2016 19th International Conference on Network-Based Information Systems

Clustering for Similar Recipes in User-generated Recipe Sites based on Main Ingredients and Main Seasoning

Akiyo Nadamoto
Konan University
Okamoto 8-9-1
Higashinada-ku Kobe, Japan
Email: nadamoto@konan-u.ac.jp

Shunsuke Hanai
Konan University
Okamoto 8-9-1
Higashinada-ku Kobe, Japan
Email: m1424007@center.konan-u.ac.jp

Hidetsugu Nanba
Hiroshima City University
Ozukahigashi 3-4-1
Hiroshima 731-3194 Japan
Email: nanba@hiroshima-cu.ac.jp

***Abstract*—Today, many people are constantly using user-generated recipe sites when they prepare their meals. Users search for recipes for their meals from the recipe sites. However, when they search for a recipe using an ingredient name, numerous similar pages which are coincidentally similar or which have been plagiarized are found for them. Moreover, a user who searches for a recipe usually does not select a high-ranking recipe from the search results, reacting better than one might with usual web searches. Given many similar recipes included in the search results, it appears to be difficult for them to compare multiple recipes. Actually, when users compare similar recipes, they must better understand the different points of similar recipes. That need for comparison imposes a great burden on users. Therefore, a system classifying the results of user searches according to similar pages in real time would be beneficial for users. In this paper, we propose a clustering method for user-generated recipe sites based on page structure and main ingredient and main seasoning of the food. It provides a means of classifying the user search results according to similar pages. We conducted an experiment to measure the benefits of our proposed method. The experiment results presents the benefits of our proposed method, which classifies similar recipes based on the main ingredients and main seasonings.**

I. Introduction

Today, many people are constantly using user-generated and commercially generated recipe sites, such as Food.com[1](U.S.), Mis Recetas[2](Hispanic), Beitai Chufang[3](China) and Cookpad[4](Japan), when they prepare their meals. Users search for recipes for their meals from the recipe sites. When a user searches for a recipe, the user poses queries of two types, typically incorporating food names such as beef stew or lasagna, and ingredient names such as chicken, cabbage, or onions. Maruha–Nichiro Holdings investigated which query is more used when users search recipes posted on recipe sites[1]. Results show that ingredient keywords are used more often than food name keywords. In fact, ingredient names account for 75% of all keywords. Therefore, when users use recipe sites, they input an ingredient name as a query. However, when they search for a recipe using an ingredient name, they are often deluged by similar recipes which use the same ingredients. Consequently, they become confused. Such similar recipes impede a user's recipe searches. For instance, when a user inputs "Chicken and Onion" in Cookpad, a famous Japanese user-generated recipe site, the search results extend to more than 46,000 pages. Numerous similar pages are found. The similar pages which have become coincidentally similar or which are plagiarized. The resultant " information overload" confuses users.

Moreover, a user who searches for a recipe usually does not select a high-ranking recipe from the search results, reacting better than one might with usual web searches[2]. Data show that users typically compare multiple recipes. Given many similar recipes included in the search results, it would be difficult to compare multiple recipes. After all, when users compare similar recipes, they must better understand different aspects of similar recipes. That need for comparison imposes a great burden on users. A system classifying the results of user searches according to similar pages in real time would be beneficial for users.

Therefore, we propose a method that classifies user search results according to similar pages based on the page structure and the types of important words. Clustering tools of many kinds have been proposed, but it is difficult to classify similar recipes using existing clustering tools because recipe sites have a unique page structure that includes the dish title, with ingredients, directions (preparation instructions), and comments. These passages of pages differ in their roles, importance, and meaning. We have already proposed clustering methods for some recipe sites[3][4]. Based on results of our earlier studies[3], we demonstrated the following four points for the clustering of similar recipes.

1) No image is necessary to judge recipe similarity.
2) Food names, ingredients, seasonings, and cooking methods in a title are important words. Especially, users care about food names and cooking methods in the title when judging similar recipes.
3) Main ingredients and the main seasoning in an ingredient list are important.
4) Sizzle words in a title are not necessary to judge the recipe similarity.

