Semi-automated creation of regulation rule bases using generic template-driven rule extraction

Deepali Kholkar
TCS Research
54B, Hadapsar Industrial Estate
Pune, Maharashtra 411013
deepali.kholkar@tcs.com

Sagar Sunkle
TCS Research
54B, Hadapsar Industrial Estate
Pune, Maharashtra 411013
sagar.sunkle@tcs.com

Vinay Kulkarni
TCS Research
54B, Hadapsar Industrial Estate
Pune, Maharashtra 411013
vinay.vkulkarni@tcs.com

ABSTRACT

Formal approaches to checking compliance manually encode individual obligations from the regulation text as rules. Automated extraction approaches identify key elements in regulatory text, and create annotated, in some cases structured, representations of regulation text. It is desirable to combine the two approaches to automate creation of a regulation rule base that can be used for inferencing and reasoning about compliance. In this paper we present a semi-automated approach that uses a generic semantic model of regulations to guide automated extraction of rule suggestions. The suggestions help domain experts author rules in Structured English using the generic model as template. Rules are translated automatically into a Semantics of Business Vocabulary and Rules (SBVR) model and defeasible logic rules, creating a hierarchical knowledge base that reflects the regulation structure and enables querying and reasoning about compliance.

KEYWORDS

Information extraction, knowledge extraction, knowledge base, automated compliance checking, defeasible logic, Structured English, controlled natural language, SBVR, semantic model

1 INTRODUCTION

Enterprises need to comply with a plethora of regulations. The process of compliance is made more complex by the fact that regulatory bodies publish guidelines pertaining to a single legislation in multiple forms and regulatory documents such as directives, regulations, and annexures containing supporting information such as reporting formats, data descriptions, and example cases. It is a great deal of effort for domain experts to manually compile, correlate, and interpret information from all of these sources and translate it into implementation of compliance.

Several approaches exist for automated legal information extraction that identify patterns and classify information available in unstructured natural language text, annotate it, and in some cases convert it into structured representations such as XML [10, 21, 28, 29]. The resultant rules are however, not in a logic form that can be rigorously reasoned with. Formal compliance checking approaches on the other hand use logic formalisms to represent regulation rules, however, these need to be encoded manually by human experts [26, 27].

Formal approaches in research usually encode a small subset of rules from regulation text and demonstrate compliance to individual rules. Encoding rules manually from the entire natural language text of a regulation is a complex endeavor due to the volume of text, legal language, and abstract nature of guidelines described. A bigger knowledge engineering problem is creating an isomorphic rule base structured such that it is an accurate representation of the regulation, necessary for it to be usable by its users, and also easier to maintain [4]. Getting the rule hierarchy, predicates, and arguments of each rule right, necessary for correct inferencing, presents the greatest complexity in manual rule creation. All of these require the person(s) writing the rules to be an expert in the regulation domain as well as formal logic, which is hard in practice. Building a rule base of an entire regulation thus becomes a daunting task. With multiple regulatory document sources to be considered for compliance, (semi-) automated information extraction become desirable [26, 27]. Even then, building and structuring the rule base from extracted information remains a major challenge.

We carried out case study experiments of building rule bases for two large real-life regulations, viz. MiFID-2 (Markets in Financial Instruments Directive) and KYC (Know Your Customer) regulations [18, 26, 27]. Although we were helped by domain experts in understanding the business domain of each regulation, it was a complex task to encode formal rules from the natural language regulation text, both with and without the help of (semi-) automated extraction [26, 27]. Following our own approach of (semi-) automated extraction [26, 27], it took several iterations before we got the rule hierarchy and parameters of predicates right. Most importantly, it was hard to pinpoint rules modularly in the regulation text. By modularly, we mean that our rule extraction process [27], based on the generation of a domain model and a dictionary [26], enabled us to classify the legal text sentences into those that pertain to regulatory rules and those that do not. Unaware of the greater structure in which a regulatory body organizes the regulations, we ended up overlooking some critical rules that relate to such organization. It is in this context that the work presented in this paper becomes relevant. Examples of these problems are provided in the case study section.

This paper presents an approach to address the challenge of building structured rule bases for large regulations guided by a generic semantic model. We use our approach for (semi-) automated extraction of a domain model, dictionary, and rule suggestions to get to the rules [26, 27]. The generic model also serves as a hierarchical template for creating the rule base, wherein the domain expert fills in extracted information to create rules in a controlled natural language. The template helps create a coherent knowledge base of rules with an inference hierarchy that makes reasoning about higher-level goals

In: Proceedings of the Second Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2017), June 16, 2017, London, UK.

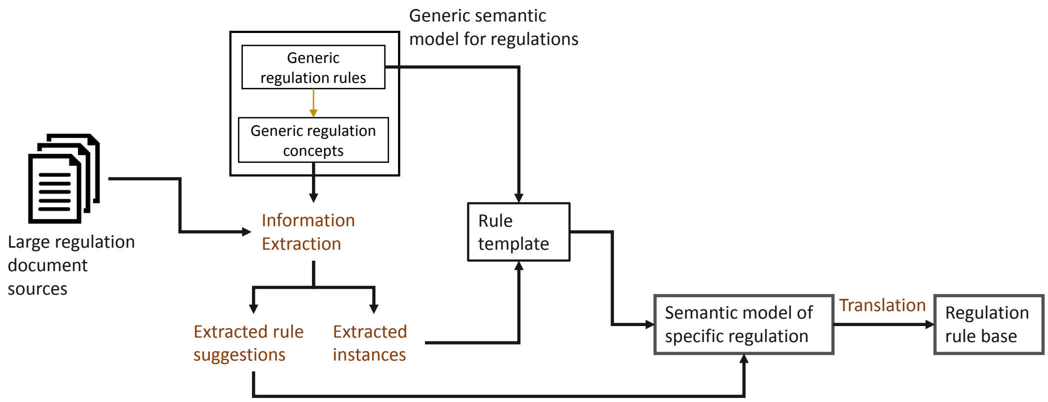

Figure 1: Our method for automated rule extraction and regulation rule base creation

of the regulation possible. Most importantly, the template gives a skeletal structure that ensures inclusion of principal categories of rules. The rules are translated automatically into a Semantics of Business Vocabulary and Rules (SBVR) model[1] and further into a defeasible logic formalism DR-Prolog[1] as we detailed in [18].

Our overall approach is depicted in Figure 1. We first briefly review our (semi-) automated extraction approach in Section 2, followed by the description of the generic semantic model in Section 2. Section 3 describes rule base creation and querying for compliance, Section 4 discusses the utility of our approach, Section 5 describes related work, and Section 6 concludes the paper. We illustrate our approach using a real-life case study from the MiFID-2 regulation applicable in the European Union (EU).

2 OUR APPROACH FOR RULE EXTRACTION

In this section, we first review the generation of the domain model and dictionary. We also go over the creation of a classifier that uses these artifacts to classify legal text sentences into those that contribute to rules and those that do not. We then proceed to elaborate on our approach for (semi-) automated creation of hierarchical rule bases. Note that we only expound the key ideas without restating the results already published in [27]. We proceed by revisiting the motivation behind the domain modeling first.

