registration process to get full access to the features of Peri Nomou system and then to log in with his credentials and redirecting to members' protected page. Furthermore, there is the option to click on "Recover Password" link, in case of a forgotten password. A unique URL is being generated by User Authentication mechanism and sent to the user's email.

5.1 Upload a Greek Law

The next step is to upload the law document at the appropriate publishing period option. Afterward, the user has to fill a simple form, with law's unique number and the file selection of his personal computer, and then he has to click on "Analysis" option to start the analysis process. Nevertheless, the analysis time depends on the length of the document, the CPU usage and the analysis process which follows, as we mentioned in the previous section. Peri Nomou informs the user either about the process completion with the corresponding message "File Uploaded and Analyzed Successfully" or about the existence of the law document at the database with another message.
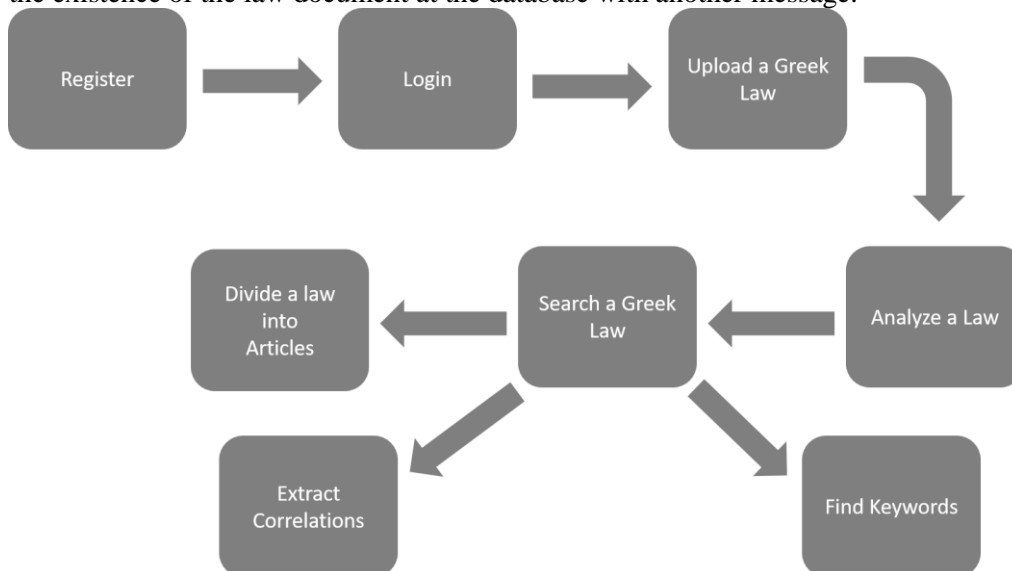

Figure 5: Application Scenario

5.2 Presentation of an Analyzed Law – Search options

As we described above, Peri Nomou provides three presentation types for a legal document analysis after a successful legal document upload. The user has to click on "Search for a Law" button, which is a Main Menu option, and automatically is redirected to Search engine page.

(a) Search for a law based on a unique number: In this mode, the user fills in a field with the law document's unique number, such as "3409/2005", and he clicks on "Search" button. At this time, Peri Nomou retrieving the necessary data from the database through Data Management process and thus prepares them for presentation using the Application Layer. Afterward, the user is redirected to the results page which consists of three components: Law document's unique number with the pdf file of the law is presented at the top of the page, the law's tags are located in the middle of the page and all law documents correlations are displayed in a table after the tags. Furthermore, each correlation is accompanied either by its file which is viewable, printable, and downloadable or by a search in google for the corresponding legal document option (Figure 6).

Figure 6: 3409/2005 Correlations and Tags

(b) Search for a law based on a keyword: Our system has, also, the ability to search for a legal document without knowing the unique number. The Peri Nomou system presents the results of the user's keyword search on a table including the unique numbers of legal documents and their files. This search mode option depends on the files' existence in the database.

(c) Search for a law divided into its articles: This consists an important option because of the legal documents' length and the difficulties it can cause. The Peri Nomou system can present a fully responsive table of a law separated into its articles by searching his unique number in the corresponding search engine. Thus, professionals and citizens can read a law in an easy accessible structured format.

