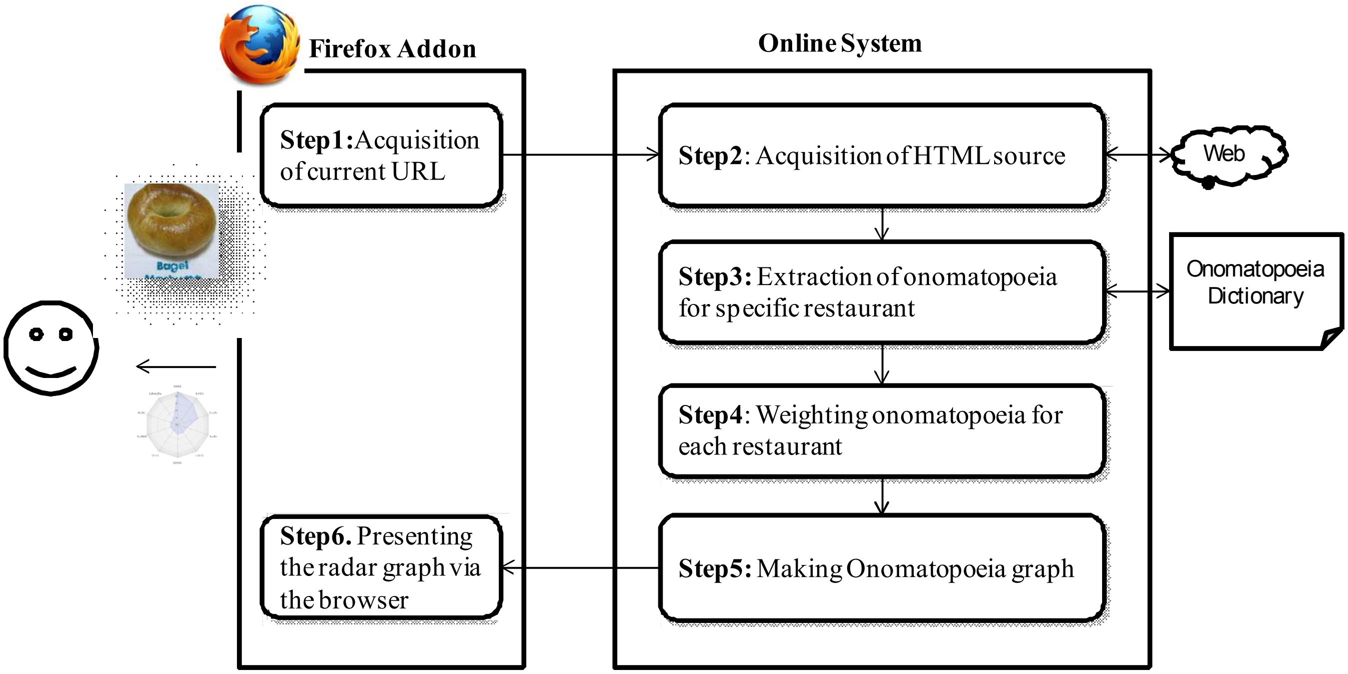

Fig. 6:The detailed procedure of online system of Fig. 3. This process displays onomatopoeias that characterize restaurants.

Step4. Weighting onomatopoeia for each restaurant

In order to provide onomatopoeias that will help the user to evaluate the restaurant, we use TFIDF to extract the onomatopoeias that represent the characteristics of restaurants, and calculate the weights of each onomatopoeia. There are two candidates for TFIDF calculation as shown in equation (1) and (2). As described in the next section, we found that equation (1) is more effective in selecting characteristic onomatopoeia than equation (2). Therefore, equation (1) is used for TFIDF calculation. In equation (1), we treat $tf_{i,j}$ as "the number of times onomatopoeia i is used in the reviews of restaurant j", in other words, it corresponds to the number counted in Step3 for each onomatopoeia i. $df_{i,j}$ is the number of times onomatopoeia i is used in all reviews in Tabelog.

