Suggesting Model Fragments for Sentences in Dutch Law

EMILE DE MAAT, RADBOUD WINKELS*


1. INTRODUCTION

A main issue in the field of artificial intelligence and law is the transformation of sources of law that are written in natural language (and therefore rather informal) into formal models of law that computers can reason with. This is a time and effort consuming process, error prone and different knowledge engineers will arrive at different models for the same sources of law. Moreover, these models should be closely linked to the original sources (and at the right level of detail, i.e. isomorphic) since these sources tend to change over time and maintenance of the models is a serious problem. This calls for tools and a method for supporting this modelling process and increasing inter-coder reliability.

We have been researching a method to create isomorphic models semi-automatically, focusing on (Dutch) laws. This article presents a next step in this creation process.

1.1. General Approach

In order to achieve (semi-)automatic modelling of sources of law, we follow a number of steps, as shown in Fig. 1. The process starts with the source document, written in natural language (Dutch). Currently, we focus on laws, though we hope to expand to other types of sources of law later on. We first make the structure of the document explicit, by marking up the different parts, such as chapters, paragraphs and sentences, and assigning identifiers to each part. We then proceed to mark all references within the source to

* E. de Maat is Researcher at the "Leibniz Center for Law" (LCL); R. Winkels is Associate Professor in Computer Science and Law at LCL of the Faculty of Law of the University of Amsterdam.
other sources of law, using a parser based on patterns for references\(^1\). This structure and reference information is stored in CEN/MetaLex XML\(^2\).

![Fig. 1 – Steps in automatic modelling of legal texts](image)

The next step is to create models for each individual statement in the text. In most cases, each sentence in Dutch law forms a complete statement (though possibly part of a bigger construct), so we are, in fact, creating a model for each sentence in the text. In the last step, these individual models are integrated with each other to come to a complete model. In order to create the models, we start by classifying each sentence in the text as a specific provision, such as a definition, a duty, or a modification of an earlier law. In total, we recognise ten different main categories. As with the references, this is done by automatic recognition of certain patterns in the text\(^3\).

For several types of sentences, these patterns, together with some added features, are sufficient to extract all information needed to create a model of the sentence. This is usually the case with sentences that are about the law itself, instead of the subject matter of the law. These sentences are discussed in Section 2. Other sentences, such as obligations, do focus on the subject matter, and can vary wildly. Simple patterns will not suffice to deal with these sentences, and to extract information from these types of sentences, we

---


Continua ...