

## Daily News on Japanese Legislation toward Global Sharing of Japanese Legal Information

KAWACHI Gen<sup>1</sup>, NAKAMURA Makoto<sup>2</sup>, OGAWA Yasuhiro<sup>1,3</sup>,  
OHNO Tomohiro<sup>1,3</sup>, TOYAMA Katsuhiko<sup>1,3</sup>

<sup>1</sup>*Graduate School of Information Science, Nagoya University*

<sup>2</sup>*Japan Legal Information Institute, Graduate School of Law, Nagoya University*

<sup>3</sup>*Information Technology Center, Nagoya University*

**Abstract.** Our purpose in this paper is to develop a system for the international sharing of Japanese legal information. We plan to promptly provide the Outlines of all the newly promulgated Japanese statutes in English, using machine translation. Structured documents in XML format made it possible to achieve various functions concerning machine translation, text search, and display style. We developed the Document Type Definition for the Outlines based on our investigation of their structure. We also developed a rule-based annotator for the Outlines, which adequately marked up 87.2% on the Outlines of Japanese acts.

**Keywords:** Machine translation, XML annotation, Online access to law, Legal informatics

### 1. Introduction

Recently, social demand for disseminating up-to-date legal information on Japan has arisen for various reasons related to social and economic globalization. This demand has activated the international sharing of trends in Japanese legislation. Currently, Japan does not provide an English version of statutes at the time of promulgation. One of the reasons is that it is quite difficult in terms of time to disseminate Japanese statutes in English on the same day as the Japanese.

The Japanese Law Translation Database<sup>1</sup> (hereinafter, JLT) (Toyama et al., 2011) was released under the Japanese government's leadership in 2009. JLT provides legal information in English such as statutory laws including acts and bylaws related to them like cabinet orders, ordinances of ministries, etc. It can be regarded as a static disseminator, providing English translation of Japanese statutes that are currently effective and much in demand. However, we cannot obtain the latest legal information in Japan from JLT due to spending too much time for release of translations. It is fresh information that makes the world go around. Therefore, we need a dynamic disseminator for simply but expeditiously reporting everyday activities of Japanese legislation. The

---

<sup>1</sup> <http://www.japaneselawtranslation.go.jp/>

combination of these static and dynamic systems would aid in deep and wide understanding of Japanese legislation.

Our purpose in this paper is to develop a dynamic system for delivering daily news on Japanese legislation, which hereinafter we call “LegiViewJP.” LegiViewJP embodies the following three concepts: “transparency,” “internationality,” and “promptness.” The system will provide all the statutes newly promulgated for the transparency of Japanese legislation. Contents will be translated into English toward global sharing of Japanese legal information. The machine translation technique (Koehn, 2010) makes it possible for the system to promptly release an English translation.

We focus on the Outlines of statutes to establish the dynamic disseminator. When an act or cabinet order is promulgated, we can read its Outline in the official gazette at the same time. The Outlines are shorter and simpler than the original statutory documents because some sentences in the original are brought together or omitted. These characteristics aid in the translation and understanding of legislation.

LegiViewJP consists of five modules: (1) Downloader of the Outlines, (2) XML annotator, (3) Japanese-English Machine translator, (4) Posteditor for translation, and (5) Uploader to the LegiViewJP server. LegiViewJP makes use of XML markup and machine translation for prompt dissemination of Japanese statutes in English. In particular, the XML markup plays central roles in both the simple translation process and the user-friendly interface. In this paper, we deal with developing the Document Type Definition (DTD) and the automatic XML annotation system for the Outlines.

This paper is organized as follows. In Section 2, we present the concepts of LegiViewJP, comparing with existing systems. In Section 3, we introduce the Outlines of Japanese statutes. Then, we propose LegiViewJP in Section 4. We briefly explain the DTD for the system and the automatic annotation system in Section 5 and show markup experiments with it in Section 6. Finally, Section 7 summarizes this paper.

## 2. Related Works

We explain related works from the viewpoints of internationality, transparency, promptness and fineness in granularity on sharing legal information.

Since 1998, EUR-Lex has provided information on EU laws, which now includes regulations, treaties, case laws, and so on (Düro, 2009). These documents are currently translated into 24 official languages in

EU. In addition, most statutes become available for reading a few days after promulgation. EUR-Lex achieves internationality, transparency, and promptness at the same time. Although the contents of EUR-Lex are different from the ones we need, this multilingualism motivates our system.

Japanese legal information in English is partly disseminated at the Japanese Law Translation Database. JLT emphasizes fineness in granularity, that is, it translates full sentences of statutes. However, the problem of the small number of translated statutes has been pointed out since its release (Sekine et al., 2012). In fact, the number of statutes included in JLT was 394 as of August 2014, which is only 4.9% of currently effective statutes. This means that non-Japanese speakers are unable even to be aware of promulgation in terms of the remaining 95.1% of statutes. Although the concept of “transparency” implies that all the legal information is provided, unfortunately, JLT does not seem to satisfy this. Moreover, JLT does not provide prompt translation. It takes three years on average for a statute to be released on JLT since it appears on the list of the translation plan.