2.1 Domain Modeling for Regulatory Compliance

In our engagements and interactions with the domain experts from enterprises active in the banking and financial services, we found that the domain experts would encode their knowledge in the form of descriptive artifacts, within which they would establish some form of traceability. But in most cases, the backbone of this activity was a mental model of the regulation, which the domain experts had to somehow corroborate with the artifacts, that the governance, risk, and compliance (GRC) frameworks or the in-house solutions would let them create. However, the solutions did not offer the domain experts a way to formalize their knowledge.

A domain model and a dictionary of the concepts in the model could be used as the central artifacts to drive the compliance process, giving the domain experts a more principled way of managing compliance. Such a domain model would be also helpful, if one were to introduce the benefits of formal compliance checking in an industry setting [25]. This observation led us to come up with a method and a tool for generation of a domain model and a dictionary, detailed in [27], revisited below.

Using Distributional Semantics for Building Domain Model and Dictionary Instead of using natural language processing (NLP) for syntactic analysis of legal text, we chose to use NLP to implement *distributional semantics* in the process of building the domain model and the dictionary. Most of the state of the art NLP approaches in creating domain models or ontologies rely on syntactic features of the tokens in the text [26, 27]. These approaches tend to use heuristics, for instance, every noun phrase is a candidate for a concept, every verb phrase is a candidate for a relation, and every adjective is a candidate for a characteristic of a concept, etc. In our experience, such approaches are feasible, when a) the sentences in the given text are small[2], (b) the sentences possess simple phrasal and clausal structures that do not lead to multiple parses, and c) the overall number of sentences in the text under consideration is few hundreds of sentences. For several hundreds of long and complex sentences[3], which is the usual case in business domains like banking and financial regulations[4,5], we needed to use techniques that did not specifically depend on the syntactic features for constructing the domain models.

We chose to use *distributional semantics* hypothesis [14] to help the domain expert discover the domain model and the dictionary of concepts. The distributional semantics hypothesis states that words that occur in the same contexts tend to have similar meanings. Since

[1] Semantics of Business Vocabulary and Business Rules, http://www.omg.org/spec/SBVR/1.2/

[2] Examples from most of these approaches contain sentences with 5-15 tokens (words). The Penn TreeBank, on which the statistical parsers like Stanford PCFG parser and Malt parser are trained, has sentences with average length of 25.6 tokens [12].

[3] In our *Know Your Customer* (KYC) for Indian Banks case study, we found that average length of the sentences was 31.7 tokens. For the MiFID-2 text, it is 38.27 tokens.

[4] The number of sentences in the text offered online for KYC is 526, while in the MiFID-2 is 4069. The sentences are obtained using heuristics based sentence detection model and do not consider additional text from relevant documents that an enterprise may have to consider for enacting compliance to these regulations.

[5] The KYC and MiFID-2 links are provided at the end of this section.

this hypothesis is independent of syntactic features, the length or the phrasal or clausal complexity of the sentences do not restrict either the scope or the scale of its application. Following observations helped us in designing the implementation of the distributional semantics hypothesis for domain modeling:

- All regulations constrain the interaction of domain concepts in some manner [27]. To do so, the text of the regulations uses mentions of domain concepts. By getting handle on concepts and their mentions, it becomes intuitively easy to understand what the regulation is trying to do and how to specify it [26].
- *Fact-orientation* (FO), a domain modeling method used for constructing vocabularies in SBVR [13, 22], uses the same principle as the distributional semantics hypothesis in its conceptual schema design procedure (CSDP). The very first step in CSDP is *transform familiar information examples into elementary facts*. When performed manually, a modeler essentially strives to check whether the contexts of familiar examples contain some hints to obtain concepts and relations [26].

We refer to occurrences of the instances and the synonyms of the domain concepts as *mentions*. Based on above observations, we compute the spans of texts of a configurable length, around (both to the left and to the right of) the mentions of the domain entities. We cluster the contexts of each concept discovered so far, so as to find its other mentions and the mentions of the concepts, to which it is likely related [26, 27].

The domain expert has the option to provide a seed set of domain concepts and their mentions to the system, generally found in the *definitions* section of the most industry regulations[6], or build the domain model from scratch, starting with a single concept and its mention

Using Informed Active Learning for a Rule Classifier Our choice of active learning technique was motivated by the fact that the active learning process aims at keeping the domain expert annotation effort to a minimum, only asking for advice where the training utility of the result of such a query is high [24].

For the purpose of classification of legal text sentences, it is possible to use features based on various n-grams (n items like letters or words), and part of speech classes like verbs, modal auxiliaries, word couples and so on. Such features do provide acceptable results for detecting arguments in legal text [21]. Instead of such features, we make use of the domain model and the dictionary obtained previously. A dictionary-based feature is activated whenever a mention of a domain concept is found in a given sentence. During the active learning sessions, the role of the domain expert is essentially to provide a judgment over classification suggested by the active learner. The domain expert is queried for the top-k sentences one by one in each session in a console-based application, whereby the domain expert inputs the true class of the sentence queried by the active learner [27].

[6] See the *definitions* section in European MiFID-2 regulations, *Article 4 Definitions* at http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32014L0065&from=EN and the definitions section of Indian Know Your Customer regulations, *Section 2 Definitions* at https://rbi.org.in/scripts/BS_ViewMasCirculardetails.aspx?id=9848

The interested reader is invited to refer to [26] and [27] for experimental results in both the generation of the domain model and dictionary as well as the rule classifier.

From Classified Rules to Structured Regulations In the following section, we describe how this approach of building a domain model and a dictionary, and then constituting a rule classifier on top of these, helps in hierarchical structuring of the regulations. We describe the structure of a regulation and the regulatory compliance problem context, then briefly outline the SBVR standard by Object Management Group (OMG), that we use for creating a semantic model of regulation rules. The rules in SBVR are translated to logic form, that can be used for querying and checking compliance. SBVR allows rules to be defined in its variant of controlled natural language called Structured English (SE). We first create a generic rule model for regulations that serves as a template for the domain expert to construct the rule base for a specific regulation using extracted information, as depicted in Figure 1. The detailed process is described in the next few sections.

2.2 The Regulatory Compliance Context

Regulatory bodies introduce legislation to mitigate *risks* faced by individuals or enterprises. Introduction of new legislation often involves issue of a *directive* that gives abstract guidelines, followed by a *regulation* that makes concrete recommendations for the guidelines. The regulatory body usually also makes available other supporting documents such as *regulatory technical specifications* (RTS), and consultation papers giving guidelines for implementation through data and reporting formats, explanatory use case scenarios, etc.[7].

The directive and regulation, both define *goals* that aim to mitigate risks. The regulation typically applies in conjunction with the parent *directive* if it exists. Regulations always have a well-defined *scope* within which they are applicable. They include detailed *scope rules* defining this scope such as, entities to which the regulation applies, conditions under which it applies, and exemption conditions. They lay down *obligations* for entities that fall within the scope. Obligations are individual regulatory rules that apply to enterprises. Obligations are usually grouped into sections based on the domain functions that they govern. In the prevalent manual practice of regulatory compliance, enterprises that need to comply with the regulation, legal and compliance experts, auditors, and even regulators spend huge effort in understanding and interpreting the contents of regulations in the context of enterprise compliance.