5.3 Peri Nomou System Evaluation

The evaluation of the Peri Nomou System was conducted using the evaluation model described in Alexopoulos et al. (2016) for the evaluation of open government data systems. The model was adapted to the needs of the current study excluding the first layer of efficiency metrics and keeping the effectiveness and future behavior ones. This was done in order to conduct a first quick evaluation towards the measurement of the system usefulness, ease of use, user satisfaction and future behavior.
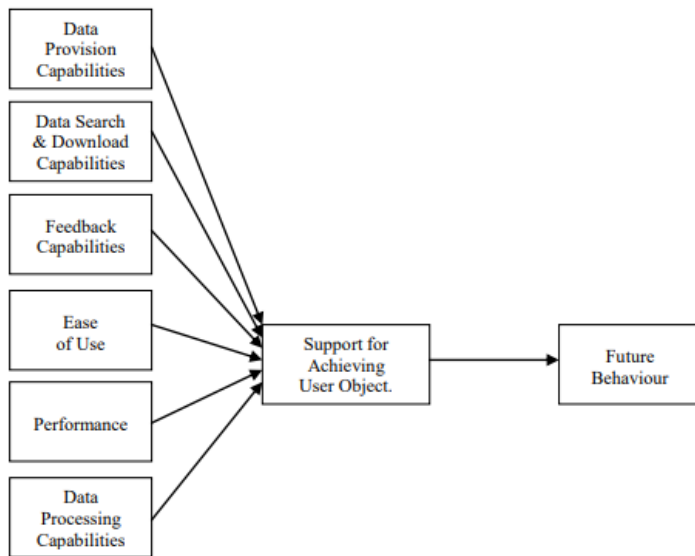
Figure 7 :Original Evaluation Model (Alexopoulos et al., 2016)

A questionnaire was formed in Google Forms using a 10-point scale with 0 indicating "strongly disagree" and 10 indicating "strongly agree" for measured questions. We have also included open-ended questions to help us with suggestions for improvement. More specifically, part one asked professionals for their age and gender, part two asked for the user experience using Peri Nomou, part three asked for the content quality of suggesting solution and part four for system response time.

The targeted sample consists of 254 lawyers, four from each Bar Association (63 Bar Associations in Greece) of which we received 84 answers. The users were asked to first read the manual and proceed with the execution of the application scenario presented in section 5 before they fill in the questionnaire.
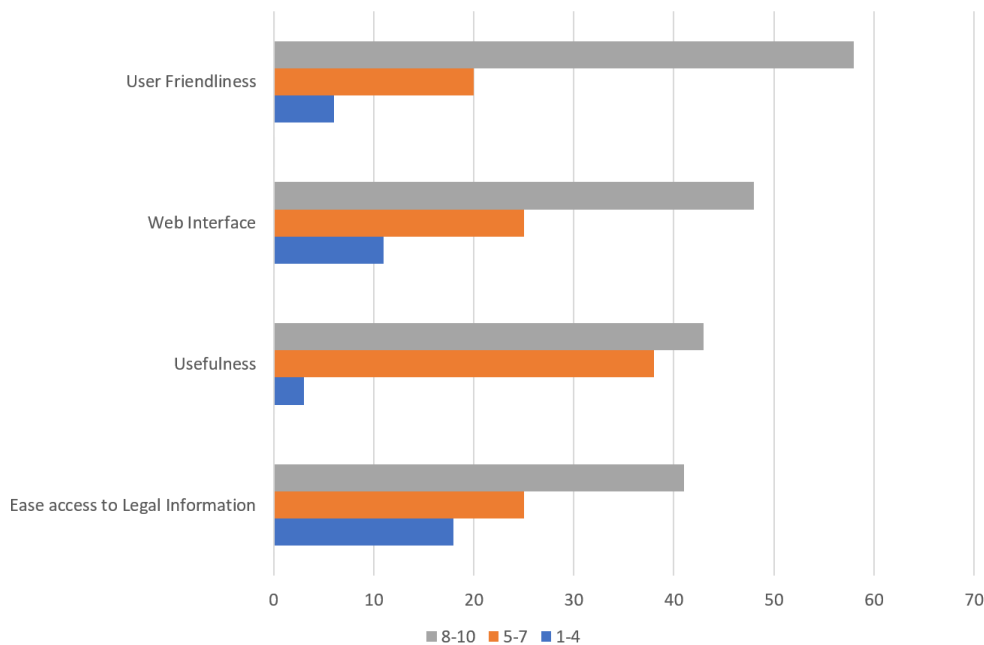

Figure 8: System Evaluation

In conclusion, users have appreciated Peri Nomou system as they referred to the easy understanding of the system without any need of help, the easy access and they also liked the features that Peri Nomou

provides (Figure 7). More specifically, about usefulness, 51% of the lawyers believe that the the system is very useful at their every day work assigning 8 to 10 in a 10-point likert scale; 45% believe that the system is in a medium usefulness level assigning 5 to 7 and a very low percentage (3%) believe that the system provides no or small added value to their every day work assigning 1 to 4 in a 10-point likert scale. The web interface design, 57% of the lawyers believe that the web interface is very well designed assigning 8 to 10 in a 10-point likert scale; 29% believe that the system is well designed assigning 5 to 7 and 13% believe that the system is badly-designed. Furthermore, about the ease access to Legal Information, 48% believe that the system provides a very easy access to legal information; 29% believe that the system provides an easy access and 21% that the system provides a difficult access. Lastly, 69% of the lawyers believe that the system is very friendly; 24% believe that the system is friendly and a very low percentage (7%) believe that the system is not friendly to the user.

