Extraction of Onomatopoeia Used for Foods from Food Reviews and Its Application to Restaurant Search

Ayumi Kato
Graduate School of Interdisciplinary Information Studies, the Univ. of Tokyo
7-3-1 Hongou, Bunkyo-ku, Tokyo, 113-0033 Japan
+81 90-7277-4295
kato@ics.t.u-tokyo.ac.jp

Yusuke Fukazawa
School of Engineering, the Univ. of Tokyo
7-3-1 Hongou, Bunkyo-ku, Tokyo, 113-0033 Japan
+81 90-5346-0342
fukazawayuu@nttdocomo.co.jp

Tomomasa Sato
Graduate School of Information Science and Technology, the Univ. of Tokyo
7-3-1 Hongou, Bunkyo-ku, Tokyo, 113-0033 Japan
+81 3-5841-6441
tomomasasato@jcom.home.ne.jp

Taketoshi Mori
Graduate School of Medicine, the Univ. of Tokyo
7-3-1 Hongou, Bunkyo-ku, Tokyo, 113-0033 Japan
+81 3-5841-360
tmoriics-tky@umin.ac.jp

ABSTRACT

Onomatopoeia is widely used in food reviews about food or restaurants. In this paper, we propose and evaluate a method to extract onomatopoeia including unknown ones automatically from food reviews sites. From the evaluation result, we found that we can extract onomatopoeia for specific foods with more than 46 % precision; we find 18 unknown onomatopoeia, i.e. not registered in an existing onomatopoeia dictionary, in 62 extracted onomatopoeia. In addition, we propose a system that can present the user with a list of onomatopoeia specific to a restaurant she is interested in. The evaluation results indicate that an intuitive restaurant search can be done via a list of onomatopoeia, and that they are helpful for selecting food or restaurants.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services, I.2.7 [**Natural Language Processing**]: Language parsing and understanding

General Term

Algorithms, Design, Human Factors,

Keywords

Onomatopoeia, TFIDF, Foods and restaurants, Word of mouth, Food review

1. INTRODUCTION

1.1 Background

Recently, user generated media (UGM) such as blogs, Social Network Service and Twitter are widely used in many kinds of domains. In these sites, food-related articles are one of the main topics. Users tend to present their thoughts or reviews about what and where they ate dinner or lunch. There are also user-generated sites specific to restaurants such as “Tabelog [1]” in Japan and “Toptable [2]” in Europe and USA. A screen capture of a restaurant summary on the Tabelog site is shown in Fig. 1.

Fig. 1:Restaurant rating summary of Tabelog. From left to right, ratings of taste, service, atmosphere, Cost Performance (CP) and alcohol are shown.

In Tabelog, the users can read, write and share their thoughts and reviews freely about what and where they ate. The reviews totaled 2,908,529 as of Oct. 16, 2011, and this information is useful in selecting food or restaurants. The reviews often contain onomatopoeia to describe the taste, texture and look as indicated in [3]. Fig. 2. shows set of review sentences that include onomatopoeia inside. Onomatopoeia can express the status of objects or atmosphere more directly and delicately with better realism than is possible with general words[4]. Fujino explained that when people use onomatopoeia, it means they have an image that cannot be expressed in any direct way [5]. In addition, Ohashi said that onomatopoeia is increasing dramatically these days. For example, to express the feeling of deliciousness, existing expressions are “hagotae” or “hazawari”. However, we use onomatopoeia such as “Mochimochi”(chewy texture) or “Sakusaku”(crunching sound) instead [6].

According to a dictionary of Japanese Onomatopoeia [7], there are 4,500 Japanese onomatopoeias. However, based on our observation on UGM, users create new onomatopoeia according to what they feel. In this paper, these onomatopoeias are called “unknown onomatopoeia”. For example, we found that there are onomatopoeias such as “Mochumochu”, “Mocchuri” and “Mocchumochu”, which are not defined in the existing dictionary.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.
WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

These onomatopoeia are derived from base onomatopoeia "Mochimochi"(chewy texture). We focus on the base onomatopoeia to extract unknown onomatopoeia.

