

Title	法令文の理解及び利用を支援するための言語処理法に関する研究
Author(s)	Le, Thi Ngoc Tho
Citation	
Issue Date	2015-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/13536
Rights	
Description	Supervisor:NGUYEN, Minh Le, 情報科学研究科, 博士

氏名	LE THI NGOC THO		
学位の種類	博士(情報科学)		
学位記番号	博情第 329 号		
学位授与年月日	平成 27 年 9 月 24 日		
論文題目	Study on language processing methods for supporting understanding and using multiple legal documents (法令文の理解及び利用を支援するための言語処理法に関する研究)		
論文審査委員	主査	NGUYEN, Minh Le	北陸先端科学技術大学院大学 准教授
		東条 敏	同 教授
		白井 清昭	同 准教授
		島津 明	同 シニアプロフェッサー
		Ho Bao Tu	同 教授

論文の内容の要旨

Law plays a significant role in governing our society and business. The system of legal documents in every country is often complicated with various kinds of documents, which are modified frequently to reflex the changing in situations of society/business, or to make the law more completed. Practically, the performance of retrieving legal information is still low when using the traditional strategy. Heretofore, the best solution to improve the performance is the exploiting a knowledge-base in retrieval. Nevertheless, the resource of knowledge-bases is not at hand and manually making knowledge-base is very expensive. For that reason, there is a requirement of automatic constructing of a knowledge-base to improve the performance of legal information retrieval. In addition, the contents and structures of legal documents are often complicated. Therefore, searching and reading legal documents is not easy for both normal citizens and legislators. We motivate to support the retrieving task by constructing the legal knowledge-base automatically; and, to help the readers by providing a hierarchical structure of legal indices which structurally yields the important information of legal documents. We divided the generation of the hierarchical structure into two main tasks: extracting legal indices and discovering relations among these indices.

The first task, extracting the indices which yield the main contents of legal documents, is treated as the problem of keyphrase extraction. We explored this extraction problem on two languages:

Japanese and English. In the Japanese legal context, the legal indices are words, phrases and clauses. Since Japanese keyphrases are found in chunks and clauses, we approach index extraction using structural information of Japanese sentences, i.e. chunks and clauses. In English text, however, the chunk information does not really help improving the extraction performance because English chunks include words that cause noise in keyphrases. In the literature, current studies often extract English keyphrases by collecting adjacent important adjectives and nouns. Analysis on the data shows that keyphrases also contain other kinds of words. Hence, we proposed a solution to improve the extraction performance by involving new kinds of words to keyphrases.

The second task, constructing the relations among the indices, is treated as the problem of legal ontology construction. We proposed an approach to extract the super/subordinate relation between each pair of concepts individually based on directional similarity. The relations among a set of legal indices are represented in a directed graph and the hierarchical structure of indices is simply exported from this graph. We adopted this proposal to the Japanese National Pension Act document. The resulted hierarchical structure is compared to an annotated legal ontology on the number of correct relations.

In this dissertation, there are two main contributions: novel approaches to extract keyphrases from Japanese and English text and novel approach to discover relationships among legal concepts in the construction of Japanese legal ontology. Our study serves as the necessary steps to construct the knowledge-based for legal information retrieval. In addition, the hierarchical structure of legal indices also serves as a structural summary of the main concepts, which enables the readers understand the relations among the legal concepts.

Keywords: legal engineering, unsupervised approach, keyphrase extraction, hierarchical index, ontology construction.

論文審査の結果の要旨

Le Thi Ngoc Tho studies the research topic named “Study on language processing method for supporting understanding multiple legal documents”. This research is challenging and difficult in the

sense that we do not have enough training data for the legal domain. She deals with this difficulty by considering two major components as follows: (1) the first component deals with the key-phrase extraction for the legal domain. (2) The second task shows a method for constructing a conceptual structure based on extracted indexes. (1) In the first task, she develops a simple and efficient method that does not need training data. The reported results on the benchmark corpora showed that this method is better than the state of the art results. The second task shows a very good way of using similarity measures for constructing conceptual structures. The study from law documents is new and the result is promising. Le Thi Ngoc Tho successfully defended her thesis with the agreement of the committee members. The committee concludes that the results contribute to Legal Engineering are valuable and significant. Beside, the candidate also shows her ability in working as an independent researcher for dealing with legal documents. She published two international journal articles, in which one paper is accepted. She also published four international conference papers. In conclusion, this is an excellent dissertation and we approve awarding a doctoral degree to Ms. Le Thi Ngo Tho.