

Semi-automatic Ontology Construction for Improving Comprehension of Legal Documents

Bojan Cestnik^{1,2}, Alenka Kern³, and Helena Modrijan³

¹ Temida d.o.o., Ljubljana, Slovenia

² Jozef Stefan Institute, Ljubljana, Slovenia

bojan.cestnik@temida.si

³ The Housig fund of the Republic of Slovenia, a public fund, Ljubljana, Slovenia
{alenka.kern, helena.modrijan}@stanovanjskisklad-rs.si

Abstract. In this paper we present a new method based on semi-automatic ontology construction that can be used to improve the understandability of legal documents. Legal documents typically extensively define requirements and procedures in a specific legislative area; usually, they are hard to comprehend for citizens without a proper legal knowledge. However, a vast majority of today's e-Government activities for citizens (G2C) are governed by legal documents. Therefore, by improving the citizen's comprehension many intricacies that occasionally occur during G2C activities can be avoided. Our method first divides a legal document into several paragraphs. From the paragraphs it semi-automatically constructs ontology of the field by using a tool OntoGen. Ontology concepts are then used to classify each paragraph and the resulting classification is visualized in a simple matrix, where rows represent paragraphs and columns represent top-level ontology concepts. Based on the visualization, paragraphs that need revising are identified; they can be relocated to more suitable context within the document or rewritten using more appropriate wording. We demonstrate the presented method on the document defining the tender for selling flats at favorable prices at the Housing Fund of the Republic of Slovenia, a public fund. We argue that by using the new method we were able to substantially improve the comprehension of the document. In addition, the constructed ontology helped the Fund's officers improve the structure of their knowledge about the underlying business process.

Keywords: ontology construction, knowledge management, legal documents.

1 Introduction

For decades, scientists have used ontologies as a means to systematize scientific information and to provide a common vocabulary of concepts for exchanging information. Ontologies have the capability to contribute to common understanding of problem domains. Therefore, they are capable of supporting research with the ability to reason over and to analyze the information at stake [1]. As a result, ontologies are most commonly used as a form of knowledge representation. Typically, ontologies consist of descriptions of objects, concepts, attributes and relations between objects.

Until recently, ontology construction relied mostly on manually identifying some interesting concepts and organizing them in an appropriate hierarchy. In the process, some sort of language was used to represent manually extracted common sense knowledge from various sources. Lately, several programs that support manual ontology construction have been developed, like for example Protégé [2]. However, since manual ontology construction is a complex and demanding process, there is a strong tendency to provide more active computerized support for the task.

With the emergence of new knowledge technologies, ontologies can be constructed semi-automatically by processing textual data at hand. In recent years, many tools that help constructing ontologies from texts in a given problem domain were developed and successfully used in practice [3]. Therefore, the process of ontology construction can be made more effective and feasible in practice. Based on text mining techniques that have already proven successful for the task, OntoGen [4] is a tool that enables interactive construction of ontologies from text documents in a selected domain. A user can create concepts, organize them into topics and also assign documents to concepts. With the use of machine learning techniques OntoGen supports individual phases of ontology construction by suggesting concepts and their names, by defining relations between them, and by automatic assignment of documents to the concepts [5].

Nowadays, citizens are faced with vast amounts of legal documents that govern their everyday activities. Most of such documents are difficult to read and comprehend for an average person. Usually, misunderstandings that arise are handled within or after the process and require substantial amount of human resources. In the field of e-Government services to citizens (G2C), improving the comprehension of legal documents can thus be considered as a task that both improves citizens' satisfaction with the services and saves time and money at the same time. Note that a formal representation of knowledge in the form of ontology can be effectively used to communicate the key concepts from the responsible officers to citizens.

As an important Slovenian public institution, the Housing Fund of the Republic of Slovenia, a public fund, has been actively involved in several tasks with the intention of improving the housing conditions for the citizens of Slovenia. In the first period, the Fund started offering the loans under favorable conditions to the individual housing investors [6]. Afterwards, its focus changed to encourage individual housing savings by offering attractive Housing Savings Schema to citizens. The motivation for this action was to direct considerable amount of financial resources from consumption to savings, which has, among other affirmative things, turned out to have a positive effect on Slovenian economy [7]. Even though the Fund operates most of its businesses as a part of the public sector, the underlying requirements for its activities demand organizational form of an enterprise. The Fund has a unique position within Slovenian economy, which puts it under considerable media attention.

In the past years, effective public communication has been one of the key priorities of the Fund's management. Even though the task of preparing legal documents governing the Fund's business activities was every time carried out with a great care, there were cases that evidently reveal the lack of public understanding of the Fund's requirements and intentions. Sometimes, especially in the case of public tenders for selling flats to citizens, media contributed to confusions about the issues at stake. Therefore, the task of introducing a method for improving the understandability and clarity of presentation is well supported by the Fund's management.