Towards Integrating the FLG Framework with the NLP Combinatory Framework

Amin Rabinia | Sepideh Ghanavati | Mauro Dragoni

ABSTRACT

Automatic modeling of privacy regulations is a highly demanded goal in requirements engineering. The FOL-based Legal-GRL (FLG) is a semi-automated modeling framework for extracting and representing the legal requirements of IT systems. One limitation of the FLG framework, however, is its manual extraction process. Manual extraction of legal requirements is cumbersome, error-prone, and time-consuming. To overcome this shortcoming, we integrate the FLG with the Combinatory framework which combines several natural language processing (NLP) approaches. The Combinatory framework exploits NLP techniques and tools to automate the extraction of rules from legal texts. The integration of the two enables us to fully automate the FLG framework in order to propose a comprehensive framework for modeling privacy regulations. This poster outlines the two frameworks and their integration process.

BACKGROUND

- The FLG framework as a formal approach aims at automating legal requirements modeling. It has three phases:
  - A. manual requirements extraction;
  - B. storing the requirements in a database;
  - C. requirements models generation.
- The NLP Combinatory framework converts natural language statements to rules. It has two sets of NLP techniques:
  - Stanford Parser and C&C/Boxer, for grammatical analysis of the sentences,
  - WordNet, for handling the deontic language of the legal texts.

INTEGRATION OF THE FRAMEWORKS

The integrated framework has three phases:

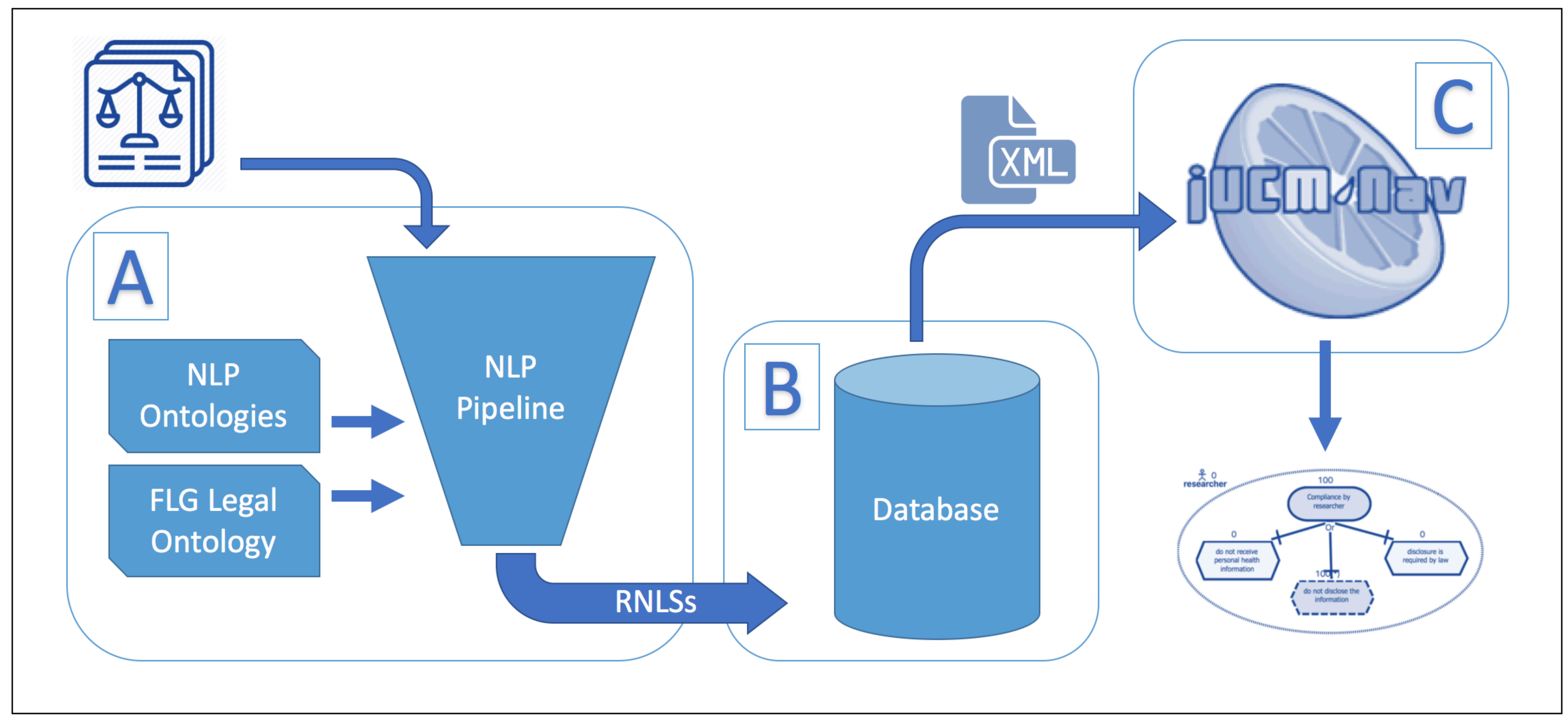
A. Legal requirements extraction, where the legal statements turn into the RNLSs. The NLP Combinatory applies several ontologies along with the NLP techniques upon the input statements.

