Comparison of tourism data using double ranking

Hirokawa, Sachio
Research Institute for Information Technology, Kyushu University

Zeng, Jun
Graduate School of Information Science, Kyushu University

https://hdl.handle.net/2324/833903

出版情報：Proceedings - 1st ACIS International Symposium on Software and Network Engineering, SSNE 2011, pp.67-72, 2011-12-01
バージョン：
権利関係：

Comparison of Tourism Data using Double Ranking

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

Jun Zeng
Graduate School of Information Science,
Kyushu University
Fukuoka, Japan
zeng.j.000@s.kyushu-u.ac.jp

***Abstract*— The prompt discovery and mining of user's reputation, opinion and complaint are becoming a hot topic in the field of search engine for Blog and Twitter. This paper proposes "Double Rank" method to analyze the search result using two viewpoints, where the polarity degree of keywords in the search results are evaluated from the viewpoints. Most previous researches concern mainly in positive and negative evaluation as for the polarity degree. Two viewpoints are specified as the search condition in the present paper. The blog articles related to sightseeing are analyzed as case studies, where the characteristics and the sightseeing situation are compared for two prefectures.**

Keywords- Ranking, Sentiment Analysis, Tourism blog, Feature words

I. INTRODUCTION

With the increasing of Web Blogs, more and more researchers try to discover the blog users' opinion and comment, in order to solve some problems such as handling the customer complaints, developing or improving the new products. The traditional researches just pay attention to sorting the documents into two classes: positive and negative. Although, recently some researchers consider to analyze the object, attribute and evaluation to sort the documents. However, these methods also try to use user's subjective expression to sort the documents into two opposite classes such as good and bad, interesting and boring. Therefore these methods can not compare some concept pairs which can not be sorted into two opposite classes absolutely, such as Japan and America, noodle and rice. In these methods, the degree of the two opposite classes is not considered at all.

In this paper, we extract the feature words of targets of comparison pair (A and B) from documents, and rank the feature words of A and B separately. According to the ranking of the feature words of A and B, we can find out the difference between A and B. We call this comparison method as Double Ranking. In order to evaluate the usability of Double Ranking, we realized a Double Ranking Analysis System, and collected 1303 blog entities from "Kyushu seifuku Blog"[1], which is a tourism blog site, as the experiment data. Using the search engine, we can compare the tourism data. For example, we compared the difference between Ramen and Udon. Figure 1 shows the pictures of Ramen and Udon, both of which are typical Japanese noodles. When we limit the feature words to the names of prefectures in Kyushu area, we find that Miyazaki, Nagasaki and Saga appear in the feature word list of Udon, meanwhile, Kakoshima, Kumamoto and Fukuoka appear in the feature word list of Ramen. This result is not just a positive or negative classification, but shows the fine distinction of two concepts.

"Ramen" "Udon"

Both Ramen and Udon are Japanese noodle dishes.

Figure 1. "Ramen" and "Udon"

II. RELATED WORKS

With the development of web blog, the analysis of both public facts and the private comments or opinions has become more and more popular [13,14]. Recently, some researchers try to extract and compare the positive or negative comments for the article, as well as the evaluations for the attribute of commodity [6,8,10]. Moreover, the purpose of such researches is various, such as discovering the similarity of user's trend [7], analyzing the bloger's degree of enthusiast [9], analyzing the feature of regionality, and analyzing by onomatopoeia [2,4].

For example, B.J. Jansen et al. [1] found that 25 percent of queries can not be classified into a single category of intent based on a manual coding of 400 queries. As for the display method of search result, besides the mainstream of result ranking, K. Hashimoto et al. [3] and T. Seki et al. [11] proposed a facet interface to display the search result. Y.Takama et al. [12] demonstrated the validity of multiple axes in an exploratory analysis of spatiotemporal trend information. C. Yin et al. [16] proposed a method to change the rank of search results by choosing different feature words in the search results. T. Seki [11] proposed a multiple viewed search engine which retrieves documents of an indicated search area and displays a matrix of the distribution of the clustering from two aspects of the retrieval result. M. Kato et al. [5] also proposed a display method of two dimensions.

