Original Article

An evaluation method for travel journals as tourism data resources:
A case study of "Noto yuki"

Ran Kamiyama (Department of Civil and Environmental Engineering, Toyo University, kamiyama@toyo.jp)

Abstract

This research evaluates a travel journal regarding as data source potential for discovering tourism resources. Text data mining analysis is applied for discovering common and distinctive features from the travel journal, and locational analysis is implemented to specify the areas where the common and distinctive features appeared. As a result, the distribution of extracted landscape terms is differently appeared in each district by cross tabulation analysis. The distribution suggests that each district has distinctively different characteristics. Moreover, the result indicates that cluster analysis can reveal the relationship among extracted landscape terms and help to understand the characteristics in districts.

Keywords

travel journal, landscape, text data mining, Noto, Kaneko Yuhi

1. Introduction

Generally, tourist resources and attractions are discovered from scenery, nature, culture. Subsequently, the awareness of community resources and the utilization of these resources are an essential process for developing tourism. However, finding attractive scenery, nature, and culture is sometimes difficult for people who live in the community, because these resources are attached to their life, and often too ordinary for their everyday life. In that case, tourism resources are often discovered from the standpoint of tourists and visitors. Alternatively, tourism resources are discovered from a wide variety of media such as a guide book, newspaper, magazine, map, painting, or poem. These media are also useful for finding historical and cultural tourism resources. A travel journal is one of the media and concerned a valuable resource for discovering tourism resources because of valuable information based on the actual observation and experience by visitors and including unknown new factors.

One of the oldest Japanese travel journals is Tosa Nikki by Kino Tsurayuki which is written about a journey from Kyoto to Tosa province in 935. After that, the journal written by intellectuals became popular particularly in the early modern times, because the improvement of major roads made individual trips more available. Consequently, many travel journals were published. These travel journals are written in classic style Japanese or written in classic Chinese styles. To understand these traditional formatting texts requires a special knowledge and reading skills to understand in detail. As a result, the analysis method of travel journals is highly depended on the reader's knowledge and skills. Similarly, the evaluation method of discovered key words or sentences and understanding process of the relationships among discovered words are relied upon a reader's judgement. As a result, classic travel journals are not entirely qualified as a tourism data resources. Therefore, this research is to consider statistical and quantitative methods of utilizing travel journeys in order to find tourism resources.

1.1 Research scope and study area

The Noto province is located in the middle of Japan, and several travel journals were published during the Edo period. "Noto Yuki", a travel journal written by Kaneko Yuhi (Kakuson) during March 21 through May 22 in 1816, is one of the notable travel journals. The author was a famous confucian as well as a painter, and served the feudal lord of Kaga during the Edo period (1603-1868). He travelled from Kanazawa castle in Kaga to the east side of Noto province for two months, and recorded his travel with time series (Figure 1). The travel journal consists of classical Chinese styled texts, fourteen poems, and twenty seven sketched drawings. He was mainly concerned about various sceneries of the Noto province comparing with other journals. In addition to the content, this travel journal is feasible to identify the visited places and routes. Thus, "Noto

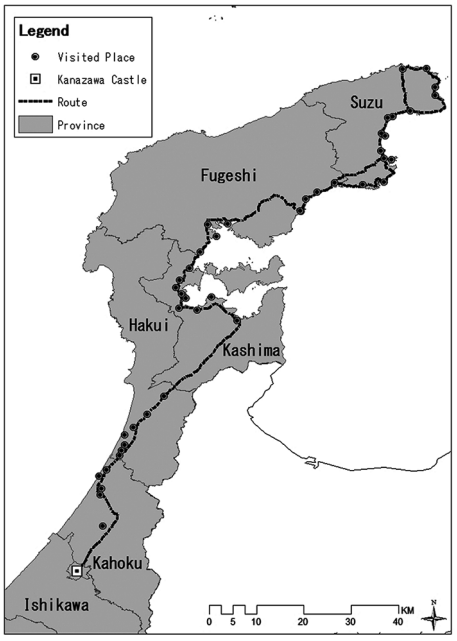

Figure 1: Visited places and travel route

yuki" is considered to be desirable for exploring tourism resources in a certain place and condition.