B. Storing the requirements, where the RNLSs would be stored and finally retrieved as an XML/GRL file.

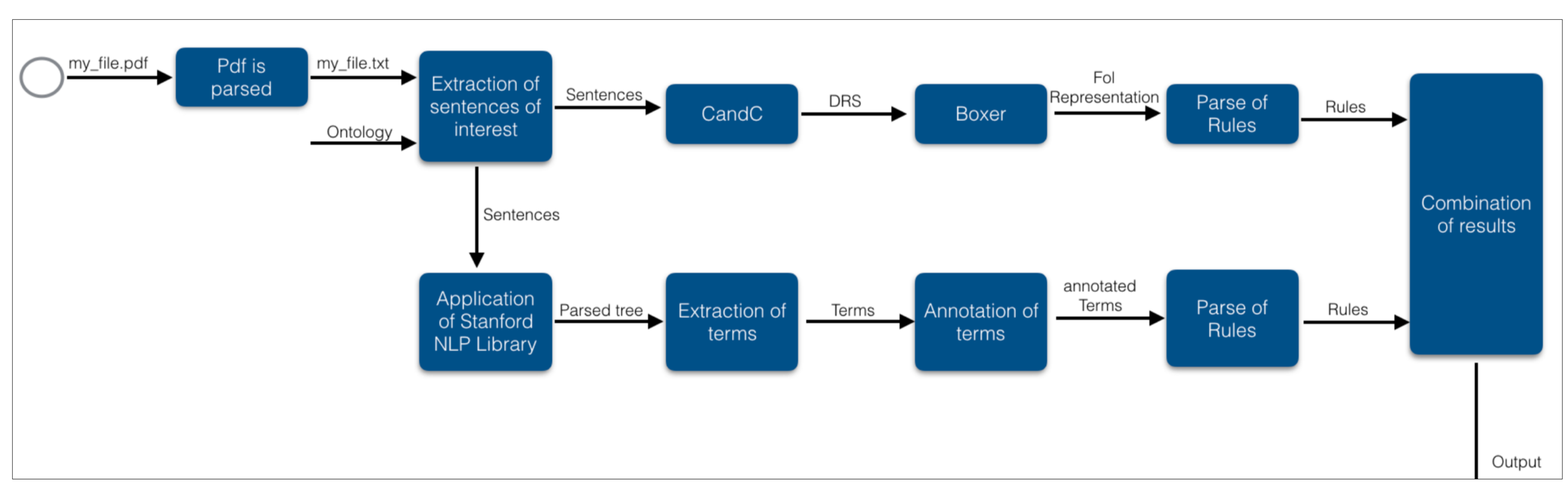
C. Goal model generation, where the XML file is imported into jUCMNav to generate goal models.

Within the integrated framework, the entire process of modeling would be automated.