It was remarkable that fifty-seven professionals asked to include case laws in the future version. Furthermore, they asked for a community mode in order to take part in uploading their case laws.

The evaluation concerns both the first and the second version of the Peri Nomou System. As we' ve mentioned above, the first version of the system was based on our personal research while the second one is the result of the first evaluation. Specifically, by the use of Google Scholar and Scopus libraries (see Chapter 2), the first version of Peri Nomou System was created, which was capable of retrieving and visualizing law's correlations in an automated way including the keywords functionality. After the first version's evaluation, we proceeded to the second version of our System, including some additional functionalities, based on the results of the above questionnaire. Particularly, the generation of law tags, the division of each law into its articles, the modification of more types of legislative documents (such as presidential decrees and ministerial decisions) and the creation of a pdf file for any law before 2000, were the functionalities which were added in the second version of the system. Finally, as we also mentioned above, the same questionnaire is being used for the evaluation of this system's version.

6 Conclusion and further research

This paper presents "Peri Nomou", a system capable of extracting correlations automatically from Greek legal elements, tags of legal documents, as well as, of dividing a law into its articles. The proper implementation of the system functionalities includes the deployment of text mining techniques the use of an OCR tool and the development of unique font. All of the above were of major importance in order for the Peri Nomou system to be capable of indexing and processing all Greek legislation documents. Finally, a questionnaire has been used in order to validate our system and identify possible improvements. The results of each step of the methodology have been clearly reported in the different sections of this paper. On the other hand, all existing Greek Legal Information Systems offers either a non-automatic solution to identify correlations and to divide a law into its articles or a semi-automatic solution like OpenLaws.gr. Furthermore, all legal documents, which were published before 2001, were converted to documents from images via manual typing.

Our research revealed that mining techniques are convenient for the creation of an automated legal system. From a technical point of view, we can conclude that the creation of a legal system, capable of processing any legal document, at any level, is possible, even in the Greek language. The challenge we are facing lies more in the Greek language rather than in any tool or technique, due to the complexity of legal terminology and, as far as Greece is concerned, in any potential error, due to the original writing system. Moreover, we can easily conclude that in Greece the legal systems are focus more on legal terminology rather than in correlations and that most of them cannot index all the Greek legislation because of the need of an OCR tool. Semantic and syntactic techniques are one of the core components of analyzing a legal document, but information retrieval methods are also important and consist the second basic component (Wagh, 2013). Despite the fact that all systems can reveal the interrelation of Greek legal elements, there is a lack of a detailed reference to them, and as a result, it is still difficult for someone to verify which component of a law is in force.

Even if the evaluation of the system was short, great improvements have been gathered for the next version. An evaluation using the full model of Alexopoulos et al. (2016) is foreseen as a next step including more stakeholders such us accountants and citizens. Further developments already spotted are: (a) to extend our model by including syntactic and semantic techniques; while syntactic techniques are used in order to provide well-defined data structures (for example by morphologically analyzed and lemmatized), semantic techniques will provide information about the nature and scope of inter-chunk dependencies; (b) information retrieval methods augmenting with semantic and syntactic analytic techniques is going to pave the way for the development of an integrated legal system; (c) adding advanced visualization techniques as a very useful extension of a law system and (d) adding case laws in our system which is of major importance for professionals.

References

Alexopoulos, C., Loukis, E., & Charalabidis, Y. (2016). A methodology for determining the value generation mechanism and the improvement priorities of open government data systems. *Computer Science and Information Systems, 13*(1), pp. 237-258.

Bhardwaj, R. K., & Madhusudhan, M. (2016). Online legal information system (OLIS) leveraging access to legal information recourses in Indian environment. *DESIDOC Journal of Library & Information Technology, 36*(1).

Bui, T. D., & Ho, Q. B. (2014, October). An approach for automatically structuring vietnamese legal text. *In Asian Language Processing (IALP), 2017 International Conference on*, pp. 187-190.

Bui, T. D., Nguyen, S. T., & Ho, Q. B. (2014). Towards a Conceptual Search for Vietnamese Legal Text. *In IFIP International Conference on Computer Information Systems and Industrial Management* (pp. 175-185). Springer, Berlin, Heidelberg.