Step5. Making Onomatopoeia graph

In order to help the user understand the distribution of onomatopoeia used in the restaurant reviews more easily, we use a radar chart to visualize the onomatopoeias. To show the differences in weights of several onomatopoeias, we adopt two different weighting metrics of onomatopoeia; one is frequency of the onomatopoeia used in the reviews of the restaurant and the other is TFIDF value of the onomatopoeia. We show a radar chart that adopts frequency as the weighting metric in Fig. 7, and the radar chart that adopts TFIDF as the weighting metric in Fig. 8. The scales in both Fig. 7 and Fig. 8 represent frequency because TFIDF values are difficult for users to understand.

Step6. Presenting the radar graph via the browser

We present the radar charts created in Step5 to the user in a pop up window.

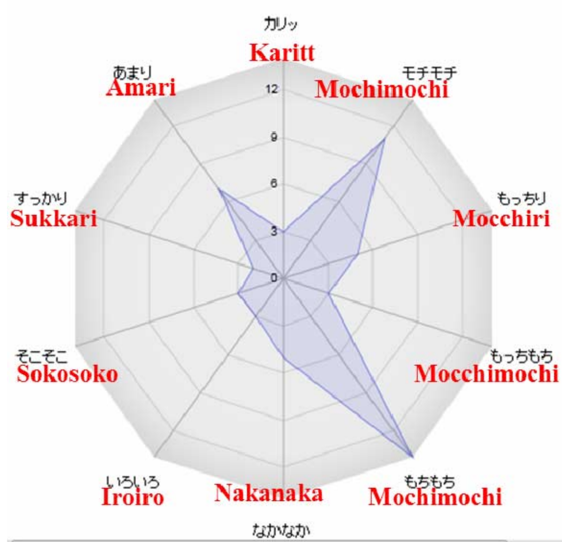

Fig. 7:Radar chart of the top 10 onomatopoeias sorted by frequency

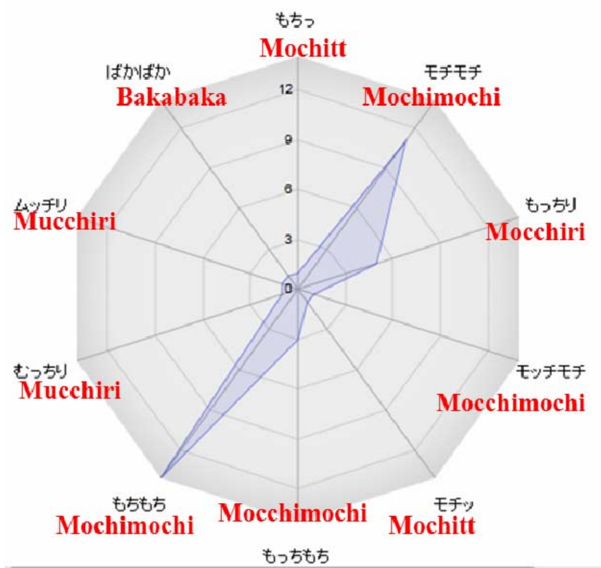

Fig. 8: Radar chart of the top 10 onomatopoeias sorted by TFIDF

4. Experimental Evaluation

This section describes our evaluation of the proposed system. We conducted two evaluations; one evaluated the coverage and precision of onomatopoeia extracted from the Tabelog site, and the other evaluated the user's reaction. 26 people participated, and three food categories were examined: "Bagel", "Hamburger" and "Ramen".

4.1 Evaluation of food onomatopoeia dictionary

We determined how many onomatopoeias are specific to food from the onomatopoeias extracted from Tabelog. There are five method for sorting the extracted onomatopoeias; A) order of frequency i.e. tf, B) order of TFIDF calculated by equation (1), C) order of TFIDF calculated by equation (2), D) onomatopoeias whose frequency exceeds 600 are selected from those sorted by B and E) onomatopoeias whose frequency exceeds 600 are selected from those sorted by C. We evaluated the top 30 onomatopoeias from those output by A) to E). In order for participants to keep their concentration on the experiments, we only evaluate the top 30 onomatopoeias for each food category. We assume that ordinary people and food specialists use different onomatopoeias, so we evaluated the extracted onomatopoeias by both ordinary people and food specialists.