[1]Food.com http://www.food.com/
[2]Mis Recetas http://www.misrecetas.com/
[3]Beitaichufang http://www.beitaichufang.com/
[4]Cookpad http://cookpad.com/

2157-0426/16 $31.00 © 2016 IEEE
DOI 10.1109/NBiS.2016.49

Our previous proposed methods are twice clustering, which calculates by food name, and which calculates by ingredient and seasoning name. In our previous methods, we consider (1) and (2) above, but we did not consider (3). Then, as described herein, we first describe how to extract the main ingredient and main seasoning. Next we propose modification of our clustering method, particularly considering the main ingredient and main seasoning. We define that main ingredient as the most important ingredient in the food, and the main seasoning as determining the food taste.

II. RELATED WORK

The purpose of our research is extraction of similar recipes. For the purpose of our research, we propose method is twice clustering, with specific examination of the page structure and types of important words. This section describes studies on recipe recommendation system and extraction of recipe information as related work.

Recipe Recommendation System

Along with the growth of recipe-sharing services, many studies have examined recipe recommendation processes. Wagner et al. [5] proposed a system that tracks a user's cooking activities with sensors in kitchen utensils and which recommends healthy recipes that might increase the user's cooking competence. Svensson et al. [6] applied their idea of social navigation for recipe recommendation. Specifically, they assigned users to groups based on their explicit preference information such as ingredients, fat level, and time to cook. Geleijnse et al. [7] designed a prototype of a personalized recipe recommendation system, which suggests recipes to users based on their past food selections and nutrition intake. Freyne et al. [8] presented the suitability of recipe recommendation algorithms based on food preferences. Ueda et al. [9] proposed a personalized recipe recommendation method based on user preferences. Their method estimated user preferences from a user's past actions, such as their recipe browsing and cooking history. Karikome et al. [10] proposed a system that helps users plan nutritionally balanced menus and which visualizes their dietary habits. Lawrence et al. [11] proposed a product recommendation system for grocery shopping using collaborative filtering to rate products based on the user's prior purchase behavior. Harvey et al. [12] explained factors of selecting recipes by assessing the results of long-term research to recommend recipes that match a user's preferences. However, we particularly examine recipe similarity, not personalization. We aimed to be helpful for user's search recipes by extracting spam recipes.

Teng et al. [13] proposed ingredient recommendation systems using ingredient networks. They constructed networks of two types to capture the relations among ingredients: ingredient complements and substitute ingredients. Pinxteren et al. [14] identified important features and extracted them from the recipe texts. They calculated the degree of similarity among recipes, and changed to healthy recipes. Shidochi et al. [15] proposed a method to identify replaceable ingredients by matching the cooking actions that correspond to ingredient names from recipe texts. Forbes et al.[16] apply a matrix factorization method to recipe recommendation. Experimental results demonstrated that the algorithm not only improves the recommendation accuracy; it is also useful for swapping ingredients and creating new recipes. These studies has focused on ingredients. However, these studies specifically examined substitute ingredients and substitute recipe recommendation systems: the extraction of similar recipes is not the goal.

Extraction of Recipe Information

Kuo et al. [17] proposed a method for constructing a recipe graph to capture the co-occurrence relations among recipes of social recipe sites. Wang et al. [18] created a graph-based model of recipes using ingredients and cooking processes, showing that a similar subgraph exists between two recipes. Li et al. [19] proposed a method using a graph-based model of recipes to extract suitable recipes to match a user's preference. Yamakata et al. [20] proposed a method that extracts typical cooking processes from multiple recipes by creating flow graph of recipes. Our research differs in that we extract page structure and type of important words without using a graph-based model.

Mori et al. [21] proposed the application of a machine-learning approach to a recipe text processing problems, aiming at converting a recipe text to a cooking process. Chung et al. [22] proposed an efficient method that finds related words in a recipe domain based on data structures using user-generated recipe data. They found that people usually write the main ingredient in the first position of ingredient lists of each recipe. For that reason, that ingredient is strongly related to the categories in which recipes belong. However, these studies do not examine recipe recommendation systems.