However, in all of these papers, the authors just displayed the similar results in the same cell or some nearby cells in the result table, but did not consider the disposition of the result table. In this paper we propose a novel comparing method called Double Ranking. For a given concept pair, we extract

[1] http://www.welcomekyushu.jp/

the feature words of the two concepts from documents. By ranking the feature words from two viewpoints, we can find the fine distinction of two concepts. We also develop a Double Ranking Analysis System to visualization of Double Ranking.

III. DOUBLE RANKING ANALYSIS SYSTEM

A. *Design Goal*

The goal of Double Ranking Analysis System is to realize the visualization of Double Ranking, and help user to analyze and compare the difference between two given concepts. This method does not just sort the two concepts into a positive or negative classification, but shows the fine distinction of two concepts.

B. *Sentence-based Index*

Before developing Double Ranking Analysis System, we need to create an index. Index is design to optimize speed and performance in finding relevant web page for a search query. Without an index, the search engine would scan every document in the corpus, which would require considerable time and computing power. The existing search engines focus on the full-text index of web pages. However, in this paper, we well not only build an index of full-text but also an index of each sentence.

For building a sentence-based index, we need to create a sentence-based frequency file. First of all, we divide every article into sentences. Next, we use "ChaSen", which is a morphological parser for the Japanese language, to analyze morpheme of every word. As a result, we get a sentence-based frequency file as shown in Figure 2.

```
…
@ 1-12
1 h:1
1 what
1 can
1 one
1 …
@ 1-13
1 h:1
1 most
1 hotel
1…
…
```

Figure 2. Fragment of Sentence-based Index

The lines beginning with "@" present the identifier of each sentence. For example, "@ 1-12" means the 12th sentence of the page whose id is "1". The other lines present the index of keywords and their frequencies. The keyword beginning with "h:" is the id of the HTML file. Finally, we use "GETA", which is a generic engine for transposable association, to build an index of each sentence based on the frequency file mentioned above.

As for the full-text index, we will not introduce the process for creating a full-text index, that is because the process is similar with the process for creating a sentence-based index.

C. *Extraction for Feature Words*

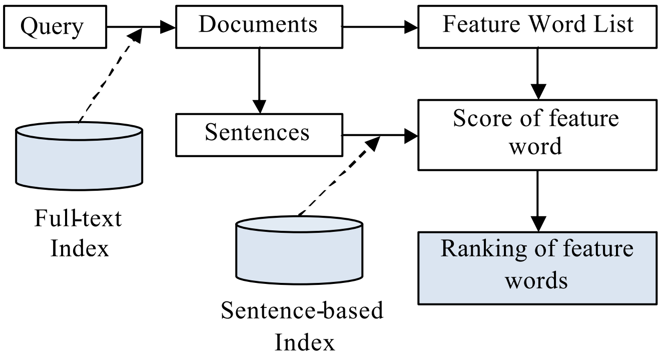


Figure 3. Extraction of Feature Words

Figure 3 shows the outline of the process for extracting the feature words. There are many systems that display the documents and related words with ranking. Conventional search engines display the snippets that contain user's query word. The novelty of the present system is to use the sentence-based index as well as the full-text index.

First of all, given a query (keyword *A* and keyword *B*), the system generates the list of document in which the keywords appear. Next, the system extracts feature word list from the documents of search result and chooses sentences that contain the feature words.