South Korean statutes are also provided in English by the Korea Legislation Research Institute (KLRI)<sup>2</sup>. Similarly to JLT, not all statutes go public in English, and it often takes more than one year for translation. In contrast with Japan and South Korea, Taiwan has succeeded in prompt release of the English and Chinese versions of statutory laws<sup>3</sup>. It takes only a couple of weeks to translate a statute by professional law firms or translation company. Nevertheless, an administrator in each of the authorities decides statutes to be translated, which implies some statutes are not translated. In this sense, this system does not satisfy the condition of “transparency,” either.

The Japanese government provides an electronic edition of the official gazette<sup>4</sup> on the website. The official gazette boasts transparency due to its abundant contents, for example, all the statutory laws including acts and cabinet orders, ministerial ordinances and so on. Promptness of the official gazette is also the best in all media because, by definition, publishing in the official gazette regards statutes as promulgated. Despite these facts, the official gazette does not play a role as an international disseminator of Japanese legislation, because it is provided only in Japanese.

Table I shows a comparison matrix among legal information systems. These examples show the difficulty for East Asian jurisdictions in constructing legal information systems that achieve international-

<sup>2</sup> [http://elaw.klri.re.kr/eng\\_service/main.do](http://elaw.klri.re.kr/eng_service/main.do)

<sup>3</sup> <http://law.moj.gov.tw/eng/>

<sup>4</sup> <http://kanpou.npb.go.jp/>

Table I. Comparison among legal information systems

	transparency	promptness	internationality	fineness
EUR-Lex	✓	✓	✓	✓
JLT			✓	✓
KLRI			✓	✓
Taiwan		✓	✓	✓
Japanese Official Gazette	✓	✓		✓
LegiViewJP	✓	✓	✓	

ity, transparency, and promptness at once. Our system places greater emphasis on these three concepts than on the fineness of information. Thus, we discard the publication of full text in Japanese statutes, but employ the Outlines, which are shorter and simpler than the original statutes. Accordingly, LegiViewJP comes to have these three concepts.

The Global Legal Information Network (GLIN)<sup>5</sup>, which provided nearly 191,000 legal resources from 60 jurisdictions all over the world as of December 2010, set eligibility rules for participation, one of which was to provide a summary of all the enacted laws in English (Medina et al., 2004; Hyde et al., 2010). Although Japan does not currently satisfy this condition, we expect the products of LegiViewJP to complement the English summary.

### 3. Outlines of Japanese Statutes

Toward prompt translation of Japanese statutes, we focus on the Outlines of Japanese statutes. The government has published the Outlines of Japanese statutes since 1973 to serve the general public’s understanding of legislation (Ogawa et al., 2013). We can read them in an official gazette when their original statutes are promulgated.

Figures 1 and 2 show an act and its Outline, respectively. Acts and Outlines are originally published in Japanese and the figures show their English translation. While Article 3 of the original act has six paragraphs, the Outline contains only two sentences, where the first one corresponds to paragraphs (1) and (2) of the original and the second corresponds to paragraphs (3) and (4). Paragraphs (5) and (6) of the original are omitted in the Outline. Since the Outlines are shorter and simpler than their original statutes in this way, translating them into English will be easier.

<sup>5</sup> As of mid-2014 GLIN is no longer available from the Law Library of Congress.

- (Inspection)  
Article 3
- (1) The Commandant of the Japan Coast Guard may order Coast Guard Officers to undertake the following measures, if there are reasonable grounds to believe that a vessel in the internal waters of Japan is carrying Specified North Korea Cargo:
    - (i) to stop the vessel for inspection;
    - (ii) to enter the vessel, inspect the cargo, documents, and other items, or question the crew and other relevant persons of the vessel;
    - (iii) to collect samples, limited to the minimum amount necessary for inspection;
    - (iv) to unload or reload the cargo or to instruct the captain of the vessel, etc. to unload or reload the cargo, within the limit necessary for inspection.
  - (2) The Commandant of the Japan Coast Guard may order Coast Guard Officers to undertake the following measures, if there are reasonable grounds to believe that a vessel in the territorial waters of Japan or on the high seas (including the Exclusive Economic Zone set forth in the United Nations Convention on the Law of the Sea; the same shall apply hereinafter) is carrying Specified North Korea Cargo:
    - (i) to request the captain of the vessel, etc. to stop the vessel for inspection;
    - (ii) to take the measures listed in item (ii) or item (iii) of the preceding paragraph, with the consent of the captain of the vessel, etc. ;
    - (iii) to unload or reload the cargo with the consent of the captain of the vessel, etc., or to request the captain of the vessel, etc. to unload or reload the cargo, within the limit necessary for inspection.
  - (3) The Director-General of Customs may order Coast Guard Officers to undertake the following measures, if there are reasonable grounds to believe that a vessel or an aircraft (excluding military aircraft and aircraft owned or operated by a foreign government for non-commercial purposes; the same shall apply hereinafter) is carrying Specified North Korea Cargo at a port or an airport in Japan:
    - (i) to enter the vessel or aircraft and inspect cargo, documents, and other items or question the crew and other relevant persons;
    - (ii) to collect samples, limited to the minimum amount necessary for inspection;
    - (iii) to unload or reload the cargo, or to order the captain of the vessel, etc., the captain of the aircraft, or the person acting on behalf of the captain (referred to in paragraph (2) of the following Article as the “captain of the aircraft, etc.”) to unload or reload the cargo, within the limit necessary for inspection.
  - (4) The Director-General of Customs may order Coast Guard Officers to undertake an inspection of cargo, documents, and other items, to question the owner, possessor, manager, or other relevant persons, or to collect samples, limited to the minimum amount necessary for inspection, if there are reasonable grounds to believe that the cargo placed in a bonded area (meaning an area as prescribed in Article 29 of the Customs Act, including those designated by the Director-General of Customs pursuant to the provisions of Article 30, paragraph (1), item (ii) of the Customs Act; the same shall apply in paragraph (2) of the following Article).
  - (5) Pursuant to Ordinance of the Ministry of Land, Infrastructure, Transport and Tourism and Ordinance of the Ministry of Finance, Japan Coast Guard and Customs officials shall wear uniforms and carry identification cards when conducting the inspections prescribed in each of the preceding paragraphs.
  - (6) The authority pursuant to the provisions of paragraphs (1) through paragraph (4) may not be construed as having been granted for a criminal investigation.

