SmaRT Visualisation of Legal Rules for Compliance

Selja Seppälä, Marcello Ceci*, Hai Huang, Leona O'Brien, and Tom Butler

Governance, Risk, and Compliance Technology Center
University College Cork
13 South Mall, Cork, Ireland
{selja.seppala,marcello.ceci,hai.huang,tbutler,leona.obrien}@ucc.ie
http://www.grctc.com

Abstract. *This paper presents a visualization technique to assist legal experts in formalising their interpretation of legal texts in terms of regulatory requirements. (Semi-)automation of compliance processes requires a machine-readable version of legal requirements in a format that enables effective compliance assessment. The use of a semi-structured controlled natural language as an intermediate step of the translation from a human-readable text to a machine-readable and understandable format ensures that the process of interpretation of those requirements is as simple as possible. However, it does not ensure that the formal representation resulting from the interpretation faithfully represents the intended semantics provided by the legal expert. Visualization techniques such as property graphs in Neo4j could fill this gap, allowing legal experts to understand and control the formal representation of the result of their act of interpretation.*

Keywords: SBVR, RegTech, Controlled Natural Languages, Neo4j

1 Introduction

Ensuring compliance with regulatory requirements represents a considerable challenge for industries that deal with large amounts of data across different jurisdictions. This is particularly true for safety-critical industries such as the international financial industry, driven by the proliferation and complexity of the financial regulatory environment in the aftermath of the global financial crisis. The wide acceptance in the industry that traditional Governance, Risk, and Compliance (GRC) information systems are in deficit is leading to a growing interest in semantic technologies as a solution [1, 4]. By using such technologies, compliance of companies' business policies and processes – and even of their activities represented by company data – could be assessed automatically. Even before achieving such an integration, exploration of the regulatory space could be performed in a very effective way, e.g., by querying the relevant regulatory

* Corresponding author.

knowledge base to quickly retrieve all relevant obligations for a certain business activity.

One of the problems with representing the interpretation of legal rules in a machine-readable format lies in the lack of understanding and control of the formal models by legal experts. Some research approaches [9, 17, 10, 13] rely on formal representations of rules that, despite being very expressive and thus allowing powerful inference, are not provided in a format that is understandable and manageable by a legal expert. To overcome this, other approaches [16, 12] suggest the use of controlled natural languages to translate the semantics of a legal text into a machine-readable representation. Similarly, our approach aims to define a controlled natural language that has complete coverage of the relevant legal effects of requirements while at the same time constraining the natural language as little as possible.

The research behind the present paper follows an approach that allows legal experts to interpret a legal statement by rewriting it in a semi-structured controlled natural language using a dedicated software called SmaRT. The core translation process relies on the markup and annotation of strings in the rewritten text in terms of given vocabulary elements. However, this is not sufficient to put legal experts in control of all the elements that compose the logical formulation of the legal rule. The paper thus presents a possible solution that relies on property graphs to visualize the interpreted rule and all the relevant elements in a format that a lawyer can understand and manipulate. This should allow to ensure that the human-readable text and the machine-readable output of a legal rule carry the same semantics, thus achieving a semi-automatic translation of legal requirements for compliance purposes.

The rest of the paper is structured as follows. In the next section, we introduce the Regulatory Interpretation Methodology (RIM) and the controlled natural language. In Section 3, we look into relevant details of the RIM and the rule-editing tool SmaRT to understand the visualisation needs for the rule interpretation task. In Section 4, we present a possible visualisation solution that addresses these needs.

2 The Regulatory Interpretation Methodology

Understanding regulations is a complex task and legal experts face a number of challenges in interpreting a regulatory text, including: following and fleshing out references and citations; identifying, delimiting, and disambiguating definitions; making sense of complex sentences; clarifying ambiguities resulting from legalese; accounting for exceptions [2]. Our research aims to bridge the gap between the legal expertise required to interpret the regulatory text and the modelling skills required to build a semantic knowledge base. The goal is to foster compliance in the financial sector by supporting corporate lawyers, risk practitioners and compliance professionals in their role of subject matter experts in making law more readily consumable and comprehensible by the industry.

The translation of regulatory text into machine-readable information is articulated around a methodology that defines a process for transforming a regulatory text into a formal representation using an intermediate human-readable representation in semi-structured English. The process (see Fig. 1) is designed as a collaborative one, involving the legal expert as a subject matter expert (SME) and the modeller as a semantic technology expert (STE) through multiple iterations [15].