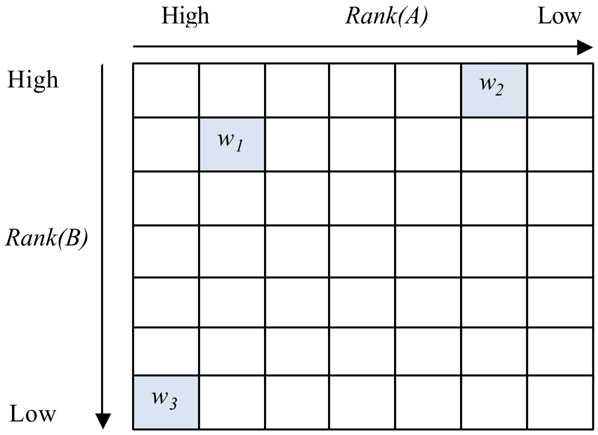


Figure 4. Example of Double Ranking

After we get the feature word list, we use SMART [17], which is algorithm to calculate the similarity between two words, to calculate the similarity between a feature word and keyword *A (B)*. In other words, a feature word *w* will have *SMART(A, w)* and *SMART(B, w)* at the same time, where *SMART(A, w)* is the similarity between *A* and *w*, and *SMART(B, w)* is the similarity between *B* and *w*. If word *w* have a higher *SMART(A, w)* or *SMART(B, w),* it means that *w* is more similar with *A* or *B*. Because GETA support the computation of SMART, we can get *SMART(A, w)* and *SMART(B, w)* of all the feature words, when we get the feature word list.

According to *SMART(A, w)* (or *SMART(B, w)*), we sort the feature word list as *Rank(A)* (or *Rank(B)*). A feature word will be ranked twice, so we call the method as Double Ranking. Figure 4 shows an example of Double Ranking. w_1, w_2 and w_3 are three feature words of keyword *A* and *B*. w_1 appears in high rank of both *Rank(A)* and *Rank(B)*, it means w_1 is common point between *A* and *B*. w_2 appears in high rank of *Rank(B)*, but low rank of *Rank(A)*. It implies that w_2 is the distinctive feature of *B*. Similarly, w_3 appears in high rank of *Rank(A)*, but low rank of *Rank(B)*. It implies that w_2 is the distinctive feature of *A*.

Finally, we extract the top 5 feature words of *Rank(A)* as the feature words of keyword *A*, and extract the top 5 feature words of *Rank(B)* as the feature words of keyword *B*.

D. Design of Double Ranking Analysis System

The present paper proposes a method to analyze the target of comparison pair by ranking the feature words from two viewpoints. A user specifies his two viewpoints as two queries A and B. The system retrieves the sets of documents that satisfy each query. A viewpoint is represented as a score vector of words in the search results. Using the two scoring, the ranking *Rank(w, A)* and *Rank(w, B)* of the word w is calculated. The feature words of the viewpoint A is obtained as whose has high ranking in *Rank(w, A)* and has low ranking in*Rank(w, B)*. Thus the difference *Rank(w, A)*-*Rank(w, B)* represents the polarity of the word w with respect A and B. If we restrict the words to be adjectives, we can compare the two viewpoints in terms of adjectives.

Figure 5 shows the interface of the system. The introduction of ①~⑦ is as following:

1) ①is a textbox to enter a keyword w to limit the range of feature words of target comparison.

2) ②is used to enter the comparison pair u and v. In Figure 5, 福岡(Fukuoka) and 長崎(Nagasaki) are chosen as the comparison target.

3) ③is used to select the set of feature words W. We have prepared 4 sets of feature words: ordinary words, adjective, onomatopoeia and names of prefectures in Kyushu area.

4) ④ is used to select the number of feature words, the default number of feature words is 5.

5) ⑤is used to select the scoring method fn by "weight" or "DF_d".

6) ⑥shows the feature word list of comparison target.

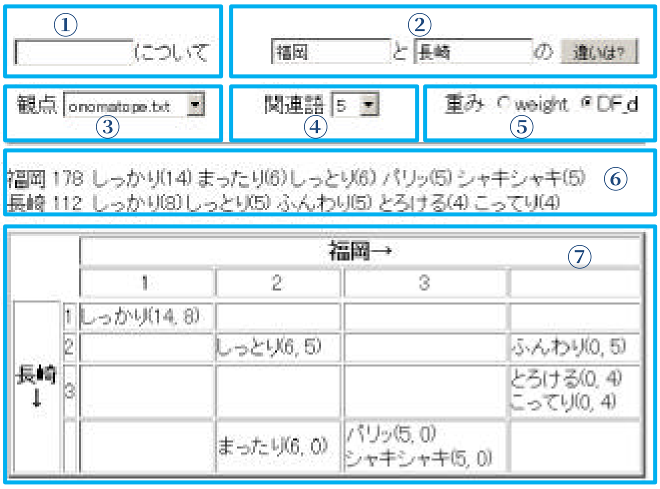
7) ⑦ shows the double ranking of feature words by two viewpoint.