If a knowledge base that encoded all the obligations of a regulation were available, the various stakeholders would be able to query the same, for the purpose of implementing compliance, or to ascertain enterprises' compliance to the regulation. Queries could include: *'What are the goals of the regulation?'*, *'What are the risks it aims to mitigate?'*, *'What is the scope of the regulation?'*, *'What sort of entities does it apply to?'*, *'What are the broad groups of obligations it describes?'*, *'What are the obligations impacting enterprises of type X?'*, *'Given a certain set of data from the enterprise, is it compliant?'*. A knowledge base would make it possible to query compliance to goals or sub-goals at various levels, to the regulation as a whole, or to groups of obligations, as also to individual obligations, at various stages in the compliance process.

[7] MiFID documents

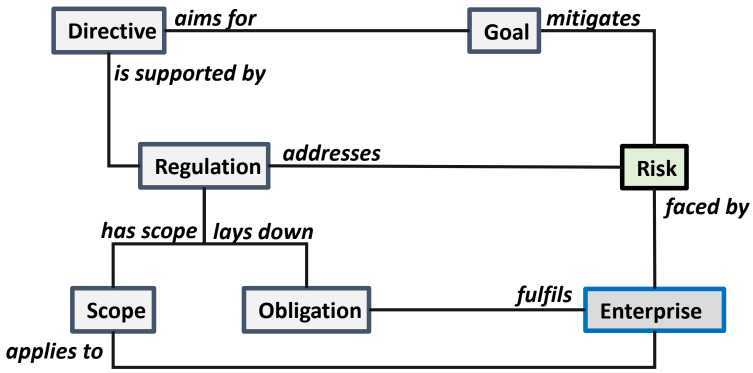

Figure 2: Generic set of regulation concepts

The key elements *goals*, *risks*, *directives*, *regulations*, *scope*, *obligations*, and their relationships define the semantic structure of a regulation, depicted in Figure 2. We define a generic set of compliance rules based on this structure, and term it our generic semantic (rule) model for regulations. We use this generic model for both extraction of rules from the regulation text as well as a template for creating a regulation rule base. The next section describes SBVR and SBVR SE, used to capture our semantic rule model.

2.3 Semantic modeling of rules using SBVR

SBVR is an OMG standard that helps define a semantic model of *rules*, where rules are defined as compositions of *fact types*. Fact types are relations between concepts. This is called a semantic model because the *meaning* of the rule is explicated through its component facts and concepts.

Since SBVR is intended to capture the vocabulary and rules of a *business domain*, OMG provides a controlled natural language notation for specifying the model, called SBVR Structured English (SE)[8]

Rules in SE are written by imposing modalities such as *obligation* and *necessity* onto compositions of fact types. e.g. *It is obligatory that* account *has* balance *if* customer *holds* account. Here, customer, account and balance are *concepts*, and 'customer *holds* account' and 'account *has* balance' are fact types. SBVR SE being a restricted subset of natural language, can be understood and used with ease by domain experts. We use SBVR SE to define some generic rules for regulations, detailed in the next subsection.

2.4 Generic Semantic Model for Regulations

The key elements of a regulation and their relations depicted in Figure 2 are concepts and fact types in SBVR terminology, in other words, the conceptual model of a regulation. We define generic rules for checking compliance based on these concepts and fact types, depicted in Listing1.

Listing 1: Generic rules for compliance

```
rule gen001 It is obligatory that enterprise
    complies_with directive if directive
    is_supported_by regulation && enterprise
    complies_with regulation
rule gen002 It is obligatory that enterprise
    complies_with regulation if enterprise
    falls_within_scope_of regulation && regulation
    lays_down regulation_obligations && enterprise
    fulfils regulation_obligations
rule gen003 It is obligatory that enterprise
    falls_within_scope_of regulation if regulation
    has_scope regulation_scope && regulation_scope
    applies_to enterprise
rule gen004 It is necessary that regulation addresses
    risk if regulation aims_for goal && goal mitigates
    risk
rule gen005 It is obligatory that enterprise
    falls_within_scope_of directive if directive
    is_supported_by regulation && enterprise
    falls_within_scope_of regulation
```

Rules gen001 and gen002 define compliance for an enterprise to the directive and regulation respectively. Rule gen003 evaluates whether the enterprise falls within the scope of the regulation. Rule gen004 defines the relation between goals and risks. These rules define a generic template that can be *instantiated* to create a rule template for a specific regulation, by substituting generic concepts with their instances from the regulation text. The specific rule template is then filled in with rules from the regulation text to create its rule base. Instances of generic concepts, and rule suggestions are found through automated extraction from the regulation text, using the techniques detailed in earlier sections.

We illustrate our approach of rule base creation using a subset of the MiFID-2 regulation. The next section gives a brief description of the regulation.

2.5 MiFID-2 example

MiFID-2 is a directive introduced in the European Union to regulate the functioning of financial markets and bring in greater transparency in their operation, for safeguarding the interests of customers of investment firms. It mainly lays down obligations on investment firms to report trading transactions carried out on secondary markets, to the appropriate authorities, to enable oversight. The directive is

[8]Semantics of Business Vocabulary and Business Rules: Annex A: SBVR Structured English, http://www.omg.org/spec/SBVR/1.2/

supported by the MiFIR regulation, an RTS, and several consultation papers[9].

Broadly, the level of detail increases from directive to regulation to RTS. We used a subset of the text from each regulatory document for our case study. Elements of regulatory information, their document sources, and the criteria applied in picking the subset of document text for the case study are listed below.

- Risks, goals, scope, definitions, and high-level obligations from Directive. The Introduction and Article 1 (Scope and Definition section) from the Directive were used as source text.
- Scope and definitions from regulation. Article 1 (Scope and Definitions) was used as source text.
- Obligations from regulation. Here, we scoped the text by selecting one high-level obligation from the directive and picking the corresponding sections from directive, regulation, and RTS, viz. Article 26.
- Detailed specification of obligations from RTS (Sections with references to Article 26).
- Data definitions from regulation appendix

Sample text from the directive document is reproduced here to illustrate rule and non-rule text. The first paragraph is non-rule text while the second paragraph gives a scope rule.

The financial crisis has exposed weaknesses in the functioning and in the transparency of financial markets. The evolution of financial markets has exposed the need to strengthen the framework for the regulation of markets in financial instruments, including where trading in such markets takes place over-the-counter (OTC), in order to increase transparency, better protect investors, reinforce confidence, address unregulated areas, and ensure that supervisors are granted adequate powers to fulfil their tasks......

Article 1: Scope

1. This Directive shall apply to investment firms, market operators, data reporting services providers, and third-country firms providing investment services or performing investment activities through the establishment of a branch in the Union....

The next section describes automated extraction of these elements from the document sources.