2. Research method

This research consists of text data mining analysis and locational analysis. Text data mining is applied for discovering common and distinctive features from the text. In addition to the text data mining analysis, locational analysis enables to specify the areas where the common and distinctive features appeared.

2.1 Text data mining analysis

(1) Digitalizing data

The full text of the documents of "Noto Yuki" were transferred into the digital format [(1)] to operate text data mining software. The original journal of "Noto yuki" (Kaneko, 1816) is written in classical Chinese style. The revised edition of the Noto yuki (Kaneko, 1993) which is written in Japanese grammar format is used as a data analysis in this study, because generally text data mining analysis is designed to analyse natural language text.

(2) Extraction of frequently appeared Landscape Terms

Text data mining software called, KH Coder is used to discover words which are expressed certain scenery and the relevant words in the text. The analysed words are listed by Part of Speech (POS) classification. The morphological analyser by Mecab was selected to extract words in this study. The frequently used common nouns and names of a place are selected to narrow down the list. The extracted frequently used words from the above processes are called "Landscape Terms" in this study.

2.2 Locational analysis

In this analysis, previously extracted Landscape Terms were used as reference words. The actual geographic location is specified and then, determined Landscape Terms appearance in the text. Lastly, the distinctive features and the characteristics of the districts in the Noto province are identified by cross tabulation analysis and cluster analysis.

(1) Identification of the visited places

First, the visited places and the route of the journey are identified from the travel journal. The name of a place such as an old town name is identified through a geographical dictionary called "Ishikawa-ken no Chimei" (Heibonsya Chiho Shiryo Senta, 1991), and the geographical place is confirmed according to a revised map of "Shusei Nijyuman bun no ichi zu (1/200000 map)" [(2)] (Figure2). The original map was published in 1888-1889 and the most reliable surveyed map at that time.

(2) Geographic categorization

To specify the geographic location of the Landscape Terms, the full text of the travel journal are divided by five sections according to the geographical districts, which are Kahoku, Hakui, Kashima, Fugeshi, Suzu (Figure 3). This category is based on the former boundary of the Noto province on the previously introduced map.

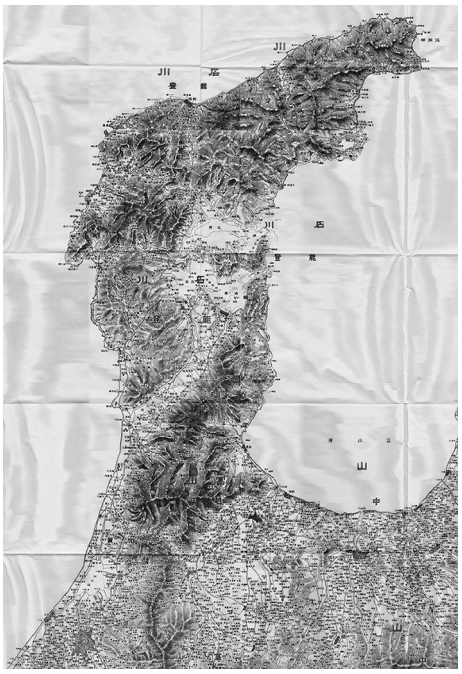

Figure 2: Revised map of "Shusei Nijyuman bun no ichi zu
Source: Heibonsya, 1991.

Figure 3: Former rural district of Noto province

(3) Cross tabulation of Landscape Terms and geographic locations

To analyze the relationships between the extracted Landscape Terms and geographic location, the frequency of Landscape Terms in each section were calculated by the cross tabulation tool of KH Coder.

3. Results

The total number of words in the text file was 15,859 words, and the total number of KH Coder recognized as analysis

words was 8,322 out of 15,859 words. The number of types of those words was 3,536 words and recognized as analysis words was 3,158 words.