This paper proposes to extract both known and unknown onomatopoeia for foods from the reviews on UGM sites specific to food and a method that uses these onomatopoeia to search for restaurants.

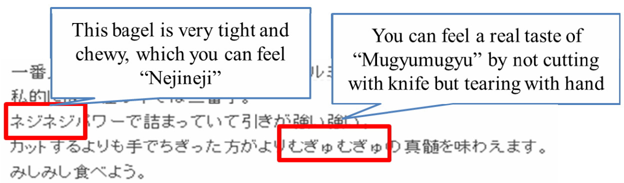

Fig. 2 Set of review sentences that include onomatopoeia inside, which are collected from the bagel community of a Japanese social networking site Mixi!. Both "Mugyumugyu" and "Nejineji" are frequently used onomatopoeias about bagels, which make the user imagine that the bagel is very tight and chewy.

1.2 Related works

In terms of extracting onomatopoeia from the web, Asaga proposed a method to extract automatically from the Web corpus sentences that use onomatopoeia [9]. However, the number of onomatopoeias covered by the dictionary is only 80, which is not enough to cover all Japanese onomatopoeia. Uchida collected snippets of blog posts that actually used onomatopoeia [10], and made a database that links emotions to onomatopoeia. This research focused only on 299 existing onomatopoeia and did not deal with unknown words.

With regard to using onomatopoeia for food applications, Asaga et al. proposed "Onomatoperori": Food Recommendation System using onomatopoeia [11], which enables users to find a recipe for cooking from among recipes posted on the "Cookpad" site using onomatopoeia [12]. However, as in the above research, the onomatopoeia used in the research was fixed in advance, and unknown onomatopoeia were not considered. Hirata et al. studied how onomatopoeia is used by Twitter users [13]. In these studies, Hirata asked those who followed the account of Hirata to answer the question "what do you think is the onomatopoeia that expresses Japanese noodles?". In response to the question, the follwers answer onomatopoeia "Tduru-tsuru" that expresses Japanese noodles. This was not cost-free to the followers and coverage of onomatopoeias was not assured if number of the active followers are few.

Apart from onomatopoeias, various methods have been proposed to extract unknown terms. Zhang proposed a method to identify unknown Chinese terms based on the Conditional Random Field model [14]. They construct a probabilistic field that expresses what word would come after another word. If a word appears with unexpected probability, it is regarded as an unknown word. Chen et al. proposed a method to extract unknown Chinese words [15]. They use the N-gram method to extract candidate words, and they identify unknown words by using mutual information about the occurrence of candidate words. Qiu proposed a model to annotate appropriate parts-of-speech (POS) to unknown Chinese words [16]. It consists of two parts; first they use a supervised machine-learning method to predict the POS of an unknown onomatopoeia. For low credibility words, they use context words (words surrounding the target word). Adler et al. introduce an algorithm to acquire unknown words that is language independent. It disambiguates unknown words after being trained over the known words observed in the corpus and the distribution of the unknown words in known tag contexts. [17]. Morlanen et al. proposed a method to judge whether unknown sentiment words trigger "Good" or "Bad" emotion in people [18]. They proposed a method that uses five classifiers to assign prior sentiment polarities to unknown words based on known sentiment carriers.

Existing methods try to assign POS or polarities (such as "Good" or "Bad") to unknown words found during morphological analysis. However, the extraction of unknown onomatopoeias, which range over several POSs (adverb and nouns) and are created on a daily basis has not been proposed.

1.3 The purpose of this research

In this paper, we propose a method to automatically extract both known and unknown onomatopoeia that characterize different food categories. In addition, we propose a system that can help the user understand a restaurant well by presenting a list of onomatopoeia that characterize the restaurant.