2.6 Rule Extraction from Regulatory Documents

In the first iteration, the generic concepts of Figure 2 and their mentions are given as seed concepts for extraction. These are shown in Listing2. Concepts and mentions are given as *mention:CONCEPT*, and relations as *CONCEPT>relation>CONCEPT*.

Listing 2: Generic concepts and mentions input to rule extractor

```
directive:DIRECTIVE directives:DIRECTIVE risks:RISK risk:RISK regulation:REGULATION regulations:REGULATION
aim:GOAL goal:GOAL aims:GOAL goals:GOAL need:OBLIGATION necessary:OBLIGATION necessity:OBLIGATION requirements:OBLIGATION requirement:OBLIGATION policy:POLICY policies:POLICY regulatory technical standards:RTS
scope:SCOPE obligations:OBLIGATION obligation:OBLIGATION definitions:DEFINITION definition:DEFINITION rule:RULE controls:CONTROL
competent authority:REGULATOR legal framework:REGULATORY FRAMEWORK regulatory framework:REGULATORY FRAMEWORK
enterprise:ENTERPRISE enterprises:ENTERPRISE entity:ENTERPRISE entities:ENTERPRISE organization:ENTERPRISE organizations:ENTERPRISE institution:ENTERPRISE institutions:ENTERPRISE firm:ENTERPRISE firms:ENTERPRISE
DIRECTIVE>aims for>GOAL DIRECTIVE>is supported by>REGULATION GOAL>mitigates>RISK REGULATION>has scope>SCOPE REGULATION>lays down>OBLIGATION ENTERPRISE>fulfils>OBLIGATION SCOPE>applies to>ENTERPRISE RISK>faced by>ENTERPRISE REGULATION>addresses>RISK
```

The rule extractor is run on all the available documents, viz. directive, regulation, and RTS. This brings up rule suggestions from the regulatory texts that contain mentions of these key concepts. Instances of concepts can be found in these, e.g. *MiFID* as instance of directive, *transparency in financial markets* as instance of goal. Examples of rule suggestions that come up in the first iteration are scope rules and high-level obligations, due to the seed concepts *scope*, *enterprise*, *requirements* and their mentions given as input. These are illustrated in Listing 3.

Listing 3: Extracted rule suggestions from MiFID-2 documents

```
// From Directive
rule a5697 It is obligatory that This Directive applies to investment firms market operators data reporting services providers third country firms providing investment services or performing investment activities establishment of branch in Union
rule a1292 This Directive establishes requirements to authorisation operating conditions investment firms provision of investment services or activities third country firms.
// From Regulation
rule a7790 This Regulation establishes uniform requirements to disclosure of trade data to public reporting of transactions to competent authorities trading of derivatives organised venues discriminatory access to clearing discriminatory access to trading in benchmarks product intervention powers of competent authorities ESMA EBA powers of ESMA position management controls position limits provision of investment services or activities third country firms or branch
```

The suggestions are in a format very close to Structured English. The domain expert can use them to write SE rules with very little editing, illustrated in the next section. In subsequent iterations, specific concepts from obligations that need to be explicated further are given to the rule extraction engine, to extract detailed obligations. The next section describes the steps to create the rule base using extracted information.

3 RULE BASE CREATION

Rule base creation using the template and extracted rule suggestions is described here as a set of steps, illustrated in Figure 3.

Step 1: Identify instances of generic concepts Instances of concepts found in the rule suggestions are listed by the experts as instances in the rule base, using is_a facts, as shown in Listing 4.

[9] MiFID-2: http://ec.europa.eu/finance/securities/isd/mifid2/index_en.htm

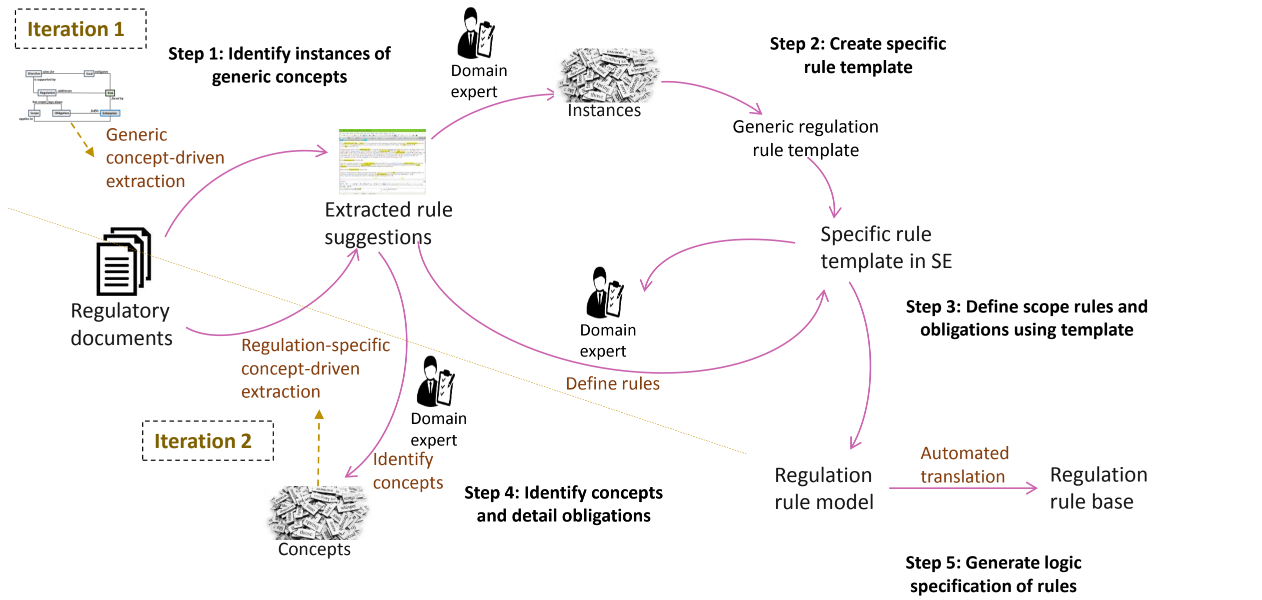

Figure 3: Steps in our rule base creation process

Listing 4: Instances of generic regulation concepts extracted from MiFID-2 documents

```
rule um001 MiFID is_a directive
rule um002 MiFIR is_a regulation
rule um003 MiFID is_supported_by MiFIR
rule um004 transparency_in_financial_markets is_a goal
rule um005 weakness_in_functioning_of_financial_markets
     is_a risk
rule um006 regulation_of_financial_markets is_a goal
rule um007 MiFIR_scope is_a regulation_scope
rule um008 MiFIR aims_for
     transparency_in_financial_markets
rule um009 MiFIR has_scope MiFIR_scope
rule um010 transparency_in_financial_markets mitigates
     weakness_in_functioning_of_financial_markets
rule um011 regulation_of_financial_markets mitigates
     weakness_in_functioning_of_financial_markets
rule um012 MiFIR_obligations is_a regulation_obligations
rule um013 MiFIR aims_for
     regulation_of_financial_markets
rule um014 MiFIR lays_down MiFIR_obligations
rule um015 MiFIR has_scope MiFIR_scope
```

Step 2: Create specific rule template
The generic rule template of Listing 1 is *instantiated* by replacing concept names with instance names in the rules to generate the specific rule template for the regulation shown in Listing5.

