九州大学学術情報リポジトリ
Kyushu University Institutional Repository

Analysis and Visualization of Tourism Blog

Nakatoh, Tetsuya
Research Institute for Information Technology, Kyushu University

Yin, Chengjiu
Research Institute for Information Technology, Kyushu University

Hirokawa, Sachio
Research Institute for Information Technology, Kyushu University

https://hdl.handle.net/2324/1500372

出版情報：Procedingsof IIAI International Symposium on Applied Information (IIAI ISAI 2012). 2012, pp.26-27, 2012-05. International Institute of Applied Informatics(IIAI)
バージョン：
権利関係：

Analysis and Visualization of Tourism Blog

Tetsuya Nakatoh, Chengjiu Yin, Sachio Hirokawa
Research Institute for Information Technology, Kyushu University
{nakatoh,yin,hiroakwa}@cc.kyushu-u.ac.jp

***Abstract*—In recent years, with the development of the Internet, and the rapidly increasing number of tourism portal sites and blogs, we can obtain a variety of tourist information on the Internet. If we have a specific need, we can obtain the required information through checking the retrieved results one by one. However, in order to see the whole trend of the results, it is necessary to analyze and visualize the document group. In this paper, we show the two visualization methods for blog articles. First, we show how to clarify the difference among one pair of words about tourism by visualizing document groups with the words. Next, we propose the method of extracting a place-name from a blog article automatically. Furthermore, we try clarification of a place-name.**

I. Introduction

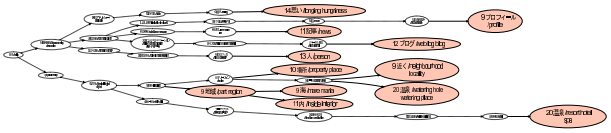
Generally, a tourist resort has each special feature. In order to draw tourists, hotels and tour companies are making an effort to make many people know such special features through media, campaign, etc. Moreover, in recent years, much tourism information is offered in tourism portal sites to those who gather tourism information on WWW. Probably, tourists are actually enjoying the special feature of each tourist resort investigated by WWW.

On the other hand, the tourist who actually went to the tourist resort may discover the special feature which generally is not known, and may enjoy it. Or they may have evaluation which is different from information about the known special feature. Such the special feature and information that were experienced by the tourist were shared only among very close persons. However, they are exhibited by blog articles, such as a travel record, now in many cases. Such information is profitable also for people who want to choose the tourist resort which should be visited, the tourist agent who wants to improve the quality of service, and the local government which wants to find out the special feature of new tourism.

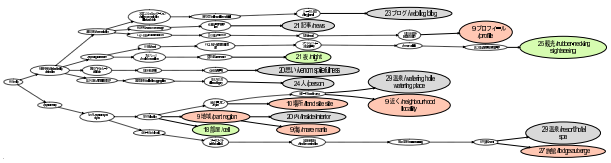
We are researching about the extraction and analysis of information in tourism blogs [1]–[14]. In this paper, we introduce about visualization of the information in tourism blogs.

II. Feature Words on WordNet

Although we are going to understand the feature of an area from document groups, such as a tourism blog, there is a limitation in the information acquired from each obtained feature word. We proposed the system visualized by mapping the obtained feature word on WordNet which is a conceptual thesaurus. In this system, the contents of the document group also including the relation of the feature word can be grasped at a glance. Moreover, it becomes possible by mapping two document groups on the same thesaurus to visualize the difference clearly. Fig. 1 shows the difference between hotels and japanese inn as an example of such visualization.



(a) Japanese inn Only



(b) Japanese inn and Hotel

(c) Hotel Only

Figure 1. Comparison of Tourism Information about Japanese inn and Hotel

III. Place Name Disambiguation

In order to extract useful information from tourism blogs, it is necessary to clarify the area. It is relatively easy for human to understand the place name, since we have a lot of background knowledge and we can grasp the context. However, automation of the processing is not easy. This section reports a trial of place name extraction from tourism blogs and disambiguation of location names.

Extraction of the name of a place from a tourism blog was performed by the following method. First, a dependency analysis restricts the context of tourism. Next, the dictionary of a name of a place candidate extracts the name of a place from the pair of a dependency analysis. By making the extracted name of a place into an attribute, blog reports are visualized using a classification network.

Fig. 2 is the classification network of the 25,113 blog articles that contain “TENJIN”. Each different Tenjin is classified into two or more partial graph.