III. EXTRACTION OF MAIN INGREDIENT AND MAIN SEASONING

A. Extraction of main ingredient

As described in this paper, we regard the main ingredient as the most important ingredient of the food, which means that we cannot change the main ingredient when preparing the food. Chung[22] describes almost all main ingredients as written in the first position on the ingredient list on the recipe page. Their proposed method, called FI, is not capable of extracting multiple main ingredients. However, some dishes do have multiple main ingredients, such as Braised Meat and Potatoes (Niku-jaga). We consider not only that the position of the ingredient list is important, but also that the quantity of the ingredient is also important. Therefore, we propose main ingredient degree MI_i, which incorporates the position of the ingredient list and the ingredient quantity.

$$MI_i = PA_i * \frac{q_i}{(\sum_{j=1}^{n} q_j)} \tag{1}$$

$$PA_i = t_i + \frac{|L| - (m+1)}{|L|)} \tag{2}$$

$$t_i = \begin{cases} 1 & \text{if } i \text{ is in the title} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

In that equation, q_i stands for a quantity (g) of the ingredient in the list. n denotes the number of recipe ingredients Also, $|L|$ signifies the number of ingredients in an ingredient list on a recipe page. m represents the position of the ingredient in an ingredient list. When MI_i is greater than a threshold value, ingredient i is inferred as the main ingredient.

B. Extraction of main seasoning

We investigate the location of the main seasoning using a recipe website with about 35,000 recipes. Results show that the seasoning in a title becomes the main seasoning. The ingredient list rank can imply a main seasoning, but the seasoning quantity is important to infer the main seasoning. Therefore, we propose the main seasoning degree MS_k of a seasoning k as follows:

$$MS_k = t_k + \frac{q_k}{(\sum_{l=1}^{n} q_l)} \tag{4}$$

$$t_k = \begin{cases} 1 & \text{if } i \text{ is in the title} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Therein, q_k signifies a quantity (g) of seasoning in the list and n stands for the number of seasonings included in the recipe When MS_k is greater than some threshold value, seasoning k is inferred as the main seasoning.

IV. Similar Recipe Clustering

We propose similar recipe clustering methods to extract similar recipes. We specifically examine four points of results of our experiment in our previous research[3]. We propose a similar recipe clustering method.

- We do not use images.
- We classify twice based on the recipe page structure and important word types.
  - We first classify recipe pages based on the food name and cooking method in the title.
  - We extract the main ingredient and the main seasoning from each recipe page.
  - We next classify the results of first clustering based on ingredient and seasoning which are considering feature value.
- We ignore sizzle words on recipe pages.

A. First Clustering Based on Food Names and Cooking Methods in the Title

Our experiment demonstrated that the food name and cooking method in the title is the most important element to judge similar recipes. We first classify recipe pages based on food names and cooking methods in the title.

The target data are search results of recipe pages using user input multiple ingredients such as "pork and onion," and "chicken and tomato." We extract titles and extract food names and cooking methods using our food database, which includes food names and cooking methods. When we classify recipe pages, we use Repeated Bisection[23] which is a method used for bayon[24] and CLUTO[25]; it is a kind of K-means method. Repeated Bisection is suitable for short sentences. We therefore divide recipe pages into passages to classify them. It therefore becomes a set of short sentences. After the first clustering, the recipe pages are clustered by food names or cooking methods included in the title area.

B. Second Clustering Based on Ingredients and Seasoning

First clustering uses only the food name and cooking method. There are many different meals but they are in the same cluster. For example, one cluster is related to curry, which is a food name. The elements of the cluster include "Vegetable curry," "Tomato curry," "Pork curry," "Soft pork curry," "Chicken curry," and "Chicken spicy curry." They are not similar recipes. Next, we classify them again based on ingredients and cooking methods.

