Classification network of tourism information

Yin, Chengjiu
九州大学情報基盤研究開発センター

Ito, Eisuke
九州大学情報基盤研究開発センター

Nakatoh, Tetsuya
九州大学情報基盤研究開発センター

Hirokawa, Sachio
九州大学情報基盤研究開発センター

出版情報：Proceedings of iCAST 2011 (3rd International Conference on Awareness Science), pp.171-174, 2012-09-06
バージョン：
権利関係：

Classification network of tourism information

A smart phone-based system for supporting "Petit Trips"

Chengjiu Yin
Research Institute for Information Technology
Kyushu University
Fukuoka, Japan
yin@cc.kyushu-u.ac.jp

Tetsuya Nakatoh
Research Institute for Information Technology
Kyushu University
Fukuoka, Japan
nakatoh@cc.kyushu-u.ac.jp

Ito Eisuke
Research Institute for Information Technology
Kyushu University
Fukuoka, Japan
itoh@cc.kyushu-u.ac.jp

Sachio Hirokawa
Research Institute for Information Technology
Kyushu University
Fukuoka, Japan
hirokawa@cc.kyushu-u.ac.jp

***Abstract*— We are developing a concept dictionary for Petit Trips (Short Time Trips). Based on the concept dictionary, we propose a smart phone-based system to support the Petit Trips. This system provides travelers a Classification Network of tourist spots according to their locations. While travelers select their preferred tourist spots, the system will recommend the suitable tourist spots which have similar features. Through the Classification Network of tourist spots, travelers can go to their preferred tourist spots one after another.**

Keywords- Mobile Environment; Classification Network; Petit Trips; Tourism Concept Dictionary

I. INTRODUCTION

In recent years in Japan, the local tourism information, which includes the location (latitude and longitude) information, is shown on Japanese website. Currently, as smart-phones are widely used and the number of smart-phones in use globally expands, at the same time the latitude and longitude for the Location is beginning to be recorded, while we public the personal views and thoughts on the website. We can get some information of reputable tourist spots by extracting and analyzing this information.

It is possible to develop more accurate and efficient knowledge based on the included text mining, which includes the location information that will available in the future. For example, if the location information of tourist spots is developed and there is a dictionary, which includes the relationship between the words, then we can identify those blog articles which are from a personal blog website, or a news blog website, or a tourism blog website.

Nowadays, a lot of attention has been paid to the amounts of information sent by blog users, which is also included in the tourist information. The value of travel blogs on the tourist information has been evaluated. We paid attention to not only the information provided by travel agents, but also reputation information from public users, such as stories of one's experience.

Also, many of the people who travel for the business, many of them want to go on a Petit Trips (Short Time Trip) by taking advantage of their free time. However, they do not know the tourist information around hotels, and it is not easy to know something like, 1) "Is there anything special and interesting around our hotel?", and 2) "Where can we find anything special around our hotel?"

In this paper, we are proposing a smart phone-based system for supporting Petit Trips. In this system, the words describe tourist spots as well as attributes which are extracted and classified from the website, according to the similar attributes. These words are classified as dynamic Classification Network, which is put on the map. By providing information spots on the map, not only can the system navigate the trip route, but also recommend the similar tourist spots based on travelers' preferences, travelers can visit the various tourist spots one after another. In easier words, this system can recommend travelers most suitable tourist spots according to the users' preferences during their Petit Trips.

II. CLASSIFICATION NETWORK

Classification is one of the most basic methods to analyze data. The results of the analysis would vary according to the focused attributes. Formal concept analysis is an analysis method based on the mathematical notion of formal concept lattice [1, 2], where a concept is defined as a pair of an object set O and an attribute set A that characterize each other. The set inclusion relation forms a complete lattice and is used as a hierarchy of attributes. In our case, we consider the words as attributes, and the documents as objects. Thus, we can obtain hierarchy words from the formal concept lattice to classify documents according to keywords.

Classification network is a simplified form of the concept lattice [3] defined as follows.

A context is a triple (D,W,M) of a set D of documents, a set W of words and a relation $M \subseteq D \times W$ of documents and words. Given a document d in D and a word w in W, $(d,w) \in$

M means that the word occurs in d. Given X ⊆ D an Y ⊆ W, doc(Y) and word(X) denotes the set of documents that contain a word in Y, and the set of words that occur in a document in X, respectively. A concept is a pair (X,Y) such that doc(Y)=X and word(X)=Y. The set of all concepts with respect to (D,W,M) is denoted by CL(D,W). The order relation between concepts (X,Y) and (X',Y') is defined by doc(Y)⊇doc(Y'). This relation forms a complete lattice and is called the formal concept lattice of (D,W,M). The classification network CL(D,W) is a sub-graph of CL(D,W) whose elements are the set of concepts determined by single words.