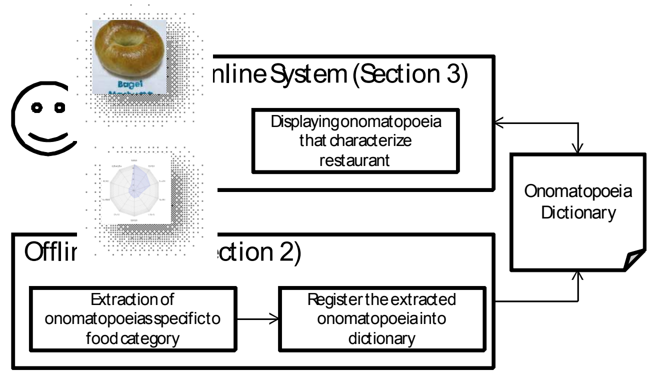

Fig. 3:The flow of whole structure

1.4 Approach

We show the overall flow of the proposed method in Fig. 3. It consists of two modules:

1. Offline system: Extract onomatopoeia specific to food categories and register them into dictionary

2. Online system: Display onomatopoeia that characterize a restaurant in response to user's request.

The first module focuses on the fact that most unknown onomatopoeia are derived from the repetition of a specific word, which we call the root onomatopoeia, the smallest set of characters. For example, "Fuwa" is the root onomatopoeia of "Fuwafuwa"(softly). Therefore, first we extracted the root onomatopoeia from existing onomatopoeia and generate unknown onomatopoeia by applying derivative patterns. For example, we can extract the root onomatopoeia "Fuwa" from the existing onomatopoeia "Fuwafuwa", and use it to generate new unknown onomatopoeia such as "Fuwwafuwa".

The second module shows the user a chart consisting of extracted onomatopoeia and their frequencies to make searches for food and restaurant reviews both easy and intuitive. The charts allow the user to understand the characteristics of a restaurant at a glance without reading the reviews in detail.

For each module, we evaluate following points:

In order to evaluate the first module, we compare the number of unknown onomatopoeia extracted by the proposed module to that present in an existing onomatopoeia corpus. In addition, we evaluate the precision of the unknown onomatopoeia extracted, which means whether the proposed module can extract known and unknown onomatopoeia that are suitable for specific food categories, by conducting user surveys.

In order to evaluate the second module, we verify whether the proposed onomatopoeia are useful to users who are looking for restaurants.

1.5 International applicability

The target language of the proposed method is Japanese but many foreign languages also have many onomatopoeias. We believe that the system will also support foreigners in Japan. When we recommend traditional and original foods, which are unknown to the foreigners, they cannot imagine the food taste from the name of the food. By showing onomatopoeias which express the food, the foreigner will be able to understand the food's taste.

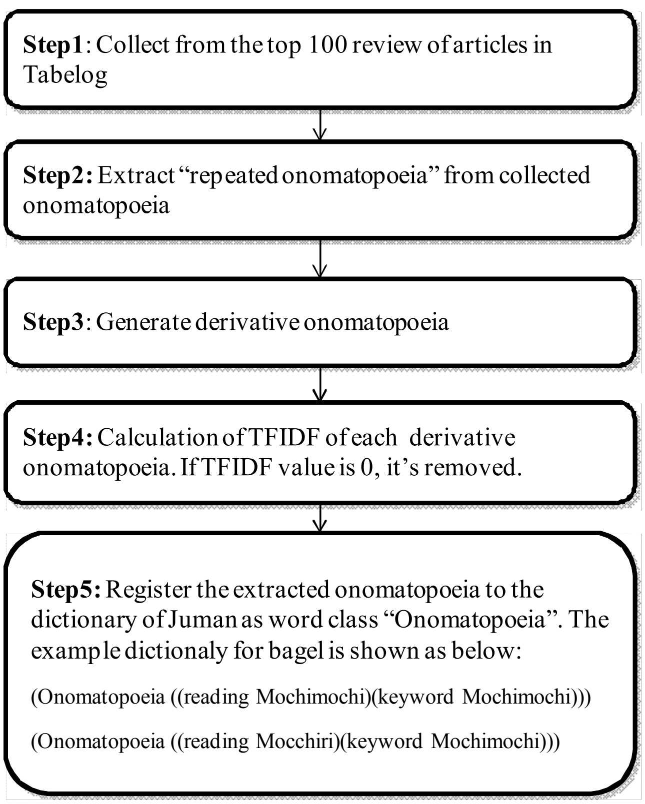

Fig. 4: **The detailed procedure of offline system of Fig. 3. This process extracts onomatopoeias specific to food category, and registers the extracted onomatopoeia into dictionary**