Figure 1. Original of Article 3 of Act on Special Measures concerning Cargo Inspections etc. Conducted by the Government Taking into Consideration United Nations Security Council Resolution 1874, etc. (Act No. 43 of 2010)

<p>3 Inspection</p> <p>(i) If there are reasonable grounds to believe that a Ship in Japan's inland waters or territorial waters or the high seas is loaded with North Korea Specified Cargo, the Commandant of the Japan Coast Guard may, with the consent of the Captain, etc. of the ship if it is in Japan's territorial waters or the high seas, have a coast guard officer(s) take such measures as an inspection of the relevant cargo, documents and/or other articles. (The point of Article 3, Paragraphs 1 and 2)</p> <p>(ii) If there are reasonable grounds to believe that a Ship, etc. located at a port in Japan is loaded with North Korea Specified Cargo or that any North Korea Specified Cargo is included in the cargo placed in a bonded area, the Director-General of the relevant Custom-House may have a customs official(s) take such measures as an inspection of the relevant cargo, documents and/or other articles. (The point of Article 3, Paragraphs 3 and 4)</p>
---

Figure 2. Outline of Article 3 of Act on Special Measures concerning Cargo Inspections etc. Conducted by the Government Taking into Consideration United Nations Security Council Resolution 1874, etc. (Act No. 43 of 2010)

*The act*

<p>Part of the Act on Limitation on Number of Court Officials shall be revised as follows: The term "1,717" in the table of Article 1, shall be revised to "1,782," and "1,020" to "1,000."</p>
---

*The Outline*

<p>The Act for Partial Revision of the Act on Limitation on Number of Court Officials (Act No. 11) (Ministry of Justice)</p> <p>1 The number of judges shall be increased by 65. (The point of Article 1)</p> <p>2 The number of assistant judges shall be reduced by 20. (The point of Article 1)</p>
--

Figure 3. Act 1 of The Act for Partial Revision of the Act on Limitation on Number of Court Officials (Act No. 11 of 2010) and its Outline

In addition, the Outlines are sometimes not a simple summary. Particularly, in the case of an amendment statute, it consists of amendment statutory sentences that indicate how to revise the original statutory sentences, as shown in the upper part of Figure 3. We cannot understand the intention of the sentence without referring to the statute to be amended. In contrast, the Outline of the sentences is in the lower part of Figure 3. We can understand the substantive contents of the amendment only from the Outline. Therefore, the Outlines are useful for understanding the Japanese statutes and their changes.

While they are written only in Japanese, these good points of the Outlines are suitable for both translating and understanding.