A user can enter three keywords *w, u* and *v* where *u* and *v* are targets of comparison, and *w* limits the range of feature words. Feature words are extracted and displayed in an $N \times N$ matrix. The cell in $i \times j$ position displays the set of feature words *X(i, j)* which is determined as follows:

$$X(i,j) = \{ w_k \ \varepsilon \ W \mid Rank\,(fn, w_k, Search(w \text{ and } u)) = i, \quad Rank\,(fn, w_k, Search(w \text{ and } v)) = j \} \qquad (1)$$

Here, *Search(q)* denotes the set of documents obtained by a query *q*, *Rank(fn ,x, D)* denoted the rank of a word *x* in the document set *D* with respect to the scoring function *fn*. We ignore the ranking of the words below *N*.

Figure 5 also displays the result with respect to *w=empty*, *u=Fukuoka*, *v=Nagasaki*, *W* to be onomatopeia words, the scoring method *fn* is the default of GETA and the number *N* of words is 5.



福岡 178 しっかり(14) まったり(6) しっとり(6) パリッ(5) シャキシャキ(5)
長崎 112 しっかり(8) しっとり(5) ふんわり(5) とろける(4) こってり(4)

		福岡→			
		1	2	3	
長崎↓	1	しっかり(14, 8)			
	2		しっとり(6, 5)		ふんわり(0, 5)

	3				とろける(0, 4)
					こってり(0, 4)

			まったり(6, 0)	パリッ(5, 0)
				シャキシャキ(5, 0)	

Figure 5. Interface of Double Rank Analysis System

IV. EXPERIMENT AND RESULT

A. Collection of Experiment data

In this paper, we choose the tourism blogs of site "Kyushu seifuku Blog" as the experiment data. We manually collected 1,303 blog entities and saved as html files. By analyzing the blog entities we prepare 4 kinds of feature words.

1) Ordinary words: the words appear in the search result without any limitation.

2) Adjective: the adjective appear in the search result.

3) onomatopoeia: we collect the onomatopoeia manually.

4) Names of prefectures in Kyushu area: we collect the names of prefectures in Kyushu area manually.

B. Comparison and analysis by feature words

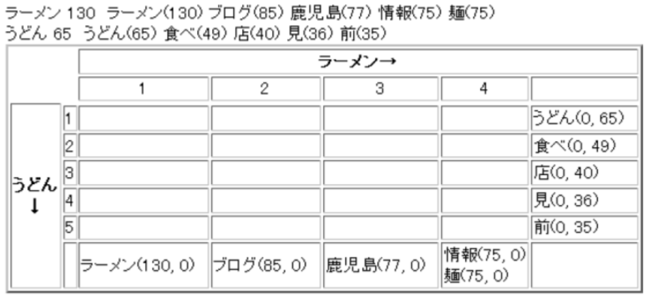
We take "Ramen" and "Udon" as an example to compare and analyze the difference of the feature words which associate with "Ramen" and "Udon". First we conduct a query for both "Ramen" and "Udon". We choose the 4 kinds of feature words mentioned above:

1) Ordinary words (Figure. 6);

2) Onomatopoeia (Figure. 7);

3) Adjective (Figure. 8);

4) The names of prefectures in Kyushu area (Figure. 9).

