Automated Knowledge Extraction from the Federal Acquisition Regulations System (FARS)

Srishty Saha, Karuna P. Joshi, Renee Frank, Michael Aebig, Jiayong Lin
Computer Science and Electrical Engineering Department (CSEE)
University of Maryland, Baltimore County
MD-21250, USA
Email:{srishty1, Karuna.Joshi, rmfrank, maebig1, jiayong1}@umbc.edu

***Abstract*—With increasing regulation of Big Data, it is becoming essential for organizations to ensure compliance with various data protection standards. The Federal Acquisition Regulations System (FARS) within the Code of Federal Regulations (CFR) includes facts and rules for individuals and organizations seeking to do business with the US Federal government. Parsing and gathering knowledge from such lengthy regulation documents is currently done manually and is time and human intensive. Hence, developing a cognitive assistant for automated analysis of such legal documents has become a necessity. We have developed semantically rich approach to automate the analysis of legal documents and have implemented a system to capture various facts and rules contributing towards building an efficient legal knowledge base that contains details of the relationships between various legal elements, semantically similar terminologies, deontic expressions and cross-referenced legal facts and rules. In this paper, we describe our framework along with the results of automating knowledge extraction from the FARS document (Title 48, CFR). Our approach can be used by Big Data Users to automate knowledge extraction from Large Legal documents.**

***Keywords*-Deep Learning, Information Retrieval, NLP, Code of Federal Regulations**

I. Introduction

Rapid adoption of Big Data analytics for decision making, especially of sensitive datasets containing private information, is increasingly coming under the radar of regulatory organizations [1] . The FTC report on Big Data Tools [2] has raised concerns about biased outcomes of big data analytics and potential of violation of the Fair Credit Reporting Act (FCRA) and Equal Credit Opportunity Act (ECOA) by organizations using Big Data tools. EUs new law on data protection - General Data Protection Regulation (GDPR) [3] will be enforced in 2018 and further regulate the usage of Big Data. Hence, it is essential for Big Data providers as well as consumers to be well versed with the regulations on data usage. However, most of these regulations are currently available as large textual documents that contain legal elements and require extensive human effort to interpret and apply. As part of our ongoing work [4][5][6][7][8] we have been looking at approaches to automate knowledge extraction from large legal texts. In this paper, we present our results on analyzing the Code of Federal Regulations (CFR) documents[9]. Regulations issued by the United States Federal Government can be intimidating for novices and veterans because many areas of regulation are voluminous and may be scattered across multiple sections making it hard to cross-reference critical procedures and rules. The Federal Acquisition Regulations, found in Title 48 [10] of the Code of Federal Regulations, are a good exemplar of these problems. Having a firm understanding of the acquisition rules is necessary for individuals and organizations seeking to sell goods and services to the Federal government as well as government employees responsible for running or overseeing government purchases. This set of rules seems an excellent candidate for automation assistance to help find all relevant information related to a particular acquisition activity as well as potentially compare differences between practices across the Federal departments and agencies and to be able to reason over the data to provide answers to questions that users of the regulations might have. The Code of Federal Regulations is available in electronic form [9] on a variety of free and pay-wall sites, but its organizational structure makes it a challenge to find all of the relevant sections that a user may need to review to answer a particular question. Keyword search capability against the text of the code can be helpful, but is dependent on choosing ideal or nearly ideal search terms. Keyword searches may also return vast numbers of possible matches requiring large amounts of human review to analyze and sort the relevant and irrelevant responses. The organizational structure of the data also makes it difficult to find and compare relevant provisions across sections and titles because indexing of the information (through sectional tables of contents) is carried out at relatively high levels within the regulatory sections. The rules involving doing business with the federal government can be complicated. Having the ability to discern the rules that govern the acquisition of goods and services by gathering the sections that describe those rules is helpful to both providers and federal customers. Understanding how acquisitions progress from “requests for proposal (RFP)” to “contract award” and sometimes, “appeal” of an award helps prospective providers participate in trade with the government. Questions that might be asked about the federal acquisition process include: “How many days at a minimum must an RFP be posted/open/available for offers;” “What is the maximum number of days that a RFP may stay posted;” “How is a bidder notified of contract award;” or “How does a vendor appeal an award decision?”. In this paper, we describe our analysis model for legal documents like Federal Acquisition System (Title 48, CFR)[10]. We have created a framework for legal knowledge base that captures vital legal elements, relationships between vital key-terms, relevant sections with respect to user query and deontic expression. This legal knowledge base will contribute in developing question and answer system. Section 3, 4, 5, 6 describes our 3 module required for building an efficient and automated legal knowledge base. Section 7 describes our results and Section 8 describes conclusion and future work