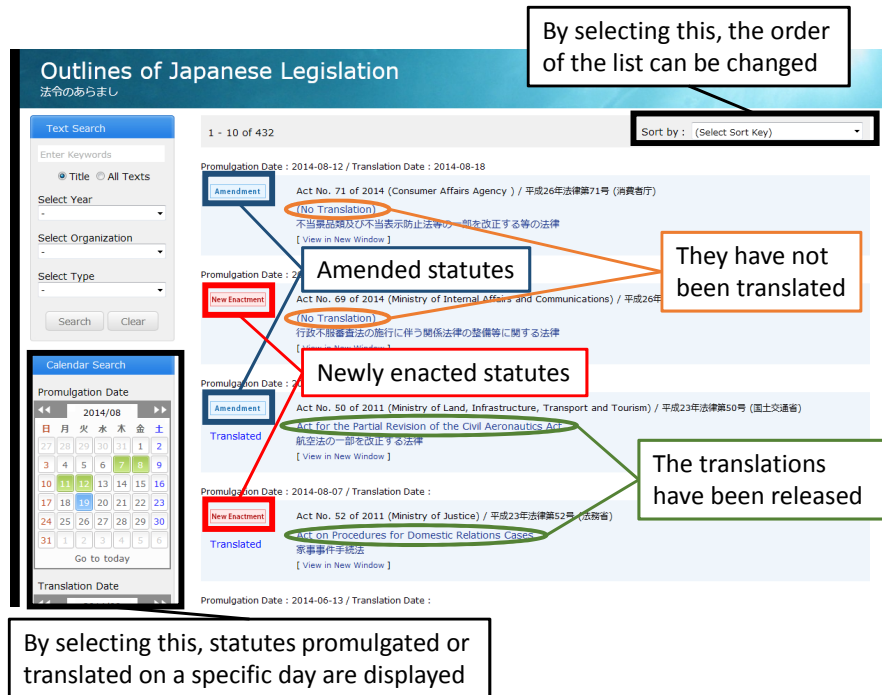


Figure 4. List of newly promulgated statutes on LegiViewJP.

#### 4. LegiViewJP, a Daily News on Japanese Legislation

We propose LegiViewJP, which translates the Outlines of all Japanese statutes and releases them on the website as soon as the statutes are promulgated.

##### 4.1. USER INTERFACE

Figure 4 illustrates the top page of the LegiViewJP website, which shows the list of newly promulgated statutes in the order of promulgation or translation date. Users can perceive what kinds of statutes are promulgated, whether the statutes are newly enacted or amended, and whether the Outlines are already translated or not on the list. In this sense, LegiViewJP is considered as a passive-type system, while JLT is an active type that requires some operations like inputting a query. Users can also find statutes promulgated or translated on a specific day by selecting a calendar. In addition, RSS feeds are available so that users can perceive all the update of the website by subscribing them.

LegiViewJP has some functions described below (Figure 5). Their design basically follows that for JLT, which also provides various func-

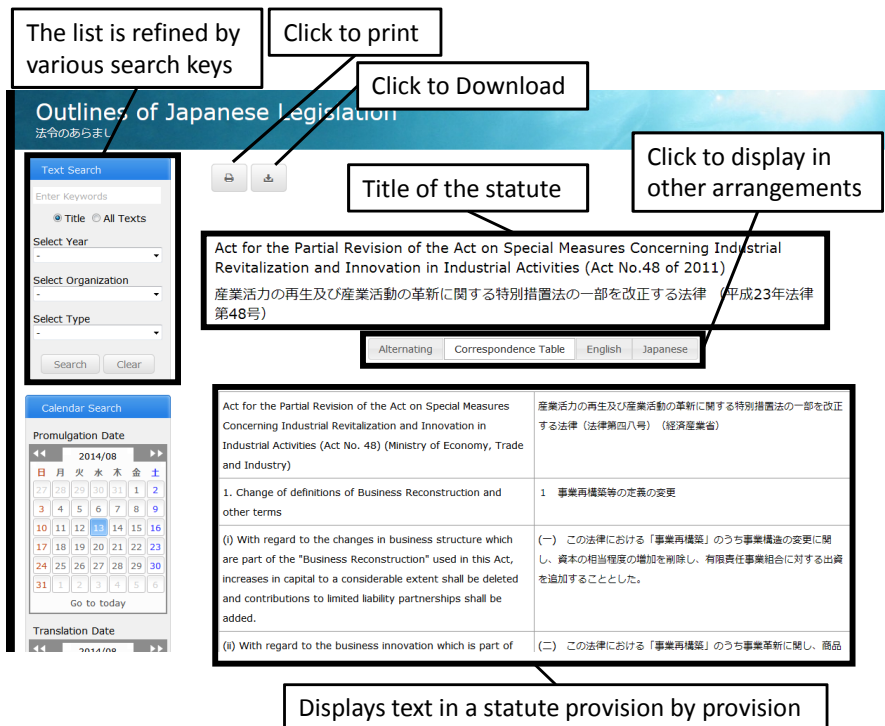


Figure 5. Display of an Outline.

tions in terms of usability and accessibility. Keyword search is available from a statute text, a statute title, promulgation year, a field of the statute, and a government organization which has jurisdiction over the statute. The contents of the Outlines chosen from the list are divided provision by provision and displayed in various arrangements that users can easily switch between: alternating Japanese-English texts sentence by sentence, a Japanese-English corresponding sentence table, Japanese text only, and English text only. Users can select and print any part of the retrieved document and can also download its data as plain text, PDF, XML, or Word document.

#### 4.2. SYSTEM ARCHITECTURE

LegiViewJP consists of the following five modules shown in Figure 6.

##### Module 1. Downloader

This module downloads the contents of the official gazette available on its website. Then, it extracts plain text of the Outlines from the gazette.