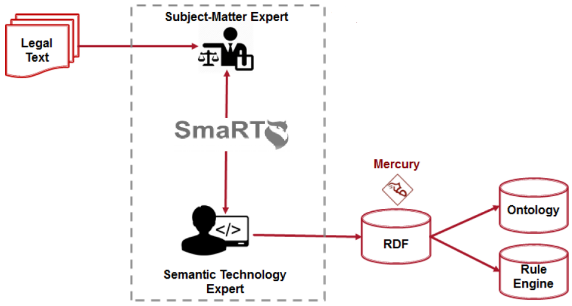

Fig. 1. The translation process outlined in the Regulatory Interpretation Methodology. The dashed box identifies the part of the process covered by SmaRT. Currently, the Semantic Technology Expert is needed to ensure the quality of the output model, but will not be needed as the software is improved.

The solution allows SMEs to represent the semantics of regulatory requirements in a machine-readable format through a SME-friendly process. This is ensured through the use of SBVR (Semantic of Business Vocabularies and Business Rules), a Object Management Group specification based on formal logics and well known to the industry [14]. In SBVR, a requirement is rewritten in Structured English, a semi-structured controlled natural language where every term used in the rules is sourced and specified in a terminological dictionary. SBVR is a powerful instrument for modelling an area of business activity and for building a business vocabulary [11], but it is not suitable – as is – for the representation of legal rules; some SBVR components are not needed or overcomplicate the task of rule representation (e.g. the logical formulation of a sentence, see [7]), and some components fall short in capturing legal concepts (e.g. constitutive rules).

To overcome this, our research group at the Governance, Risk, and Compliance Technology Center (GRCTC) has devised a resource called Mercury, composed of:

- Mercury-SE, a semi-structured English based on SBVR used as an intermediate language between the legal text and the machine-readable format, and
- Mercury-ML, a persistence model in RDF format, capturing the semantics of Mercury-SE in a machine-readable format.

Mercury represents rule statements contained in regulations and describes the concepts used in those rules in a vocabulary. The process of translation from the legal text to the Mercury-SE language is manual and tool-assisted. The software, SmaRT, is specifically designed to make the translation process as intuitive as possible, while at the same time reducing the user's time devoted to the most repetitive work. The reader is invited to refer to [6] for more details on SBVR, Mercury, and the ways in which the latter enhances the former.

Mercury relies on both (a) technologies from the Semantic Web (SW) stack at the RDF layer and (b) non-SW technologies (SBVR and the RIM). It relies on upper SW layers, particularly OWL, for advanced classification and reasoning on rules and vocabulary. The GRCTC is currently developing a set of ontologies called FIRO (Financial Industry Regulatory Ontology) to enable semantic applications such as classification, querying, and reasoning, and has devised a mapping of all SBVR elements relevant to Mercury into RDF/OWL to assist the STE in translating Mercury rulebooks and vocabularies [3].

Because of limitation of OWL and Description Logics [5], FIRO does not involve complete rule-based reasoning but only rule representation. The aim is to capture relevant information on the regulatory requirements and to be able to:

- run queries on the resulting RDF/OWL knowledge base;
- perform abstract classification and reasoning on rules and their regulated actions (e.g., detecting which rules regulate a subset of another rule's regulated action);
- validate data representing instances of regulated actions (events) as compliant or breaching one or more rules.

In order to perform these tasks, we need the legal rules – represented in Mercury-ML – to be complete in terms of semantics and computable. Mercury being a semi-structured controlled natural language, there is a need for solutions that allow SMEs to have greater control over the formal representation of their interpretation of a legal rule without having to learn any specific formalism. We propose to achieve this by providing a graphical solution that allows the SMEs to visualise rules in a more schematic way. For example, Fig. 2 shows a graphical representation of the following rule:

Rule 1. *It is obligatory that a market operator that operates a trading venue makes public credits and debts.*

The graph in Fig. 2 makes explicit the information about the logical formulation of the rule that is not immediately apparent in the textual form. For example, it shows that *market operator* is the subject of three verb concepts: *market operator operates trading venue*, *market operator makes public credit*, and *market operator makes public debt.* It also shows which verb concepts express the deontic condition of the rule – see the dashed box in red – distinguishing it from the other verb concept, which expresses the rule's applicability condition. Finally, it shows how the former and the latter verb concepts relate to each other – they are connected via the same subject, *market operator*.