Figure 1 is the classification network obtained from 19 documents of search results with "spots sea" and with characteristic words in the search result that occur more than 10 times. We can think of "fireworks held in MIYAZAKI prefecture" and someone's diary on the event. On the other hand, the word "station" and "road" would puzzle the reader. The two words combine a single word "road station". This is the service station or service area along the road where local

farmer sells their farm products directly to the guests. These types of markets are becoming popular in many rural areas not only to the tourists but also to people in city.

Figure 2 is obtained by adding this two words "road" and "station", where we find other words "Giants", "heritage", "TAKATUKI" and "saint". The first word "Giants" reminds that the baseball team Giants stay in MIYAZAKI prefecture for their camp season. From the other three words, we would guess a historical heritage "TAKATUKI saint".

III. TOURISM CONCEPT DICTIONARY

The tourism information data are extracted from articles which are from personal blogs, news websites, or tourism websites. These data are classified and stored into a concept dictionary database.

The tourism industry has changed greatly thanks to the application of the information technology. Research trends on tourism information can be classified into the following.

(a) Research that identifies document related to tourism from a large amount of Web documents [4].
(b) Research that extracts important words from Web pages and analyzes the relations between those words [5, 6]
(c) Research to use extracted information for concrete application [7, 8, 9]

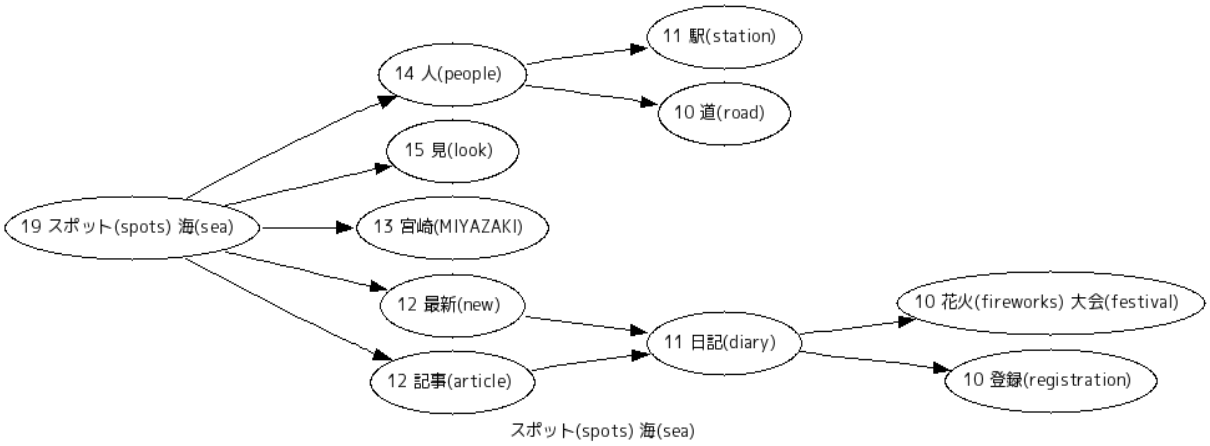

Figure 1. The retrieval result of “spots, sea”.

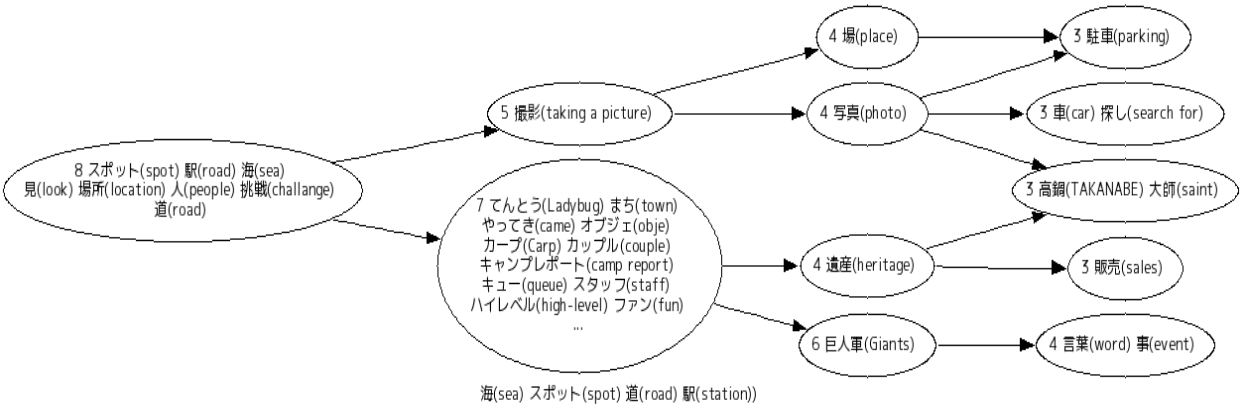

Figure 2. The retrieval result of “spots, sea, Road, Station”. ”

From these previous work, we learn that tourism information, the dictionary and the geographic information is effective and valuable if they are mutually related and used in collaborative manner. However, most of the previous researches of type (a) do not focus on particular target application. In researches of type (b) and (c), it is pointed out that creating the dictionary that can be used adjusting to real data is the key issue. Nonetheless, the present paper is the first attempt to combine these information for Mobile-based petit-tourism system, as long as authors know.