4.1.1 Evaluation of precision of extracted Onomatopoeia by ordinary people

In order to gather correct onomatopoeia from ordinary people, we conducted a user survey. The question posed is as follows:

We sort onomatopoeias which appear in the personal reviews of restaurant food according to rules A, B, C, D and E, in which we present the top 30 onomatopoeia. Please carefully read the list of onomatopoeias and check if you feel that the onomatopoeia can express the foods in the specified food category

Here, we define correct onomatopoeia as onomatopoeia that was selected by more than half of the 26 survey participants. Table 3 and Fig. 9 show the results of the survey. In Table 3, we show the number of onomatopoeia that was judged as correct onomatopoeia in the context of each food category. In addition, Fig. 9 shows the percentage of correct onomatopoeia among the top 30 for each food category and the method of sorting used i.e. A) to E). As can be seen from the figure, sorting method D) yielded the highest percentage of correct onomatopoeia, more than 46%. In Table 4, we present the result in "ordinary people" column, in which we list the onomatopoeias sorted by method D), and the checked ones are correct.

We discuss each method here. Because A) sorts onomatopoeia based on frequency, onomatopoeia which are used for several foods, not just bagels, are output; e.g., "Nakanaka"(pretty). As a result, the onomatopoeias specific to bagels are not found in the top entries. As for sorting method B), the percentage increases rather than A) in the Bagel but decrease in the Ramen, This is because by using the TFIDF, we can increase the percentage of the onomatopoeias specific for each food category, however, as we described in step 4 of section 2, a noise onomatopoeia, for example "Zakun", or "Hoyan", decrease the percentage of correct onomatopoeia. On the other hand, method D) yields a higher percentage in all the food categories than B). This is because D) uses the threshold of frequency to eliminate the noisy onomatopoeia. Although method E) employs TFIDF and a frequency-based threshold, the accuracy of E) is not higher than that of D). This is because the df value of each onomatopoeia in the Tabelog site is small, so their TFIDF values are close to each other, which means that the effectiveness of df is small. As a result, method E) yields results similar to method A).

Table 3:Questionnaire derived onomatopoeia

	Correct Onomatopoeia (piece)				
	method A	B	C	D	E
Ramen	10	5	9	19	9
Bagel	8	11	11	14	10
Hamburger	4	4	5	14	5

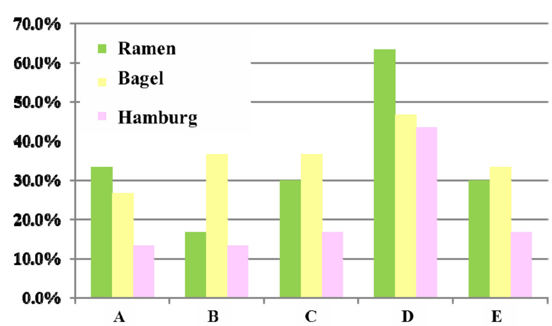

Fig. 9: Questionnaire derived onomatopoeia results(%)

4.1.2 Evaluation of precision of extracted Onomatopoeia by Tabelog User

In order to gather correct onomatopoeias from food specialists who are knowledgeable about food, we used the reviews written by Tabelog users. Here, we define the onomatopoeias that the Tabelog users actually used in writing about food as correct onomatopoeias. This definition of correct onomatopoeias differs from that used in Section 4.1.1.

Table 4: Onomatopoeia for food types output by method D Onomatopoeias for each food with method D By Tabelog Bloggers.

