



# Acquisition of hyponymy relations for agricultural terms from a Japanese statutory corpus

Makoto Nakamura <sup>a,\*</sup>, Tomohiro Ohno <sup>b,c</sup>, Yasuhiro Ogawa <sup>b,c</sup>, Katsuhiko Toyama <sup>b,c</sup>

<sup>a</sup>Japan Legal Information Institute, Graduate School of Law, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

<sup>b</sup>Information Technology Center, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

<sup>c</sup>Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

## ARTICLE INFO

### Article history:

Received 7 January 2014

Received in revised form

14 November 2014

Accepted 14 November 2014

Available online 21 November 2014

### Keywords:

AGROVOC

Japanese statutory corpus

Legal text processing

## ABSTRACT

This paper, which aims to increment the vocabulary of an existing thesaurus using hyponymy relations, focuses on an agricultural thesaurus called AGROVOC. Our main goal is to acquire AGROVOC-qualified candidates from the hyponymy relations of legal texts and tables. We propose a pattern-based approach to hyponymy relation acquisition. Our experimental result showed that 222 and 868 candidates are extracted from statutory sentences with 67.1% precision and tables with 37.0% precision, respectively.

© 2014 China Agricultural University. Production and hosting by Elsevier B.V. All rights reserved.

## 1. Introduction

Legal terms are idiomatic expressions that often describe legal matters in such documents. Since legal terms differ from daily expressions, they are defined by statutes prior to their use. The goal of our study is to construct a legal ontology based on legal terms and the hyponymy relations between them contained in large collections of legal texts consisting of the statutory sentences of acts and regulations. A hyponymy relation in this paper denotes a kind of IS-A relation between terms. In particular, this paper increments the vocabulary of an existing thesaurus using hyponymy relations. We focus on an agricultural thesaurus called AGROVOC [1].

AGROVOC, which is the world's most comprehensive multilingual agricultural vocabulary [2], and its applications were developed by many researchers [3–6]. It contains more than

40,000 concepts in over 20 languages, covering topics on food, nutrition, agriculture, fisheries, forestry, and the environment. AGROVOC is expressed in a Simple Knowledge Organization System (SKOS) and is published as Linked Data [7]. All of the terms or concepts have been added to the thesaurus by the domain experts in different languages. Such laborious human work is very time-consuming and expensive. At times, these partner organizations are unable to frequently update the terms due to unforeseen circumstances [8]. To prevent this problem, we are working toward the practical use of automatic term extraction from legal texts. We assume that the terms in the acts and the regulations are qualified for AGROVOC as long as they are related to the agricultural domain.

Our main idea is to use the hyponymy relations of legal texts to find candidates qualified for AGROVOC. Since the terms registered in AGROVOC are definitively related to it, their hypernyms or hyponyms are also considered registerable candidates. In our approach, if one of the terms in a hyponymy relation has been registered, the other may be a

\* Corresponding author.

E-mail address: [nakamura.makoto@j.nagoya-u.jp](mailto:nakamura.makoto@j.nagoya-u.jp) (M. Nakamura).  
Peer review under the responsibility of China Agricultural University.

<http://dx.doi.org/10.1016/j.inpa.2014.11.001>

2214-3173 © 2014 China Agricultural University. Production and hosting by Elsevier B.V. All rights reserved.

candidate term. New links between a hypernym and a hyponym can be acquired when both of the terms in a relation have been registered without being linked to each other in AGROVOC.

To achieve our goal, we develop two strategies for mining hidden hyponymy relations: one from statutory sentences and the other from tables described in the statutes. For the first one, a legal term, which is defined in the statutes prior to use with boilerplate expressions, often includes a hypernym of the defined term for explanation. In another case, some terms are explained with enumeration of hyponyms. We expect to extract these hyponymy relations from the explanatory sentences. Second, we analyze tables in the statutes, each row or column of which is typically constituted by a heading and its data. We recognize the heading and the data as a hypernym and hyponyms, respectively. Therefore, we propose and examine extraction methods for both of them.

Our paper is organized as follows: We first introduce the background on legal text processing in Section 2. In Section 3, we propose a method of hyponymy acquisition on the basis of the analysis of Japanese legal texts. In Section 4, we examine how the method works in the legal domain and conclude with a short summary and future research in Section 5.

