# Similarity Measure among Structures of Local Government Statute Books based on Tree Edit Distance

Yoichi Takenaka\*, Takeshi Wakao°

#### Abstract

A similarity measure between statute books of local governments that can help reveal suggestive similarities is proposed. The regulations of a local government are stored in a statute book, and they are categorized in a layered structure. The layered structure can be described as an ordered tree in computer science, and we define the similarity of statute books as the tree edit distance between two trees. We have calculated the similarities among statute books of the 47 Japanese prefectures and plotted them on a plane using multi-dimensional scaling. The results visually indicate the relationships of similarities among them, and there are several outlier prefectures and clusters. This will help find local governments with similar regulations, which will facilitate the writing or revision of statutes, especially in small local governments, which are typically short staffed.

Keywords: statute book, local government, similarity measure, tree edit distance

#### 1. Introduction

The local assemblies of local governments establish regulations, and the set of all regulations of every local government is stored in statute books. In Japan, statute books have a layered structure that is used to categorize regulations. The layered structure has a root node and taxonomic nodes. Each taxonomic node has a parent node and is reachable from the root by proceeding from parent to child nodes. This layered structure is called a "tree" in computer science. Note that each regulation belongs to one and only one taxonomic node. The number of regulations in each statute books of 47 prefectures in Japan ranges from 1030 to 4126, and the structure of the statute books facilitates effective searching of regulations.

In this study, we propose a similarity measure for statue books to reveal suggestive similarities among local governments. To comprehend the relationships among statute books visually, they are plotted on a two-dimensional surface using multi-dimensional scaling.

As laws and regulations are written in natural language, they have been the focus of natural language processing and legal informatics research. Several studies have proposed methods to process such content

<sup>\*</sup> Graduate School of Information Science and Technology, Osaka University, Osaka, JAPAN

<sup>°</sup> Faculty of Law, Dokkyo University, Saitana, JAPAN

using computers to support legal work, education, and research. Relative to research of a single regulation, Kakuta proposed a method that identifies the revision process for a target regulation (Kakuta, 2010). His research focused on reducing the amount of legal work. Takenaka et al. proposed a method to generate article correspondence tables for the comparison of local government statutes (Takenaka, 2012). Their method was intended for use in legal education. Several studies have attempted to manipulate the complete set of regulations of one or more local governments. For example, Harada et al. proposed a method that uses clusters of regulations for a given local government (Harada, 2009), and Kakuta proposed a method that considers many local governments (Kakuta, 2012). Given a local government regulation as a query, Takenaka et al. proposed a method to search similar regulations from all other local governments. Their work identified local governments that tend to enact similar regulations earliest and latest. However, to the best of our knowledge, no study has focused on the structure of statute books.

# 2. Structure of Statute Books

A statute book has a layered structure, and all regulations of a local government belong to one and only one taxonomic node of the statute book structure. Essentially, there are five layers in the statute book structure in Japan: Part, Chapter, Section, Subsection, and Subsubsection. For example, Tokyo's statute book has 17 Parts, such as Part I: Comprehensive Rules, Part II: Personnel Affairs, ..., and Part XVII: Extinguishing Fire. Part I has 10 Chapters, and Part I Chapter 1: Comprehensive Rules has four sections. Part I Chapter 1 Section 1: General Rules has no child and the regulation to determine the position of Tokyo Metropolitan Government and 21 regulations belong to this section. Fig. 1 shows a part of the structure of the Tokyo statute book. Note that we omit consecutive numbers of taxonomic nodes in the figure.

Local governments enact regulations independently; thus, the structures of their statutes books differ. TABLE III shows the names of Parts, which are the uppermost layer of the structures, from four prefectures. As can be seen, the sets of names, the order of the names, and the number of Parts differ among prefectures. For example, Tokyo and Osaka have 14 Parts, and Nagano and Fukui have nine Parts. Although Part I of these documents have the same name, i.e., Comprehensive Rules, there is no common name among the other Parts. These differences reflect the identities of different local governments.