Rank	Ramen	Ordinary people	Food specialist	Bagel	Ordinary people	Food specialist	Hamburger	Ordinary people	Food specialist
1	Tantan			Mucchiri	*	*	Ju-ju-	*	*
2	Churuchuru	*	*	Mocchiri	*	*	Juujuu	*	*
3	Washiwashi		*	Fukafuka	*	*	Tarutaru		
4	Gidogido	*	*	Sakkuri	*	*	Morimori	*	
5	Gowagowa	*	*	Fuwatt	*	*	Pekopeko		
6	Zuruzuru	*	*	Mochimochi	*	*	Fukafuka	*	*
7	Paratt		*	Gorogoro		*	Panpan		
8	Zuzutt	*	*	Paritt	*	*	Gattsuri	*	*
9	Dorodoro	*	*	Karitt	*	*	Gatsugatsu	*	*
10	Surusuru	*	*	Karikari	*	*	Gorogoro	*	*
11	Tsurutt	*	*	Fuwafuwa	*	*	Pasapasa	*	*
12	Nurunuru	*	*	Pasapasa	*	*	Batabata		
13	Parapara		*	Pakupaku			Wain		
14	Parari		*	Sakutt	*	*	Nakan		
15	Dorori	*	*	Hitotsuhitotsu			Gutsugutsu	*	*
16	Shikoshiko	*	*	Paripari	*	*	Mein		
17	Tsurutsuru	*	*	Sakusaku	*	*	Wakuwaku		
18	Kotteri	*	*	Wakuwaku			Atsuatsu	*	*
19	Dorott	*	*	Urouro			Hokuhoku	*	*
20	Kotekote	*	*	Hitotsuhitotsu			Gayagaya		
21	Fufu	*		Mattari		*	Torotoro	*	*
22	Gatsun	*	*	Tsuitsui			Purin		*
23	Piripiri		*	Purin		*	Sakkusaku		*
24	Ziwaziwa		*	Madamada			Maamaa		
25	Dekadeka		*	Chokuchoku			Mattari		*
26	Niyari			Wazawaza			Kibikibi		
27	Zarazara		*	Sukkari			Fuwafuwa	*	*
28	Betobeto	*	*	Matamata			Hokahoka	*	*
29	Cari			Attari			Girigiri		
30	Syakitt	*	*	Motomoto			Karikari		*
# of onomatopoeia		19	26		14	17		14	17
% of onomatopoeia		63.3	86.7		46.7	56.7		46.7	56.7

Note that as our method extracts onomatopoeias from the reviews in Tabelog, it is certain that all the extracted onomatopoeias exist in the review site. However, some onomatopoeias are not used to discuss food. For example, the onomatopoeia "tantan", which is the 1st ranked onomatopoeia of Table 4, is actually part of the name of a type of noodle "tantan noodle". To eliminate those noise onomatopoeias and count only the number of correct onomatopoeias the Tabelog users actually used in writing about foods, we searched for each onomatopoeia in the review, and judged whether they were used to describe food attributes. In this evaluation, we used only the onomatopoeia sorted by method D) which shows the best performance in Section 4.1.1. To make judgment on the correct onomatopoeias, we examined 30 onomatopoeias for each of the 3 foods (90 onomatopoeias in total as shown in Table 4). We present the result in Table 4 in the "food specialist" column. The correct onomatopoeias are 26 (87%) for Ramen, 17 (57%) for Bagel, and 17 (56.7%) for Hamburger. Tabelog bloggers are expected to know of and to use more onomatopoeias. This shows that food specialists use a wider variety of onomatopoeia than ordinary people when describing food. As specialists, they try to describe in more detail their feeling for the food or express the atmosphere of food with more realism, and so employ rarely used or new onomatopoeias.

4.1.3 Evaluation of coverage of extracted onomatopoeia

In order to verify the extent to which unknown onomatopoeia could be extracted from the Tablelog site, we counted the number of existing onomatopoeia present in two other corpora. In this case, we use the onomatopoeias judged as correct by ordinary people in Section 4.1.1. One is the "Japanese dictionary of the website Goo [8]" and the other is the "Japanese onomatopoeia dictionary [7]". The former contains a dictionary on onomatopoeia, "Onomatopedia". Here, we verify the onomatopoeia extracted for the three food categories of Bagel, Ramen, and Hamburger. We show the comparison result in Table 5. As shown in the table, the dictionary of Goo held only 34 of the 62 correct onomatopoeias. However, most of these were registered as adverbs and only three were registered as onomatopoeia. (See *1 in Table 6) . On the other hand, the Japanese onomatopoeia dictionary held 44 onomatopoeias.

As can be seen, our proposed method offers wider coverage than existing corpora. It seems that synonym onomatopoeias are not present in the existing corpora. For example, in the Japanese onomatopoeia dictionary "JyuuJyuu"(sizzle) is registered, "Jyuu*1Jyuu*1" and " Jyuu*2Jyuu*2 ", in which *1 is written as the long vowel indicator and *2 is written as small character of Hiragana in Japanese, is not registered. In addition, the Japanese onomatopoeia dictionary only offers "Mochimochi",(chewy texture) but our dictionary also offers "Mochumochu". As described, a significant point of our proposed method is that we can extract unknown onomatopoeias which are similar to existing onomatopoeias by derivative patterns.

Table 5: Number of onomatopoeia found in existing dictionaries. *1 represents the onomatopoeia whose POS is registered as "onomatopoeia".