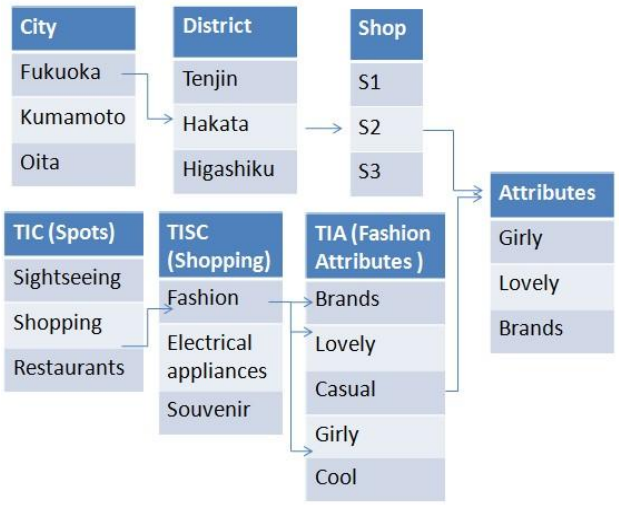

Figure 3. Structure of Concept Dictionary

In this paper, the data are classified into shopping, restaurants, sightseeing, and so on, named Tourist Information Categories (TIC). TIC are classified into subcategories, such as shopping is classified into fashion, souvenir, electrical appliances, and so on, named Tourist Information Sub-Categories (TISC). All these TISC have their own attributes. For example fashion shops have attributes like brands, lovely, causal, girly, and so on (Figure 3), named Tourist Information Attributes (TIA).

We are developing a text mining concept dictionary for analyzing tourist spots reputation, which includes location information. The dictionary here is not like an English-Japanese dictionary, it is composed of two information lists. One is a list of certain words with their related words and synonyms, such as S1 and its related words are Hakata-Fukuoka (Fukuoka->Hakata->S1). The other is a TIC-TISC-TIA list of certain words which are classified into their categories, such as S1 is classified by the category "Shopping->Fashion->Brand, Lovely, Girly". All of them are stored into a database of a concept dictionary.

IV. PETIT TRIP SYSTEM

This system is running on mobile phones, which can access the Internet anywhere at any time. As motioned above, this system provides travelers TIC and TISC of tourist spots. By these two categories, travelers can select their preferred tourist spots. Then the system will recommend a suitable Classification Network of tourist spots which have similar attributes on a map.

As shown in table1, we extract the tourist spots' attributes not only from the concept dictionary database, but also from the spots' website and blogs. Table 1 is an example of fashion shops. From table1 we can get a Classification Network which is shown on Figure 4. For example, S1and S5 have same attributes as well as S1 and S4 which are connected to "girly, pretty".

As shown in figure 5, this system is running on a smart phone. A traveler takes smart phone and selects TIC. Then TIC such as Shopping will be shown. After the traveler selects shopping, the TISC such as fashion will be shown to the traveler. After the traveler selects fashion, the TIA of fashion will be shown. After the traveler selects their preferred attributes (girly, lovely, brands), the system will generate a new Classification Networks according their location (Figure 6).

Figure 7 is an example of sightseeing. When a traveler selects sightseeing through TIC and TISC, the TIA will be shown to them. After s/he selects a preferred attributes, the system will recommend a suitable tour to her/him (Figure 7. A). After s/he clicks the "map" button, a Classification Network of tourist spots will be shown on a map (Figure 7. B). If s/he clicks the red button, the detailed information about the spot will be shown to the traveler (Figure 7. C).

TABLE I. ATTRIBUTES OF FASHION SHOPS

	Attributes in dictionary DB				Spots' website		Blogs evaluation		
	High Brands	Casual	Inner Brands	Cool Sexy	Girly	Men's	Nice	Pretty	Cheap
		●			●			●	
S1	high				girly	..	..	pretty	cheap
S2		casual				men			cheap
S3			inner	cool			nice		
S4		casual			girly		nice	pretty	
S5	high				girly			pretty	cheap
S6				cool		men	nice		cheap
...									
S128									