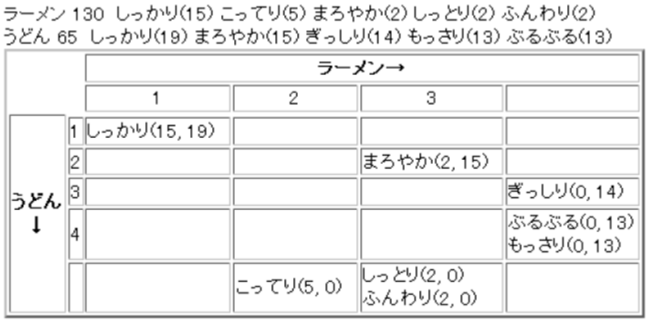


ラーメン 130 ラーメン(130) ブログ(85) 鹿児島(77) 情報(75) 麺(75)
うどん 65 うどん(65) 食べ(49) 店(40) 見(36) 前(35)

		ラーメン→				
		1	2	3	4	
うどん↓	1					うどん(0, 65)
	2					食べ(0, 49)
	3					店(0, 40)
	4					見(0, 36)
	5					前(0, 35)

		ラーメン(130, 0)	ブログ(85, 0)	鹿児島(77, 0)	情報(75, 0)
					麺(75, 0)	

Figure 6. Comparison by Ordinary words



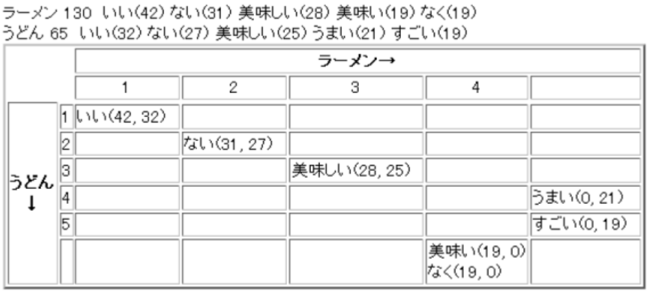
ラーメン 130 しっかり(15) こってり(5) まろやか(2) しっとり(2) ふんわり(2)
うどん 65 しっかり(19) まろやか(15) ぎっしり(14) もっさり(13) ぷるぷる(13)

		ラーメン→			
		1	2	3	
うどん↓	1	しっかり(15, 19)			
	2			まろやか(2, 15)	
	3				ぎっしり(0, 14)

	4				ぷるぷる(0, 13)
					もっさり(0, 13)

			こってり(5, 0)	しっとり(2, 0)
				ふんわり(2, 0)	

Figure 7. Comparison by Onomatopoeia

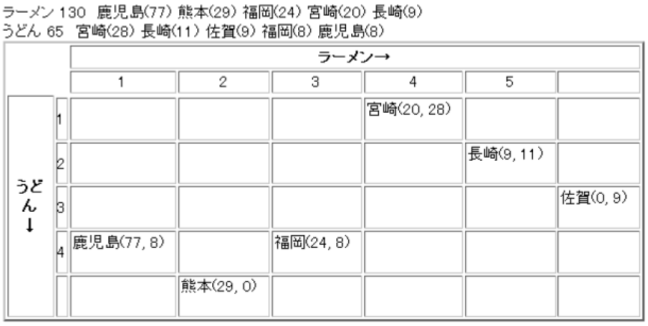


ラーメン 130 いい(42) ない(31) 美味しい(28) 美味い(19) なく(19)
うどん 65 いい(32) ない(27) 美味しい(25) うまい(21) すごい(19)

		ラーメン→				
		1	2	3	4	
うどん↓	1	いい(42, 32)				
	2		ない(31, 27)			
	3			美味しい(28, 25)		
	4					うまい(0, 21)
	5					すごい(0, 19)

					美味い(19, 0)
					なく(19, 0)	

Figure 8. Comparison by Adjective



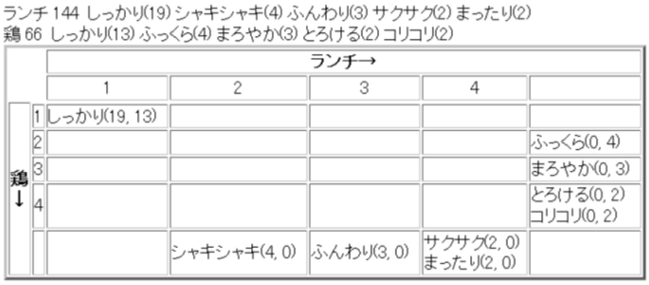
ラーメン 130 鹿児島(77) 熊本(29) 福岡(24) 宮崎(20) 長崎(9)
うどん 65 宮崎(28) 長崎(11) 佐賀(9) 福岡(8) 鹿児島(8)