Rank	Extracted Onomatopoeia	Dictionary of Goo	Japanese onomatopoeia dictionary	Rank	Extracted Onomatopoeia	Dictionary of Goo	Japanese onomatopoeia dictionary
1	Atsuatsu	*	-	33	Dorodoro	*	*
2	Gatsugatsu	*	*	34	Dorori	*	*
3	Gattsuri	*	*	35	Nikuniku	-	-
4	karikari	*	*	36	Nurunuru	*	*
5	Karitt	*	*	37	Necchinechi	-	-
6	Gidogido	*	*	38	Pasapasa	*	*
7	Gutsugutsu	*	*	39	Parripari	-	-
8	Kotteri	*	*	40	Paritt	-	*
9	Kotekote	*	*	41	Paripari	*	*
10	Gorogoro	*	*	42	Piripiri	*	*
11	Gowagowa	*	*	43	Fu-fu-	-	*
12	Sakusaku	*	*	44	Fukafuka	*	*
13	Sakutt	*	*	45	PutsunPutsun	-	-
14	Sakkuri	*	*	46	Puripuri	*	*
15	Shikoshiko	*	*	47	Fuwatt	**1	*
16	Shittori	*	*	48	Fuwafuwa	*	*
17	Syakisyaki	*	*	49	Betobeto	*	*
18	JuuJuu	-	-	50	Hokahoka	**1	*
19	Ju-ju-	-	*	51	Mashimashi	-	-
20	JuuJuu	-	-	52	Mugyutt	-	*
21	Ju-tt	-	*	53	Mugyumugyu	-	-
22	Zuzutt	-	*	54	Mucchimuchi	-	-
23	Surusuru	*	*	55	Mucchiri	-	*
24	Zuruzuru	*	*	56	Mochimochi	*	*
25	Tappuri	*	*	57	Mochumochu	-	-
26	Churuchuru	-	*	58	Mocchuri	-	-
27	Churunchurun	-	-	59	Mocchiri	*	*
28	Tsurutt	-	*	60	Moccchiri	-	-
29	Tsurutsuru	*	*	61	Mofutt	-	-
30	Dorrodoro	-	-	62	Mofumofu	*	*
31	Dorott	-	*	# of onomatopoeia found		34	44
32	Torotoro	**1	*	% of onomatopoeia found		54.8	71.0

4.2. Evaluation of user interface

This section evaluates the effectiveness of the radar charts when displaying the results output by method D.

4.2.1. Evaluation of effectiveness of displaying onomatopoeia

In order to verify whether onomatopoeias are useful in understanding restaurants, we choose one famous restaurant in each food category, Ramen, Bagel, and Hamburger, and presented radar charts of onomatopoeias for each of the 3 restaurants to subjects who are the same as the participants of 4.1.1 (# of participants are 26). The subjects observed the radar chart and indicated how many onomatopoeia were useful to them. 46% of the subjects indicated that 7 of the 10 onomatopoeias were helpful in understanding the restaurant. Onomatopoeias that has were not indicated as helpful included "Kibikibi"(rapidly) and "Nainn". The former is indicative of the service provided not the food. The latter is associated with restaurant attributes other than food. Most of those onomatopoeias can be eliminated by fine tuning the frequency threshold. It is possible that thresholding may eliminate all noise and we should deal with this in the future.

4.2.2. Comparison of restaurants

Our system can also be used to compare restaurants, and this functionality was examined.

Fig. 10 and Fig. 11 show the radar charts for two ramen shops. They were presented to the subjects who then indicated if they understood the differences in the restaurants intuitively or not. As a result, 69% of the subjects understood the differences very well, 23% indicated some understanding.

According to the opinions collected from the subjects, the ramen shop of Figure 10 is probably oily and rich while that of Figure 11 looks plain and light. We also asked which onomatopoeias were most helpful in differentiating the restaurants. As a result, most subjects answered that "Assari"(plain), "Kotteri"(oily) and "Mochimochi"(chewy texture) were helpful.

Next, we showed the subjects a brief description of the restaurants and pictures of its food. After that, we asked whether the impression acquired from the radar chart matched that acquired from the description and the pictures. As a result, 50% of subjects indicated full agreement, 42% indicated general agreement, and 4% indicated disagreement. This indicates that the radar charts of onomatopoeias are very useful in helping users determine the difference between restaurants.