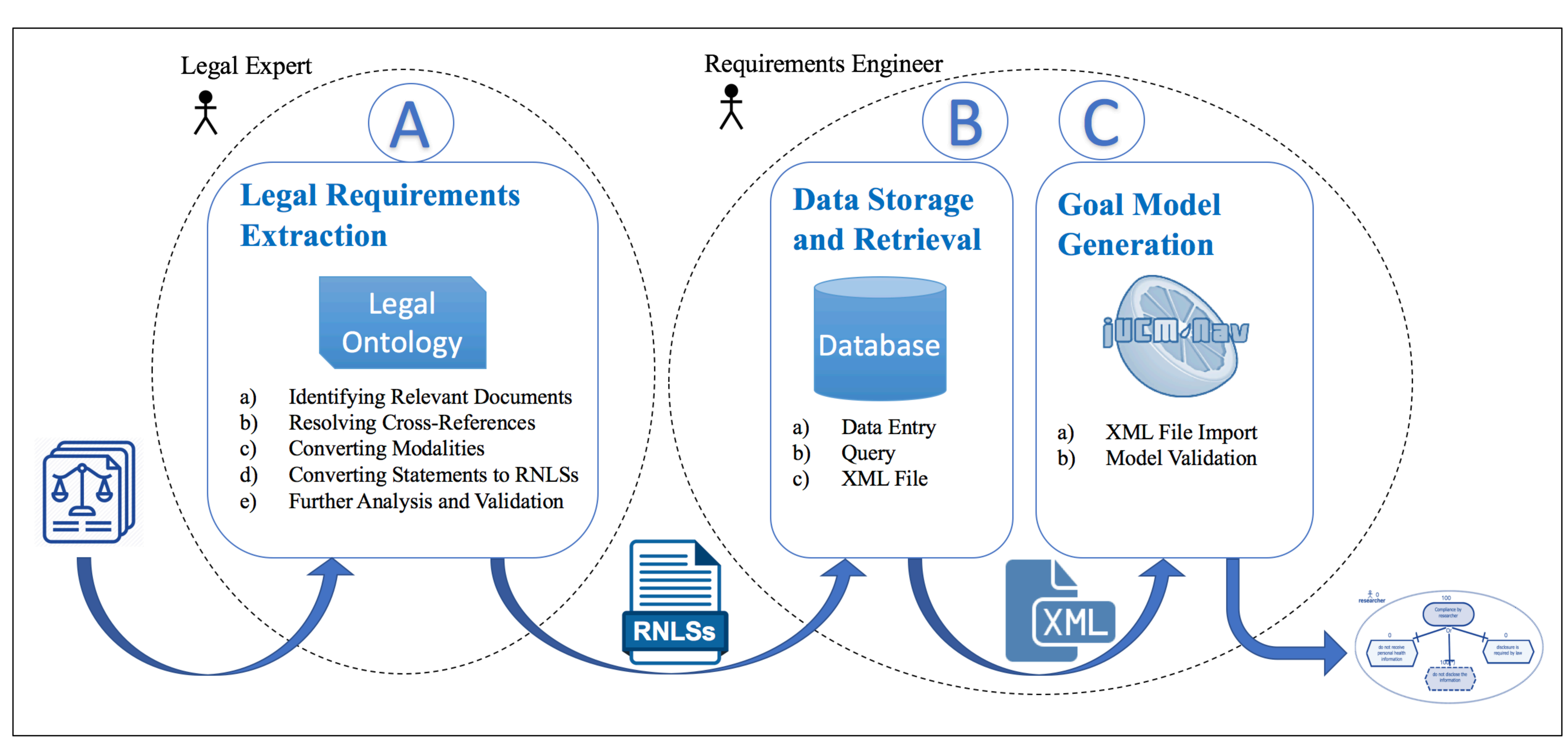
THE FLG FRAMEWORK



THE NLP COMBINATORY FRAMEWORK



THE INTEGRATED FRAMEWORK



UNIVERSITY OF MAINE, SCHOOL OF COMPUTING AND INFORMATION SCIENCE
FONDAZIONE BRUNO KESSLER

Towards Integrating the FLG Framework with the NLP Combinatory Framework

Amin Rabinia[1], Mauro Dragoni[2], Sepideh Ghanavati[1]
[1] School of Computing and Information Science, University of Maine, Orono, ME, USA
{amin.rabinia, sepideh.ghanavati}@maine.edu
[2] Researcher at Fondazione Bruno Kessler, Trento, Italy
dragoni@fbk.eu