Iwasaki conducted an experiment that investigated whether foreigners can understand the meaning of onomatopoeia from their sound [19]. They asked both native Japanese speakers and English speakers to listen to Japanese mimetic words for laughing. As a result, English speakers construed many of the features of laughing in a similar manner as Japanese speakers. From this result, we assume that foreigners might be able to understand Japanese mimetic words and onomatopoeia from their sound. Although laughter is different from taste, some food onomatopoeias are derived from the sound that occurs when the food is eaten, therefore their conclusion has some applicability to our method. Nishinari et al. compare the texture terms of foods in different languages; English, French, Japanese and Chinese languages [22].They confirmed that, after discussion, people agree on the fundamental features common to the different languages. Strauss presents a cross-cultural/cross-linguistic analysis of taste terms in food commercials from Japan, Korea, and the United States [23]. By the result of much of the discussion, they noted nuanced commonalities between the taste terms of different languages. From the above discussion, some of the Japanese onomatopoeias can be understood from their sound, nuance or other features.

Section 2 describes the first module; it extracts unknown onomatopoeia and builds a dictionary for onomatopoeia. Section 3 introduces the second module; it helps the user understand restaurants easily by using unknown onomatopoeia. Section 4 details the experiments conducted and we conclude the paper in Section 5.

2. Extraction of onomatopoeia specific to food

The flow of first module is shown in Fig. 4 with the example of "Mochimochi"(chewy texture). As can be seen, the process consists of the 5 steps described below.

Step1. Acquisition of food reviews

In this Step, we collect a set of reviews from the restaurant review site Tabelog. The system automatically collects reviews from articles on the 100 most popular restaurants for each food category. Tabelog sets several food categories such as curry, Japanese food, and Chinese food. These categories are open to the public at Tabelog site. We use the search functionality of Tabelog that allows the user to search restaurant reviews for a specific category.

Step2. Acquisition of onomatopoeia

We extract the "repeated onomatopoeia" from the reviews. For example "Fuwafuwa"(softly) consists of the repetition of the root onomatopoeia "fuwa". In order to extract repeated onomatopoeia, first we morphologically analyze the texts and collect words that are judged as either adverbs or nouns. Here, we use Juman [20] for Japanese morphological analysis. Juman outputs words with unknown POS that consist of repeated characters such as "FuwaFuwa" as adverbs. This means that if we collect nouns and adverbs, Juman outputs both known and unknown "repeated onomatopoeia".

Step3. Generation of derivative onomatopoeia

According to Suwa, in many cases, Japanese onomatopoeias are derived from root onomatopoeia [21]. Following Suwa, we use the six patterns shown in Table 1 to generate unknown onomatopoeia candidates, "derivative onomatopoeias".

Table 1:Derivative patterns with example of "Fuwa"

Derivative patterns	Example of "Fuwa"
Root + Root	Fuwafuwa
Root + "Ri"	Fuwari
Root + "N"	Fuwan
Root +double consonant	Fuwatt
Root +double consonant+ "Ri	Fuwwari
Root +double consonant+ Root	Fuwwafuwa

Step4. Calculation of TFIDF

In order to extract the frequently used onomatopoeia for specific food categories, we use the TFIDF method to evaluate each derivative onomatopoeia[24]. Our proposal uses two methods of calculating TFIDF. Here, $W_{i,j}$ represents the weight of onomatopoeia i for food category j. $W_{i,j}$ is calculated by equation (1) or (2).