		ラーメン→					
		1	2	3	4	5	
うどん↓	1				宮崎(20, 28)		
	2					長崎(9, 11)	
	3						佐賀(0, 9)
	4	鹿児島(77, 8)		福岡(24, 8)			
			熊本(29, 0)				

Figure 9. Comparison by name of prefecture

In Figure 6, we notice that the feature word “**鹿児島**(Kakoshima)” appears at top 3 of the feature word list of “Ramen”, but it does not appear at the feature word list of “Udon”. In Figure 7, the feature words “**しっかり**(eat well)” and “**まろやか**(taste smooth)” appear at the high ranking of the feature word list of both “Ramen” and “Udon”. However, the feature word “**こってり**(a filling dish)” appears at top 2 of the feature word list of “Ramen”, but does not appear at the feature word list of “Udon”. That is because “**こってり**(a filling dish)” is the distinctive feature word of “Ramen” but “Udon”. The feature word “**ぎっしり**(packed like sardines)” appears at the top 3 of the feature word list of “Udon”, but does not appear at the feature word list of “Ramen”. In a word, “**こってり**(a filling dish)” is the feature of “Ramen” and “**ぎっしり**(packed like sardines)” is the feature of “Udon”. In Fig. 4, we can find the adjective feature word of both “Ramen” and “Udon” is nearly the same. According to Fig. 5, we can suppose that in Kakoshima, Kumamoto and Fukuoka, people prefer Ramen, while in Miyazaki, Nagasaki and Saga “Udon” is more popular.

C. Comparison and analysis by quantification of similarity and dissimilarity



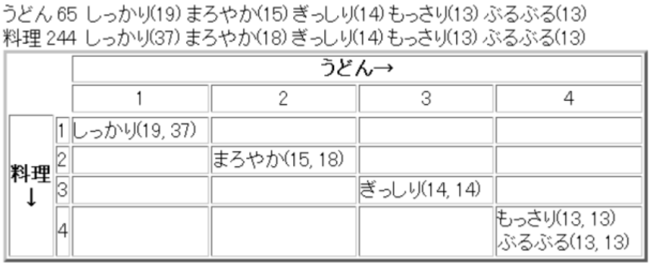
ランチ 144 しっかり(19) シャキシャキ(4) ふんわり(3) サクサク(2) まったり(2)
鶏 66 しっかり(13) ふっくら(4) まろやか(3) とろける(2) コリコリ(2)

		ランチ→				
		1	2	3	4	
鶏↓	1	しっかり(19, 13)				
	2					ふっくら(0, 4)
	3					まろやか(0, 3)

	4					とろける(0, 2)
						コリコリ(0, 2)

			シャキシャキ(4, 0)	ふんわり(3, 0)	サクサク(2, 0)
					まったり(2, 0)	

Figure 10. Comparison of Chiken and Lunch



うどん 65 しっかり(19) まろやか(15) ぎっしり(14) もっさり(13) ぷるぷる(13)
料理 244 しっかり(37) まろやか(18) ぎっしり(14) もっさり(13) ぷるぷる(13)

		うどん→			
		1	2	3	4
料理↓	1	しっかり(19, 37)			
	2		まろやか(15, 18)		
	3			ぎっしり(14, 14)	

	4				もっさり(13, 13)
					ぷるぷる(13, 13)

Figure 11. Comparison of Udon and Dish

We can also compare and analyze two viewpoint by quantification of similarity and dissimilarity of comparison targets. In the Matrix of Double Ranking, if two words are similar with each other, the feature words are more likely appear at the diagonal line of the Matrix. If two words are dissimilar with each other, the feature words will appear at the right side or bottom of the Matrix, which are out the rank. Therefore, we can calculate the distance from the position of each feature word to the diagonal line, and calculate the sum of all the distances in order to determine the quantification of similarity and dissimilarity of comparison targets.