Abstract

Automatic modeling of privacy regulations is a highly demanded goal in requirements engineering. The FOL-based Legal-GRL (FLG) is a semi-automated goal-oriented modeling framework for extracting and representing the legal requirements of IT systems. One limitation of the FLG framework, however, is its manual requirements extraction process. Manual extraction of legal requirements is cumbersome, error-prone, and time-consuming. To overcome this shortcoming, we integrate this requirements modeling framework with another framework that combines several natural language processing (NLP) approaches. This Combinatory framework specifically exploits NLP techniques, such as Part-Of-Speech tagging and syntactic parsing, along with NLP tools, such as C&C and Boxer, to propose an automated approach for extraction of rules from legal texts. This approach enables us to fully automate the FLG framework in order to propose a comprehensive framework for modeling privacy regulations and other legal requirements. This paper outlines the two frameworks and their integration process.

Introduction

Privacy regulations enforce a set of legal requirements for any IT system. Developers need to capture and satisfy such requirements in their products. Legal-GRL (Ghanavati et al. 2014) is a modeling language for extracting and representing privacy regulations. One drawback of Legal-GRL is its manual modeling process. This limitation has motivated an alternative modeling framework, the FOL-based Legal-GRL (FLG) (Rabinia and Ghanavati 2017 and 2018). The FLG framework suggests a new approach based on First-order Logic (FOL) to facilitate the automation of modeling process. The FLG covers two main tasks of legal requirements extraction and representation. The representation task, is fully automated (Rabinia and Ghanavati 2018). The requirements extraction task, however, is still manual. To automate this task, we suggest integration of FLG with the NLP Combinatory framework (Dragoni et al. 2016). This framework combines several natural language processing (NLP) techniques and tools in order to extract rules from legal texts. While the Combinatory framework performs the extraction task, the FLG handles the representation part of legal requirements modeling. In this paper, we aim to sketch a brief outline of the two frameworks and the theoretical/practical possibilities for integration of both in one unified framework.

Background

The FLG framework consists of three phases: A. manual requirements extraction; B. storing the requirements in a database; and, C. requirements models generation. The extraction phase (A) starts with collecting and analyzing natural language legal statements. This is a manual task that legal experts have to accomplish using a legal ontology. The result of this process is a set of atomized and structured statements called Restricted Natural Language Statements (RNLSs). In the second phase (B), the RNLSs, which are in FOL notation format, would be stored in an SQL database.[1] This database is capable of exporting the data (or the formalized legal requirements) as a formatted XML file. In the third phase (C), the XML file is imported into the modeling tool support, jUCMNav (Amyot et al. 2011), to automatically generate the requirements goal models.

To automate the first phase of the FLG framework, we integrate it with the NLP Combinatory framework. This framework combines two sets of NLP techniques: first, Stanford Parser and C&C/Boxer, for grammatical analysis