3.1 Extraction of frequently appeared Landscape Terms

Landscape Terms were selected from the word list by POS (Part of speech) classification. "Term Frequency" for the number of occurrences of each analysis word in the entire data is shown in Figure 4. As a result, when the number of occurrence is once, the frequency is about 62 %. About 16 % of words appeared twice in the text. Three times of occurrences is about 7%, and four times is about 4 %. The average of the occurrences is about 3 %. In this study, more than 4 times frequently appeared common nouns including verb formed nouns, which consist of more than two Chinese characters, in the entire text are selected to be Landscape Terms (Table 1).

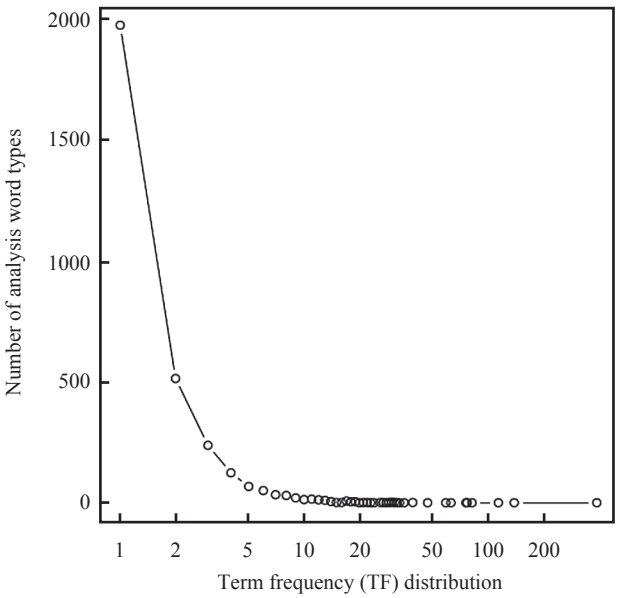

Figure 4: Term Frequency (TF) distribution

As a consequence, a total 33 words were extracted to be Landscape Terms. Additionally, the 33 words are classified based on the similarity and context (Table 1). On the whole, selected Landscape Terms are roughly distinguished by the words which represent; "Scenery" "Mountain", "Sea", "Name of place", "People", and "Other".

3.2 Identification of visited places and locational classification

The visited places are identified through the travel journal text and referenced map (Figure 2) and listed in Table 2. Moreover, the entire text is divided by five sections of rural districts, which are Kahoku, Hakui, Kashima, Fugeshi, and Suzu. Additionally, outward and homeward trips are organized in different sections. The district of Suzu is subdivided by four sections, because the trip pattern differs from the other districts. The author stayed at Ushizu in Suzu district and traveled around from Ushizu (Table 2). "Shinro-ki" is included as a part of the text, which is a story about a mirage phenomenon explained by one of villagers. This additional story is also included for the analysis.

Table 1: Landscape terms frequency

ID	Frequency	Landscape Terms	Classification
1	8	Mountain and sea	Scenery
2	8	Scenery	
3	6	Vista	
4	5	Scenic view	
5	4	Mountain and water	
6	4	Sequence	
7	8	Seashore	Sea
8	5	Sea surface	
9	4	On the sea	
10	4	Ship	
11	8	Pine forest	Mountain
12	7	Mountains	
13	5	Green mountain	
14	4	Mountain path	
15	6	Ecchu	Name of Place
16	6	Tateyama	
17	20	Master	People
18	8	People	
19	6	Local people	
20	6	Villager	
21	5	Traveller	
22	5	Teacher	
23	4	Load	
24	13	Kanazawa castle	Other
25	5	Conversation	
26	4	Traffic	
27	4	Hot spring	
28	4	Hometown	
29	4	Private house	
30	4	House	
31	4	Recreation	
32	4	Poem	
33	4	Homeward journey	

3.3 Cross tabulation of Landscape Terms and geographic locations

Landscape Terms are categorized into the geographic locations, and the frequency of Landscape Terms in each district is shown in Figure 5. The following are the relationship between Landscape Terms and locations described based on the classification of Landscape Terms (Table 1). The terms which delineated human characters in the text and terms which express people are excluded, because the human characters are assumed less relevant to the geographic location.