Our experiment results show that words in a title are more important than other words in a passage. Calculating the weight of words is necessary based on their position of appearance on a page. Furthermore, a surprising degree of ingredients in a food name is important because unusual ingredients for the food should be regarded as recipe characteristics. For example, for the recipe title "Lasagna," some ingredients are tomato, pork, onion, cheese, and tofu. In this case, tofu is an unusual lasagna ingredient. Therefore, that unusual ingredient becomes a recipe characteristic. Then we calculate the surprising degree of ingredients using our surprising degree S-RF-IIF. Our previous proposed S-RF-IIF[4] does not consider the main ingredient and main seasoning. For the purposes of this paper, we regard the main ingredient and main seasoning in S-RF-IIF and we call new S-RF-IIF as S-RF-IIF2. Expression of new S-RF-IIF2 is the following:

$$S\_RF\_IIF2^i_m = \alpha \log \frac{|R_m|}{|R^i_{t,m}|} + \beta \log \frac{|R_m|}{|R^i_{o,m}|} + \gamma \tag{6}$$

$$\gamma = \begin{cases} 0.5 & \text{if } i \text{ is the main ingredient or seasoning} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The given i is an ingredient name and m is a food name. In addition, R_m represents the number of recipes of food m, $R^i_{t,m}$ denotes the number of recipes of food m, which includes ingredient i in the title of food m. $R^i_{o,m}$ denotes the number of recipes of food m, which includes ingredient i in passages except for the title of the recipe page of food m. In that equation, α signifies a weight that is dependent on a term's appearance position in title, and β signifies a weight that is dependent on a term's appearance position in ingredient list. As described in this paper, we regard α as 1.0 because the title is the most important passage used to judge similar recipes.

Then we classify the results of first clustering using Repeated Bisection based on ingredients and seasonings, which are feature values calculated using S-RF-IIF2.

V. Experiments

We conducted three experiments. First, we conducted an experiment to ascertain whether Repeated Bisection is suitable for clustering of the recipe, or not, based on comparison of LDA and Repeated Bisection. Next, we measure the accuracy of our proposed method.

A. Experiment 1: Comparing Clustering Method

Our proposed method is designed to classify them twice using Repeated Bisection. We measured whether Repeated Bisection is suitable for clustering of the recipe based on

TABLE I. CONDITIONS OF EXPERIMENT 1.

Query	Number of recipe pages	N in GibbsLDA+	N in Repeated Bisection
Chicken ∩ Cabbage	8,562	250	247
Pork ∩ Lettuce	3871	200	183
Avocado ∩ Tomato	5643	250	233

comparing GibbsLDA+[27] and Repeated Bisection. When we compare two clustering tools, we want only to compare a Repeated Bisection to ascertain its suitability for clustering recipes. Then we only use first clustering methods in our proposed method.

Condition

We extract recipe pages from Cookpad using ingredients as a query. The datasets are in Table I. GibbsLDA+ and Repeated Bisection each have some parameters. In GibbsLDA+, we set α = 50/K, β = 0.1, and iteration count as 1,000. Parameter N, which is the number of topics, is shown in Table I. In Repeated Bisection, the dividing threshold is 1.0. Dividing number N of topics is also shown in Table I. We fix the number N of topics using bayon[24].

Results and Discussion

Table II presents the top three topics of each query. The results give the GibbsLDA+ almost identical values of the top three topics of each query. Actually, few words are both food names and ingredients. Therefore, we can not obtain a cluster of each food using GibbsLDA+. However, in the Repeated Bisection results, the value of top topic and other two topics differ. The Repeated Bisection can classify results into each food. The results demonstrated that Repeated Bisection is suitable for clustering the recipes. Therefore, we use Repeated Bisection to classify similar pages.

B. Experiment 2: Usefulness of the Proposed Method

We assess the usefulness of our proposed method. We extract recipe pages from Cookpad using ingredients as a query. The datasets are presented in Table III. We fix the number of clusters using bayon[24]. The cluster threshold in bayon is 1.0. The number of clusters is presented in Table III. We discuss the results as divided into first clustering and second clustering.

First Clustering Based on Food Name and Cooking Method

First, we classify recipes based on the food name and cooking method. Table IV presents examples of the results for each of the top three clusters. The results demonstrated that we can find that the clusters are divided into the food name or cooking method. However, we can find that a cluster divided into food or cooking methods includes various main ingredients or seasonings. For example, the query of "Chicken and Onion" in the first cluster includes coconut curry, green curry, and vegetable curry. In this case, the cluster is divided into the food name "curry," but the main ingredients differ; they are not similar recipes. Therefore, we must classify smaller clusters using ingredients and seasonings.