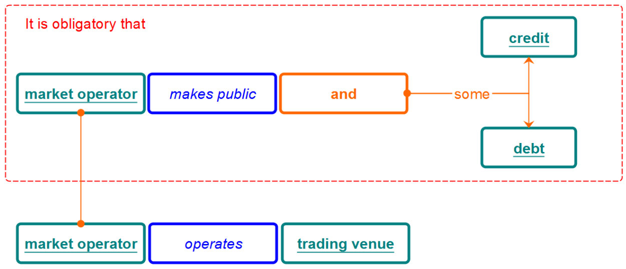

Fig. 2. Example of a graphical representation of a legal rule

In the next section, we describe the tasks of the Regulatory Interpretation Methodology in more detail to specify further our visualization needs.

3 Visualisation Needs

To create a rule from a legal document, SMEs follow the methodology defined in the RIM. First, they identify a unit of analysis within a given legal document. Second, they rewrite the text of the unit of analysis, fleshing out references. Third, they identify and define the noun concepts (i.e., entities) and verb concepts (i.e., actions with their participants and other attributes) and mark the keywords (e.g., the logical operators) composing the rule. Finally, they specify the logical formulation of the rule by linking verb concepts to each other. This logical formulation follows specific patterns for constitutive rules [8] and more generic ones for regulative rules [6]. Here, we focus on the third and fourth steps and, more specifically, on the subtasks of (i) creating verb concepts within a rule and (ii) linking verb concepts to each other to form rules.

All these steps are carried out using SmaRT, our web-based rule editing tool. Fig. 3 shows the editor window at the first step of the methodology. The original regulatory text is displayed in the upper box and the text rewritten by the SME is displayed in the lower box.

3.1 Visualising Noun Concepts and their Links to Verbs

The task of creating verb concepts consists of identifying all the noun concepts – usually domain-specific terms – and linking them to the appropriate verbs or verb phrases using one of the roles that specify the semantic relation between a noun concept and the action denoted by the verb, such as *hasSubject*, *hasObject*, *hasIndirectObject*, and *hasLocation*. A verb concept is created when all the relevant noun concepts are linked to the verb concept's verb with one (and only one) of these relations. The verb of a verb concept should be captured in its active form and should account for any of its variant forms, e.g., its passive form.

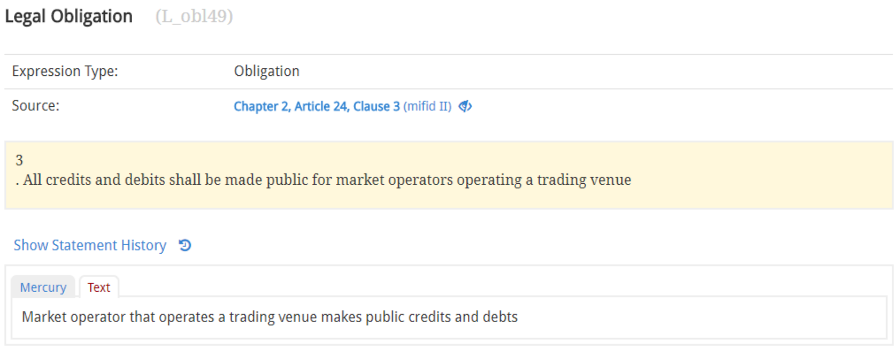

Fig. 3. SmaRT editor view displaying the original and the rewritten text

Therefore, the first visualisation requirement to assist the SMEs in the task of creating verb concepts is to display (i) the building blocks of a verb concept – the verb and the noun concepts – and (ii) the links representing the roles relating noun concepts to the verb in a verb concept.

3.2 Visualising Verb Concepts and their Links to Each Other

The second task of connecting verb concepts to each other to form the rules brings about further requirements for useful visualisation aids. This task consists of two subtasks:

- Identifying the logical operators (*and*/*or*) forming a conjunction or disjunction of two or more noun concepts that play the same role with respect to a verb concept. This results in the creation of two or more verb concepts denoting two or more actions with their participants and other attributes (see Fig. 4).
- Identifying noun concepts that play a role in more than one verb concept (see Fig. 5).

Therefore, the second visualisation requirement to assist the SMEs in the task of linking verb concepts to each other is to display these larger building blocks and their links in a user-friendly manner that captures and gives a clear understanding of the logical structure of a rule (see Fig. 6).