Casanovas, P., Palmirani, M., Peroni, S., van Engers, T. M., & Vitali, F. (2016). Semantic Web for the Legal Domain: The next step. *Semantic Web, 7*(3), pp. 213-227.

Chalkidis, I., Nikolaou, C., Soursos, P., & Koubarakis, M. (n.d.). Modeling and querying greek legislation using semantic web technologies. *In European Semantic Web Conference*, pp. 591-606.

Charalabidis, Y., Alexopoulos, C., & Loukis, E. (2016). A taxonomy of open government data research areas and topics. *Journal of Organizational Computing and Electronic Commerce, 26*(1-2), pp. 41-63.

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Marijn, J., Ferro, E., & Lampoltshammer, T. (2018). The World of Open Data: Concepts, Methods, Tools and Experiences.

Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. *34*(3), pp. 1707-1720.

Erhardt, R. A., Schneider, R., & Blaschke, C. (2006). Status of text-mining techniques applied to biomedical text. *Drug discovery today, 11*(7-8), pp. 315-325.

Feldman, R., & Dagan, I. (1995). Knowledge Discovery in Textual Databases (KDT). *In KDD*, *95*, pp. 112-117.

Fulton, C. (2011). Web accessibility, libraries, and the law. *Information Technology and Libraries, 1*, pp. 34-43.

Garofalakis, J. D., Plessas, K., & Plessas, A. (2016). A semi-automatic system for the consolidation of Greek legislative texts. *PCI*.

Gupta, V., & Lehal, G. S. (2009). A survery of text mining techniques and applications. *Journal of emerging technologies in web intelligence, 1*(1), pp. 60-76.

Harris, J. W. (1979). Law and legal science: an inquiry into the concepts legal rule and legal system.

Hearst, M. (2003). What is text mining. *SIMS.* UC Berkeley.

Koniaris, M., Anagnostopoulos, I., & Vassiliou, Y. (2017). Network Analysis in the Legal Domain: A complex model for European Union legal sources. *Journal of Complex Networks, 6*(2), pp. 243-268.

Koniaris, M., Papastefanatos, G., Meimaris, M., & Alexiou, G. (2017, September). Introducing Solon: A semantic Platform for Managing Legal Sources. *In International Conference on Theory and Practice of Digital Libraries*, pp. 603-607.

Krippendorff, K. H. (2012). Content analysis - An introduction to its methodology.

Lampoltshammer, T. J., Sageder, C., & Heistracher, T. (2015). The openlaws platform - An open architecture for big open legal data. *In Procceedings of the 18th International Legal Informatics Symposium IRIS, 309*, pp. 173-179.

Merkl, D., & Schweighofer, E. (1997). En route to data mining in legal text corpora: Clustering, neural computation, and interational treaties. *In Database and Expert Systems Applications, 1997. Proceedings., Eighth International Workshop on* (pp. 465-470). IEEE.

Moens, M. F. (2001). Innovative techniques for legal retrieval. *Artificial Intelligence and Law, 9*(1), pp. 29-57.

Nakamura, M., Nobuoka, S., & Shimazu, A. (2008). Towards translation of legal sentences into logical forms. (K. Satoh, A. Inokuchi, K. Nagao, & T. Kawamura, Eds.) *JSAI 2007. LNCS (LNAI), 4914*, pp. 349-362.

Nguyen, L. M., Bach, N. X., & Shimazu, A. (2011). Supervised and semi-supervised sequence learning for recognition of requisite part and effectuation part in law sentences. *In Proceedings of the 9th International Workshop on Finite State Methods and Natural Language Processing* (pp. 21-29). Association for Computational Linguistics.

Pereira, G., Charalabidis, Y., Alexopoulos, C., Mureddu, F., Paryced, P., Ronzhyn, A., . . . Wimmer, M. A. (2018). Scientific foundations training and entrepreneurship activities in the domain of ICT-enabled Governance. *dg.o 2018: 19th Annual International Conference on Digital Government Research.* Delft.

Stranieri, A., & Zeleznikow, J. (2005). Information retrieval and Text Mining. *In Knowledge Discovery from Legal Databases*, pp. 147-169.

Tan, A. H. (1999). Text mining: The state of the art and the challenges. *In Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases, 8*, pp. 65-70.

Wagh, R. S. (2013). Knowledge Discovery from Legal Documents Dataset using Text Mining Techniques. *International Journal of Computer Applications, 63*(23).

Wass, C. (2017). Openlaws.eu - Build Your Personal Legal Network. *J. Open Access L., 5.*

Winkels, R. (2015). The OpenLaws project: Big Open Legal Data. *In Proceedings of the International Legal Informatics Symposium (IRIS 2015)*, (pp. 189-196).