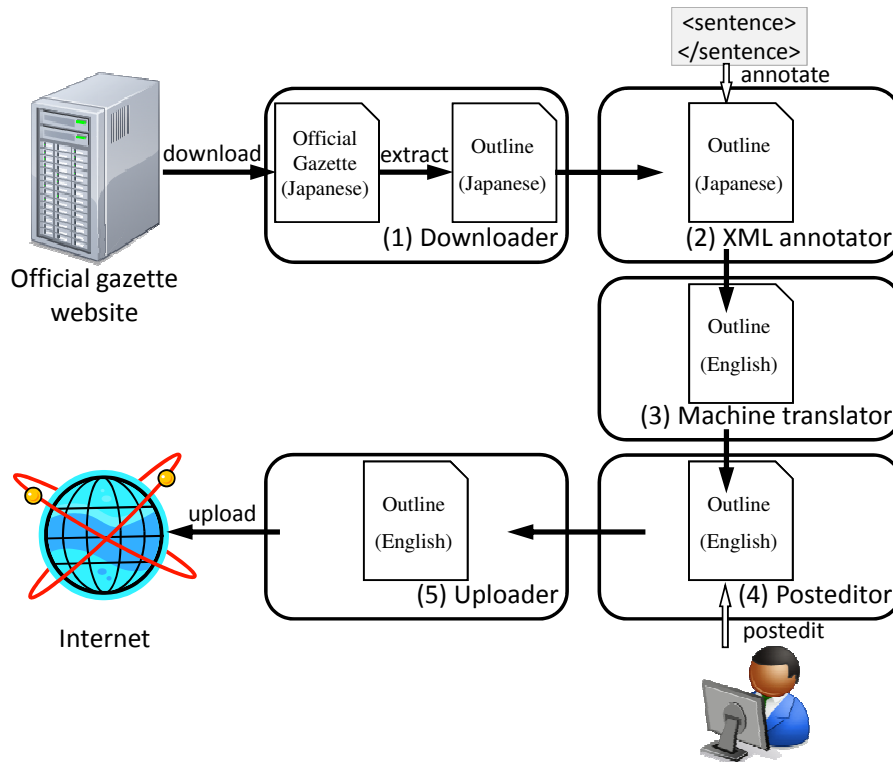


Figure 6. Workflow of LegiViewJP

#### Module 2. XML annotator

This module analyzes the logical structure of the Outlines, and then marks them up in XML format.

#### Module 3. Machine translator

This module translates sentences of the Outlines from Japanese to English.

#### Module 4. Posteditor for translation

The output of the machine translator is manually edited in this process.

#### Module 5. Uploader

The translated Outlines are finally uploaded to the server of LegiViewJP.

Module 3 employs the previous study on the statistical machine translation tuned for the Outlines (Inagi et al., 2013). Figure 7 shows an example of a sentence translated by this translator, where Reference denotes a correct translation given by professional translators.

Japanese	: この法律は、公布の日から <b>施行</b> することとした。
Inagi	: this Act shall come into <b>effect</b> as from the day of promulgation has.
Google translator*	: This Act, decided to come into <b>force</b> from the date of promulgation.
Reference	: This Act shall come into <b>effect</b> as of the day of promulgation.

\* <http://translate.google.com/toolkit/>

Figure 7. Comparison of translations (Inagi et al., 2013)

Automated Japanese-English translation tends to fail due to difficult word order, while fragments of words are correctly translated despite difficult legal texts. In fact, Inagi’s system exceeds Google Translator in BLEU<sup>6</sup> score, that is, 32.33 by the former against 26.34 by the latter. Although the postediting module is required, the cost of postediting would be much less than that of the whole manual translation.

## 5. Structuralizing the Outlines

The Outlines extracted from the Official Gazette are formatted in plain text, from which it is not easy to extract necessary information for LegiViewJP. Structuralizing the Outlines in XML format makes it possible to provide various functions as mentioned in Section 4. In this section we discuss the design of Document Type Definition (DTD) for the Outlines, the document structure of the Outlines, and the development of an automatic annotation system.

### 5.1. POLICY OF DTD FOR LEGIVIEWJP

All the contents of the Outlines should easily be reused for various purposes, such as displaying the Outlines on the website, converting document format, and so on, in accordance with the policy of Open Data. Therefore, we designed the DTD for LegiViewJP that XML documents retain all the contents of the Outlines in plain text as elements. An XML document conforming to this can be restored to the original plain text by simply removing XML tags except layout characters such as white spaces and indentations. This policy of the DTD follows that

<sup>6</sup> BLEU is an automatic evaluation metric that calculates the degree of agreement with the reference translations by dividing the maximum number of  $n$ -gram shared between the result of machine translation and the reference by the total number of candidate  $n$ -grams (Papineni et al., 2012).

of JLT since LegiViewJP is similar to JLT in that both of them provide legal documents in Japan.

Since the machine translator deals with a single sentence for a translation, we need to extract sentences one by one from documents. For the smooth translation process, we designed our scheme that facilitates extraction of sentences. The DTD for LegiViewJP guarantees extraction of a single sentence with the element named “Sentence.” The same holds true for the DTD of JLT as a result, although it was designed to facilitate automatic consolidation of amendment in Japanese legislation. In this point of view, it is different from METALex (Boer et al., 2002) and Akoma Ntoso (Vitali et al., 2007), the famous schemes designed for statutes. Although both of them also allow to mark up a single sentence, its element is necessarily not guaranteed for a sentence.