We extract 22 high-frequency words related to food from blog data as following:

Noodle (488), Soup (347), Ramen (331), Dish (282), Lunch (165), Chashu (157), Vegetable (142), Meat (141), Pig (128), Onion (120), Wine (112), Udon (112), Gourmet (107), Chicken (102), Fish (101), Soy (100), Egg (98), Restaurant (97), Dinner (91) Cattle (87), Rice (86), Pig bone (84)

Here the numbers behind the words are the frequencies of the words that appear.

TABLE I. DISSMILARITY OF WORDPAIRS (DF)

	Dissimilarity	Word pair
Dissimilarity Word Pairs	2.9	Lunch : Chicken
	2.8	Lunch : Dish
	2.7	Lunch : soup
		Lunch : Chashu
		Noodle : Dish
	2.5	Pork bone : Ramen
		Soup : Chicken
		Soup: Dish
		Lunch : Noodle
	2.4	Lunch : Ramen
		Ramen : Chashu
		Fish : Rice
Similarity Word Pairs	1.5	Dinner : Rice
		Cattle : Pig
		Onion : soy
	1.4	Onion : Cattle
		Ramen : Egg
	1.3	Onion : Rice
		Onion : Chashu
	1.2	Dish : Noodle
	1.1	Udon : Dish

TABLE II. DISSMILARITY OF WORDPAIRS (WEIGHT)

	Dissimilarity	Word pair
Dissimilarity Word Pairs	3.0	Gourmet:Pig Gourmet:Chashu
	2.9	Noodle : Soy
	2.8	Dish : Rice
	2.7	Soy : Chashu
		Lunch : Chashu
		Pork bone : Ramen
		Egg : Dish
		Ramen: Egg
		Rice : Chashu
		Lunch : Noodle
		Onion : Fish
Similarity Word Pairs	1.5	Udon : Restaurant
		Chashu : Restaurant
		Udon: Fish
		Udon : Onion
	1.4	Dinner : vegetable
	1.2	Restaurant : Wine
		Cattle : Dish
		Fish : Wine
	0.7	Meat : Dish
	1.1	Udon : Dish

We compare the words by 328 onomatopoeias, and analyze the dissimilarity and similarity of every word. Table I shows the result. It may be unexpected that Lunch & Chicken is the most dissimilar word pair. Figure 10 shows the detail comparison of Lunch & Chicken by Double Ranking. We notice that only "しっかり(eat well)" is the common feature word. Because their common feature words are few, they are considered as dissimilarity word pair. Figure 11 also shows that Udon & Dish is the most similar word pair. Figure 11 shows the detail of comparison of Udon & Dish. We find that Udon & Dish have same feature words, and all of the feature words are on the diagonal line. Therefore, they are considered as the most similar words.

If we choose the weight as the sort method, the rank of feature words will change, and the similarity and dissimilarity will change as well. TABLE III shows that Gourmet & Pig, and Gourmet & Chashu are the most dissimilar words. Therefore, we analysis the feature words of Gourmet & Chashu.

TABLE III. FEATURE WORDS OF GOURMET AND CHASHU

Gourmet	
Ordinary words	Gourmet (17.82), Grade (7.73), Log (6.71), My List (6.44)
Onomatopoeia	KOTTERI (a filling dish) (2.83), SAKUSAKU (crunchy) (2.35), KARIKARI (crunch-crunch) (2.31) KARI (crunch) (1.54) SHAKISHAKI (be crisp to eat) (1.35)
Adjective	Delicious (6.36), nice (4.04), Quadrangle (3.8), few (3.28)
CHASHU	
Ordinary words	CHASHU (9.61), Ramen (3.37), Relation (3.12), Soup (2.76), Stress (2.68)
Onomatopoeia	KOTTERI (a filling dish) (2.83), MAROYAKA (taste smooth) (0.88) SHIKKARI (eat well) (0.71) FUKKURA (soft) (0.47)
Adjective	Thin (2.34), few (1.86), Thick (1.84), deep (1.84), lonely (1.81)

TABLE III shows the result of analysis where the numbers in the brackets are the SMART scores of feature words. We sort the feature words into three parts: ordinary words, onomatopoeia and adjective. The ordinary words of Chashu are almo st the ingredient of Ramen. The onomatopoeia of Gourmet are the words describing mouth feel, and the onomatopoeia of Chashu are the words describing the ingredient. The adjectives of Gourmet almost describe the taste of food, and the adjectives of Chashu describe the quantity or satisfy feel of food. By analyzing the difference of feature words, we can compare the dissimilarity of word pairs.