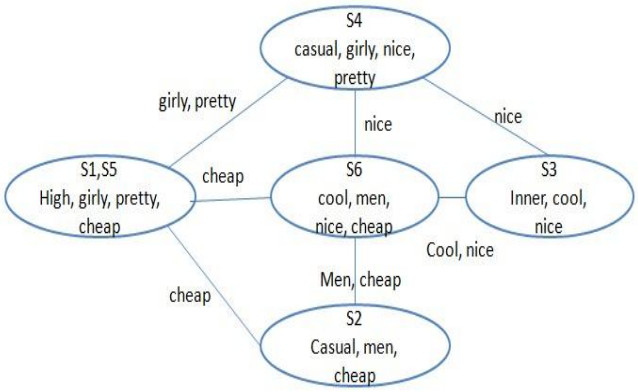

Figure 4. Classification Network

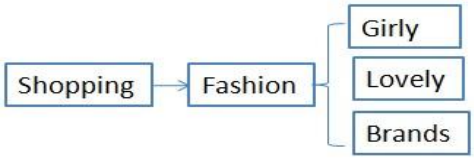

Figure 5. Selects the TIC, TISC and TIA

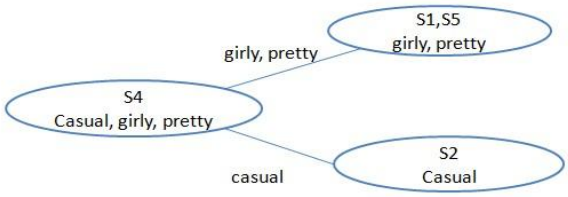

Figure 6. A new Classification Network according to preferences

A. Tourist trip B. Classification Network C. The detail of spots

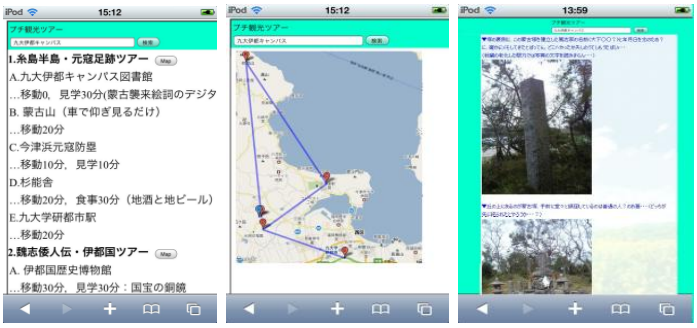

Figure 7. Example of spots

V. Conclusion and Future Works

In this paper, we proposed a smart phone-based system to support the Petit Trips. A concept dictionary for this system is also described. The feature of this system is to provide traveler a suitable Classification Network of tourist spots according to their locations and preferences.

In the future, we are planning to develop the concept dictionary and the system. The evaluation will be arranged to verify the effectiveness of the system.

References

[1] C. Carpineto, G. Romano, Concept Data Analysis Theory and Application, John Wiley and Sons, 2004

[2] G. Ganter, R. Wille, C. Franzke, Formal Concept Analysis Mathematical Foundation, Springer, 1999

[3] T. Baba, T. Nakatoh, S. Hirokawa, Text Mining of Bankruptcy Information using Formal Concept Analysis, Proc. CSIE (2nd World Congress on Computer Science and Information Engineering), to appear, 2011

[4] A. Ishino, H. Nanba, T. Takezawa, Automatic Compilation of an Online Travel Portal from Automatically Extracted Travel Blog Entries, Proceedings of the 18th International Conference on Information Technology and Travel & Tourism 2011/01

[5] H. Saito and A. Ohuchi, A Study of Visualizing Method of WWW Documents to Construct the Concept on Sightseeing Information, TECHNICAL REPORT OF IEICE (DE2001-70), Vol.101, No.192, pp.261-267, 2001.

[6] Q. Hao, R. Cai, Ch. Wang, R. Xiao, J.-M. Yang, Y. Pang, L. Zhang, Equip Tourist with Knowledge Mined from Travelogues, Proc. WWW2010, pp.401-410, 2010

[7] S. Esparcia, V. Sanchez-Anguix, E. Argente,A. Garcia-Fornes, V. Julian, Integratiing Information Extraction Agents into a Tourism Recommender System, Proc. EAIS2010, Springer LNAI 6077, pp.193-200, 2010

[8] J. M. Ruiz-Martinez, D. Castellanos-Nieves,R. Valencia-Garcia, J. T. Fernandez-Brieis, F. Garcia-Sanchez,P. J. Vivancos-Vincente, J. S. Castejon-Garrido, J. B. Camon,R. Martinez-Bejar, Accessing Touristic Knowledge Bases through a Natural Language Interface, Proc. PKAW2008, Springer LNAI 5465,pp.147-160, 2009

[9] H. Ozaku, M. Utiyama, H. Isahara, Y. Kono, M. Kidode, An Event Information Retrieval Method Using Features of Keyword Appearance in Newspaper Corpora, Transactions of the Japanese Society for Artificial Intelligence : AI 19, 225-233, 2004-11-01