$$\boldsymbol{W}_{i,j} = \frac{\boldsymbol{tf}_{i,j}}{\boldsymbol{df}_i} \tag{1}$$

$$\boldsymbol{W}_{i,j} = \boldsymbol{tf}_{i,j} \times \log\left(\frac{N}{df_i}\right) \tag{2}$$

Here, $tf_{i,j}$ is the total number of restaurants in specific food category j whose reviews use onomatopoeia i, df_i is the total number of restaurants whose reviews use onomatopoeia i, and N is the number of restaurants (in Tabelog). We threshold $W_{i,j}$ and extract the onomatopoeias whose $W_{i,j}$ exceed the threshold.

The derivative onomatopoeias include noise in that some have no accepted usage. For example, "Gapegape" has the highest TFIDF value for the food category "Ramen". However, "Shikoshiko" (chewy texture for noodles), which has a lower TFIDF value, is frequently used in the context of Ramen. As seen from Table 2, "Gapegape" was used in only one review.

Table 2: **Example for onomatopoeia as noise**

Onomatopoeia	TF	DF	TFIDF
Shikoshiko	4452	7314	**0.61**
Gapegape	1	1	**1**

We identify and eliminate noisy onomatopoeias as those that have extremely low tf and df values but high TFIDF. We determined the threshold for each food category heuristically by observing the top onomatopoeias. For example, in the food category ramen, set the tf threshold to 600. We evaluate the effectiveness of eliminating onomatopoeias of low tf by comparing the method without elimination (method B) with that with elimination (method D).

Step5. Registration in Juman

We register the onomatopoeia extracted by Step4 in a food-specific onomatopoeia dictionary. Each food category has its own dictionary. Concretely, we make a new part of word class, "onomatopoeia" in Juman and use it to register the found onomatopoeia. Therefore, Juman can output onomatopoeia as word class "onomatopoeia".

3. Displaying onomatopoeias that characterize restaurants

Fig. 6 shows the process of displaying onomatopoeias that characterize restaurants. As shown in the figure, the system consists of two parts: the client implemented as a Firefox add-on and the server implemented as a Tomcat application and a C-language application. We describe each step below.

Step1. Acquire URL

We get the URL of the page being browsed by the Firefox add-on and send it to the server. Concretely, a contextual menu is displayed by right click on the top page. By choosing the menu item "Show Onomatopoeia" from the contextual menu, the URL of the top page is sent to the server (Fig. 5).

Fig. 5: Firefox Add-on for getting URL. This a screen capture of the contextual menu and the web page acquired after selecting one of the contextual menu items. The sentences in the red circle mean "Show the onomatopoeia (a code of letter is Shift_JIS or UTF_8)".

Step2. Acquisition of HTML source

The server takes the URL sent by the client, acquires the HTML source, and preserves it as a text file. We eliminate machine readable parts in the HTML source such as tags to minimize noise in natural language processing.

Step3. Extraction of onomatopoeia for specific restaurant

We extract onomatopoeias from the text by morphological analysis against the updated Juman. Note that the server chooses the dictionary whose food category matches that of the restaurant on the top of the page.