3.3 Limitations of the Current Visualisation Capabilities

To help the SMEs in these tasks, the current version of SmaRT integrates display functions that address only part of the visualisation needs. The tool allows SMEs to visualise the different elements of a verb concept as shown in Fig. 7. These are displayed using color-coded tags that encapsulate the respective text spans using

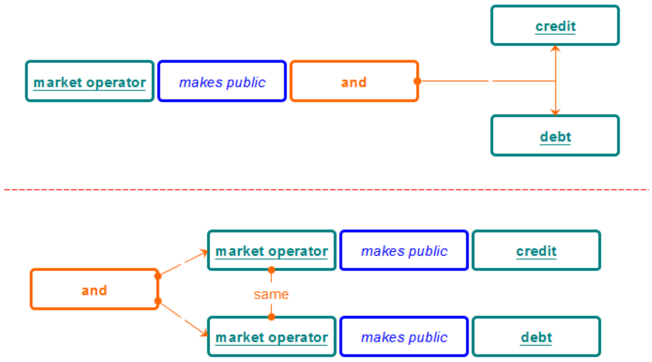

Fig. 4. The graph at the top shows a logical operator (*and*) connecting two noun concepts. The original sentence is *market operator makes public credits and debts*. The semantics is *market operator makes public credits and same market operator makes public debts*. The graph at the bottom shows the semantics of this connection: the logical operator is really connecting two verb concepts, each of which has a different noun concept (credit and debt) in the role of object, and the same instance of the same noun concept (market operator) in the role of subject.

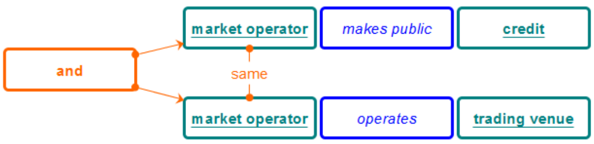

Fig. 5. This graph shows the keyword *same* connecting two noun concepts each playing a role in a different verb concept. The original sentence is *market operator operating a trading venue makes public credits*. The semantics is *market operator makes public credits and same market operator operates trading venue*.

distinct colors for noun concepts (in green) and verbs – verb parts of speech and verbal phrases – (in blue).

Selecting a verb highlights the entire verb concept, that is, the verb and the related noun concepts. This is illustrated in Fig. 8, where the verb *operates* in the verb concept *market operator operates trading venue* is marked in blue and the related noun concepts, *market operator* and *trading venue*, in red. However, the current solution is limited in that the roles played by each noun concept within a verb concept are not graphically represented in the main window. They are specified and displayed in a separate pane (see the text box on the right-hand side of the editor in Fig. 8).

Similarly, the current implementation does not support the visualisation of more complex relations between verb concepts resulting from conjunctions or

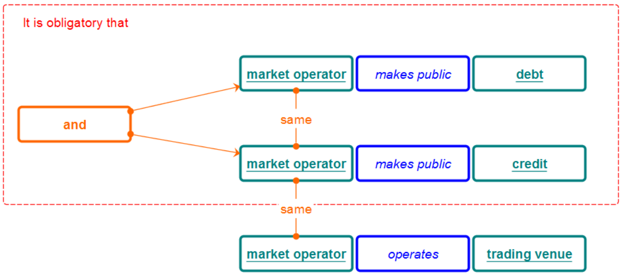

Fig. 6. This graph shows the entire Rule 1 in a more verbose graph than the one contained in Fig. 2. The graph was obtained by joining the graphs in Figs. 4 and 5, and by distinguishing the deontic condition from the applicability condition. In turn, the applicability condition loses the logical operator *and* connecting it to the deontic conditions.

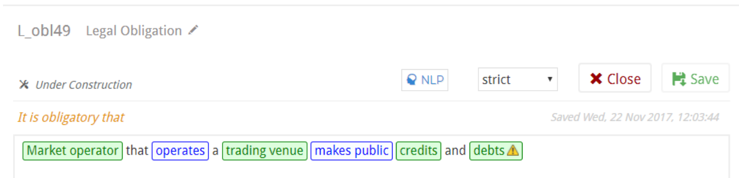

Fig. 7. Visualisation of noun concepts and verbs in SmaRT

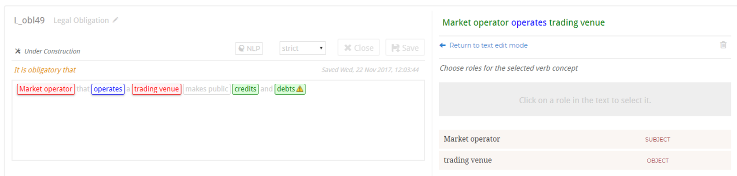

Fig. 8. Visualisation of verb concepts in SmaRT

disjunctions of noun concepts – "make public *credits* AND *debts*", nor of how two or more verb concepts are linked to each other – *market operator* is the subject of *makes public* and of *operates* and thus links the corresponding verb concepts.