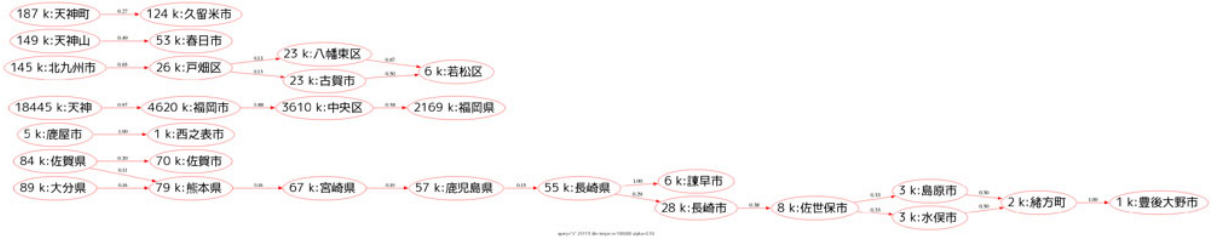


Figure 2. Classification Network of “TENJIN”

REFERENCES

[1] Sachio Hirokawa, Chengjiu Yin, Tetsuya Nakatoh, Component-Based Search Engine for Blogs, Proc. FUZZ-IEEE, pp.1074–1078, 2011

[2] S. Hirokawa, C. Yin, K. Hashimoto, Multidisciplinary Multi-Faceted Search Engine of Literatures on Tourism, Proc. 2nd International Symposium on Applied Informatics (ISAI 2011)

[3] Sachio Hirokawa, Chengjiu Yin, Kiyota Hashimoto and Kazuhiro Takeuchi, Search and Analysis of Gourmet Blogs with a Particular Reference to Onomatopoeia, ICIC Express Letters, Vol.5, No.8(B), pp.2971-2978, 2011

[4] Sachio Hirokawa, Jun Zeng, Comparison of Tourism Data using Double Ranking, Proc. SSNE, International Workshop on Innovative Tourism Informatics pp.67-72,2011

[5] T. Nakatoh, C. Yin, H. Matsuura, S. Hirokawa, Visualization of Tourism Information using WordNet, Proc. iCAST2011, pp.413–418, 2011

[6] Tetsuya Nakatoh, Chengjiu Yin and Sachio Hirokawa, Characteristic Grammatical Context of Tourism Information, ICIC Express Letters, Vol. 6, No. 2, pp.563-568, 2012

[7] Tetsuya Nakatoh, Chengjiu Yin, Sachio Hirokawa Extraction and Disambiguation of Name of Place from Tourism Blogs SSNE, International Workshop on Innovative Tourism Informatics pp.73-78, 2011

[8] X. Wu, S. Hirokawa, C. Yin, T. Nakatoh, and Y. Tabata, Extraction and Comparison of Tourism Information on the Web, Proc. AROB16 (16th International Symposium on Artificial Life and Robotics), pp.228–231. (2011)

[9] Xiaobin Wu, Sachio Hirokawa, Chengjiu Yin, Tetsuya Nakatoh, Yoshiyuki Tabata, Extraction and Comparison of Tourism Information on the Web, Proc. AROB16 (16th International Symposium on Artificial Life and Robotics), pp.228–231, 2011

[10] C. Yin, X. Wu, S. Hirokawa, and T. Nakatoh, A Proposal of “TOIEBA” Search Engine for Tourism Event, IEICE technical report. Artificial intelligence and knowledge-based processing 110(301), pp.43–47. (2010) (in Japanese)

[11] C. Yin, T. Nakatoh, S. Hirokawa, X. Wu, and J. Zeng, A proposal of search engine XYZ for tourism events, Proc. JCAI (International Joint Conference on Artificial Intelligence) Vol.1, pp.178–181. (2010)

[12] Chengjiu Yin, Tetsuya Nakatoh, Ito Eisuke, Sachio Hirokawa, Classification network of tourism information – A smart phone-based system for supporting ”Petit Trips”, Proc. iCAST2011, pp.171–174, 2011

[13] Chengjiu Yin, Tetsuya Nakatoh, Sachio Hirokawa, Xiaobin Wu, Jun Zeng, A proposal of search engine XYZ for tourism events, Proc. JCAI (International Joint Conference on Artificial Intelligence) Vol.1, pp.178–181, 2010

[14] Jun Zeng, Toshihiko Sakai, Chengjiu Yin, Takahiko Suzuki, Sachio Hirokawa, Automatic Generation of Tourism Quiz using Blogs, 7th International Symposium on Artificial Life and Robotics(AROB2012), pp.97-100, 2012