Listing 5: Specific rule template for the MiFID-2 regulation

```
//Instantiated rules
rule gen001 It is obligatory that enterprise
     complies_with MiFID if MiFID is_supported_by MiFIR
     && enterprise complies_with MiFIR
rule gen002 It is obligatory that enterprise
     complies_with MiFIR if enterprise
     falls_within_scope_of MiFIR && MiFIR lays_down
     MiFIR_obligations && enterprise fulfils
     MiFIR_obligations
rule gen003 It is obligatory that enterprise
     falls_within_scope_of MiFIR if MiFIR has_scope
     MiFIR_scope && MiFIR_scope applies_to enterprise
rule gen004 It is necessary that MiFIR addresses risk if
     MiFIR aims_for goal && goal mitigates risk
rule gen005 It is obligatory that enterprise
     falls_within_scope_of MiFID if MiFID
     is_supported_by MiFIR && enterprise
     falls_within_scope_of MiFIR
```

Step 3: Define scope rules and obligations using the template The specific template contains placeholders for scope rules and obligations, in rules *gen003* and *gen002*, in the predicates *MiFIR has_scope MiFIR_scope*, *MiFIR_scope applies_to enterprise*, *MiFIR lays_down MiFIR_obligations*, and *enterprise fulfils MiFIR obligations* respectively. These need to be further detailed in order to complete the definition of the rule base. Their details are obtained from the extracted rule suggestions illustrated in Listing3. Rules written using the suggestions can be seen in Listing6, with the same rule numbers.

Listing 6: Rule base with scope rules and obligations

```
rule a5697 It is obligatory that MiFIR_scope is_for
     enterprise if enterprise is_a investment_firm ||
     enterprise is_a regulated_markets || enterprise
     is_a reporting_firm || enterprise is_a
     third_country_investment_firms_operating_in_EU &&
     enterprise has_established branch_in_EU
rule a7790 It is obligatory that enterprise fulfils
     MiFIR_requirements if enterprise fulfils
     requirements_for disclosure_of_trade_data_to_public
     &&
enterprise fulfils requirements_for_reporting_of
     transactions && enterprise fulfils
     requirements_for_trading_of_derivatives_on
     organised_venues &&
```

```
enterprise fulfils requirements_for_non_discriminatory
    access_to_clearing && enterprise fulfils
    requirements_for
    non_discriminatory_access_to_trading_benchmarks &&
    enterprise fulfils requirements_for_product
    intervention_powers &&
enterprise fulfils
    requirements_for_activities_by_third_party_firms
```

rule a5697 defines *MiFIR_scope is_for enterprise* in terms of the scope rule *rule a5697* extracted from the directive text, that details the kinds of enterprises the regulation applies to. *rule a7790* defines *enterprise fulfils MiFIR_requirements* as a set of high-level obligations given in the regulation, again obtained from the extracted rule suggestion.

The obligation rules are detailed further, using suggestions extracted from regulation or RTS documents. e.g. *rule a8033, a9259, a8133, a8233* define *enterprise fulfils requirements_for_reporting_of transactions* using a hierarchy of rules expressing obligations, shown in Listing 7. Concepts from the obligations are then given as input to the rule extractor to extract the next set of rule suggestions, that can be used to detail obligations still further. This process is iteratively followed.

Step 4: Generate logic specification of rules The SE rules written by the domain expert are automatically translated into an SBVR model of rules. The description of this work is outside the scope of this paper. Rules in SBVR are translated to defeasible logic formalism DR-Prolog[1], using the translation mechanism described in [18]. The translated rules in DR-Prolog are shown in Listing8.

These rules can be directly checked against enterprise data facts. It is seen from the listing in DR-Prolog that the lowest-level rules contain data definitions. Our earlier work [17] deals with the process of arriving at these rules, as well as the necessary enterprise data facts, hence it is not explained here. The term *obligations* has been used throughout the text to mean rules that have any of the modalities *obligation, permission, prohibition and necessity*. Each of these modalities has been implemented using the defeasible logic metaprogram of DR-Prolog[1], which provides an implementation for each modality using the constructs available in standard Prolog. In the next section, we discuss whether the described approach meets its objectives.

4 DISCUSSION

One of the key objectives of using the template based approach is isomorphism, i.e. imparting a structure to the rule base similar to that of the original regulation document sources, discussed in [4]. The generic concept model we have used to guide the extraction refers to the regulatory sources, viz. directive and regulation, and embodies the structure of a regulation. This structure and traceability is maintained right from initial rule specification in Structured English to the model and formal rule specification in DR-Prolog. In the absence of a process such as this, the onus is on experts to ensure both structure and correctness. Structuring and achieving isomorphism becomes subjective. A manual process of rule construction goes through several iterations to achieve correctness in the rule hierarchy. Our generic model-guided extraction seeks to avoid omission of key elements and reduce errors through automated generation of the rule hierarchy, predicates, and parameters. This takes away the complexity associated with manual construction of formal logic rules. Since extracted rules depend completely on seed concepts given, which may vary from user to user, guided extraction with the generic set of seed concepts gives uniformity and assurance. An example of omission during our earlier manual rule base creation experiment is that we had missed encoding scope rules regarding enterprises to which the MiFID-2 regulation is applicable.

Reduced burden on domain experts and faster knowledge engineering seem to justify the development cost of our rule generation framework. However, empirical evidence is crucial to support this claim. We are in the process of conducting a systematic empirical evaluation of our approach. It must be mentioned here that the problems mentioned in the pioneering and extensive work on encoding regulations in formal logic [5, 23] such as need for simplification when encoding legislations and handling cross-references within regulations, remain. We have dealt with some of these such as simplification of complex sentence constructs, bulleted lists, and cross-references in our work on extraction.

The important objective of having a formal rule base is being able to answer queries about the regulation. Using our resultant rule base structure, we are able to answer the queries listed in Section 2.2 with regard to goals addressed by the regulation, the kinds of enterprises it applies to, as well as compliance of all or specific entities whose ground data is provided, to all or specific rules.

Pros of our approach include automated extraction of necessary supporting rules such as *investment_firm executes transaction* and *transaction trades financial_instrument*, that were not retrieved even in non-guided extraction. We are currently testing this hypothesis on larger examples. We have not so far encountered any problems of using automated rule generation. A difference from the manual encoding approach is that the user needs to familiarize himself with generated rule names and some indirections in the generated rules when tracing rule execution during compliance checking. Rule expressiveness in our approach is adequate and scales well, since SBVR has a very rich meta-model with a direct correspondence to SBVR Structured English. Being fact-oriented, SBVR maps directly to DR-Prolog, which scales for large datasets as well.