Second Clustering Based on Ingredient and Seasoning

Next, we classify the results of first clustering based on ingredients and seasonings by consideration of the appearance

TABLE II. CENTER WORD OF EACH CLUSTER

	GibbsLDA++		Repeated bisection	
Query : Chicken AND Cabbage				
Cluster1	Noodle	0.13136	Soup	0.99987
	Cutlet	0.11073	Curry	0.00754
	Ginger pork saute	0.09697	Roll	0.00660
Cluster2	Curry	0.19656	Pot-au-feu	0.99999
	Soup	0.13123	Oden	0.00386
	Noodle	0.08967	Gratin	0.00386
Cluster3	Soup	0.13484	Salisbury steak	0.99983
	Steam	0.11810	Teriyaki	0.00920
	Saute	0.10130	Paste	0.00920
Cluster4	Chinese dumplings	0.18139	Stew	0.99994
	Soup	0.18139	Borscht	0.00566
	Pot-au-feu	0.08768	Chicken meatball	0.00566
Query : Pork AND Lettuce				
Cluster1	Soup	0.10287	Saute	0.99991
	Pork dumplings	0.03541	Rice	0.00523
	Fried chicken	0.03541	Plate noodle	0.00262
Cluster2	Saute	0.11051	Deep fry	0.97670
	Soup	0.11051	Pork cutlet	0.10310
	Roll	0.09686	Grill	0.10310
Cluster3	Noodle	0.09064	Sauteed vegetable	0.99871
	Roll	0.09064	Cut	0.03318
	Nanban-style	0.07578	Japanese Omelette	0.02709
Cluster4	Salad	0.13066	Simmer	0.99927
	Chinese dumplings	0.07891	Boiled dishes	0.02498
	Sauteed vegetable	0.05304	Jasmine	0.02040
Query : Avocado AND Tomato				
Cluster1	Salad	0.10818	Pasta	0.95634
	Dressing	0.09742	Carpaccio	0.00263
	Dip	0.05436	Smoothie	0.00263
Cluster2	Dip	0.07013	Dip	0.81649
	Cheese	0.05942	Pizza	0.00828
	Gratin	0.04872	Pasta	0.06296
Cluster3	Sandwiches	0.07581	Sandwiches	0.81649
	Sauce	0.04687	Cutlet	0.01424
	Japanese style	0.03530	Salad	0.00282
Cluster4	Salad	0.27663	Salad	0.94721
	Sauce	0.05966	Hamburger	0.00472
	Gratin	0.02515	Saute	0.00472

TABLE III. CONDITION AND RESULTS OF EXPERIMENT 2.

Query	Number of recipe pages	Number of cluster in clustering
Chicken ∩ Egg Plant	2,216	68
Pork ∩ Onion	10,635	193
Tofu ∩ Onion	10,359	185
Carrot ∩ Radish	10,403	202
Tomato ∩ Cheese	13,584	164

points. Table V presents examples of the results of the top three clusters of each cluster 1 in Table IV. The results demonstrate that the same food names with different main ingredients are classified into different clusters. For example, in the first clustering at query for "Chicken and Eggplant", a different main ingredient in the same cluster was found, such as coconut curry and green curry. However, as a result of second clustering, these recipes are classified into different clusters, which are coconut curry cluster and green curry cluster. Almost all similar recipes are classified into the same cluster. We can then classify similar recipes using our proposed method.