In sum, the visualisation needs for a seamless execution of these tasks can be broadly categorized into two types: grouping textual spans at different levels of granularity and linking these spans to each other in a user-friendly manner that clarifies and makes the logical structure of a rule explicit. Ideally, the rule-editing tool would allow the SMEs not only to visualise the rules, but also to interact with the visual environment to build the rules. Finally, the visualisation solution should also address end-user needs by providing them with additional query functions that would allow them to retrieve and display relevant rules for enhanced regulatory compliance verification. To address the current shortcomings of the tool and improve its user-friendliness, we have explored a visualisation solution that would meet these needs and assist the SMEs in the rule-editing task.

4 Graphical Representation of Legal Rules

Graphs are helpful to visualise relationships. In this section, we present a method to represent vocabularies and legal rules in property graphs and store them in the graph database management system Neo4j[1] that allows for native graph storage and processing. The property graph contains connected entities (the nodes) which can hold any number of attributes (key-value pairs). Nodes can be tagged with labels representing their different roles in a domain. In addition to contextualizing node and relationship properties, labels may also serve to attach meta data, index, or constraint information to certain nodes.

Neo4j uses Cypher[2], a declarative, SQL-inspired language for visually describing patterns in graphs using an ascii-art syntax. Here, Cypher is used to represent and query graph data.

4.1 Graphical Representation of Noun and Verb Concepts

Graphical Representation of Noun Concepts. Noun concepts constitute building blocks of legal rules; they can be financial concepts. Each noun concept includes attributes, such as *label*, *concept type*, *context type*, and their corresponding values. They also include object properties such as *hasGeneralConcept*, which indicates a subclass relationship between two noun concepts. We represent each noun concept as a graph node, with key-value pairs specifying its attributes and their corresponding values.

Fig. 9 shows an example of a noun concept, *investment firm*, with some of its attributes and corresponding values (shown in the bottom bar), and a relationship *hasGeneralConcept* between *investment firm* and the noun concept *company*, indicating that the concept *investment firm* is a subclass of the concept *company*.

Graphical Representation of Verb Concepts. Usually, verb concepts denote actions contained in rules and describe basic relationships between noun concepts. Each verb concept includes a subject and a verb or verbal phrase called *verbSymbol*. It may also include an object, and indirect object or other types of relations such as *hasLocation*. We represent each verb concept as a graph node that has edges (relationships) such as: *hasSubject*, *hasVerbSymbol*, *hasObject*, and *hasIndirectObject*. Fig. 10 shows an example of the verb concept *market operator operates a trading venue*.