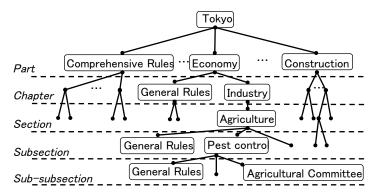


Figure 1. Tokyo Statute Book Structure

The lower layers of the structure also differ by prefecture. TABLE IV shows the positions of *Outdoor Advertisement Goods regulation*, which 46 of 47 Japanese prefectures have enacted<sup>1</sup>. TABLE IV indicates that the names of *Parts* and *Chapters* demonstrate low variation, where the words "Construction," "Planning," and "Public Works" dominate. Although these structures differ among prefectures, TABLE IV indicates they also have some similarities. The purpose of this study is to define a similarity measure to reveal similarities among local government regulations in Japan.

### 3. Similarity Measure

The structure of a statute book can be described as a tree in computer science. The names of items in a statute book comprise nodes with labels of correspondence items. The layered structure of a statute book comprises the edges between items. As a statute book has no explicit root, we set the root of the tree as a node that has no name label. In the statute book, items on the same layer have consecutive numbers, i.e., there is order among siblings. Therefore, we can represent the structure as an ordered tree.

Let ordered trees  $T_A$  and  $T_B$  represent the structures of the statute books of local governments A and B, respectively. Here, we define the similarity between A and B as the tree edit distance between  $T_A$  and  $T_B$ .

The tree edit distance  $TED(T_A, T_B)$  is defined as the minimum number of edit operations required to transform  $T_A$  to  $T_B$ , or the minimum cost of the sum of weights of the edit operations. Note that there are three edit operations.

<sup>&</sup>lt;sup>1</sup> The exception is Ishikawa prefecture.

#### Deletion

Delete a node from  $T_A$  and connect its children to its parent.

#### Insertion

Insert a node as a child of  $T_A$  and move a subsequence of one's consecutive children to its children.

#### Substitution

Rename the node label.

The first algorithm to calculate edit distance was proposed by Tai in 1979 (Tai, 1979). This algorithm requires  $O(n^6)$  time, where n is the number of nodes. Subsequently, Zhang proposed an  $O(n^4)$  time algorithm that uses dynamic programming (Zhang, 1989). Demaine et al. (Demaine, 2009) proposed an  $O(n^3)$  time algorithm, which is presently the fastest edit distance algorithm.

# 4. Experiments

We used statute books from the 47 Japanese prefectures and calculated the tree edit distances among them. We then visualized the relationships among the statute books using multi-dimensional scaling (MDS). MDS was used to plot prefectures on a plane from the distance matrix. Note that we used two distance measures.

(a) [Node Label Match]

Deletion and Insertion: 1

Substitution:

0 if the node labels match,

1 otherwise.

(b) [Edit Distance of Node Labels]

Deletion and Insertion: Number of letters in the label

Substitution:

0 if the node labels match,

number of letters in the label otherwise.

In our implementation, we used the Python zss library to calculate tree edit distance. This library uses Zhang-Shasha's algorithm (Zhang, 1989). The visualization was executed using the R language with the classical metric multidimensional scaling function (cmdscale).

The distributions of the two distance measures among the 47 prefectures are shown in Figures 2 and 3, and the statistics are shown in TABLE I. The table gives the maximum and minimum distances, averages, and standard deviations for two distance measures. Figures 4 and

(a) Match or Non-match Distance Measure (b) Edit Distance Mimimum Distance 96 (AK,NA) 378 (NA,FI) Maximum Distance 379 (HK,HR) 1794 (HK,HR) 239 1007 Average Standard Deviation 49 247

Table I. Statistics of Tree edit distances among the statue book structures of 47 prefectures in Japan

5 shows a plot of the prefectures by MDS. The two-letter abbreviations and font colors for the prefectures used in TABLE I and Figures 4 and 5 are shown in Fig. 6.

Note that the distributions of two types of tree edit distances shown in Figures 2 and 3 are similar. The Pearson's correlation coefficient between them is 0.964 (p-value of 2.2E-16). The maximum values were recorded for the same prefectures, i.e., Hiroshima and Hokkaido, which are outliers in Figs. 4 and 5.

Tokyo (TY) and Kanagawa (KN) are also outliers in both figures. This indicates that both prefectures have dissimilar structures from the others. Note that Aichi (AI), Toyama (TM), Fukui (FI), Shiga (SI), Nara (NR), Tokushima (TK), Okayama (OK), and Fukuoka (FO) form a cluster in both figures. These prefectures are in the Kinki region and adjacent areas (with the exception of Fukuoka); thus, the clustering implies geographical locality.

