Semantic Data Web Technologies/EIS Lab

Summer Semester 2016

Software Requirements Specification

Ontology-Driven Summarization and Visualization of Legal Documents

Project Team

Ahmad Hemid

Lars Möllenbrok

Isunge Mwangase

Md Saiful Islam Faisal

Project Mentor

Najmeh Mousavu Nejad

Table of Contents

1. Introduction

While using computers, we all have to deal with end user license agreements (EULA). This happens when you install software or sign up for a web service and typically consist of several pages. But because of the length and the jargon of these documents most of us have gotten into the habit of simply accepting these agreements without looking at them which may lead to unexpected or unwanted consequences. In our project we tackle this problem by developing a system that uses ontology-based information extraction to quickly summarize a license, classify the different terms and conditions and visualize the results. But what is an ontology?

There have been many attempts to define what constitutes an ontology, perhaps the best known (at least between computer scientists) being to Gruber:"an ontology is an explicit specification of a conceptualization". They constitute a formal conceptualization of a particular domain of interest that is shared by a group of people. When building ontologies into information systems, it is possible to modularize many software aspects mostly related to the domain (e.g., taxonomic structures) from ones mostly related to the processing (e.g., querying) and visualization (e.g., layouting) of data [1,2].

1.1 Purpose

This document provides a detailed overview of the software requirements used in the implementation of this project. This requirement document is to be used in the software development process by stakeholders, developers, testers and project leaders.

1.2 Scope

The scope of the project is defined as follows:

- Understand the content of vocabularies of legal document
- Extract structures and semantics of vocabularies related to legal terms
- Summarize the legal document based on a predefined ontology
- Classify the sentences based on the semantics of vocabularies

- Present the output in an easy way to understand web GUI to the customer
- Work as a cross-platform software

1.3 Definitions, Acronyms, and Abbreviations.

EULA – End User License Agreement

URL – Uniform Resource Locator

FR – Functional requirement

NFR – Non-functional requirement

2. The Overall Description

2.1 Product Perspective

A lot of people accept EULAs without even reading the content or understanding the impacts of accepting to the terms of such documents. From this vantage point, we intend to have a software that can summarize the content of EULAs and clearly show the important legal terms it contains. However, a few of recent research scientists have put effort on such a problem, a master thesis [3] related to our project has touched the extraction part of our project in the license domain. Our product delivers a quick, scalable and fast interface which is highly needed, showing the legal terms contained in EULAs with their summaries. Figure 1 briefly describe the main functions, planned to be implemented to fulfill the project requirements.