The generic model described here is a little basic, but can and should be altered as required, if the regulation being worked on has a different structure or important sections or elements that need to be incorporated into the structure of the rule base. We plan to also enhance the model using the learning from our experience of applying the approach to several regulations. The next section reviews related work.

5 RELATED WORK

We survey work with similar objectives as ours, of structured representation of regulatory content to allow formal checking and analysis. Encoding the British Nationality Act as a logic program in Prolog [23] was pioneering work in encoding regulations in formal logic, as is [5]. A need for intermediate representation of natural language regulations was underlined in [4]. These and later formal approaches to encoding regulations [3, 11] require regulation rules to be encoded manually in the logic formalism.

The important requirement of integrating compliance checking and accessing related regulatory documents is addressed in [16].

Listing 7: Rule base with detailed obligations

```
// Detailed obligations from directive/ regulation
rule a8033 It is obligatory that enterprise fulfils requirements_for_reporting_of_transactions if enterprise is_a
      investment_firm && enterprise reports complete_accurate_details_of_transactions_to_competent_authority &&
      investment_firm executes transaction && transaction trades financial_instrument && financial_instrument
      traded_at_or_admitted_to trading_venue
rule a8037 It is obligatory that enterprise reports complete_accurate_details_of_transactions_to_competent_authority if
      enterprise reports transaction_details && enterprise reports_to trade_repository && enterprise reports
      before_close_of_working_day
rule a9259 It is necessary that financial_instrument traded_at_or_admitted_to trading_venue if financial_instrument
      traded_at trading_venue ||
  financial_instrument admitted_to trading || underlying_instrument traded_at trading_venue ||
  underlying_index_or_basket traded_at trading_venue
rule a8133 It is obligatory that enterprise reports transaction_details if enterprise reports reportable_transactions &&
  enterprise does_not_report excluded_transactions
rule a8233 It is obligatory that enterprise reports reportable_transactions if trade is_constituted_for reporting_purpose
rule a7029 It is obligatory that trade is_constituted_for reporting_purpose if trade is acquisition || trade is disposal
      || trade is simultaneous_acquisition_disposal_no_change_in_ownership
rule a2467 trade is acquisition if trade is purchase_of_financial_instrument || trade is
      entering_into_derivative_contract || trade is increase_in_notional
```

They build a compliance assistance framework that uses first-order logic (FOL) representations of regulation rules, as well as related questions and answers. The FOL rules are linked to their occurrences in regulation documents through an XML representation of tagged regulation information. Writing of the FOL rules and tagging to regulation text is done manually. Our endeavour is semi-automated creation of a rule base of the entire regulation, and logical structuring of the rule base to be able to reason about compliance to the higher-level goals of the regulation.

Most of the current state of the art in legal rule extraction contains an implicit step of rule identification. This step often encompasses several other constituent steps, like identifying segmentation of regulations, ascertaining modality of the regulations such as whether the rule is an obligation or a permission, and so on [7, 28–30]. The final constituent step is often writing the chosen logical specification of NL rule. We believe that by separating these concerns from rule identification, it is possible to defer their treatment until we obtain logical specifications, albeit partial ones. We believe that the logical specification language itself, such as for instance DR-Prolog [1, 2] can be used to take care of the segmentation and cross referencing, because at that level of abstraction, we already have access to schema of the information required for regulations.

In contrast, the current state of the art often focuses on the treatment of legal syntactical specifics early on. This is evident in the *governance extraction model* which manually classifies and attempts to extract regulations as legal requirements in terms of procedural, declarative, ontology statements [15]. Further fine level classification includes access-rights statements and delegation of authority rights. Another work proposes to group legal sentences into few categories referred to as *juridical natural language constructs* (JNLCs). JNLCs are proposed to be parsed using unification grammars [28]. In a similar vein, legal concepts are proposed to be classified into rights, obligations, privileges, no-rights, powers, liabilities, immunities, and disabilities using a production rule model in [20]. Another work proposes to use a categorization of provisions and an ML classifier trained to identify the provisions in [6].

GaiusT, a tool based on Cerno framework and related research presents semi-automated rule extraction with precision and recall numbers similar to ours [7, 8, 19, 30, 31]. In contrast to our approach, this tool and the framework use a number of intermediate artifacts, namely form simplification through semantic parameterization [7, 8], structural comprehension [30], as well as semantic annotation [30]. Their approach seems to be restricted in applicability as well, since all of the above activities have to be performed for any new regulation to which they apply the framework. It is likely that when the regulation is large like MiFID-2 or FATCA[10], the interaction between various components becomes hard to handle. At the same time, this approach presents some ideas around a conceptual (meta-) model of deontic concepts and a rules generator which could be of use to us.

An approach presented in [9] contrasts ML-based text classification with knowledge engineering-based (KE-based) text classification. The idea behind KE is that definitions of legal terms are formulated using specific phrases and presuming that only a few clear and easily observable pattens were used for each type of legal sentences or provisions, then these so called classification patterns could be used for classification. It was found that such approach is susceptible to the same complexities of legal sentences which also affect ML-based classification negatively. Some of these complexities are classification keywords appearing in auxiliary sentences rather than the main sentence to be classified, missing standard phrases, syntactical and lexical variation in the standard phrases and so on. In our own experiments, we too included certain phrases as indicating definitional regulations (those regulations which define domain entities and their specializations) as well as rules in the approximate dictionary chunker. Like the results in [9], our own experiments indicated that there are no perceptible differences in the performance of the classifier when these additional phrases are considered as well. In contrast to [9], if we liken our approach to KE-based classification, then we go one step further and actually make use of the domain model entities and the dictionary in the classification.

[10] FATCA: Foreign Account Tax Compliance Act, https://www.irs.gov/businesses/corporations/foreign-account-tax-compliance-act-fatca