Landscape Terms relevant to "scenery" frequently appeared in the districts of Kashima, Fugeshi, and Suzu. These areas are closely located to mountains and sea and provide a wide vari-

Table 2: Visited places and rural districts

Trip	Date	Rural District	Place
Outward	3/21	Kahoku	Kanazawa Castle
			Katahata
	3/22		Kasashima Shinmura
			Nanakubo
			Kizu
			Takamatsu St.
		Hakui	Futaya
			Menden
			Kawajiri
			Imahama St.
	3/23		Honjyuku
			Shiho
			Iiyama
	3/24	Kashima	Tokoroguchi
	3/25		
	3/26		Tazurugahama
			Shirahama
			Otu
			Shiozu
			Kasashi
			Kawamura
			Nakajima
			Komaki
	3/27	Fugeshi	Sobuku
			Otogasaki
	3/28		Anamizu
			Ukawa
			Nakai
			Kawajiri
	3/29		Ukawa
	4/1		
	4/2		Yanami
			Hanami
	4/3		Ushizu
	4/4	Suzu 1	Tokinaga
			Yukinobe
			Matsunami
			Koiji
			Ukai
			Uedo
	4/5		Ukai
	4/6		

Trip	Date	Rural District	Place
Outward	4/7	Suzu 2	Koiji
			Matunami
			Nunoura
			Koshizaka
	4/8		Ogi
	4/9		Mawaki
	4/10		
	4/11		
	4/12		Ushima
	4/13		Yamanaka
	4/14		Orito
	4/15		
	4/16		
	4/17		Misaki
			Kawaura
			Noroshi
	4/18		Jike
			Misaki
			Honmura
			Shoin
			Iida
Homeward	4/19	Suzu 3	Ushizu
	4/20		Ushima
	4/24		Uedo
	4/29		Tako-shima
			Eizen-ji
	5/1		Kousyou-ji
	5/2		Uedo
		Suzu 4	Ushizu
	5/10		Mituke-jima
	5/17		Kinpo-ji
	5/18		Mawaki
	5/19	Fugeshi	Ushizu
			Ukawa
			Kawajiri
			Nakai
	5/20		Anamizu
			Ninzaki
		Kashima	Wakuura
			Tokoroguchi
	5/21		Kasajima Shinmura
		Kahoku	
	5/22		
	5/23		Kanazawa Castle

ety of scenic views (Figure 5).

To focus on the Landscape Terms relevant to “sea”, Kashima, Fugeshi, and Suzu areas are strongly connected with the sea. In particular, Kashima and Fugeshi areas are characterized by “seashore”. On the other hand the characteristics of the Suzu area are “sea surface” and “on the sea”. Especially, the Suzu area is famous for a mirage phenomenon, and open view to the sea is one of the most attractive scenic resources.

The Landscape Terms relevant to “mountain” appeared in most of the districts. Among all, Fugeshi area is the most frequently appeared, because there is a distinctive mountain called Futagoyama (Figure 6).

Two Landscape Terms relevant to “Name of place” are extracted. “Ecchu” is the name of the adjacent province of Noto, and “Tateyama” is located in the Ecchu province and one of the famous mountains in Japan. Tateyama is viewed especially from Fugeshi and Suzu area. The scenic beauty is often sketched in the travel journal (Figures 7 and 8).

“Other word” which is not categorized into a particular group contributes to the unique tourism resources. For example, The extracted Landscape Term of “Hot spring” in Fugeshi to Kashima district indicates a famous hot spring in Waku(u)ra (Figure 9). The term of “Recreation” in Fugeshi district is extracted because the author took the recreational boat from the Tsukumo Bay area. Another term of “Poem” in the Suzu district suggests the famous poems by Otomo no Yakamochi

