

Title	Study on Extracting Conceptual Structures from Legal Texts
Author(s)	ファム, アンカム Jr
Citation	
Issue Date	2012-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/10414
Rights	
Description	Supervisor: Prof. Akira Shimazu, 情報科学研究科, 修士

Study on Extracting Conceptual Structures from Legal Texts

by

Pham Anh Cam (1010057)

School of Information Science

Japan Advanced Institute of Science and Technology

February 06, 2012

Keywords: *Japanese laws, conceptual model, UML class*

Abstract

Law is a system of rules and guidelines which are enforced through social institutions to govern behavior, wherever possible. Therefore, to understand and to check the correctness and the consistence of a law is very important but normally very difficult because of long and difficult texts. Our research aims to take a hand in this work by modeling Japanese laws. Such laws can be seen as specifications of social systems. Our purpose is to clarify a method for extracting a model (conceptual model) of social system which such laws represent.

Our research is conducted as the study of legal engineering, which aims to exam and verify whether a law has been established appropriately according to its purpose, whether the law is consistent with related laws, and whether the law has been modified, added, and deleted consistently. There are two important goals of legal engineering which are to help experts make complete and consistent laws; and to design information systems which work based on laws.

To model the social system represented by Japanese laws, we use a conceptual model. The term conceptual model may be used to refer to model which is represented by concepts or related concepts which are formed after a conceptualization process in the mind. In order to express the result of conceptual modeling, we need a modeling language. A modeling language can be graphical or textual. There are several elaborate graphical notations that could be used for conceptual modeling, ranging from those used in the past for Entity-Relationship Model [Chen 1975], conceptual graphs [Sowa

1984] to those presently used for object modeling [OMG 1997].

Recently there are some researches about using UML class diagrams to model natural language. However, all of them are applied to English language only and all methods used a semi-formal language as a medium. Different from these above methods, our method is used for Japanese language, concentrate on Japanese legal texts; and we directly extract UML classes from dependencies of Japanese legal terms which are extracted based on Cabocha parse output. Besides, because of a huge part of extracted classes, we don't represent classes graphically but textually.

Our model is built by the following way: the common nouns are extracted to become the UML classes. These classes have attributes, methods and relations extracted correspondingly from the dependencies of legal terms as below. The verbs are extracted to become methods of classes. The modifier nouns, verbs and adjectives are extracted to become attributes of classes. The relations between nouns are extracted to become relations between classes. These classes then are represented textually. There are 3 steps to transform Japanese legal texts to UML class components. First, we parse Japanese law text corpus using Cabocha. Second, we extract dependencies of legal terms from Cabocha output. Third, we transform the dependencies to UML class components.

Our research contribution is that our model can be used as a tool to understand the law easily. Instead of reading a difficult law, people can take a look to the model to find out the related terms, actions as well as the relations between the objectives of the law – legal terms. Moreover, the relations between legal objectives are modeled; the logic of the law is also clearly expressed so that we can check the correctness and consistence of a law easily. We can also design the information systems based on laws using our conceptual model.

The challenges of this research are how to extract correctly and completely relations between Japanese legal terms and how to represent these terms with their characteristics and relations clearly. The prior problem is solved by careful analysis and extraction of the most important Japanese dependencies. The later problem is solved by representing classes with their attributes, methods and relation textually. Graphical representation is not used because there is a huge part of classes are extracted from each law.