II. Related Work

Representation of legal documents has been an active area of research. The legal documents like Code of Federal Regulations System (CFR) [9] is a long and complex document. The analysis of CFRs requires legal expertise and is a time consuming and labour intensive process. Developing cognitive assistant for such long and complex documents will help novice users as well as legal experts to analyze the legal elements easily and efficiently. Traditional techniques of Natural Language Processing and Informational Retrieval techniques like Bag of words model or vectorized model alone cannot automate the analysis process of legal documents. This is because, it fails to captures the semantic relationships between various legal elements spread across the deep hierarchical structure of legal documents. The Code of Federal Regulations comprises of 50 titles and each title has an average of 8000-10000 chapters, each of those chapters have various sections and sub-sections and each those section/ sub-sections have parts and subparts. This complex structure of CFRs make the analysis of legal elements very difficult. There has been work done in representing CFRs in organizational structure like in XML format. Gloria T. Lau et. al. [11] created a repository in XML format for several accessibility regulations as well as environmental regulations based upon CFRs in a tree hierarchical structure for extracting feature information such as definitions and measurements. But, it did not capture the mismatches between provisions that use same phrases with different meanings in similarity analysis and did not capture relationships between various chapters and sections. Dealing with heterogeneous legal facts and rules in semi-structured format like XML is difficult in terms of answering user queries and performing analysis on various legal element. Hence, building ontologies for legal documents is one of the possible efficient solutions to capture various facts and rules of legal documents in order to perform analytics and answer queries. In our previous work, we have developed a semantically rich ontology for Service Level Agreements and Privacy Policies for cloud based services [4][5][30][31]. The Semantic Web deals primarily with data instead of documents. It enables data to be annotated with machine understandable meta-data, allowing the automation of their retrieval and their usage in correct contexts. Semantic Web technologies include languages such as Resource Description Framework (RDF) [12] and Web Ontology Language (OWL) [13] for defining ontologies and describing metadata using these ontologies as well as tools for reasoning over these descriptions. These technologies can be used to provide semantic relationships between various legal elements of Code for Federal Regulations. Information extraction from text documents have been active area of research. Rusu et. al. [14] used parse trees to generate triplets as subject-predicate-object and author in the paper [14] showed how to perform “Noun Phase Extraction” . Etzioni et. al. [15] used pattern learning to generate to extracts facts from lrage documents in an unsupervised manner. Various textual information extraction and retrieval systems have been proposed in [16][17][18][19]. Another important NLP technique used for information extraction from unstructured text is Noun Phrase Extraction. Use of automated techniques for extracting permissions and obligations from legal documents, such as text mining and semantic techniques have been explored by researchers in the past [20][21][22]. In our previous work, we extracted key SLA definitions and measures from these documents using pattern-based rules using the Stanford PoS Tagger [23] and CMU Link Parser [24] and also used pattern based rules for extracting permission and obligation [6][7]. CFRs titles such as Federal Acquisition Regulation Systems(Title 48, CFR) [10] are much longer and complex documents than Service Level Agreements or Privacy Policies of cloud services, we need to improve and redefine our existing approach for developing an ontology for CFRs in order to capture various facts and rules spread the documents. In Section 3A, we have described our methodology to capture vital key-terms and their semantic relationships. along with this, We have developed a module to retrieve all the sections which are applicable to answering user queries from various chapters. Section 3B describes this module. Also, we have explained the method to categorize retrieve sentences which are relevant to answering a query into Permission and Prohibitions in section 3C.