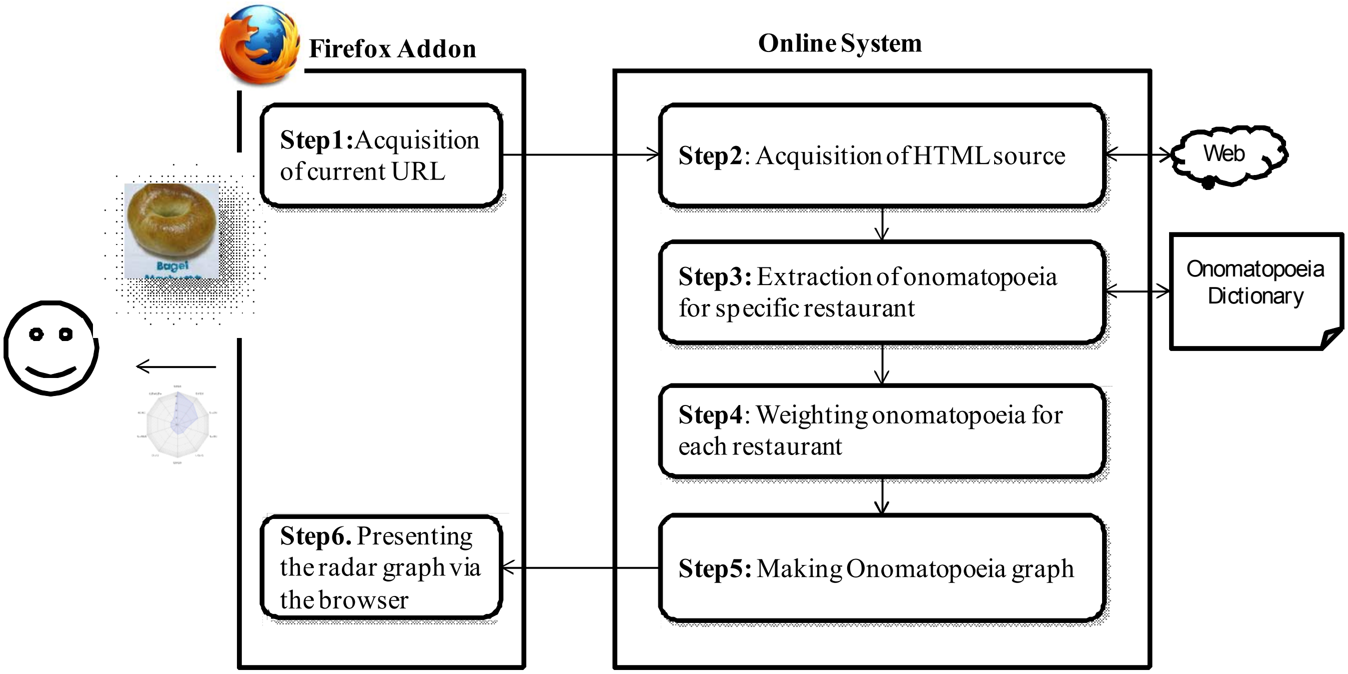

Fig. 6:The detailed procedure of online system of Fig. 3. This process displays onomatopoeias that characterize restaurants.

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.