At the end of the survey, we ask the subjects whether they wanted to use the system or not. 19% answered very positively, 65% answered positively, 11% answered unsure, and 3% answered slightly negatively. Our experiments verified the onomatopoeia for just 3 food categories. We should check the proposal's applicability to other food categories in future work.

5. Conclusions

We have proposed and evaluated a method that can automatically extract onomatopoeias, including unknown onomatopoeia, specific to different food categories, and build a comprehensive onomatopoeia dictionary. Evaluations confirmed that the method can extract onomatopoeias not covered by existing onomatopoeia corpora. In addition, since it combines TFIDF based weighting and frequency thresholding, the method can extract onomatopoeia appropriate for specific food with precision greater than 46 %. Moreover, we introduced a system that uses the proposed method to present onomatopoeia for specific restaurants. Evaluation results indicate that by displaying a radar chart of onomatopoeias, the system helps users understand restaurants intuitively, and is helpful for comparing and selecting restaurants.

In the future, we plan to improve the accuracy of extracting unknown onomatopoeia and coverage. In addition, we plan to publish the system to web users, and verify the effectiveness of using onomatopoeias to help users understand restaurants by real user.

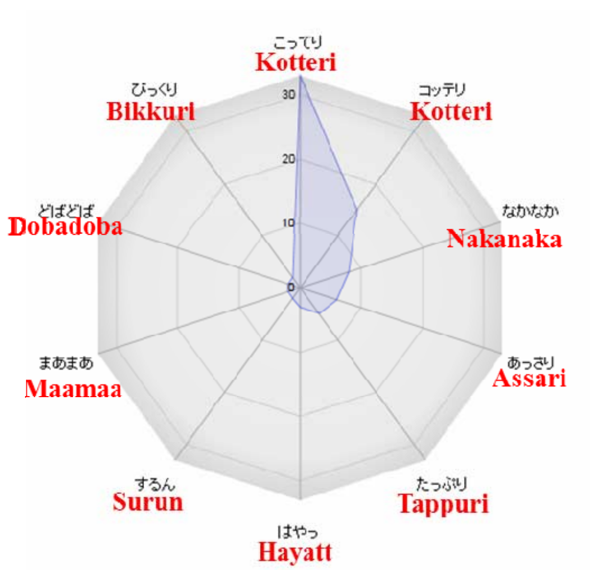

Fig. 10:Radar chart of Ramen shop A

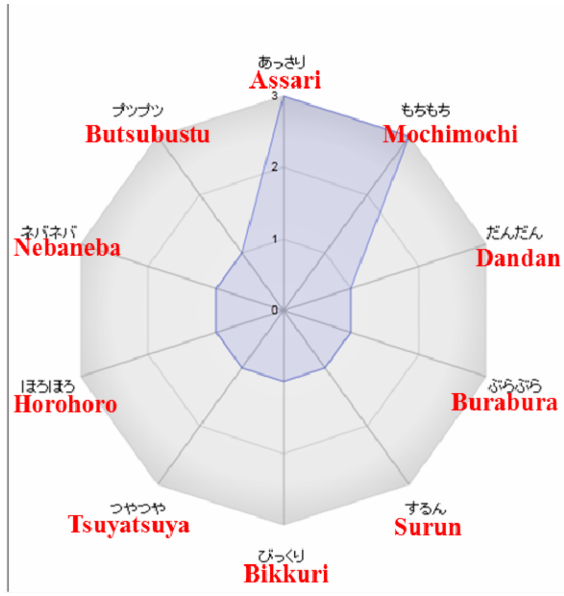

Fig. 11:Radar chart of Ramen shop B

6. REFERENCES

[1] "Tabelog," http://tabelog.com/ (accessed at 16th, Jan. 2012).

[2] "toptable," http://www.toptable.com/ (accessed at 16th, Jan. 2012).

[3] Katsuyoshi Nishinari, "Texture and Rheology in Food and Health,", Food Science and Technology Research, Vol. 15, pp.99-106, 2009.

[4] Hiromi Akiyama, Takanori Komatsu "Comprehension System of Onomatopoeias for Assisting Users' Intuitive Expressions," Human–Agent Interaction Symposium,2008.