Our scheme attaches great importance to mark up of the document structure. We do not mark up semantic-related tags unlike the XML scheme used in Eunomos (Boella et al., 2012), a legal document management system that focuses on document classification.

## 5.2. DOCUMENT STRUCTURE FOR OUTLINES OF JAPANESE STATUTES

Since the DTD for JLT is not designed for the Outlines of Japanese statutes, we need to design a new DTD for them. For this purpose, we investigated the document structure of the Outlines of acts, statutes enacted by the National Diet. As a result, we defined 44 elements in our DTD, in which there are six required elements and the others are optional. Table II shows the names of elements, some of which are defined based on the following remarkable characteristics in terms of the document structures in the Outlines.

### 1. Hierarchical structure

Both original statutes and their Outlines have logical hierarchical structure. We named classes in the hierarchy of the Outline “Chapter,” “Paragraph,” “Item,” “Subitem1,” “Subitem2,” “Subitem3,” and “Subitem4” in descending order as listed in the *Provisions* in Table II. The hierarchical structure is expressed with sequential numbers or letters attached to the beginning of each provision. In the original statutes, the typeface of labels corresponds to a certain level in the hierarchy because the format of the whole document is strictly specified. On the other hand, the format specification for the Outlines is ambiguous, which results in fluctuation in expressing the hierarchical structures. For example, in Figure 8, the Arabic numeral “2” and the Roman numerals “(i)” and “(ii)” represent upper and lower class in the hierarchy, respectively. Some Outlines

Table II. Elements defined in DTD for Outlines (\*: required elements).

<i>Entire document</i>			
*StatuteOutline			
<i>Title of the document</i>			
*OutlineTitle	TitleMark	*StatuteName	*StatuteNum
*Jurisdiction			
<i>Body of the document</i>			
*OutlineBody			
Introduction			
<i>Provisions</i>			
Chapter	ChapterNum	ChapterTitle	ChapterSentence
Paragraph	ParagraphNum	ParagraphTitle	ParagraphSentence
Item	ItemNum	ItemTitle	ItemSentence
Subitem1	Subitem1Num	Subitem1Title	Subitem1Sentence
Subitem2	Subitem2Num	Subitem2Title	Subitem2Sentence
Subitem3	Subitem3Num	Subitem3Title	Subitem3Sentence
Subitem4	Subitem4Num	Subitem4Title	Subitem4Sentence
<i>Sentences</i>			
Sentence			
RelevantProvisions			
<i>Tables</i>			
Table			
Row		Column	
<i>Notes</i>			
Note			
NoteMark		NoteSentence	

use these labels in reverse order, and others use different typefaces of labels. We have to develop an annotation system that deals with the fluctuation. We discuss this matter in the next section.

## 2. Elements unique to Outlines

A sentence in the Outlines often includes a citation of provisions in the original statute relevant to the sentence, which is parenthesized and located at the end. Examples can be seen with the expressions “(The point of Article 3, Paragraphs 1 and 2)” and “(The point of Article 3, Paragraphs 3 and 4)” in Figure 2. We defined the element “RelevantProvisions,” for such citations.

## 3. Elements seen in the provisions

Each provision in the Outlines is constructed by some of the fol-

2 Securing of stable supply (The point of Article 2)
(i) As a foundation for securing a stable supply of energy, measures shall be taken to diversify energy sources, increase energy self-sufficiency, and achieve security in the energy sector.
(ii) Measures shall be taken regarding the supply of energy that is difficult to substitute for or store, so that the reliability and stability of its supply can be guaranteed.

Figure 8. Paragraph 2 of the Outline of Basic Act on Energy Policy (Act No.71 of 2002)

lowing elements. We defined these elements in the DTD for Legi-ViewJP.

- “ChapterNum,” “ParagraphNum,” ..., or “Subitem4Num” as sequential numbers or letters that express the hierarchical structure.
- “ChapterTitle,” “ParagraphTitle,” ..., or “Subitem4Title” as titles of the provisions.
- “Sentence” as sentences.
- “Table” as tables used for comparison of the prior and amended article and so forth.
- “Note” as notes related to the previous sentence.

Conforming to the DTD, the paragraph in Figure 2 is marked up as in Figure 9.

### 5.3. AUTOMATIC ANNOTATION SYSTEM FOR OUTLINES OF JAPANESE STATUTES

We implemented an automatic annotation system for the Outlines. This system converts documents in plain text of the Outlines as input to that of XML format as output.

The overall structure is analyzed by checking numbers or letters attached to the beginning of each provision. Elements such as title, table, and so on can also be obtained by using simple pattern matching since these elements have characteristics in terms of appearance.

As we mentioned above, the Outlines have several types of numbering rules. Thus, this system has to identify the numbering rule of each document.

We used Racc, a Ruby<sup>7</sup>-implemented LALR parser generator, to analyze document structure, and REXML, the XML parser, to mark up the documents. Input documents are generated by removing HTML

<sup>7</sup> Ruby is an object-oriented language (Flanagan et al., 2008).