[1] https://neo4j.com/

[2] https://neo4j.com/developer/cypher-query-language/

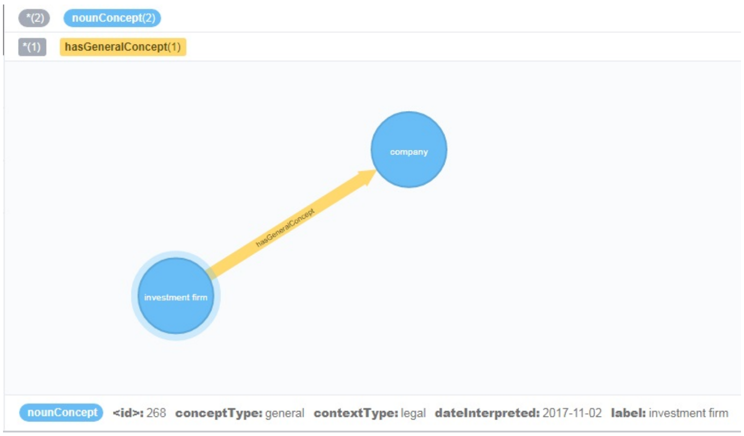

Fig. 9. Example of graphical representation of noun concepts

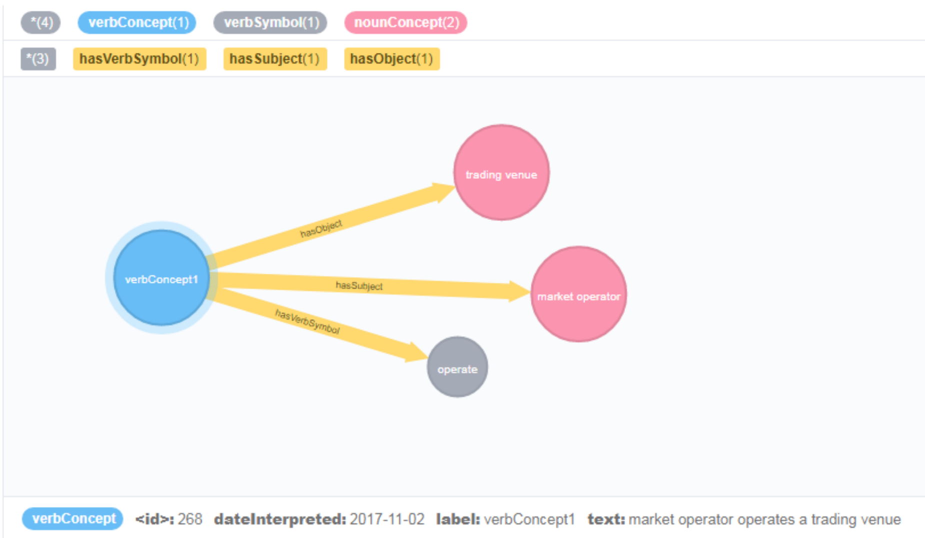

Fig. 10. Example of graphical representation of verb concepts

4.2 Graphical Representation of Rules

Legal rules are more complex than noun and verb concepts. Each rule includes one or more applicability conditions that determine whether a given action is relevant to a given rule or not, and one or more deontic conditions that determine whether a relevant action complies with or breaches a rule. In our current implementation, we represent a legal rule as a graph node that is connected to the graph nodes of applicability condition and deontic condition with the edges *hasApplicabilityCondition* and *hasDeonticCondition*. The nodes of applicability condition and deontic condition have edges *associatedWith* connecting them to action (verb concept) nodes.

Fig. 11 shows an example of graphical representation of the rule used throughout the paper (Rule 1). This rule has key-value pairs containing some basic information such as the original regulatory text shown in the bottom bar. The $rule_1$ node connects two deontic condition nodes and one applicability condition node. Each of them is connected to a verb concept node by the edge *associatedWith*.

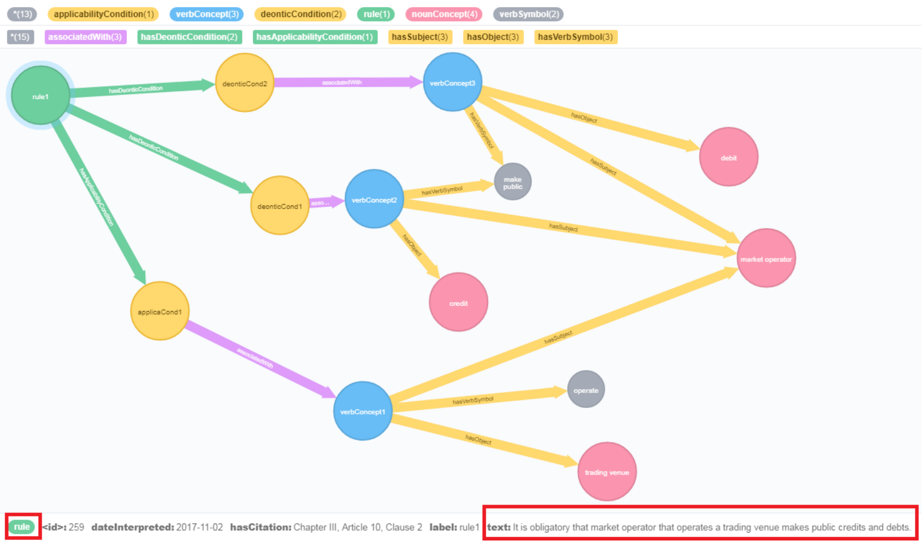

Fig. 11. Example of graphical representation of a rule

4.3 Querying Graph Data

The rules and vocabularies represented in property graphs can be queried and retrieved by users. Queries over graph databases are often graph patterns. Here, we use the Cypher language of the Neo4j graph database to formulate graph patterns and pose the Cypher queries in the graph database to retrieve graph data. Note that queries can also be visualised since they are graphs with variables. We also formulate some common queries as query templates which can help the users who have no knowledge of Cypher to formulate their queries.