III. Technical Approach and Challenges

This section describes various modules for our proposed framework for automating the analysis of Federal Acquisition Regulation System [10]. We have used Deep Learning, Natural Language Processing (NLP) and Semantic Web techniques to build a framework for automatic knowledge extraction from Federal Acquisition Regulations System (Title 48, CFRs)[10] for answering basic legal questions. Figure 1 shows our proposed architecture. First,

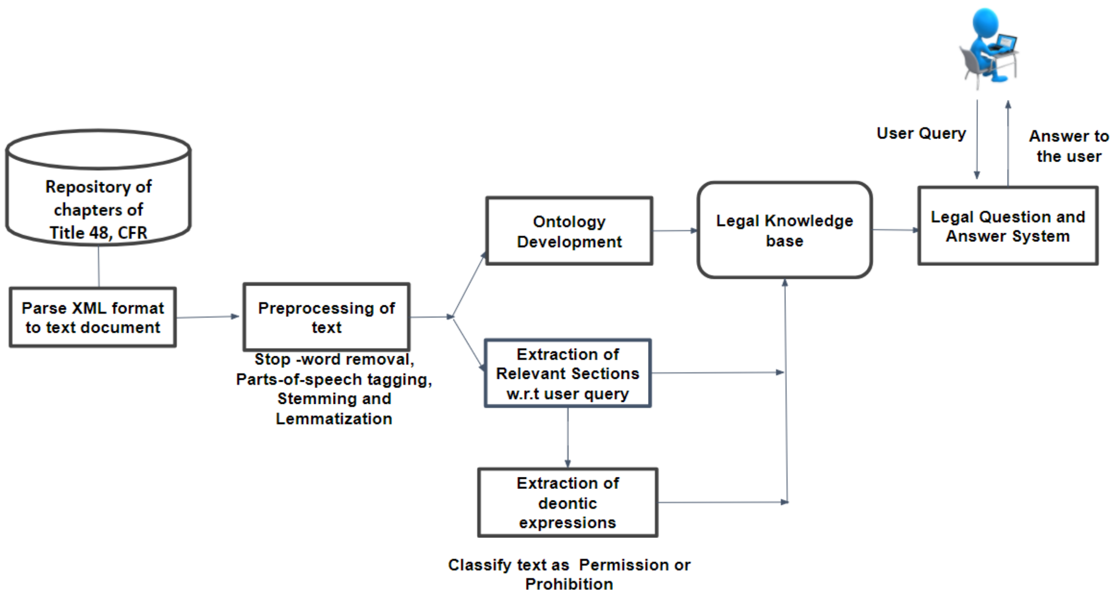

Figure 1: System architecture

we created a repository of various chapters of Title 48 describing the Federal Acquisition Regulation System within the Code of Federal Regulations(CFRs) in machine readable form. Legal documents like CFRs are often long and semi-structured data containing tables and gures. The text portion from these documents is extracted using ElementTree python library [25]. Then, we preprocessed the extracted text using NLP techniques such as conversion to lowercase,removal of stop words, lemmatization and parts of speech tagging. For our analysis, we do not remove certain stop-words like "should" or "must" from the corpus as these might semantically refer to words like "prohibition", "permission" or "authorization" rule which could be useful in resolving the issue of context disambiguation. After preprocessing the text, we used three step approach towards building a framework for an efficient and automated Legal Question and Answer system. Following is the description of our three step approach:

- **Building semantically rich Ontology for Legal Documents** The complex structure of CFRs pose problem in manually figuring-out important key-terms and drawing relationships between those key-terms from the document. We aim to automate the process of extracting vital key-terms from legal documents like CFRs and build an ontological representation of relationships between those key-terms. This section is described in detail in Section 4.
- **Extraction of relevant sections for a user query** Federal Acquisition Regulations System (Title 48, CFR) has disparate multiple sections containing facts and rules that might be applicable to answering a question. We intend to extract all those sections where there could possibly be answers to a question. Section 5 provides detailed explanation of this method.
- **Extracting Deontic expressions from the text** In our previous work, we used text mining techniques to extract deontic rules from cloud SLA documents [6]. We will use this technique to extract deontic expressions from the Federal Acquisition Regulations System[10]. For this initial phase, we will classify the deontic expression into basic two categories: Permissions and Prohibitions.This section is described in detail in Section 6.