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE StatuteOutline SYSTEM "statuteoutline.dtd">
<StatuteOutline Era="Heisei" Lang="en" Num="043" Year="22">
  <OutlineTitle>
    <StatuteName>Act on Special Measures concerning Cargo Inspections
      etc. Conducted by the Government Taking into Consideration
      United Nations Security Council Resolution 1874, etc.</
      StatuteName>
    <StatuteNum Num="43">(Act No. 43 of 2010)</StatuteNum>
    <Jurisdiction>(Ministry of Land, Infrastructure, Transport and
      Tourism)</Jurisdiction>
  </OutlineTitle>
  <OutlineBody>
    <Paragraph Num="1">
      (snip)
    <Paragraph Num="3">
      <ParagraphNum>3</ParagraphNum>
      <ParagraphTitle>Inspection</ParagraphTitle>
      <Item Num="1">
        <ItemNum>(i)</ItemNum>
        <ItemSentence>
          <Sentence>If there are reasonable grounds to believe that a
            Ship in Japan's inland waters or territorial waters or the
            high seas is loaded with North Korea Specified Cargo, the
            Commandant of the Japan Coast Guard may, with the consent
            of the Captain, etc. of the ship if it is in Japan's
            territorial waters or the high seas, have a coast guard
            officer(s) take such measures as an inspection of the
            relevant cargo, documents and/or other articles.</Sentence>
        </ItemSentence>
        <RelevantProvisions>(The point of Article 3, Paragraphs 1 and
          2)</RelevantProvisions>
      </Item>
      <Item Num="2">
        <ItemNum>(ii)</ItemNum>
        <ItemSentence>
          <Sentence>If there are reasonable grounds to believe that a
            Ship, etc. located at a port in Japan is loaded with North
            Korea Specified Cargo or that any North Korea Specified
            Cargo is included in the cargo placed in a bonded area, the
            Director-General of the relevant Custom-House may have a
            customs official(s) take such measures as an inspection of
            the relevant cargo, documents and/or other articles.</
            Sentence>
        </ItemSentence>
        <RelevantProvisions>(The point of Article 3, Paragraphs 3 and
          4)</RelevantProvisions>
      </Item>
    </Paragraph>
  </OutlineBody>
</StatuteOutline>
```

Figure 9. XML markup of provisions shown in Figure 2

tags from the corresponding part of the official gazettes available online. Some typographical errors or unnecessary white spaces, which may interfere with the process of analysis, are removed in advance.

## 6. Experiments

We examined how properly the automatic annotation system marks up the Outlines. In this section, we limited our experiments to the Outlines of acts.

### 6.1. EXPERIMENTAL METHODS

We inspected the outputs from the automatic annotation system from two points of view: whether the DTD validates them and whether they express the structure of input documents adequately. Experiment 1 addresses the former question, and experiments 2 and 3 the latter. Whereas the only complete evaluation is comparison between outputs and manually marked up documents, it is too hard to carry out for all the Outlines. Therefore, we examine a limited number of samples chosen at random for thorough comparison. We also conducted an additional experiment for a visual inspection. We examined the automatic annotation system with an input set of the 4,615 Outlines corresponding to 4,703 acts<sup>8</sup> promulgated in Japan since 1973 to 2012.

#### Experiment 1. Validation

We inspected the validity of all outputs to the DTD we designed with an XML processor, xmllint<sup>9</sup>.

#### Experiment 2. Comparison with handwritten XML documents

We validated twenty randomly selected outputs by comparing them with manually marked up documents.

#### Experiment 3. Visual inspection

We made a visual inspection of 250 randomly selected XML documents to examine whether the system marked up the Outlines properly. We designed an XSLT style sheet, which forms the XML-structured documents of the Outlines, so that we could read their structure at a glance using metadata.

---

<sup>8</sup> There are cases that one Outline rounds up several related acts which are promulgated at the same time.

<sup>9</sup> <http://xmlsoft.org/xmllint.html>

Table III. Result of visual inspection of 250 Outlines

	# of Outlines	percentage
Proper markup	218	87.2
Improper markup	32	12.8
Total	250	100.0

Table IV. Reasons for improper markup

Reasons	# of Outlines
(1) Unusual structures	3
(2) Inconsistent numbering rules	2
(3) Input errors	11
(4) Structural ambiguity	16
Total	32

## 6.2. EXPERIMENTAL RESULTS

The results of the experiments are as follows.

Result 1. XML documents of all 4,615 Outlines are generated properly, and validity of all documents to the DTD is confirmed.

Result 2. We found a structural difference in one of twenty Outlines.

Result 3. Table III shows that 87.2% of the Outlines were judged as properly marked up.

Improper markup was caused by the four reasons below.

(1) Unusual structures

The set of parsing rules did not cover some structures. This pattern of errors can be solved, adding rules corresponding to the structures.

(2) Inconsistent numbering rules

The annotation system failed to analyze the structure due to inconsistent numbering rules. When the system finds such inconsistent rules, it goes to a special routine for processing this sort of input. In this case, the system can send a warning to the administrator, who then modifies that incorrect markup.