[5] Yoshitaka Fujino, "Sports Onomatopoeia –Why do the top athletes utter voice," Shogakukan, Tokyo, 2008.

[6] Masafusa Ohashi, Ayaka Muto, Makoto Yamamoto, Masako Tamekuni, Akiko Kumida, Fumiaki Sibusawa and Yuko Ogawa, "A sense and word of "Delicious" and a generation of food texture," BMTF publisher, Tokyo, 2010.

[7] Masahiro Ono, "Onomatopoeia and Mimetic word 4500 Japanese Onomatopoeia Dictionary," Shogakukan, Tokyo, 2007.

[8] "Onomatopedia – goo Dictionaly," http://dictionary.goo.ne.jp/onomatopedia/ (accessed at 16th, Jan. 2012).

[9] Chisato Asaga, Mukarramah Yusuf, Chiemi Watanabe, "Onomatopedia: Onomatopoeia Online Example Dictionary System Extracted from Data on the Web", The 10th Asia Pacific Web Conference (APweb), 8b-1, 2008.

[10] Yuzu Uchida, Junpei Watanabe, Jun Yoneyama and Kenji Araki, "Affect Analysis of Onomatopoeia Sentences Extracted from Blog Entries", The 17th association for Natural Language Processing conference, pp.332-335, 2011.

[11] Kanwipa Lertsumruaypun, Chiemi Watanabe and Satoshi Nakamura, "Onomatoperori : Recipe Recommendation System

Using Onomatopoeic Words," IEIEC 2nd technical committee document , 2010.

[12] "COOKPAD," http://cookpad.com/ (accessed at 16th, Jan.2012)

[13] Sachiko Hirata, Daiki Sawai, Hiroki Fuji and Shinichi Kita, "Gathering System of Onomatopoeia Description Data using Twitter," The 25th Annual Conference of the Japanese Society for Artificial Intelligence, 2011.

[14] Hai-Jun Zhang, Wei-Min Pan, Shu-Min Shi and Chao-Yong Zhu, "Application of Conditional Random Fields model in unknown words identification," International Conference on Machine Learning and Cybernetics (ICMLC), pp.11-14, 2010.

[15] Hao Chen, Bo Xiao and ZhiQing Lin, "Unknown Word Recognition Based on Maximal Cliques," Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2011 International Conference, pp.10-12, 2011.

[16] Likun Qiu, Changjian Hu and Kai Zhao, "A Method for Automatic POS Guessing of Chinese Unknown Words," COLING '08 Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1, 2008.

[17] Meni Adler, Yoav Goldberg, David Gabay and Michael Elhadad, "Unsupervised Lexicon-Based Resolution of Unknown Words for Full Morphological Analysis," Proceedings of ACL-08: HLT, pp.728–736, 2008.

[18] Karo Moilanen and Stephen Pulman, "The Good, the Bad, and the Unknown: Morphosyllabic Sentiment Tagging of Unseen Words," HLT-Short '08 Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, 2008.

[19] Iwasaki Noriko, Vinson David P. and Vigliocco, Gabriella, "What do English speakers know about gera-gera and yota-yota?: A cross-linguistic investigation of mimetic words for laughing and walking.," Japanese-Language Education around the Globe, 17 . pp. 53-78. 2007.

[20] "Juman," Graduate School of Informatics, Kyoto University Kurohacshi and Kawahara Laboratory, http://nlp.ist.i.kyoto-u.ac.jp/ (accessed at 16th, Jan.2012).

[21] Akio Suwa, "Phenomenon Condition Required Formation of Onomatopoeia," Phonetic Society of Japan, 1995.

[22] Katsuyoshi Nishinari, Fumiyo Hayakawa, Chong-Fei Xia, L. O. N. G. Huang, Jean-Francois Meullenet, Jean-Marc Sieffermann, "Comparative study of texture terms: English, French, Japanese and Chinese," Journal of Texture Studies, 39(5), pp.530-568, 2008.

[23] Susan Strauss, "The linguistic aestheticization of food: a cross-cultural look at food commercials in Japan, Korea, and the United States," Journal of Pragmatics, 37, pp.1427-1455, 2005.

[24] Gerard Salton, C. Buckley, "Term-weighting approaches in automatic text retrieval," Information Processing & Management 24 (5), pp.513–523, 1988.