As of January 1, 2014, there were 1742 cities, towns, and villages in Japan. Among them, 27 cities had populations greater than 500,000; however, there are many small local governments. The populations of 489 of these cities, towns, and villages are less than 10,000, 232 are less than 5,000, and 24 are less than 1,000. All these local governments have their own local assemblies and enact their own statutes. As small local governments have limited number of staffs, they rarely enact purely original statutes. In many cases, they tend to follow other local governments and make minor changes to established statutes and regulations to enact their own. Thus, a system such as the proposed method that can help identify similar statutes and regulations would be beneficial to small local governments.

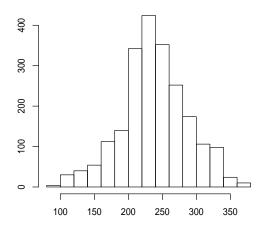
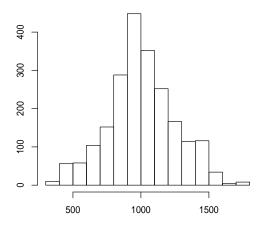


Figure 2. Distributions of Tree Edit Distance among statute book structures of 47 prefectures in Janpan [Node Label Match]



 $\it Figure~3.$  Distributions of Tree Edit Distance among statute book structures of 47 prefectures in Janpan [Edit Distance]

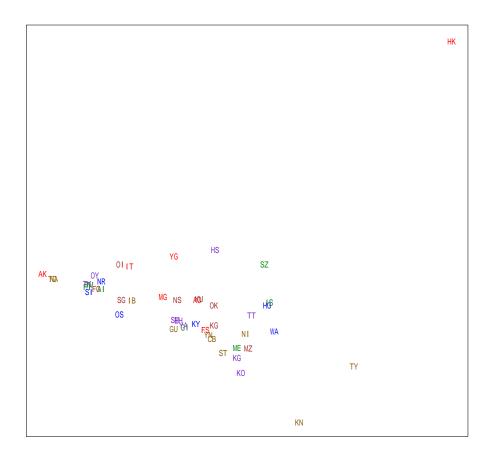


Figure 4. Similarity of the statue book structures of 47 prefectures in Japan by MDS [Node Label Match]

# 5. Conclusion

We have proposed a measurement to quantify similarity between the structures of local government statute books. We have plotted the 47 prefectures in Japan on a plane according to similarity. Two types of edit operation weights were used, and we found that their measurements were similar. The figures in which prefectures are plotted show the relationships of the similarity. Note that some clusters were observed. This research can shed light on the similarity of local government statute books. Our experimental results indicate that the

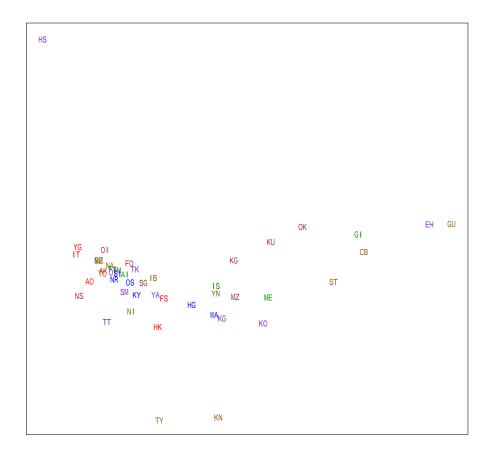


Figure 5. Similarity of the statue book structures of 47 prefectures in Japan by MDS [Edit Distance]

proposed method can be used to find local governments with similar regulations. Such information is valuable when writing or revising statutes. Therefore, we believe that the proposed method will be useful when employed by small municipal organizations, such as cities, towns, and villages, because such organizations are typically short staffed.

In addition, the information generated by the proposed method can improve the accuracy of methods proposed by previous studies, such as finding similar regulations among different local governments (Takenaka, 2015) or generating clusters of similar regulations (Harada, 2009; Kakuta, 2012). Such methods use article titles of regulations and the body

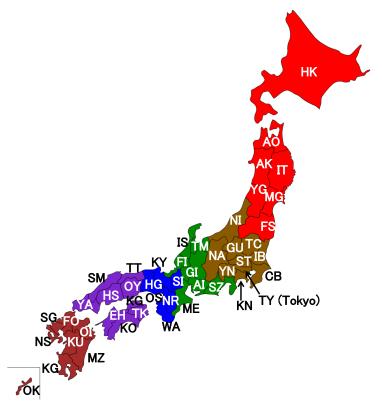


Figure 6. Prefectures in Japan. Two-letter abbreviations and colors are used in Figures 4,5 and TABLE I. The names of prefectures are shown in Table II

Table II. Two-Letter Abbreviations for 47 Prefectures in Japan.