(3) Input Errors

Wrong numbers or letters were typed in the source documents, which leads to incorrect analysis of the structure. They may come from OCR errors or typographical errors which appeared when the electronic data of the official gazettes were created. Like (2), the administrator can modify incorrect markups, receiving a warning for failure of the structural analysis. Otherwise, the system just generates incorrect markups without warning, which leads to a low performance of the system. However, the visual inspector in Experiment 3 would be useful for finding improper markups, even when LegiViewJP is in operation.

(4) Structural ambiguity

Some expressions cannot have their structural ambiguity solved without semantic analysis. If the parser finds structural ambiguity, which results in a number of outputs, the system should send a warning to the administrator, who then chooses the correct one out of all the candidates. The current system does not find ambiguity but derives only one structure.

Table IV categorizes the documents whose markups were improper in Experiment 3. The failure in markup in Experiment 2 is categorized into (4).

## 7. Conclusion

We introduced a system called “LegiViewJP” that disseminates daily news in English on Japanese legislation via the Internet. We also analyzed document structures for the Outlines of Japanese statutes and defined their DTD. In addition, we developed an automatic annotation system for the Outlines and confirmed that it contributes to our system.

Structuralizing the Outlines is an important task for our purpose. The annotation system properly structuralized 87.2% of the Outlines. Connecting it with machine translation contributes to the internationality of legal information, prompt daily dissemination, and transparency of legislation even in non-English-speaking nations.

We are in the final step of launching the website of LegiViewJP. For the moment, we concentrate our efforts on its release. In the future work, we will improve the machine translator and the automatic annotation system. The framework of LegiViewJP would also be applicable to other countries. Although some countries may not have outlines of statutes, many researchers have studied automatic summarization for

legal documents (Grover et al., 2004; Farzindar et al., 2004), which can complement the downloader module in our system.

### Acknowledgements

The authors would like to give special thanks to Prof. Amy Shee of National Chung Cheng University in Taiwan for providing information on the Laws & Regulations Database of the Republic of China. This research was partly supported by the Japan Society for the Promotion of Science KAKENHI Grant-in-Aid for Scientific Researches (S) No.23220005 and (A) No.26240050.

### References

- Boella, G., Humphreys, L., Martin, M., Rossi, P., and van der Torre, L. "Eunomos, a Legal Document and Knowledge Management System to Build Legal Services." *AI Approaches to the Complexity of Legal Systems, LNAI 7639, Springer* 2012: 131-146. Print.
- Boer, A., Hoekstra, R., and Winkels, R. "METALex: Legislation in XML." *Legal Knowledge and Information Systems, JURIX 2002: The 15th Annual Conference, IOS Press* 2002: 1-10. Print.
- Düro, M. *Crosswalking EUR-Lex: a Proposal for a Metadata Mapping to Improve Access to EU Documents*. Office for Official Publications of the European Communities, 2009. Print.
- Farzindar, A. and Lapalme, G. "LetSum, an Automatic Legal Text Summarizing System." *Legal Knowledge and Information Systems, JURIX 2004: the 17th Annual Conference, IOS Press* 2004: 11-18. Print.
- Flanagan, D. and Matsumoto, Y. *The Ruby Programming Language*. O'Reilly Media, 2008. Print.
- Grover, C., Hachey, B., and Hughson, I. "The HOLJ Corpus: Supporting Summarisation of Legal Texts." *Proceedings of 5th International Workshop on Linguistically International Corpora* 2004: 47-54. Print.
- Hyde, J., Dove, C., and Macías, F. (eds.) "Year in Review 2010." *GLIN Global Journal*. 2010, Vol. 2, No. 3. p. 7. Web.
- Inagi, D., Ogawa, Y., Nakamura, M., Ohno, T., and Toyama, K. "Statistical Machine Translation for Outlines of Japanese Statutes." *Proceedings of 7th International Workshop on Juris-informatics* 2013: 37-49. Print.
- Koehn, P. *Statistical Machine Translation*. Cambridge University Press, 2010. Print.
- Medina, R., Hyde, J., and Rodriguez-Ferrand, G. "The Global Legal Information Network: Fostering Civic Education and the Rule of Law." *Proceedings of World Library and Information Congress: 70th IFLA General Conference and Council* 2004: 11 pages. Print.
- Ogawa, Y., Inagi, D., Nakamura, M., and Toyama, K. "Translation for Outlines of Japanese Acts." *Law via the Internet 2013* 2013: 12 pages. Print.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. "BLEU: a Method for Automatic Evaluation of Machine Translation." *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics* 2012: 311-318. Print.

- Sekine, Y., Ogawa, Y., Toyama, K., and Matsuura, Y. "Development of Translation Memory Database System for Law Translation." *Law via the Internet 2012* 2012: 21 pages. Print.
- Toyama, K., Saito, D., Sekine, Y., Ogawa, Y., Kakuta, T., Kimura, T., and Matsuura, Y. "Design and Development of Japanese Law Translation Database System." *Law via the Internet 2011* 2011: 12 pages. Print.
- Vitali, F. and Zeni, F. "Towards a Country-independent Data Format: the Akoma Ntoso Experience." *Proceedings of the 5th legislative XML workshop* 2007: 67-86. Print.