## 2. Background on legal text processing

### 2.1. Basic explanation of Japanese laws

In general, laws are roughly divided into written and unwritten categories. Although unwritten laws include local customs and judicial precedents, we do not deal with these in this paper. Written laws are also called statutes, which are further divided into acts and bylaws. While acts are enacted by the National Diet (Parliament), bylaws consist of orders enacted by the cabinet and ordinances and regulations enacted by various ministries.

In this paper, we focus on statutory texts. A statute consists of a number of articles, each of which may be further subdivided into a number of paragraphs or items. Articles, paragraphs, and items have sequential numbers with different typefaces. A provision denotes an independent article or a paragraph.

### 2.2. Basic explanation of the Japanese legislative system

The rational nature of the legislative system of Japanese law maintains the notation of expressions of statutes.<sup>1</sup> Although the Cabinet and Diet members can submit a bill to the National Diet, most bills are introduced by the Cabinet. In this case, the proper authority for that law basically makes a draft of the bill. Once this is accomplished, the authority negotiates with other authorities. The Cabinet Legislation Bureau then closely examines the draft in terms of inconsistency with other statutes, expressions, formats, and so on. As a result, even the usage of commas and periods is maintained. When

a Diet member submits a bill, it is reviewed by the Legislation Bureau of the House of Representatives or Councilors.

Not every country's legislative system is similar to that of Japan. In the United Kingdom, the legislature's description check is not as strict, as in most cases the bill is drafted outside of the ministry. In the United States, there is no organization or system for the legislature's description check. In Asian countries other than Japan and Korea, often each ministry independently prepares a draft of a bill without coordinating with other ministries. As a result, the notation of bills differs among ministries. Moreover, in some countries bills are often modified during deliberation in the national assembly, while bills mostly pass the National Diet in Japan as drafted.

Since this political process results in inconsistencies in notation, this strict wording style may be an idiosyncratic feature of Japanese statutes. This suggests that simple text processing is sufficient to locate important terms or phrases with conventional expressions.

### 2.3. Previous works on legal text processing

Thus far, researchers of natural language processing have studied legal text processing using surface pattern recognition, while some advanced studies on legal text processing employ machine learning methods [9,10]. Surface pattern rules are typically described in regular expressions, which provide a concise and flexible means to match and extract strings of text. Despite the presence of high-quality dependency parsers for Japanese, we cannot rely on their performance with legal texts. Since legal sentences are designed to avoid ambiguity, they are likely to be long and syntactically complicated, which often leads to parsing failure [11,12]. Therefore, we do not rely on a syntactic parser and minimize the use of natural language processing (NLP) tools.

Kimura et al. [13] acquired knowledge from itemized expressions in legal texts. Their experimental result showed that only a few surface patterns following itemization successfully extracted itemized expressions and represented semantics. Höfler et al. [14] detected legal definitions to support domain-specific style verification in legislative drafts. Both studies show that surface pattern recognition is sufficient for legal text processing because legal documents are often written with boilerplate expressions. For this point of view, the target of this study is different from probabilistic models for learning ontologies that expand existing ontologies taking into account both corpus-extracted evidence and the structure of the generated ontologies [15,16].

It is possible for surface pattern rules to extract hyponymy relations as well as legal terms. For example, the following expression, “y is a (kind of) x,” in which both x and y are noun phrases, implies that x is a hypernym of y as well as “such x as y” [17,18]. This approach is also applicable to Japanese, for example, see Ando et al. [19] who proposed a set of Japanese surface patterns. These studies suggest that legal ontologies can be automatically constructed from legal texts containing boilerplate expressions.

Hyponymy relations can also be extracted from listings or tables. Shinzato et al. [20,21] proposed a method to acquire hyponyms for given hypernyms from HTML documents, assuming that the heading of an itemization in an HTML

<sup>1</sup> This section is based on our discussion with Prof. Matsuura of the Graduate School of Law, Nagoya University.

Download English Version:

<https://daneshyari.com/en/article/4493319>

Download Persian Version:

<https://daneshyari.com/article/4493319>

[Daneshyari.com](https://daneshyari.com)