V. Conclusion and Future Works

In this paper, we proposed a comparison method called Double Ranking. By ranking the feature words from two viewpoints, we can find the fine distinction of two concepts. We also developed a Double Ranking Analysis System to visualization of Double Ranking. We collected 1303 blog entities as experiment data. By compare the feature words (such as onomatopoeia, adjective and name of prefecture), we can analyze the similarity and dissimilarity of a given word pair effectively.

In the future we will consider the method of evaluation, analysis of questionnaire documents, and application of time series analysis.

References

[1] B.J. Jansen, D.L. Booth, A. Spink, Determining the user intent of web search engine queries, 16th International World Wide Web Conference, WWW2007 , pp. 1149-1150, 2007

[2] K. Hashimoto, K. Takeuchi, C. Yin, S. Hirokawa, Extraction of Subjective Context-Sensitive Evaluation of Japanese Onomatopoeic Expressions and its Applications, ICIC Express Letters, Vol.5, No.10, pp.3755-3760, 2011

[3] M. A. Hearst, Clustering versus faceted categories for information exploration Communications of t he ACM 49 (4), pp.59-61, 2006

[4] S. Hirokawa, C. Yin, K. Hashimoto, K. 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

[5] Motohide Kato, Chikara Yonemori, Tsutomu Matunaga, A quantification method of enterprise characteristics using documents, 4th The Japanese Society for Artificial Intelligence, 2007

[6] KOBAYASHI Nozomi, INUI Kentaro, MATSUMOTO Yuji, TATEISHI Kenji, Collecting Evaluative Expressions by A Text Mining Technique, Information Processing Society of Japan NL154-12,77-84, 2003

[7] Kumamoto Tadahiko, Tanaka Katsumi, A Web Retrieval System Based on Lexical Paraphrasing Using Two Kinds of Co-occurrence Dictionaries. Transactions of the Japanese Society for Artificial Intelligence, Vol.23 No. 5 pp.355-363, 2008

[8] Takeshi KURASHIMA, Katsuji BESSHO, Toshio UCHIYAMA, Ryoji KATAOKA, Ranking Method using Comparative Relations extracted from CGM, DEWS2007,L1-5, 2007

[9] Shinsuke NAKAJIMA, Yoichi INAGAKI, Tomoaki KUSANO, Blog Ranking Method Based on Bloggers' Knowledge Level for Providing Trustable Information, Journal of the DBSJ Vol.7, No.1, pp.257-262, 2008

[10] Toshinori SATOU, Manabu OKUMURA, Extraction of Comparative Relations from Japanese Weblog, Information Processing Society of Japan, 2007(94), pp.7–14, 2007

[11] T. Seki, T. Wada, Y. Yamada, N. Ytow, S. Hirokawa, Multiple Viewed Search Engine for an e-Journal - A Case Study on Zoological Science, HCI (4), pp.989-998, 2007

[12] Takama Yasufumi, Karino Shinji, Exploratory Analysis Support of Spatiotemporal Trend Information Focusing on Comparative Analysis, Transactions of the Japanese Society for Artificial Intelligence,Vol. 26 No. 4 pp.494-503, 2011

[13] Takamura Hiroya, Inui Takashi, Okumura Manabu, Latent Variable Models for Semantic Orientations of Phrases, Information Processing Society of Japan, 2005-NL-168, 2005

[14] Peter D. Turney, Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, Proc. 40th Annual Meeting of the Association for Computational Linguisitics(ACL), 417-424, 2002.

[15] X.Wu, S. Hirokawa, C. Yin, T. Nakatoh, 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

[16] C. Yin, T. Nakatoh, S. Hirokawa, X. Wu, 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

[17] Gerard Salton, Michael J. McGill: Introduction to Modern Information Retrieval, McGraw-Hill, 1983.