[1] The database is accessible at https://github.com/PERC-Lab/FLGFramework

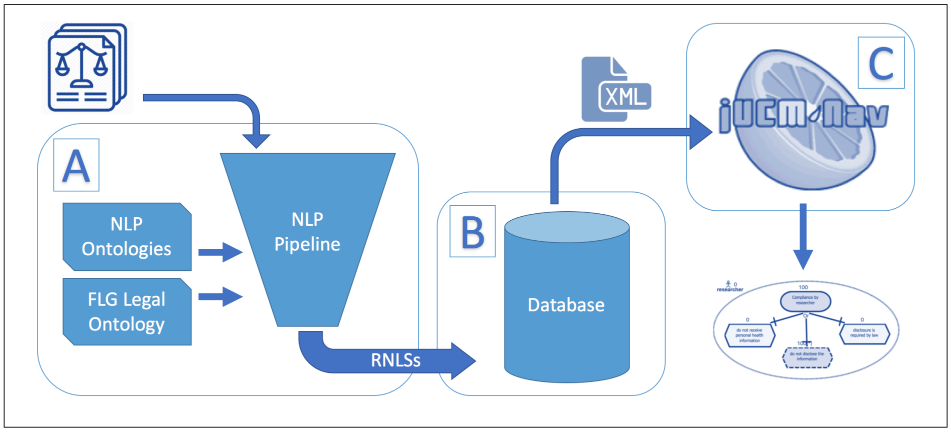


Fig. 1. Outline of the Integrated Framework

of the sentences, and second, WordNet, for handling the deontic language of the legal texts. The input data, consisting of natural language statements from a legal text, pass through two separate pipelines that perform the NLP processes, mentioned above, to extract the rules from the original statements. The result from both branches of the pipeline will be combined to confirm each other.

Integration of the Frameworks

Fig.1. shows a sketch of the integration of the two frameworks. The integrated framework would be organized in three phases: A. Legal requirements extraction, where the legal documents are imported into the NLP Combinatory pipeline and the RNLSs would be exported. The process inside the pipeline entails application of several ontologies along with the NLP techniques upon the input statements. B. Storing the requirements, where the RNLSs would be stored and finally retrieved as an XML/GRL file. C. Goal model generation, where the XML file is imported into the modeling tool support, jUCMNav, to generate the legal requirements goal models. Within the integrated framework, the entire process of modeling would be automated. This possibility, however, demands further modifications/improvements of the two frameworks, as they are still in progress. The integration process takes place in three layers:

1) A unified format of output/input needs to be set for a proper dataflow between the two frameworks. Since the FLG framework has an SQL Server database, it can import data from a variety of file formats including a plain text file. The RNLSs, then, can be stored in the NLP Combinatory framework as a text file and imported in the FLG database.

2) The FLG framework requires a finer-grained analysis of the rules than what the NLP Combinatory framework provides as its output. The general format of such output is: `Term1 => [O] Term2`, where terms are parts of a sentence or statement. However, the final goal models of the FLG need further details of the requirements, e.g. actors. These elements of the statements are accessible within the NLP Combinatory framework, as it exploits the NLP techniques..

3) The FLG framework has also a legal ontology that helps simplifying the legal requirements models. During the integration process, the ontologies of the two frameworks should be integrated in order to handle syntactical modifications of the statements.

We perform the integration procedure in an iterative, non-sequential fashion. The integration of ontologies (3) should be the first priority, since it contributes to the second layer. Running the dataflow (1) between the two frameworks also can be done afterwards, when the data format is matched during 2 and 3.

The outcome of this integration is a fully automated framework for modeling of privacy requirements.

References

Amyot, D., Mussbacher, G., Ghanavati, S. and Kealey, J., 2011. GRL Modeling and Analysis with jUCMNav. *iStar*, *766*, pp.160-162.

Dragoni, M., Villata, S., Rizzi, W. and Governatori, G., 2016, December. Combining NLP approaches for rule extraction from legal documents. In 1st Workshop on MIning and REasoning with Legal texts (MIREL 2016).

Ghanavati, S., Amyot, D. and Rifaut, A., 2014, June. Legal goal-oriented requirement language (legal GRL) for modeling regulations. In Proceedings of the 6th international workshop on modeling in software engineering (pp. 1-6). ACM.

Rabinia, A. and Ghanavati, S., 2017, September. FOL-Based Approach for Improving Legal-GRL Modeling Framework: A Case for Requirements Engineering of Legal Regulations of Social Media. In 2017 IEEE 25th International Requirements Engineering Conference Workshops (REW) (pp. 213-218). IEEE.

Rabinia, A. and Ghanavati, S., 2018, August. The FOL-Based Legal-GRL (FLG) Framework: Towards an Automated Goal Modeling Approach for Regulations. In 2018 IEEE 8th International Model-Driven Requirements Engineering Workshop (MoDRE) (pp. 58-67). IEEE.