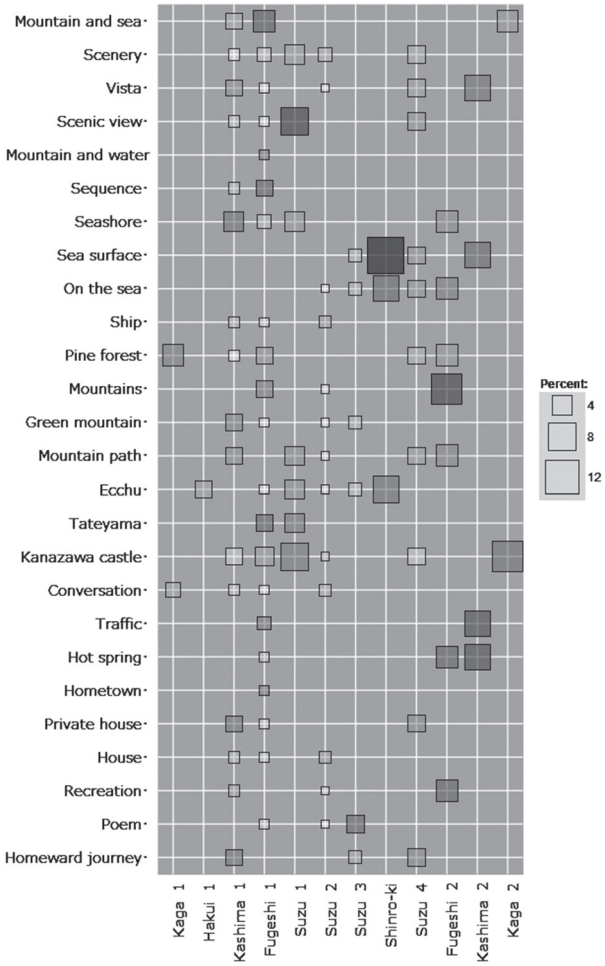

Figure 5: Landscape Terms relevant to "Scenery"

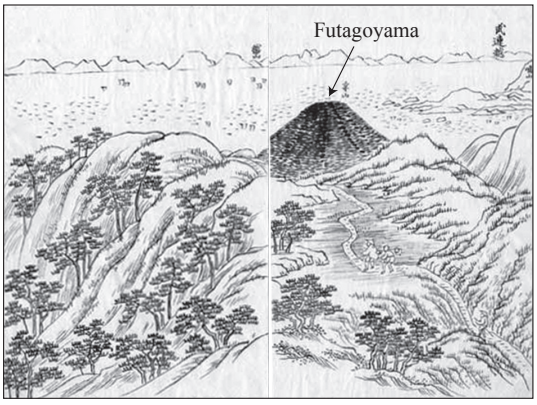

Figure 6: Illustration of Futagoyama from "Noto Yuki"
Source: Kaneko, 1816a. Added text by author.

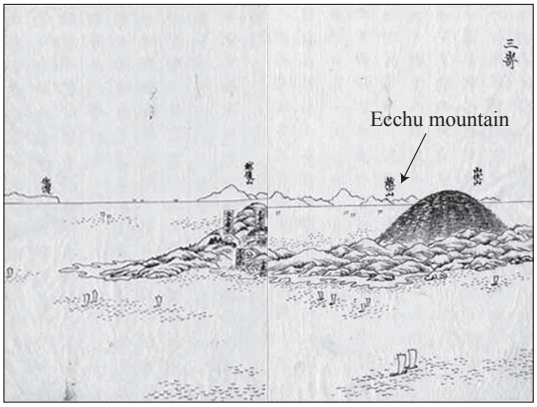

Figure 7: Illustration of Echhu from "Noto Yuki"
Source: Kaneko, 1816c. Added text by author.

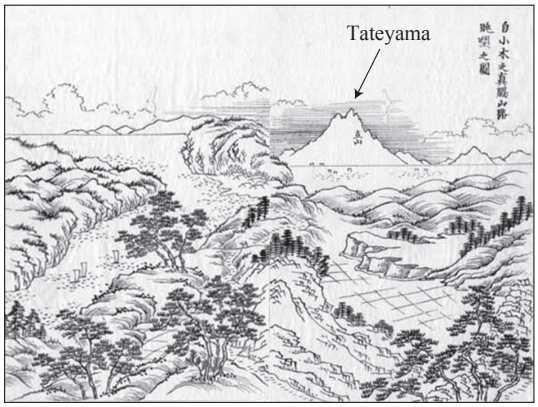

Figure 8: Illustration of Tateyama from "Noto Yuki"
Source: Kaneko, 1816b. Added text by author.