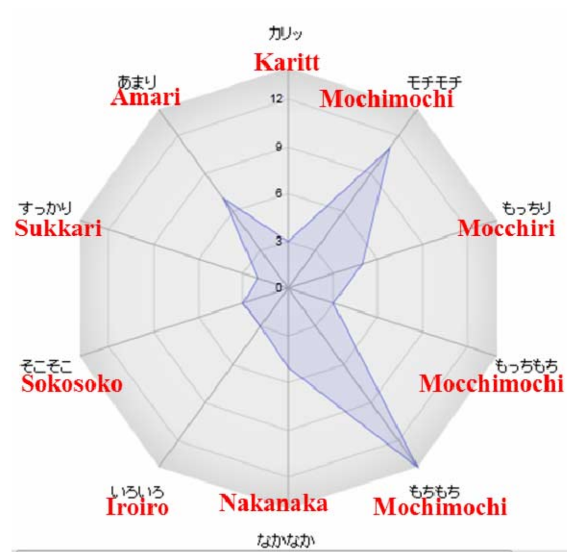

Fig. 7:Radar chart of the top 10 onomatopoeias sorted by frequency

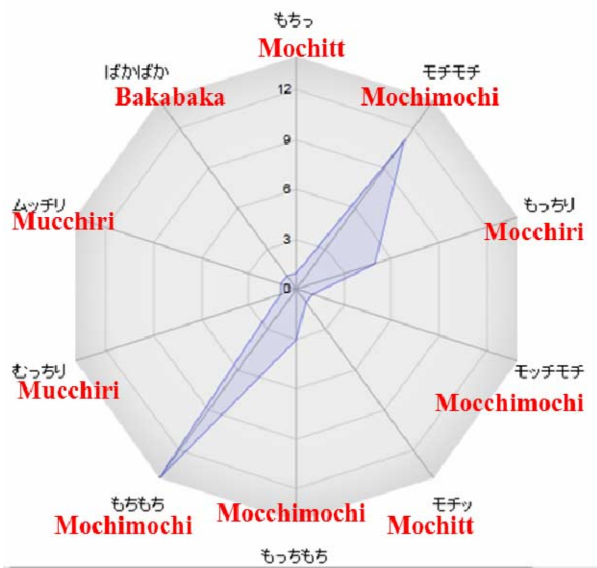

Fig. 8: Radar chart of the top 10 onomatopoeias sorted by TFIDF

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

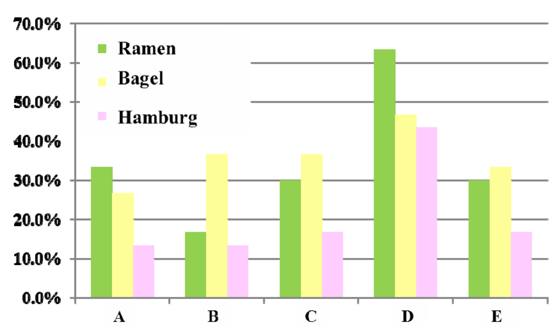

Fig. 9: Questionnaire derived onomatopoeia results(%)

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.

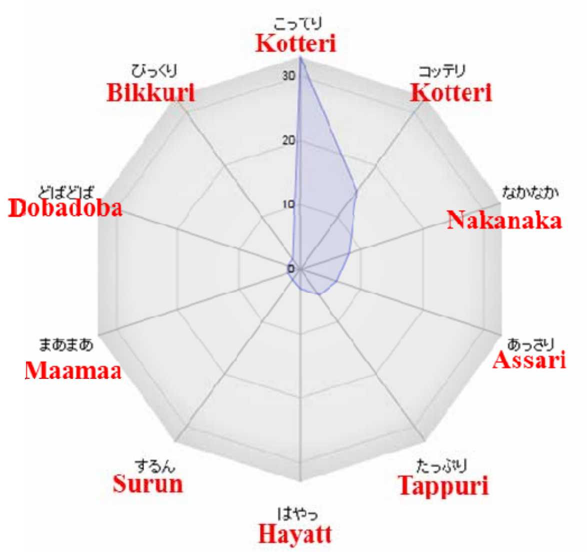

Fig. 10:Radar chart of Ramen shop A

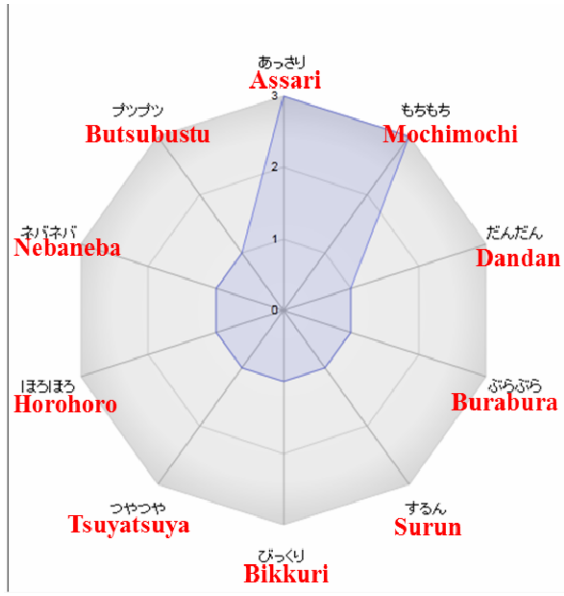

Fig. 11:Radar chart of Ramen shop B

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.