Listing 8: A section of the translated rules in DR-Prolog

```
defeasible(gen001, obligation, complies_withMiFID(LEI), [rule210(LEI)]).
defeasible(gen002, obligation, complies_withMiFIR(LEI), [rule212(LEI)]).
defeasible(gen003, obligation, falls_within_scope_ofMiFIR(LEI), [rule215(LEI)]).
defeasible(a5697, obligation, miFIR_scopeApplies_toEnterprise(LEI), [rule223(LEI)]).
defeasible(a7790, obligation, fulfilsMiFIR_obligations(LEI), [rule225(LEI)]).
defeasible(a8033, obligation, fulfilsRequirements_for_reporting_of_transactions(LEI), [rule232(LEI)]).
defeasible(a8037, obligation, reportsComplete_accurate_details_of_transactions_to_competent_authority(LEI), [rule237(LEI)
    ]).
defeasible(a8133, obligation, reportsTransaction_details(LEI), [rule240(LEI)]).
defeasible(a8233, obligation, reportsReportable_transactions(LEI), [rule242(LEI)]).
defeasible(a7029, obligation, is_constituted_forReporting_purpose(Transref), [rule302(Transref)]).
fact(rule211(LEI)) :- fact(miFIDIs_supported_byMiFIR(LEI)), fact(complies_withMiFIR(LEI)).
fact(rule210(LEI)) :- fact(rule211(LEI)).
fact(rule213(LEI)) :- fact(falls_within_scope_ofMiFIR(LEI)), fact(miFIRLays_downMiFIR_obligations(LEI)),fact(
    fulfilsMiFIR_obligations(LEI)).
fact(rule212(LEI)) :- fact(rule213(LEI)).
fact(rule216(LEI)) :- fact(miFIRHas_scopeMiFIR_scope(LEI)), fact(miFIR_scopeApplies_toEnterprise(LEI)).
fact(rule215(LEI)) :- fact(rule216(LEI)).
fact(rule223(LEI)) :- fact(is_aInvestment_firm(LEI)).
fact(rule223(LEI)) :- fact(is_aMarket_operator(LEI)).
fact(rule223(LEI)) :- fact(is_aReporting_firm(LEI)).
fact(rule224(LEI)) :- fact(is_aThird_country_investment_firms_operating_in_EU(LEI)), fact(has_establishedBranch_in_EU(
    LEI)).
fact(rule223(LEI)) :- fact(rule224(LEI)).
fact(rule226(LEI)) :- fact(fulfilsRequirements_for_disclosure_of_trade_data_to_public(LEI)), fact(
    fulfilsRequirements_for_reporting_of_transactions(LEI)), fact(
    fulfilsRequirements_for_trading_of_derivatives_on_organised_venues(LEI)), fact(
    fulfilsRequirements_for_non_discriminatory_access_to_clearing_and_trading_in_benchmarks(LEI)), fact(
    fulfilsRequirements_for_product_intervention_powers(LEI)), fact(
    fulfilsRequirements_for_activities_by_third_party_firms(LEI)).
fact(rule225(LEI)) :- fact(rule226(LEI)).
fact(rule233(LEI)) :- fact(is_aInvestment_firm(LEI)), fact(
    reportsComplete_accurate_details_of_transactions_to_competent_authority(LEI)), fact(
    investment_firmExecutesTransaction(LEI)), fact(transactionTradesFinancial_instrument(LEI)), fact(
    financial_instrumentTraded_at_or_admitted_toTrading_venue(LEI)).
fact(rule238(LEI)) :- fact(reportsTransaction_details(LEI)), fact(reports_toTrade_repository(LEI)), fact(
    reportsBefore_close_of_working_day(LEI)).
fact(rule237(LEI)) :- fact(rule238(LEI)).
fact(rule241(LEI)) :- fact(reportsReportable_transactions(LEI)), fact(does_not_reportExcluded_transactions(LEI)).
fact(rule240(LEI)) :- fact(rule241(LEI)).
fact(rule242(LEI)) :- fact(is_constituted_forReporting_purpose(Transref)).
fact(rule302(Transref)) :- fact(isAcquisition(Transref)).
fact(rule302(Transref)) :- fact(isDisposal(Transref)).
fact(rule302(Transref)) :- fact(isSimultaneous_acquisition_disposal_no_change_in_ownership(Transref)).
fact(rule303(Transref)) :- fact(isPurchase_of_financial_instrument(Transref)).
fact(rule303(Transref)) :- fact(isEntering_into_derivative_contract(Transref)).
fact(rule303(Transref)) :- fact(isIncrease_in_notional(Transref)).
fact(isAcquisition(Transref)) :- fact(rule303(Transref)).
fact(rule304(Transref)) :- fact(isSale_of_financial_instrument(Transref)).
fact(rule304(Transref)) :- fact(isClosing_out_of_derivative_contract(Transref)).
fact(rule304(Transref)) :- fact(isDecrease_in_notional(Transref)).
fact(isDisposal(Transref)) :- fact(rule304(Transref)).
fact(hasTradetype_buy(Transref)) :- fact(hasTradetype(Transref,'buy')).
fact(hasTradetype(Transref,Tradetype)) :- fact(trade(Transref,Tradetype)).
fact(isPurchase_of_financial_instrument(Transref)) :- fact(hasTradetype_buy(Transref)).
fact(hasTradetype_buy(Transref)) :- fact(hasTradetype(Transref,'buy')).
fact(hasTradetype(Transref,Tradetype)) :- fact(trade(Transref,Tradetype)).
```

6 CONCLUSIONS AND FUTURE WORK

We used a generic conceptual model of regulations to extract specific concepts, relations, and rules from the regulation text. This gave us a better directed approach to rule extraction and more structured rule suggestions. We used a generic model of regulation rules based on the conceptual model as a template for the regulation rule base and found it gave us a method, and a structured rule base with less rework, as well as some assurances on inclusion of vital sections of information about the regulation. It created a rule hierarchy that helps reason all the way from ground data to high-level goals of the regulation. We believe this principled approach gives us a more accurate and functional model of the regulation.

We have experimented with using this generic concept-driven extraction on sections of the KYC regulation. We plan to further test the method on the entire KYC regulation and two more regulations, and enhance the generic model and template as required. We also plan to conduct an empirical study comparing this approach of rule base construction to the manual one, as well as to the conventional industry approach to compliance.

REFERENCES

[1] Grigoris Antoniou, Nikos Dimaresis, and Guido Governatori. 2007. A System for Modal and Deontic Defeasible Reasoning. In *AI 2007: Advances in Artificial Intelligence, 20th Australian Joint Conference on Artificial Intelligence, Gold Coast, Australia, December 2-6, 2007, Proceedings*. 609–613. https://doi.org/10.1007/978-3-540-76928-6_62

[2] Grigoris Antoniou, Nikos Dimaresis, and Guido Governatori. 2009. A modal and deontic defeasible reasoning system for modelling policies and multi-agent systems. *Expert Syst. Appl.* 36, 2 (2009), 4125–4134. https://doi.org/10.1016/j.eswa.2008.03.009

[3] Ahmed Awad, Sergey Smirnov, and Mathias Weske. 2009. Resolution of Compliance Violation in Business Process Models: A Planning-Based Approach. OTM Conferences (1) 2009: 6-23.

[4] Trevor J. M. Bench-Capon and Frans Coenen. 1992. Isomorphism and legal knowledge based systems. *Artif. Intell. Law* 1, 1 (1992), 65–86. https://doi.org/10.1007/BF00118479

[5] Trevor J. M. Bench-Capon, G. O. Robinson, Tom Routen, and Marek J. Sergot. 1987. Logic Programming for Large Scale Applications in Law: A Formalisation of Supplementary Benefit Legislation. In *Proceedings of the First International Conference on Artificial Intelligence and Law, ICAIL '87, Boston, MA, USA, May 27-29, 1987*. 190–198. https://doi.org/10.1145/41735.41757

[6] Carlo Biagioli, Enrico Francesconi, Andrea Passerini, Simonetta Montemagni, and Claudia Soria. 2005. Automatic Semantics Extraction In Law Documents. In *ICAIL, June 6-11, 2005, Italy*. 133–140. https://doi.org/10.1145/1165485.1165506

[7] Travis D. Breaux and Annie I. Antón. 2005. Deriving Semantic Models from Privacy Policies. In *6th POLICY Workshop, Sweden*. 67–76. https://doi.org/10.1109/POLICY.2005.12

[8] Travis D. Breaux, Annie I. Antón, and Jon Doyle. 2008. Semantic parameterization: A process for modeling domain descriptions. *ACM Trans. Softw. Eng. Methodol.* 18, 2 (2008). https://doi.org/10.1145/1416563.1416565

[9] Emile de Maat, Kai Krabben, and Radboud Winkels. 2010. Machine Learning Versus Knowledge Based Classification of Legal Texts. In *Proceedings JURIX 2010*. IOS Press, Amsterdam, The Netherlands, The Netherlands, 87–96.