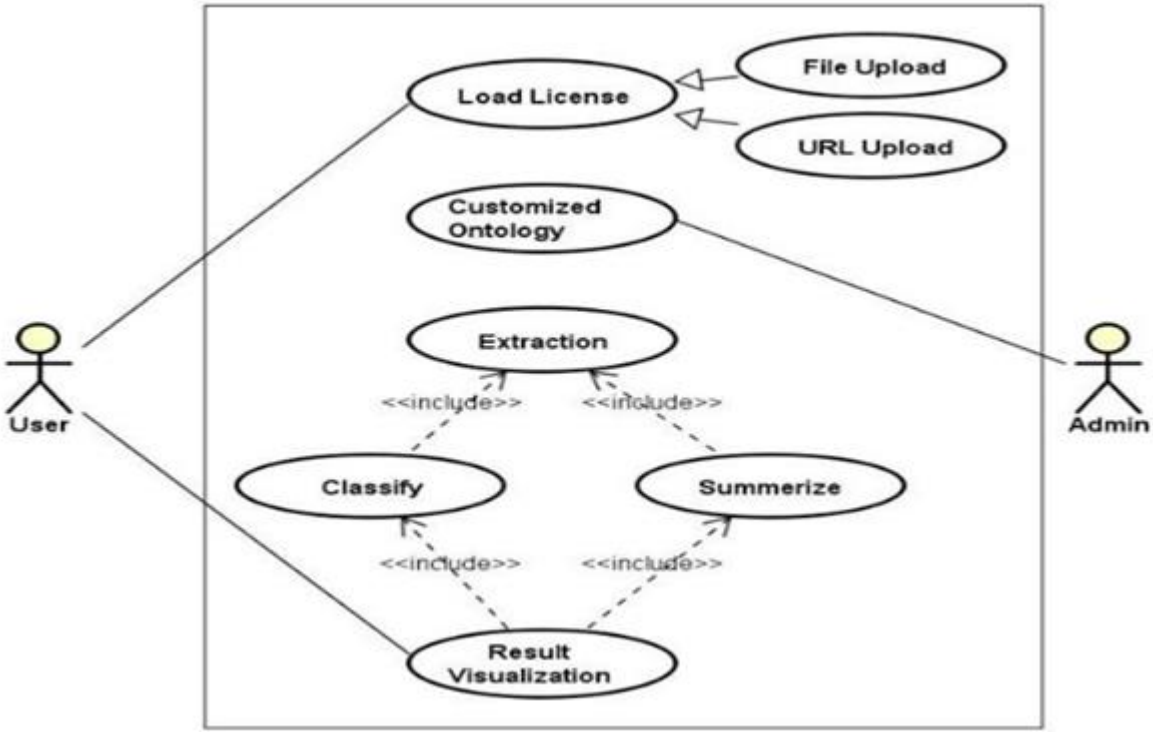

Figure 1. UML use-case diagram

2.2 Interfaces

The backend is Java and the frontend is HTML, CSS, and Javascript.

2.3 User Characteristics

The user shall be anyone interested in knowing the contents of an EULA before installing software or signing up for a web service.

2.4 Assumptions and Dependencies

- The ability to access the system with a remote access using internet or any kind of local connection between the server having the system and the user terminal
- The availability of a web browser on the user terminal
- The knowledge of uploading a file or a URL.

3. Requirement Specifications

The analysis of project requirements leads to discussion with the customer how those are formalized and tracked. The project will be built in the scope of the initial requirements agreement.

3.1 External Interfaces

The project allows the input of a legal document /License either by uploading a URL or uploading a file from a certain storage. The output will be on a web interface to give the user a summary of the uploaded document as well as represent the legal terms in such a way the user will catch the document' content with just having a look to the output. In the case of the unsuccessful upload of the document, the user will be prompted with an error message and no further processing will be done.

3.2 Functional Requirements

The following requirements describe the functions of the system. A priority notation has been used. This notation is a marker for requirements importance. A low number, such as 1, highlights a priority of high importance. A higher number presents a priority of lower importance. The system functions will be implemented in the order of priority. Requirements with priority 2 may be skipped from implementation, if time constraints do not allow it.

.

ID	NAME	PRIORITY
FR_1	**Semantification of legal document** ❖ The user shall be able to upload an EULA from a file or a URL ❖ The EULA has to be semantified using a predefined ontology	1
FR_2	**Summarize the document** ❖ The EULA has to be summarized using ontology-based extraction	1
FR_3	Classify the sentences ❖ The different legal terms have to be classified using ontology-based extraction	1

FR_4	**Present results using a web GUI** ❖ The system shall be able to present the result on a web GUI	1
FR_5	**Record event failures** ❖ The system shall rises an exception error when a failure occurred	1
FR_6	**Customize the ontology** ❖ The predefined ontology can be customized to support a wide range of legal documents' areas	2
FR_7	**Provide tool tips on some results** ❖ The system shall provide different classifications tool tips, quoting the corresponding paragraph of the original document	2
FR_8	**Visualize the results using icons** ❖ The result can be visualized used simple signs to help the user to understand the content of the EULA	2

3.3 Non-functional Requirements:

- **Performance Requirements**

ID	NAME	PRIORITY
NFR_1	The response time shall be less than one minute.	1
NFR_2	The number of simultaneous users will be supported.	2

- **Availability**

ID	NAME	PRIORITY
NFR_3	The web application will have 99% uptime.	2

- **Usability**

ID	NAME	PRIORITY
NFR_4	The user interface will be in simple English language	1
NFR_5	The user can gives feedback to enhance the quality	2

- **Security**

ID	NAME	PRIORITY
NFR_6	The system checks the uploaded file or URL for malicious damage/code.	2

- **Interoperability**

ID	NAME	PRIORITY
NFR_7	The system can be provided as a web service to operate as being part of another system.	2

4. References

[1] Franz Baader, Ian Horrocks and Ulrike Sattler. Description Logics. In Steffen Staab, Rudi Studer (2004). Handbook on Ontologies.

[2] Alexander Maedche and Steffen Staab. Ontology Learning. In Steffen Staab, Rudi Studer (2004). Handbook on Ontologies.

[3] Ken-Thomas Nilsen (2015). Ontology Based Information Extraction in the License Domain