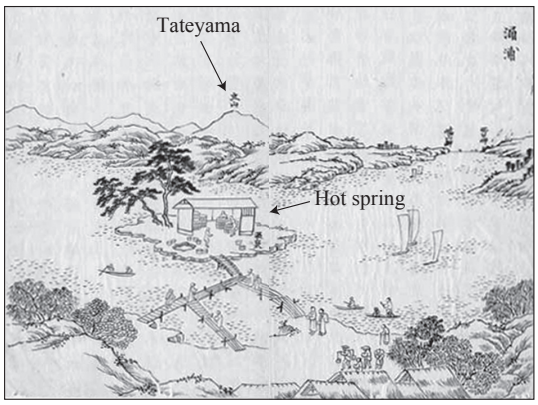

Figure 9: Illustration of Ecchu from "Noto Yuki"
Source: Kaneko, 1816d. Added text by author.

who was a Japanese statesman and famous poet. He served as a provincial governor of Ecchu province in the Nara period. He visited Noto province and created some poems about the aesthetic beauty of Noto. Apparently, poems are also important resources for tourism development in the Noto province.

3.4 Cluster analysis of Landscape Terms and geographic locations

A heat map shows the results of the cluster analysis in addition to the cross tabulation analysis of the geological location and selected Landscape Terms frequency with a gradient color scale (Figure 10). The dendrogram is shown in the left side of Figure 10, which applies the Ward method and uses the Euclidian distance (Higuchi, 2016). This analysis provides some additional contextual information by the relationships among the Landscape Terms. For example, a mirage phenomenon (Shinro-ki) is well explained by the clustered group within the extracted

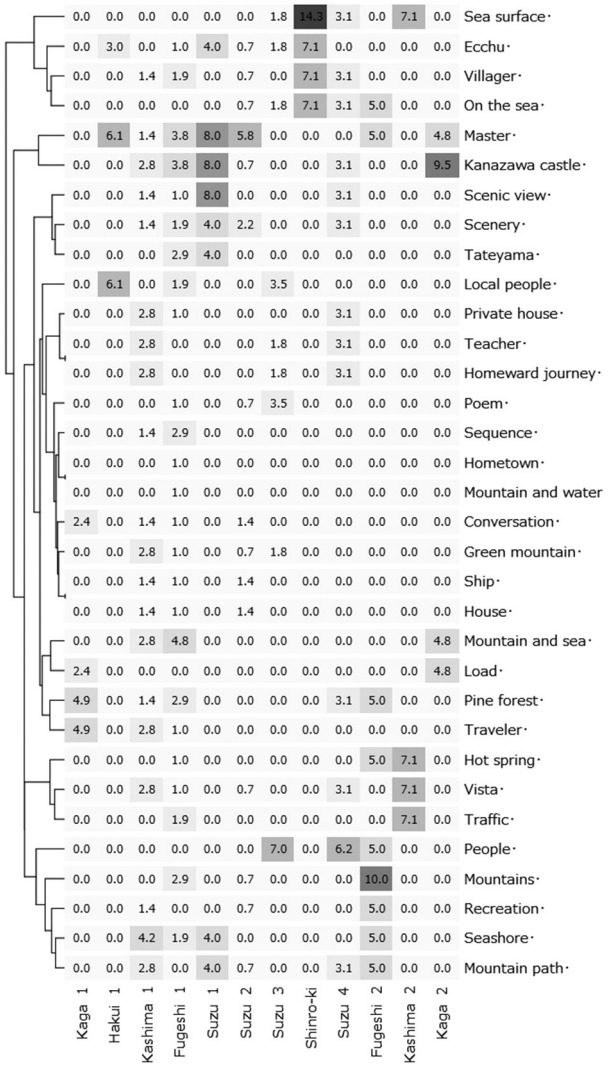

Figure 10: Heat map with cluster analysis

Landscape Terms of "Sea surface", "Ecchu", "Villager", and "On the sea". In the text, a mirage phenomenon was explained by one of villagers. According to him, a mirage can occur on the seas. And also, a mirage phenomenon was seen on the sea surface toward Ecchu province. Another example is Landscape Term "Tateyama" which is highly correlated with the scenic views and scenery. Moreover, it reveals that the vista from a hot spring is a significant scenic resource. Seashore and mountain path is regarded as recreational resources in the Noto province.