IV. Ontology representation of legal knowledge base

The Federal Acquisition Regulations System (Title 48, CFR)[10] is a lengthy and very complex document. It has 99 chapters, each chapter has various subparts and for each of those subparts, there are various sections and sub-sections. For building ontology, we need to first extract important key-terms and draw relationships between those key-terms. Development of semantically rich ontology requires legal expertise and manually extraction of key-terms is a timeconsuming and labor intensive process. In addition to this, there are various terminologies which are semantically similar to each other leading to the issue of context disambiguation. Hence, it becomes necessary to automate the process of extracting vital key-terms and semantically similar terms in order to populate and reason over the ontologies and creating

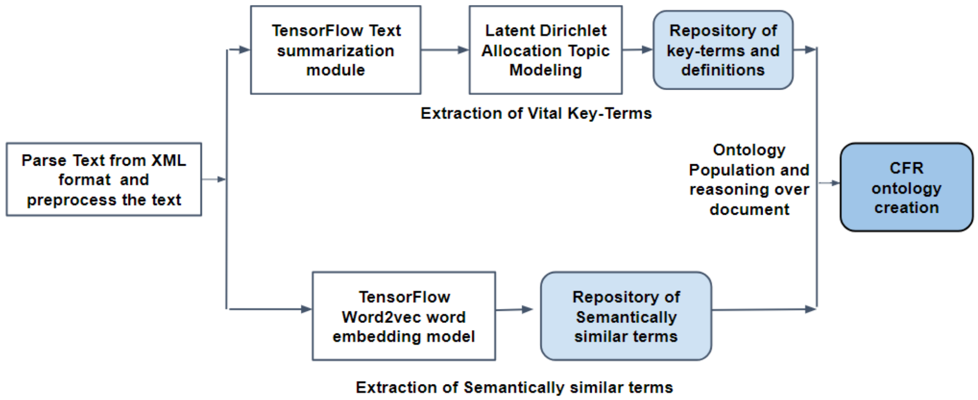

Figure 2: Methodology to extract vital key-terms and semantic relationships between legal key-terms for ontology creation and population.

knowledge graphs. Figure 2 describes the methodology for developing semantically rich ontology.

A. *Extraction of vital key-terms and definitions*

Title 48 of CFR [10] has various sections of varied length representing facts and rules of Federal Acquisition Regulations System. In order to extract important key-terms for legal ontology, we first summarized each section using abstractive Tensorflow's text summarization model[16]. Then, we implemented Latent Dirichlet Allocation model [17] to perform Topic Modeling on the summarized text to extract top k topics from the section. These top k topics will form a set of vital key-terms for our ontology. The number of topics k extracted varies with size of summarized sections. For example, if a summarized section A has 10 sentences then 5 topics will be extracted whereas another summarized section B having 50 sentences will have 15 extracted topics. The extracted topics from each section of Title 48 form a set of vital key-terms which contribute towards forming legal entities for ontology. After creating sets of vital key-terms, we extracted definitions of those key-terms from the document. In our previous work, we developed a topic descriptor system for extracting definitions of topic of importance using patter in based rules like Stanford POS Tagger [23] and CMU Link Parser [24] from Service Level Agreements of cloud based-services [5][7]. We have used the same system to extract definition of vital key-terms from Federal Acquisition Regulations System (Title 48, CFR). The definitions of key-terms will help in extracting relations between those key-terms. The set of key-terms and definitions will eventually be used to form entity-relation model for legal ontology. Section 6A describes our preliminary results.

B. *Capturing semantically similar terms*

In Federal Acquisition Regulations System (Title 48, CFR), there are various chapters which are related to each other and for a novice user, it becomes challenging to co-relate semantically similar terminologies found across various chapters. In order to resolve context disambiguation, we used TensorFlows Word2Vec deep learning architecture [28] [29] to generate word embedding model for capturing semantically similar words. This model is essentially a neural network architecture utilizing a continuous bag-of-words similar words. To build and train the skip-gram model, several parameters need to be decided, which are batch size, num skips and skip window. The skip windows represents the number of words to be considered at left and right. And num skips represents the number of output words will be picked in the span of a single word in a (input, output) tuples. The whole training process is unsupervised learning by using TensorFlow. We used the set of words we are interested in to evaluate the similarity on, every certain steps, the model is evaluated by looking the most related words of those query words. The program stops after 100000 steps, and the loss and and similar words result will be optimized. Section 3B describes the results in detail.
After creating sets of vital key-terms and definitions, with the input from our legal expert and extracted vital keyterms and semantically similar terms from the corpus, we will create an ontology representation of facts and rules contained in the Federal Acquisition Regulation Systems. This ontology will contribute in building legal knowledge base for question and answer system.