[10] Mauro Dragoni, Guido Governatori, and Serena Villata. 2015. Automated Rules Generation from Natural Language Legal Texts. In *Workshop on Automated Detection, Extraction and Analysis of Semantic Information in Legal Texts*. San Diego, USA, 1–6.

[11] Guido Governatori and Antonino Rotolo. 2013. A conceptually rich model of business process compliance. In *APCCM 2010: 3-12*.

[12] Jan Hajic, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Stepánek, Pavel Stranák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, CoNLL 2009, Boulder, Colorado, USA, June 4, 2009*. 1–18. http://aclweb.org/anthology/W/W09/W09-1201.pdf

[13] Terry A. Halpin. 2011. Fact-Orientation and Conceptual Logic. In *Proceedings EDOC 2011, Finland*. 14–19. https://doi.org/10.1109/EDOC.2011.28

[14] Zellig Harris. 1954. Distributional structure. *Word* 10, 23 (1954), 146–162.

[15] Waël Hassan and Luigi Logrippo. 2009. Governance Requirements Extraction Model for Legal Compliance Validation. In *RELAW 2009, USA*. 7–12. https://doi.org/10.1109/RELAW.2009.4

[16] Shawn Kerrigan and Kincho H. Law. 2003. Logic-Based Regulation Compliance-Assistance. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law, ICAIL 2003, Edinburgh, Scotland, UK, June 24-28, 2003*. 126–135. https://doi.org/10.1145/1047788.1047820

[17] Deepali Kholkar, Sagar Sunkle, and Vinay Kulkarni. 2016. From Natural-language Regulations to Enterprise Data using Knowledge Representation and Model Transformations. In *Proceedings of the 11th International Joint Conference on Software Technologies (ICSOFT 2016) - Volume 2: ICSOFT-PT, Lisbon, Portugal, July 24 - 26, 2016*. 60–71. https://doi.org/10.5220/0006002600600071

[18] Deepali Kholkar, Sagar Sunkle, and Vinay Kulkarni. 2017. Towards Automated Generation of Regulation Rule Bases using MDA. In *MODELSWARD 2017 - Accepted*.

[19] Nadzeya Kiyavitskaya, Nicola Zeni, Travis D. Breaux, Annie I. Antón, James R. Cordy, Luisa Mich, and John Mylopoulos. 2008. Automating the Extraction of Rights and Obligations for Regulatory Compliance. In *Conceptual Modeling - ER*. 154–168. https://doi.org/10.1007/978-3-540-87877-3_13

[20] Jeremy C. Maxwell and Annie I. Antón. 2010. The Production Rule Framework: Developing a Canonical Set of Software Requirements for Compliance with Law. In *Proceedings of the 1st ACM International Health Informatics Symposium (IHI '10)*. ACM, New York, NY, USA, 629–636. https://doi.org/10.1145/1882992.1883092

[21] Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau, and Chris Reed. 2007. Automatic Detection of Arguments in Legal Texts *(ICAIL '07)*. ACM, New York, NY, USA, 225–230. https://doi.org/10.1145/1276318.1276362

[22] Sjir Nijssen. 2007. SBVR: Semantics for business. (2007).

[23] Marek J. Sergot, Fariba Sadri, Robert A. Kowalski, F. Kriwaczek, Peter Hammond, and H. T. Cory. 1986. The British Nationality Act as a Logic Program. *Commun. ACM* 29, 5 (1986), 370–386. https://doi.org/10.1145/5689.5920

[24] Burr Settles. 2009. *Active Learning Literature Survey*. Computer Sciences Technical Report 1648. University of Wisconsin–Madison.

[25] Sagar Sunkle, Deepali Kholkar, and Vinay Kulkarni. 2015. Model-driven regulatory compliance: A case study of "Know Your Customer" regulations. In *18th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems, MoDELS 2015, Ottawa, ON, Canada, September 30 - October 2, 2015*. 436–445. https://doi.org/10.1109/MODELS.2015.7338275

[26] Sagar Sunkle, Deepali Kholkar, and Vinay Kulkarni. 2016. Comparison and Synergy Between Fact-Orientation and Relation Extraction for Domain Model Generation in Regulatory Compliance. In *Conceptual Modeling - 35th International Conference, ER 2016, Gifu, Japan, November 14-17, 2016, Proceedings (Lecture Notes in Computer Science)*, Isabelle Comyn-Wattiau, Katsumi Tanaka, Il-Yeol Song, Shuichiro Yamamoto, and Motoshi Saeki (Eds.), Vol. 9974. 381–395. https://doi.org/10.1007/978-3-319-46397-1_29

[27] Sagar Sunkle, Deepali Kholkar, and Vinay Kulkarni. 2016. Informed Active Learning to Aid Domain Experts in Modeling Compliance. In *20th IEEE International Enterprise Distributed Object Computing Conference, EDOC 2016, Vienna, Austria, September 5-9, 2016*, Florian Matthes, Jan Mendling, and Stefanie Rinderle-Ma (Eds.). IEEE Computer Society, 1–10. https://doi.org/10.1109/EDOC.2016.7579382

[28] Tom M. van Engers, Ron van Gog, and Kamal Sayah. 2004. A Case Study on Automated Norm Extraction. In *Legal Knowledge and Information Systems. Jurix 2004: The Seventeenth Annual Conference. (Frontiers in Artificial Intelligence and Applications)*, T. Gordon (Ed.). IOS Press, Amsterdam, 49–58.

[29] Adam Z. Wyner and Wim Peters. 2011. On Rule Extraction from Regulations. In *Legal Knowledge and Information Systems - JURIX 2011: The Twenty-Fourth Annual Conference, University of Vienna, Austria, 14th-16th December 2011*. 113–122. https://doi.org/10.3233/978-1-60750-981-3-113

[30] Nicola Zeni, Nadzeya Kiyavitskaya, Luisa Mich, James R. Cordy, and John Mylopoulos. 2015. GaiusT Supporting The Extraction Of Rights And Obligations For Regulatory Compliance. *Requir. Eng.* 20, 1 (2015), 1–22. https://doi.org/10.1007/s00766-013-0181-8

[31] Nicola Zeni, Luisa Mich, John Mylopoulos, and James R. Cordy. 2013. Applying GaiusT For Extracting Requirements From Legal Documents. In *Sixth International Workshop on Requirements Engineering and Law, RELAW 2013, 16 July, 2013, Rio de Janeiro, Brasil*. 65–68. https://doi.org/10.1109/RELAW.2013.6671349