We next discuss the precision of the results. Table VI presents the precision of each cluster 2 in Table V. Seven participants judged the results of cluster 2. We calculate the precision using the results reported by the participants. The value of precision having the greatest number of clusters is 0.553. The precision results are not so good because of the granularity of the main ingredient. For cheese as one example, there are many kinds of cheese such as mozzarella cheese,

TABLE IV. RESULTS OF FIRST CLUSTERING

Query : Chicken AND Eggplant		
Cluster 1, Dish name: Curry	Cluster 2, Dish name: Stew	Cluster 3, Dish name: Stir-fry
Recipe title	Recipe title	Recipe title
Coconuts Curry with plenty of summer vegetables	Simple! Chicken and tomato stew	Chicken and eggplant vinegar stir-fry
Green curry for	Chicken and tomato stew	Chicken and vegetable miso stir-fry
Coconuts Curry for Tropical Island	Chicken and vegetable stew with balsamic vinegar sauce	Eggplant and tomato hot stir-fry
Query : Pork AND Onion		
Cluster 1, Dish name: Curry	Cluster 2, Dish name: Roll	Cluster 3, Dish name: Meatloaf
Recipe title	Recipe title	Recipe title
Curry with plenty of summer vegetables	Plenty of veggies! Basic spring roll	Healthy! Meatloaf
Japanese-style tuna curry	Spring roll at home	Made with ground meat! Meatloaf
Curry flavor of pork and apple	Basic spring roll!!	For a party! Juicy meat loaf
Query : Tofu AND Onion		
Cluster 1, Dish name: Fry	Cluster 2, Dish name: Hamburg steak	Cluster 3, Dish name: Nabe
Recipe title	Recipe title	Recipe title
Agedashi tofu	Japanese style Tofu hamburg steak	Salt chicken ball Nabe
Simple agedashi tofu	Healthy tofu hamburg steak	Salted lemon and tomato Nabe
Deep fried tofu with radish	Tofu and chicken hamburg steak	Tounyu Nabe
Query : Carrot and Radish		
Cluster 1, Dish name: Boil	Cluster 2, Dish name: Soup	Cluster 3, Dish name: Salad
Recipe title	Recipe title	Recipe title
Boiled dry radish	Vegetable consomme soup	Radish salad with mayonnaise
Boiled Radish and Pork	Vegetable Tonyu soup	dry radish salad
Boiled Chicken and vegetable	Chinese vegetable soup	Prosciutto and vegetable rolled salad
Query : Tomato AND Cheese		
Cluster 1, Dish name: Salad	Cluster 2, Dish name: Pasta	Cluster 3, Dish name: Stir-fry
Recipe title	Recipe title	Recipe title
Avocado and tomato salad	Eggplant and cheese and tomato pasta	Tomato and eggplant stir-fry
Avocado salad	Tomato and mozzarella pasta	Tomato and eggplant and potato with cheese stir-fry
Basil flavor salad with mozzarella	Tomato and basil pasta	Tomato and cheese omelette

cream cheese, and cheddar cheese. However, our system judges all kinds of cheese as the same cheese, whereas participants judge these cheeses as different. We should consider the granularity of main ingredients.

VI. CONCLUSION

In our research, we propose a clustering method to extract similar recipes from user generated recipe sites such as Food.com, Mis Recetas, Beitai Chufang, and Cookpad. Our proposed method is twice clustering, with specific examination of the page structure and important word types. The first clustering classifies user search results based on the food name and cooking method in the title. The next clustering classifies the results of first clustering based on ingredients and seasonings. When calculating the second clustering, we use the ingredient weight based on appearance points. As presented herein, we first described how to extract main ingredients and main seasonings. Next we proposed modification of our clustering method, which considers the main ingredient and main seasoning. We define that main ingredient as the most important ingredient in the food, and the main seasoning as determination of the food taste. We conducted an experiment to measure the benefits of our proposed method. The results of experiments presented the benefits of our proposed method, which classifies similar recipes.

In the near future, we expect to consider the cooking flow in addition to the values of ingredients and seasonings to cluster similar recipes. We expect to produce a user interface for browsing with similarity clustering.