V. Extraction of relevant sections for a user query

In CFRs, there are disparate sections across various chapters containing facts and rules relevant to answering

Legal Terms	Definitions
acquisition	acquiring by contract with appropriated funds of supplies, services for the use of the Federal Government through purchase or lease, whether the supplies or services are already in existence must be created, developed, demonstrated, and evaluated
affiliate	associated business concerns, individuals controls one or other
claim	written demand or written assertion by one of the contracting parties
component	any item supplied to the Government as part of an end item or of another component
contract	mutually binding legal relationship obligating the seller to furnish the supplies or services and the buyer to pay for them
contracting_officer	person with the authority to enter into administer terminate contracts make related determinations, findings
conviction	judgment or conviction of a criminal offense by any court of competent jurisdiction
depreciation	charge to current operations that distributes the cost of a tangible capital asset, less estimated residual value, over the estimated useful life of the asset in a systematic and logical manner
debarment	action taken by a debarring official under 9.406 to exclude a contractor from Government contracting and Government-approved subcontracting for a reasonable, specified period
federal agency	executive agency or any independent establishment in the legislative or judicial branch of the Government
information security	protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction
servicing agency	agency that will conduct an assisted acquisition on behalf of the requesting agency

Table I: Extracted vital key-terms and definitions from Title 48, CFR

a question. We need to figure out related legal elements across various chapters automatically. So for a question queried by a user, we first perform query formulation of the question through removal of stop words and by performing lemmatization technique. For example, the query formulation of a question *"What are minimum number of days to file a request in XYZ?"* will be *"mininum, numbers, days, file, request, XYZ"*. Then, next step is to retrieve relevant section from the corpus using Term Frequency Inverse Document Frequency *(tfidf)* and text summarization technique. We used Pythons ElementTree library [25] to read relevant tags from corpus in XML format. Specifically for this approach, we searched for the XML tags such as: $< CONTENTS >$ refer to chapters; $< SECTION >$ refer to Section name; $< SECTNO >$ refers to section number; $< SUBJECT >$ refers to sub-section name; $< P >$ refers to content of the sub-section. We group all $< P >$ tags within a section together and use those groups as documents for the corpus; as we add to the corpus, we discover and maintain the term frequencies. With a specific queries retrieved after query formulation of the question, we calculated the inverse document frequency for each term and document and retrieved the necessary sections. The sections have lengthy content. So, we performed TensorFlows text summarization model on retrieved sections to present overall insight of the section to the user. As the answer to a question is spread across many chapters, so some of the retrieved sections were not directly related to the answer of a particular question. So, in order to improve this, we used the stemmed subject description text $< SUBJECT >$ as documents to the corpus, instead of the bodies of the $< P >$ sections. Section 7B describes the results of this section. The results were evaluated by our legal expert.

VI. EXTRACTION OF DEONTIC EXPRESSIONS

In previous section of this paper, we have extracted vital components of Federal Acquisition Regulations System (Title 48, CFR)such as key-terms and definitions,contextually similar terms and relevant sections applicable to answering a question. Now, we intend to classify the extracted components such as definitions of key-terms and sections into basic deontic expressions. This method is applicable in answering to questions like What are the rules applicable to acquire XYZ commodity?, the answer to such questions should clearly specify four deontic expressions, i.e, Permission (Do's), Prohibitions(Don'ts), Obligations(mandatory Do's) and Dispensation (NonMandatory conditions). For this initial phase, we have classified sentences into Permissions and Prohibitions. In our previous work, we used text mining techniques to extract deontic rules from cloud SLA documents [6]. We have used similar technique to classify sentences into Permissions and Prohibitions, i.e, we implemented the Stanford POS tagger [23] for each of the sentence of in the document comprising of vital components. Next we formulated grammatical rules based on the POS tags to obtain rules in the form of permissions and prohibition. Section 7C describes the results for classification of text into deontic expression. The following the grammar rules we used to classify text into deontic expression:

- Permissions:
  $< Noun/Pronoun >< deontic >< verb >$
- Prohibitions:
  $< Noun/Pronoun >< deontic >< negation >< verb >$

VII. RESULTS

In this section, we describe the results of our approach in automating legal document text analytics. We have developed a system for parsing, preprocessing, extracting key terms and semantically similar terms, extracting relevant sections pertaining to user query and classifying facts and rules as deontic expression, analyzing and reasoning over documents.

