

Title	参照解析と法令質問応答への適用
Author(s)	Tran, Thi Oanh
Citation	
Issue Date	2014-03
Type	Thesis or Dissertation
Text version	ETD
URL	<a href="http://hdl.handle.net/10119/12109">http://hdl.handle.net/10119/12109</a>
Rights	
Description	Supervisor: 島津 明, 情報科学研究科, 博士

氏名	TRAN THI OANH		
学位の種類	博士(情報科学)		
学位記番号	博情第 294 号		
学位授与年月日	平成 26 年 3 月 24 日		
論文題目	Reference Resolution and Its Application to Legal Question Answering (参照解析と法令質問応答への適用)		
論文審査委員	主査	島津 明	北陸先端科学技術大学院大学 教授
		東条 敏	同 教授
		Ho Tu Bao	同 教授
		白井 清昭	同 准教授
		徳永 健伸	東京工業大学 教授

## 論文の内容の要旨

Natural languages are highly related by references within them. These references bring precious information: the sentences of a discourse could not be interpreted without knowing who or what entity is being talked about. Resolving resolution, therefore, is a very important task in natural language processing research. Of all reference phenomena, the coreference is the most popular phenomenon, and is attracting much research in reference resolution. In this dissertation, we will concentrate on this challenging task – coreference resolution in general texts. Moreover, we will also focus on resolving references in a specific type of texts, i.e. legal texts. The information on reference resolution not only helps people in understanding texts, but also supports other tasks such as question answering, text summarization, and machine translation. To illustrate one of these benefits, in this thesis, we will also investigate an application of reference resolution to the task of question answering restricted to the legal domain. Most previous research proposed a pairwise approach to solve the task of coreference resolution. The drawback of this approach is that it can allow only one or two antecedent candidates to be considered simultaneously. So, it only determines how good a candidate is relative to the mention, but not how good a candidate is relative to all candidates. Our goal is to investigate another approach, which can address this drawback. While coreference resolution in general texts attracts much attention among researchers, the task in legal texts has received very little attention so far. The main reasons are mostly the complex and long legal structures and sentences, specific terms, and especially the lack of language resources (i.e. annotated corpora) in this specific

domain. Focusing on this interesting legal domain, this dissertation also aims at building a system, which can automatically extract referents for references in real time. This is a new interesting task in the Legal Engineering research. Moreover, the goal of this dissertation also includes building an application of these reference resolvers to a useful question answering system restricted to the legal domain. Particularly, the following three problems are targeted in this research:

- To realize coreference resolution in general texts, we present an empirical study on a listwise, which can address the drawback of the previous approach. This approach exploits a listwise learning-to-rank method, which considers all antecedent candidates simultaneously, not only in the resolution phase but also in the training phase. Experimental results on the corpora of SemEval-2010 shared task 1 show that the proposed system yields a good performance in multiple languages when compared to previous participating systems as well as a baseline pairwise system using the ranking support vector machine as the learning algorithm. In comparison to the best participating system SUCRE, which uses the Decision Tree algorithm with best-first clustering strategy, the proposed system achieves comparative performance.
- For the task of reference resolution in legal texts, different from previous work that only considered the referent at the document targets, this work focuses on resolving references to the sub-document targets. Referents extracted are the smallest fragments of texts in documents, rather than the entire documents that contain the referenced texts. Based on the structures of references in legal texts, we propose a four-step framework to accomplish the task: mention detection, contextual information extraction, antecedent candidate generation, and antecedent determination. We also show how machine learning methods can be exploited in each step. The final system achieves 80.06% in the F1 score for detecting references, 85.61% accuracy for resolving them, and 67.02% in the F1 score on the end-to-end setting task on the Japanese National Pension Law corpus.
- This dissertation also presents a study aimed at exploiting reference information to build a question answering system restricted to the legal domain. Most previous research focuses on answering legal questions whose answers can

be found in one document (*The term 'documents' corresponds to articles, paragraphs, items, or sub-items according to the naming rules used in the legal domain.*) without using reference information. However, there exist many legal questions, which require answers extracted from connections of more than one document. The connections between documents are represented by explicit or implicit references. To the best of our knowledge, this type of questions is not adequately considered in previous works. To cope with them, we propose a novel approach which allows to exploit the reference information between legal documents to find answers to these legal questions. This approach also uses the requisite-effectuation structures of legal sentences and some effective similarity measures based on legal terms to support finding correct answers without training data.

The contribution of this dissertation includes linguistic and computational aspects. Considering the linguistic viewpoint, our research helps in interpreting the sentences of any discourse. In the computational viewpoint, our research proposes effective solutions for linguistic problems using machine learning approaches.

## 論文審査の結果の要旨

本論文は、テキスト中の共参照関係を解析する新方法とその質問応答システムへの応用について述べている。共参照関係とは、二つの単語が共通の指示対象を持つ関係のことを言う。自然言語処理システムがテキストを正しく解釈するためには、共参照関係が正しく認識されなければならない。従来、ルールベースの方法が提案されているが一般テキストに対応できるまでになっていない。近年、一般性、頑健性を求め、機械学習による方法が研究されるようになった。本論文は、そのような研究によるもので、従来の機械学習による方法の見方を見直し、一般テキストや法令テキストを対象に新方法と新応用を提案し実験により有効性を示している。

一般テキストにおける共参照関係の解析については、本論文は、従来の機械学習による方法が共参照解析を分類問題として捉え、一時に 1 つの候補しか考慮しないのは不十分という立場から、着目単語と共参照する候補をすべて同時に考慮して適切な候補を判定する方法を提案し、共参照の研究分野の共通コーパスを用いて実験している。英語、ドイツ語などのコーパスを対象にした実験の結果について、従来法の 4 システムと提案法とを 4 つの尺度により比較している。結果は、従来の 3 システムより優れ、1 システムと同程度であ

ることから、よい方法であると評価できる。

法令テキストについては、本論文は、従来研究が条項番号の同定という簡単な問題を対象としていたのに対し、条文中の被参照表現を特定する問題を対象とし、参照表現の検出、文脈情報の抽出、先行詞候補の生成、先行詞の決定からなる 4 段階の解析法を提案している。国民年金法の実験により、最終段に上述の共参照関係解析法を用いる場合が従来法より精度がよいことを示している。この論文は、法情報学に関する一流国際会議 ICAIL において、最優秀学生論文賞を受賞している。

上記研究の応用として、質問応答システムを取り上げ、応答情報を得るのに共参照関係の同定が必要な質問の応答生成法を提案している。提案法は、共参照関係の単語がある条項の要件部と効果部を関連付け、それらと質問文の要件部と効果部との類似性を調べ、類似度の高い条項を選択し、応答を作るというものである。国民年金法を用いた実験により、従来法では適切な応答を返すことは難しく、提案法が有効であること示している。

以上、本論文は、共参照関係を解析する新方法と有効性を示し、学術的に貢献するところが大きい。よって博士（情報科学）の学位論文として十分価値あるものと認めた。