REFERENCES

[1] Investigation of cooking recipes(May, 2015), http://www.maruha-nichiro.co.jp/news_center/research/pdf/20130227_recipe_cyousa.pdf (in Japanese)

[2] Y. Sugiyama, Y. Yamakata and K. Tanaka, "Summary of similar recipe for the recipe data as a procedure information and discovery of important differences," in *the 5th Forum on Data Engineering and Information Management*, D3-5, 2013.(in Japanese)

[3] S. Hanai and A. Nadamoto, "Clustering for Similar Recipes by using cooking ingredient," TECHNICAL REPORT OF IEICE vol. 114, no. 204, DE2014-31, pp. 47-52, 2014.(in Japanese)

[4] S. Hanai, H Nanba, and A. Nadamoto, "Clustering for Closely Similar Recipes to Extract Spam Recipes in User-generated Recipe Sites," International Conference on Information Integration and Web-based Applications & Services(iiWAS2015), pp. 252-256, 2015.

[5] J. Wagner, G. Geleijnse and A. van Halteren, "Guidance and support for healthy food preparation in an augmented kitchen," in *Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation*, pp. 47-50, 2011.

[6] M. Svensson, K. Hook, J. Laaksolahti and A. Waern, "Social navigation of food recipes. ACM Trans. Comput.-Hum," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 341-348, 2001.

[7] G. Geleijnse, P. Nachtigall, van P. Kaam and L. Wijgergangs, "A personalized recipe advice system to promote healthful choices," in *Proceedings of the 16th international conference on Intelligent user interfaces*, pp. 437-438, 2011.

[8] J. Freyne and S. Berkovsky, "Intelligent Food Planning:Personalized Recipe Recommendation," in *Proceedings of the 15th international conference on Intelligent user interfaces*, pp. 2010.

[9] M. Ueda, M. Takahata, and S. Nakajima, "User' s Food Preference Extraction for Cooking Recipe Recommendation," in *Proceedings of the 2nd Workshop on Semantic Personalized Information Management Retrieval and Recommendation*, pp. 98-105, 2011.

[10] S. Karikome and A. Fujii, "A system for supporting dietary habits: planning menus and visualizing nutritional intake balance," in *Proceed-*

TABLE V. RESULTS OF SECOND CLUSTERING

Query : Chicken AND Eggplant

Cluster 1	Cluster 2	Cluster 3
Recipe title	Recipe title	Recipe title
Coconuts Curry with plenty of summer vegetables	Summer! green curry	Simple tomato curry
Coconuts Curry for Tropical Island	Basic green curry	Tomato curry
Simple Coconuts Curry	Simple green curry	Tomato and Chicken curry

Query : Pork AND Onion

Cluster 1	Cluster 2	Cluster 3
Recipe title	Recipe title	Recipe title
Thickly! Pork cartilage curry	Delicious! Pork curry	Beat the summer heat! summer vegetables curry
Low price! Pork belly and cartilage curry	In a pressure cooker! Pork curry	Plenty of summer vegetables curry
Delicious home made! Pork cartilage curry	Very easy! Pork curry in 15 minutes	Keema curry with summer vegetables

Query : Tofu AND Onion

Cluster 1	Cluster 2	Cluster 3
Recipe title	Recipe title	Recipe title
Simple Agedashi tofu	Deep fried tofu with leek	Deep fried tofu with leek miso
Mizore Agedashi tofu	Simple deep fried tofu with leek and vinegar(Ponzu)	Deep fried tofu with sesame miso
Simple!! Agedashi tofu	Deep fried tofu with leek source	Deep fried tofu with meat miso

Query : Carrot and Radish

Cluster 1	Cluster 2	Cluster 3
Recipe title	Recipe title	Recipe title
Boiled dry radish with siso	Radish and chicken stewed dish	Beef gristle stew
Simple boiled dry radish	Vegetable and chicken stewed dish	Radish and beef gristle stew
Homemade boiled dray radish	Simple radish and chicken stwed dish	Simple beef gristle stew

Query : Tomato AND Cheese

Cluster 1	Cluster 2	Cluster 3
Recipe title	Recipe title	Recipe title
Avocado and tomato with cream cheese salad	Tomato and mozzarella pasta salad	Simple Tomato and basil salad
Avocado and tomato with cheese salad	Tomato and basil pasta salad	Tomato and basil with cheese salad
Avocado and tomato with mozzarella salad	Tomato and basil with Parmigiano-Reggiano pasta salad	Tomato and basil and mozzarella salad

TABLE VI. PRECISION OF EXPERIMENT 2.