Query Keywords	**Analogous Words**
acquisition	acquisitions, procurement, subpart, department wide, purchases
certification	certifications, proprietorships, rationale, approval, balances
debarment	suspension, action, ineligibility, actions, protest, debarring, suspended
request	waiver, obtain, invite, requested, approval, submit, provide
violation	prohibited, breach
patent	invention, application, experiments
publication	document, findings, survey, certification
agency	authority, office, DHS, official
rule	guidelines, terms, provision, regulations
enforcement	memoranda, obligation, legislate

Table II: Semantically similar words extracted from word-embedding model

A. Extraction Results of Ontology Development

In this section, we extracted vital key-terms and definitions and semantically similar terms for ontology population and reasoning over it. For extracting key terms, we used TensorFlow text summarization to summarize each section and Latent Dirichlet Allocation model to extract top *k* topics from each section where k varies with size of sections. Subsequently, by using text mining techniques as mentioned in Section 4, we extracted definition/meaning for key-term. Table 1 shows some of the extracted key-terms and definitions. For extracting semantically terms, TensoFlow's word embedding word2vec model have shown some promising results. For example, for a query keyword like "acquisition", some of the words extracted from the model are "procurement" and "purchase" all of which are semantically similar to each other. Table 2 shows some of the semantically similar terms. We will use the extracted vital keys and semantic relationship between the key-terms for for ontology development. The results of this section were validated by our legal expert.

B. Classifying text into Deontic Expression

Using the grammar based rules as mentioned above, we extracted deontic expressions from text document of Federal Acquisition Regulations System (Title 48, CFR) and classified each sentences into one of the deontic modal logics (mentioned above).

Verbs in Deontic Expression
provide
acquire
issue
require
responsible
assigned
authorized
justify
approve
submitted

Table III: Most used verbs in Deontic Expressions

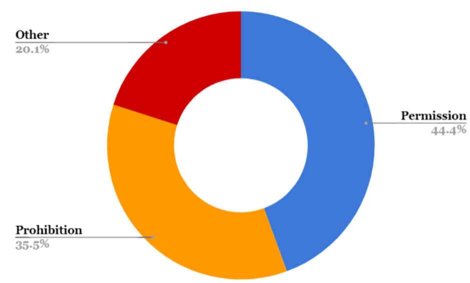

Figure 3: Distribution of deontic expressions across document

We used the following modal verbs for extraction deontic expressions:

- **Prohibition**:should not, must not, shall not
- **Permission**: can, may , could, might

For initial experiments, we tested our approach on first 50 chapters. In total, 2031 deontic expressions were extracted. With the help from our legal expert, we classified each sentences into 3 categories: Permission; Prohibition; Other. Out of 2031 sentences, 902 sentences were referred to "Permissions", 721 sentences were referred to "Prohibition" and rest were referred as Other category. Figure 4 describes the distribution of extracted deontic modal rules. Below are few examples of Permission and Prohibition rule extracted using pattern based rules:

- **Permission**:*"the contracting officer may request from CCC and any other sources whatever additional information is necessary to make the responsibility determination."* [Subpart 209.1, Part 209, Subchapter B, Chapter 2, Title 48]
- **Prohibition**:*"Matters related to legal sufficiency reviews that cannot be resolved between the respective CO and SOL Attorney-Advisor must be submitted* ..."[Subpart 1401.7001-2, Part 1401, Subchapter A, Chapter 14, Title 48]

User Question: "How many days at a minimum must a Request for Proposal (RFP) be posted open/available?"