Hokkaido	HK	Aomori	AO	Akita	AK	Iwate	IT
Yamagata	YG	Miyagi	MG	Fukushima	FS	Ibaragi	$_{\mathrm{IB}}$
Tochigi	TC	Gunma	$\operatorname{GU}$	Saitama	ST	Chiba	CB
Tokyo	TY	Kanagawa	KN	Yamagata	YM	Nagano	NA
Niigata	NI	Toyama	TM	Ishikawa	$_{\rm IS}$	Fukui	$_{ m FI}$
Shizuoka	SZ	Aichi	AI	Gifu	$_{ m GI}$	Mie	ME
Shiga	SI	Kyoto	KY	Osaka	os	Nara	NR
Wakayama	WA	Hyogo	$_{ m HG}$	Tottori	TO	Shimane	SM
Okayama	OY	Hiroshima	$_{\mathrm{HS}}$	Yamaguchi	TA	Kagawa	KG
Ehime	EH	Tokushima	TK	Kochi	IO	Fukuoka	FO
Ooita	OI	Kumamoto	KU	Saga	$\operatorname{SG}$	Nagasaki	NS
Miyazaki	MZ	Kagoshima	KG	Okinaga	OK	-	-

text to process computation. However, these methods do not use their relative positions in a statute book. As shown in TABLE IV, similar regulations tend to be located in similar layers. The proposed method can output correspondence tables of nodes using the tree edit distance calculation. Therefore, we believe that the proposed method can contribute to studies that manipulate local government regulations.

Our method requires the structures of statutes books from many local governments as it's input data. It is necessary to scrape up them from their official gazettes or web sites. To compare the two information sources, the web sites are cheaper to get the information. When the effectiveness of the legal analysis methods to the labor savings becomes recognized widely, it will motivate local governments to open their legal information public via web site. We convince that our method plays a part of the motivation, so called "free access to law movement."

We regard the structure of a statute book as an ordered tree in this study because the items have consecutive numbers. However, manipulating the structure as an unordered tree may prove beneficial because, as shown in TABLE III, the order of occurrence differs among prefectures. Although the calculation of edit distance between unordered trees is NP-hard(Zhang, 1992), we plan to include this concept in future work.

# Acknowledgements

This work was partially supported by KAKENHI 21500253.

# References

- Erik D. Demaine, Shay Mozes, Benjamin Rossman, and Oren Weimann (2009), An Optimal Decomposition Algorithm for Tree Edit Distance, ACM Trans. Algorithms, 6(1):2:1-2:19.
- Takashi, Harada, Jyunichi Aoki, and Yurika Mashima (2009), Automatic classification of Local Government Statutes based on Clustering methods, In the proceeding of the 9th Study Meeting, The Information Network Law Association, pp. 65–68.
- Tokuyasu Kakuta (2010), A Method for Supporting Legislation with Focusing on Analogy to Software Engineering (3), Journal of Law and Politics, vol. 237, chap. 2, pp. 191–252, School of Law, Nagoya University.
- Tokuyasu Kakuta (2012), An analysis of regulations of local governments using a supercomputer and the application to a regulation database, Journal of Law and Politics, vol. 246, pp. 69-91, School of Law, Nagoya University.
- Tai, Kuo-Chung (1979) The Tree-to-Tree Correction Problem, J. ACM, Vol. 26(3),pp. 422–433.

- Yoichi Takenaka and Takeshi Wakao, Automatic Generation of Article Correspondence Tables for the Comparison of Local Government Statutes, Natural Language Processing, vol. 19(3), pp. 193–212.
- Yoichi Takenaka and Takeshi Wakao, Automatic Extraction of Common Local Government Laws among Prefectures and its Applications, Transaction on Mathematical Modeling and its Application, vol. 8(1), pp. 80–86.
- Kaizhong Zhang, Simple fast algorithms for the editing distance between trees and related problems, SIAM Journal on Computing, vol. 18(6),pp. 1245–1262.
- Kaizhong Zhang, Rick Statman, and Dennis Shasha, On the Editing Distance Between Unordered Labeled Trees, Information Processing Letters, vol 42(3), pp. 133–139.