As mentioned above, the cluster analysis can reveal the relationship among extracted Landscape Terms and help to understand the context of Landscape Terms.

4. Conclusion

This research focuses on a travel journal and evaluated the potential for discovering and reconsiders the tourism resources. To determine the potential value, the statistical and quantitative research method is implemented. The research method consists of text data mining analysis and locational analysis. Text data mining analysis is applied for discovering common and distinctive features from the travel journal. In Noto province, words relevant to scenic, sea, mountain, a name of a place are extracted as common tourism resources. On the other hand, distinctive features, such as a hot spring and poems are identified through the text data mining analysis. These extracted words are considered potential tourism resources in the Noto province.

In addition to the text data mining analysis, locational analysis enables to specify the areas where the distinctive features and characteristics appeared. In consequence, the distribution of Landscape Terms are differently appeared in each district. This suggests that this method is useful to identify distinctively different characteristics in a certain area. Moreover, the characteristics in each district are mostly explained by cluster analysis. Therefore, these methods help to discover the regional characteristics and to understand the context.

Finally, this research focused on only the frequently appeared words and their locations. However, text data mining analysis is more likely to support different analysis and assist to search an individual word. The whole system is more capable to help with the further analysis of tourism resources sufficiently and effectively.

Notes

(1) OCR–Converts the characters in the photo by Gen Shinozaki is mainly used to read text and convert into digital format for this analysis.

(2) Revised map of Shusei Nijyuman bun no ichi zu (1/200000 map) by Heibonsha Chiho Shiryo Senta. Ishikawa-ken no Chimei, Nihon rekishi chimei taikei 17, Heibonsya, 1991.

References

Heibonsha Chiho Shiryo Senta (1991). *Ishikawa-ken no Chimei*. Nihon rekishi chimei taikei 17. Heibonsya.

Higuchi, K. (2016). KH Coder 3 reference manual, 74. Retrieved Nov 30, 2016 from: http://khc.sourceforge.net/en/manual_en_v2.pdf.

Kaneko, Y. (Kakuson). (1816 copied). Noto Yuki, Retrieved Nov 30, 2016 from: http://www.library.pref.ishikawa.lg.jp/toshokan/dglib/noto/noto001.htm.

Kaneko, Y. (1993). *Noto yuki*, edited by Tsurugi Choritu Hakubutukan. Tsurugi Kyoiku Iinkai.

Kaneko, Y. (Kakuson). (1816a copied). Noto Yuki, Ishikawa prefectural library, 26. Retrieved Nov 30, 2016 from: http://www.library.pref.ishikawa.lg.jp/toshokan/dglib/noto/noto027.htm.

Kaneko, Y. (Kakuson). (1816b copied). Noto Yuki, Ishikawa prefectural library, 42. Retrieved Nov 30, 2016 from: http://www.library.pref.ishikawa.lg.jp/toshokan/dglib/noto/noto043.htm.

Kaneko, Y. (Kakuson). (1816c copied). Noto Yuki, Ishikawa prefectural library, 48. Retrieved Nov 30, 2016 from: http://www.library.pref.ishikawa.lg.jp/toshokan/dglib/noto/noto049.htm.

Kaneko, Y. (Kakuson). (1816d copied). Noto Yuki, Ishikawa prefectural library, 64. Retrieved Nov 30, 2016 from: http://www.library.pref.ishikawa.lg.jp/toshokan/dglib/noto/noto065.htm.

(Received November 30, 2016; accepted December 22, 2016)