Query	Precision(%)
Chicken ∩ Egg Plant	48.4
Pork ∩ Onion	48.4
Tofu ∩ Onion	59.4
Carrot ∩ Radish	69.7
Tomato ∩ Cheese	52.4

ings of the 4th International Conference on Uniquitous Information Management and Communication, 2010.

[11] R. D. Lawrence, G. S. Almasi, V. Kotlyar, M. S. Viveros, and S. S.Duri, "Personalization of Supermarket Product Recommendations," in *Data Mining and Knowledge Discovery*, pp. 11-32, 2001.

[12] M. Harvey, B. Ludwig, and D. Elsweiler, "You are what you eat: Learning user tastes for rating prediction," in O. Kurland, M. Lewenstein, and E. Porat, editors, *String Processing and Information Retrieval*, volume 8214 of Lecture Notes in Computer Science, pp. 153-164, 2013.

[13] C. Teng, Y. Lin and L. A. Adamic, "Recipe recommendation using ingredient networks," in *Proceedings 4th International Conference on Web Science*, 2011.

[14] Y van Pinxteren, G.Geleijnse and P. Kamsteeg, "Deriving a Recipe Similarity Measure for Recommending Healthful Meals," in *Proceedings of the 16th international conference on Intelligent user interfaces*, pp. 105-114, 2011.

[15] Y. Shidochi, T. Takahashi, I. Ide and H. Murase, "Finding replaceable materials in cooking recipe texts considering characteristic cooking actions," in *Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities*, pp. 9-14, 2009.

[16] P. Forbes and M. Zhu, "Content-boosted matrix factorization for recommender systems: experiments with recipe recommendation," in *the 5th ACM conference on Recommender systems*, pp. 261-264, 2011.

[17] F. Kuo, C. Li, M. Shan and S. Lee, "Intelligent menu planning: recommending set of recipes by ingredients," in *Proceedings of the ACM multimedia 2012 workshop on Multimedia for cooking and eating activities*, pp. 1-6, 2012.

[18] L. Wang, Q. Li, N. Li, G. Dong, and Y. Yang, "Substructure similarity measurement in chinese recipes," in *the 17th International Conference on World Wide Web*, pp. 979-988, 2008.

[19] Q. Li, W. Chen, and L. Yu, "Community-based recipe recommendation and adaptation in peer-to-peer networks," in *Proceedings of the 4th International Conference on Uniquitous Information Management and Communication*, pp. 18:1-18:6, 2010.

[20] Y. Yamakata, S. Imahori, Y. Sugiyama, S. Mori, and K. Tanaka, "Feature Extraction and Summarization of Recipes using Flow Graph," in *Proceedings of the 5th International Conference on Social Informatics*, pp. 241-254, 2013.

[21] S. Mori, T. Sasada, Y. Yamakata and K. Yoshino, "A machine learning approach to recipe text processing," in *Proceedings of Cooking with Computer workshop*, pp. 1-6, 2012.

[22] Y. Chung, "Finding Food Entity Relationships using User-generated Data in Recipe Service," in *the 21st ACM international conference on Information and knowledge management*, pp. 2611-2614, 2012.

[23] Y. Zhao and G. Karypis, "Comparison of agglomerative and partitional document clustering algorithms," in *SIAM Workshop on Clustering High-dimensional Data and its Applications*, 2002.

[24] Bayon - a simple and fast clustering tool - Google Project Hosting. [Online]. Available: http://code.google.com/p/Bayon/

[25] CLUTO - Software for Clustering High-Dimensional Datasets. [Online]. Available: http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview

[26] K. Ikejiri, Y. Sei, H. Nakagawa, Y. Tahara and A. Ohsuga, "A Proposal of Calculation method of Surprising Value of Recipe Based on Ingredient," *IEICE technical report*, DE2013-33, pp. 1-6, 2013.

[27] D. Blei, A. Ng and M. Jordan, "Latent dirichlet allocation," in *Journal of Machine Learning Research*, pp. 993-1022, 2003.