5 Conclusions

This paper presents visualisation needs and enhancements for a rule-editing tool used by legal experts to achieve a machine-readable interpretation of legal requirements with a semi-structured controlled natural language. The paper presented knowledge graphs as a way to visualize an interpreted rule and all its relevant elements in a format that the lawyer can understand and manipulate. This is meant to ensure that the machine-readable output of the regulation interpretation process is semantically enriched as intended by the legal expert, thus ensuring the reliability of the interpretation stored in the machine-readable format. Next steps of the research include further investigation of graphing solutions. We anticipate that this might bring us to reconsider some basic formalization principles derived from SBVR, which in turn might lead us to reconsider some modelling choices in Mercury.

References

1. EY's 2015 global governance, risk and compliance survey: Innovating with RegTech (2015), `http://www.ey.com/Publication/vwLUAssets/EY-Innovating-with-RegTech/\$FILE/EY-Innovating-with-RegTech.pdf`
2. Abi-Lahoud, E., O'Brien, L., Butler, T.: On the road to regulatory ontologies: Interpreting regulations with SBVR. In: AI Approaches to the Complexity of Legal Systems. Lecture Notes in Artificial Intelligence, vol. 8929, pp. 188–201. Springer Berlin (2014)
3. Al Khalil, F., Ceci, M., Yapa, K., O'Brien, L.: SBVR to OWL 2 mapping in the domain of legal rules. In: International Symposium on Rules and Rule Markup Languages for the Semantic Web. pp. 258–266. Springer (2016)
4. Arner, D.W., Barberis, J., Buckley, R.P., et al.: The emergence of RegTech 2.0: From know your customer to know your data. Journal of Financial Transformation 44, 79–86 (2016)
5. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.: The description logic handbook: Theory, implementations and applications. Cambridge U. Press, Cambridge, England (2003)
6. Ceci, M., Al Khalil, F., O'Brien, L.: Making sense of regulations with SBVR. In: Proceedings of the RuleML 2016 Challenge, Doctoral Consortium and Industry Track (2016)
7. Ceci, M., Al Khalil, F., O'Brien, L., Butler, T.: Requirements for an intermediate language bridging legal text and rules. In: Workshop on 'MIning and REasoning with Legal texts' (MIREL). Nice, France (2016)
8. Ceci, M., Butler, T., O'Brien, L., Al Khalil, F.: Legal patterns for different constitutive rules. In: Workshop on 'MIning and REasoning with Legal texts' (MIREL). pp. 1–11. Kings College, London (2017)

9. Gordon, T., Governatori, G., Rotolo, A.: Rules and norms: Requirements for rule interchange languages in the legal domain. Rule interchange and applications pp. 282–296 (2009)
10. Governatori, G., Shek, S.: Rule based business process compliance. In: RuleML (2) (2012)
11. van Haarst, R.: SBVR made easy: Business vocabulary and rules as a critical asset. Conceptual Heaven, Amsterdam (2013)
12. Höfler, S., Bünzli, A.: Designing a controlled natural language for the representation of legal norms. In: Second Workshop on Controlled Natural Languages (2010)
13. Maranhao, J.S.: A logical architecture for dynamic legal interpretation. In: 16th International Conference on Artificial Intelligence and Law (ICAIL) (2017)
14. Object Management Group (OMG): Semantics of Business Vocabulary and Rules (SBVR). Specifications (May 2017), http://www.omg.org/spec/SBVR/1.4/
15. O'Brien, L., Abi-Lahoud, E., Lombard, J.: Using business natural language to capture financial regulations in a collaborative ontology development process. In: Law Via the Internet. Jersey (26-27 September 2013)
16. Siena, A., Mylopoulos, J., Perini, A., Susi, A.: From laws to requirements. In: Requirements Engineering and Law (RELAW'08). pp. 6–10 (2008)
17. Wyner, A., Peters, W.: On rule extraction from regulations. In: JURIX. vol. 11, pp. 113–122 (2011)