Table III. Names of Parts in the statute books of four prefectures

o Fukui	S Comprehensive Rules	Health and Welfare	uth, Living En- Residents livelihood	Commerce, Industry and Work	d Construction Agriculture, Forestry and Fishery	ure Civil Engineering	artment Public Sector	Education	Police	ı	ı	ı	ı	
Nagano	Comprehensive Rules	Financial Affairs	Commonwealth, Health, Living Environment	Economy	Civil Engineering and Construction	Education and Cluture	Police and Fire Department	Public Sector	Accounting	1	ı	1	1	
Osaka	Comprehensive Rules	Staffs	Financial Affairs	Environmental Conservation	Civil Administration	Sanitation	Commerce and Industry	Agriculture and Forestry	Work	Civil Engineering	Architecture	Public Sector	Education and Culture	D.1.22
Tokyo	Comprehensive Rules	Personnel Affairs	Financial Affairs	Welfare	Work	Sanitation	Economy	Urban Planning	Environmental Conservation	Housing	Construction	Harbors	Transportation	Weter and comes Courses

Table IV. Positions of the regulation of Outdoor Advertisement Goods (1/2)

Prefecture	Part	Chapter	Section		
Hokkaido	Construction	Urban Planning	Landscape, Aesthetic and Scenery		
Aomori	Land Maintenance	Urban Planning	Outdoor Advertisement Goods		
Akita	Construction	Urban Planning	-		
Iwate	Public Administration	Land Maintenance	-		
Yamagata	Public Works	Outdoor Advertisement Goods	-		
Miyagi	Public Works	Urban Planning Urban Planning	-		
Fukushima	Public Works	Urban Planning	Outdoor Advertisement Goods		
Ibaragi	Construction	Advertisement Goods	-		
Tochigi	Public Works	Urban Planning	-		
Gunma	Public Works	Advertisement Goods	-		
Saitama	Housing and Urban Area	Outdoor Advertise- ment Goods	-		
Chiba	Urban Area	Advertisement Goods	-		
Tokyo	Urban Planning	General Rules	-		
Kanagawa	Urban Area	Outdoor Advertise- ment Goods	-		
Yamagata	Public Works	Planning	Outdoor Advertisement Goods		
Nagano	Public Works and Construction	Construction	-		
Niigata	Public Works	Urban Planning	Outdoor Advertisement Goods		
Toyama	Public Works	Outdoor Advertisement Goods	-		
Fukui	Public Works and Construction	Construction	-		
Shizuoka	Urban Area and Housing	Urban Area Mainte- nance	Outdoor Advertisement Goods		
Aichi	Public Works	Landscape, Outdoor Advertisement Goods	-		
Gifu	Public Works, Construction	Advertisement Goods	-		
Mie	Public Works	Outdoor Advertisement Goods	-		

Table V. Positions of the regulation of Outdoor Advertisement Goods (2/2)

Prefecture	Part	Chapter	Section		
Shiga	Public Works, Construction and Sight-seeing	Planning and Sight- seeing	-		
Kyoto	Public Works, Con- struction	Urban Planning	Outdoor Advertisement Goods		
Osaka	Public Works	Outdoor Advertise- ment Goods	-		
Nara	Public Works	Construction and Advertisement	-		
Wakayama	Public Works	Outdoor Advertise- ment Goods	-		
Hyogo	Land Maintenance	Outdoor Advertisement Goods	-		
Tottori	Public Works	Urban Planning	Outdoor Advertisement Goods		
Shimane	Public Works, Con- struction	Advertisement Goods	-		
Okayama	Public Works	Urban Planning	-		
Hiroshima	Public Works and Construction	Planning	-		
Yamaguchi	Public Works	Urban Planning	Outdoor Advertisement Goods		
Kagawa	Public Works	Outdoor Advertisement Goods	-		
Ehime	Public Works	Advertisement Goods	-		
Tokushima	Public Works	Planning	-		
Kochi	Public Works	Outdoor Advertisement Goods	-		
Fukuoka	Public Works	Planning	Advertisement Goods		
Ooita	Public Works	Urban Planning	-		
Kumamoto	Construction	Landscape, Advertisement Goods	-		
Saga	Public Works	Urban Planning	-		
Nagasaki	Public Works	Urban Planning	-		
Miyagi	Public Works, Construction	Outdoor Advertisement Goods	-		
Kagoshima	Public Works	Advertisement Goods	-		
Okinawa	Construction	Urban Planning	-		