Query	TFIDF scores	Retrieved Section No.	Section Name	No. of Sentences in the section
many, days, minimum, must, request, proposal, rfp, posted, open, available	5.2743	15.203	Requests for proposals.	38
	3.9557	18.111	Oral requests for proposals	2
	3.2938	14.402	Opening of bids	1
	3.0187	22.1014	Delay over 60 days in bid opening or commencement of work	3
	2.7431	5.102	Availability of solicitations.	31

Table IV: Relevant sections Retrieved from various chapters of Title 48

User Question: "What are the responsibilities of a Contracting Officer?"

Query	TFIDF scores	Retrieved Section No.	Section Name	No. of Sentences
responsibilities, assigned, contracting, officer	4.4455	3.603	Responsibilities of the contracting officer	6
	4.4455	4.403	Responsibilities of contracting officers.	21
	4.4455	4.1704	Contracting officer responsibilities.	3
	4.4455	9.504	Contracting officer responsibilities.	3
	4.4455	7.204	Responsibilities of contracting officer	9
	4.4455	37.103	Contracting officer responsibilities.	14

Table V: Relevant sections retrieved from chapters of Title 48

We analyzed the list of verbs contributing to deontic logic found in legal documents like Federal Acquisition Regulations System (Title 48, CFR). Table 3 shows the list of prominent words used in deontic expressions.

C. Extraction results of Retrieving relevant sections

We have used Term Frequency and Inverse Document Frequency technique and TensorFlows text summarization model for finding related and relevant sections as part of answer to user query from the chapters/sections of Title 48 of CFR. Table 4 describes the results for a user question, **"How many days at a minimum must a Request for Proposal (RFP) be posted (open, available)?"**. Here, 5 sections were extracted from Title 48 of CFR which were applicable to answer the user question. Table 5 describes the results for a user question, **"What are the responsibilities of a contracting officer?"**. Here, there were total 6 sections retrieved from various chapters in Title 48, CFR.

The three modules: Extraction of vital key-terms and definitions, semantically similar terms; Retrieval of relevant sections with respect to user query; Classification of sentences into Permission and Prohibition, will eventually contribute to development of automated and efficient legal knowledge base as it will capture all possible facts and rules found in legal documents like CFRs. This legal knowledge base will used to answer users question related to legal documents. All results presented in Section 7 have been evaluated by our legal expert.

VIII. Conclusion and Future Work

Currently legal documents like Code of federal regulations are presented and analyzed as text documents. The Code of Federal Regulations (CFRs) [9] is a long and complex document. The analysis and retrieval of relevant information across various titles and chapters manually is a complex and time consuming process. In this paper, we presented an approach towards automating the analysis of legal documents through building an efficient legal knowledge base contributing towards legal question and answer. We focused on the Federal Acquisition Regulations System (Title 48, CFR) for our research. We developed techniques to automate the extraction of important key-terms/ definitions, semantically similar terms for ontological representation of legal knowledge base. We have a developed a module to retrieve relevant sections for a user query which would help in answering user questions. In addition to this, we classified text into deontic expressions as Permisssion or Prohibition using pattern base rules. Above mentioned three modules will help in developing a building legal knowledge base. This semantically rich legal knowledge base will be a part of legal questions and answer system.As part of our ongoing work, with the help of our legal expert and by using extracted vital key-terms and semantically similar terms of CFRs, we are in the process of creating semantically rich legal ontology for Code for Federal acquisition. This ontology will eventually be a vital part of our legal knowledge base. Currently, we evaluated the results of retrieval of relevant sections from the various chapters w.r.t user query through our legal expert. In future, we aim to use Legal ontology for evaluations of our results, thus reducing manual process. We also aim to explore other deontic expressions like Obligations and Dispensation. The long terms goal is to build an efficient and automated legal question and answer system.

References

[1] http://www.reuters.com/article/us-eu-regulations-big-data-idUSKBN14816

[2] Big Data A Tool for Inclusion or Exclusion? Understanding the Issues, FTC Report, January 2016, https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf

[3] http://www.reuters.com/article/us-eu-regulations-big-data-idUSKBN14816

[4] Aditi Guptaet.al. Streamlining management of multiple cloud services. in proceedings, IEEE International Conference on Cloud Computing

[5] Karuna P Joshi et. al. Alda : Cognitive assistant for legal document analytics. AAAI Fall symposium, 2016.

[6] Karuna P Joshi et.al. Semantic approach to automating management of big data privacy policies. InProceedings, IEEE BigData, 2016

[7] Sudip Mittal et.al. Parallelizing natural language techniques for knowledge extraction from cloud service level agreements, IEEE BigData, 2015

[8] Sudip Mittal et al., "Automatic Extraction of Metrics from SLAs for Cloud Service Management", 2016 IEEE International Conference on Cloud Engineering (IC2E 2016), April 2016

[9] Code of Federal Regulations [Online]. https://www.ecfr.gov/cgi-bin/ecfr?page=browse.

[10] Federal Acquisition Regulations System. [online]. available. https://www.ecfr.gov/cgi-bin/textidx?tpl=/ecfrbrowse/Title48

[11] Gloria T. Lau et. al. An e-government information architecture for regulation analysis and compliance assistance, 2004.

[12] O. Lassila, R. Swick and others, Resource Description Framework (RDF) Model and Syntax Specification, WWW Consortium, 1999.

[13] D. McGuinness, F. Van Harmelen, et al., OWL web ontology language overview, W3C recommendation, World Wide Web Consortium, 2004.

[14] D. Rusu, L. Dali, B. Fortuna, M. Grobelnik, and D. Mladenic, Triplet extraction from sentences, in Proceedings of the 10th International Multiconference Information Society-IS, 2007, pp. 8 12

[15] O. Etzioni, M. Cafarella, D. Downey, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates, Unsupervised namedentity extraction from the web: An experimental study, Artificial intelligence, vol. 165, no. 1, pp. 91134, 2005.

[16] F. Ciravegna, 2, an adaptive algorithm for information extraction from web-related texts, in In Proceedings of the IJCAI-2001 Workshop on Adaptive Text Extraction and Mining. Citeseer, 2001

[17] S. Soderland, Learning to extract text-based information from the world wide web. in KDD, vol. 97, 1997, pp. 251254

[18] P. Cimiano, S. Staab, and J. Tane, Automatic acquisition of taxonomies from text: Fca meets nlp, in Proceedings of the International Workshop and Tutorial on Adaptive Text Extraction and Mining held in conjunction with the 14th European Conference on Machine Learning and the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, 2003.

[19] J. Cowie and W. Lehnert, Information extraction, Communications of the ACM, vol. 39, no. 1, pp. 8091, 1996.

[20] T. D. Breaux and A. I. Anton, Analyzing goal semantics for rights, permissions, and obligations, in RE05: Proceedings of the 13th IEEE International Requirements Engineering Conference (RE05), IEEE Computer Society, August 2005, pp. 177186

[21] T. D. Breaux, M. W. Vail, and A. I. Anton, Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations, in RE06: Proceedings of the 14th IEEE International Requirements Engineering Conference (RE06), IEEE Society Press, September 2006.

[22] N. Kiyavitskaya, N. Zeni, T. D. Breaux, A. I. Anton, J. Cordy, L. Mich, and J. Mylopoulos, Automating the extraction of rights and obligations for regulatory compliance, in ER08: Proceedings of the 27th International Conference on Conceptual Modeling (ER08), Springer-Verlag, October 2008.

[23] Stanford POS Tagger. [online] http://nlp.stanford.edu/lexparser.shtml.

[24] CarnegieMellonUniversityLinkgrammarparser. [online]. available: http://www.link.cs.cmu.edu/link/

[25] Element python library. [online]. available. https://docs.python.org/2/library/xml.etree.elementtree.html.

[26] TensorFlow Text Summarization. https://research.googleblog.com/2016/08/textsummarization-with-tensorow.html.

[27] DavidM.Bleiet.al. LatentDirichletAllocation,Journal of Machine Learning Research, 2003.

[28] Tomas Mikolov et. al. Efcient estimation of word representations in vector space,https://arxiv preprint arxiv:1301.3781, 2013.

[29] Martin Abadi et. al. Tensorow: Large-scale machine learning on heterogeneous distributed systems, 2016.

[30] K. Joshi et. al Integrated Lifecycle of IT Services in a Cloud Environment, in Proceedings of The Third International Conference on the Virtual Computing Initiative (ICVCI 2009), October 2009

[31] K. Joshi. OWL Ontology for Lifecycle of IT Services on the Cloud, 2010, http://ebiquity.umbc.edu/ontologies/itso/1.0/itso.owl