A Recipe Recommendation System based on Automatic Nutrition Information Extraction

Tsuguya UETA[1], Masashi IWAKAMI[1], and Takayuki ITO[12]

[1] Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi ,Japan,
ueta@itolab.nitech.ac.jp
iwakami-almuni2011@itolab.jp
ito.takayuki@nitech.ac.jp

[2] Policy Alternatives Research Institute, University of Tokyo

Abstract. In this paper, we propose a goal-oriented recipe recommendation system that utilizes information about nutrition on the Internet. Our system enables users without knowledge about nutrition to search easily for recipes with natural language to improve specific health conditions. The natural language includes 'I want to cure my acne' and 'I want to recover from my fatigue'. To do that, we created a co-occurrence database that listed the co-occurrence of 45 common nutrients with nouns such as cold, acne, bone etc. Then we created a recipe database by collecting 800,000 recipes from www.cookpad.com the system and analyzed each recipe to calculate the amount of a nutrient in a dish. We compared the results of our system to the results we obtained by calculating the nutrient information manually. Evaluation was done on 1000 dishes. We measured the effectiveness of the system using F-Measure and the average F-measure was 0.64 respectively.

Keywords: Nutritious Information, Goal-Oriented Recipe Recommendation System

1 Introduction

The number of people who use recipe sites is increasing. This trend is especially due to the spreading use of internet-enabled mobile phones (hereafter mobile terminals) such as the iPhone. Moreover, the Japanese government is increasingly focusing on promoting healthy nutrition (e.g., The Health Japan 21 Project [1]). According to [2], life style related diseases such as diabetes account for 60% of deaths and 30% of medical costs in Japan. Therefore, improving the nutrition of a country can greatly reduce medical costs. There are recipe websites that provide detailed nutrient information, that is, the proportion of protein, vitamins, fats etc. in the recipe, thus supporting a user deciding on a menu [3]; however, this information is only useful for a professional nutritionist. A general user with a rough skin condition, for example, cannot decide whether the recipe will help her condition or not. This is because she is not a professional nutritionist. There are also websites that provide recipes designed by dieticians that target specific health conditions. Yet these are usually limited, and the chance that a user might not find a recipe for her condition is high.

We propose a system that relieves general users from the burden of having to identify the right recipes for their condition. This system will also relieve dieticians from the task of designing recipes for all the numerous health problems people suffer from. We designed our system to take in inputs such as 'want to relieve fatigue', 'want to prevent diabetes', and 'want to cure acne'. Then the system outputs recipes that address the user's health issues. With this system, a dietician does not have to make a recipe manually from scratch, and users can search for recipes to improve their health without needing a sophisticated knowledge of nutrition.

This paper is organized as follows: In Section 1, we introduce the paper, and in Section 2, we discuss the paper in relation to previous research. In Section 3, we provide and describe the configuration of the prototype system. In Section 4, we show how the system operates by giving some examples. In Section 5, we provide experimental results. In Section 6, we conclude and discuss issues for future research.

2 Related Work

We review three systems that are similar to our system. These are the Nutrition Information Retrieval Database[4], goal-oriented product recommendation systems Minami[5] and Kobayashi[6] and a recipe recommendation system based on onomatopoeia.

A Health and Nutrition Information Infrastructure Database

This database[4] was developed based on a national survey conducted through a collaborative study by the National Institute of Health and Nutrition and the Japan Science and Technology Agency. In the database, dishes that are consumed in Japan are listed. We can use the registered dish database in a meal and rearrange the information about the dish; however, there are few kinds of registered dishes in this database.

Goal-Oriented Clothes Recommendation Systems

Clothes recommendation systems such as those described by Minami[5] and Kobayashi[6] recommend clothes for a specific occasion. The occasion can be a wedding, a funeral, or a season of the year such as summer or winter. A system like this extracts an impression from the input words of a user and then matches the impression to clothes it recommends that the user wear.

A Recommendation System Based on Onomatopoeic Words for Foods

When talking about foods, Japanese people frequently use onomatopoeic words to express a vague taste or sense of a food. The system recommends food to eat based on onomatopoeic words for foods. This was shown to have higher precision than a keyword based search [7].

Relation of This Paper to Other Research on Recommendation Systems

Here we discuss the relation of this paper to other research on recommendation systems. With this type of system, the user inputs a natural sentence stating a desire to improve a certain health condition, and the system recommends a dish suited for this condition. Our system is a goal-oriented recommendation system like those of Minami[5] and Kobayashi[6]; however, it differs from both because it includes a co-occurrence dictionary and food databases. We explain the system's co-occurrence dictionary in Section 3.2 and its food databases in Sections 3.3 and 3.4. Unlike[7], our system uses nouns and the most relevant nutrient name rather than onomatopoeia in search phrases.

3 A Goal-oriented Food Recommendation System

3.1 System Configuration

We explain the system flow in five steps (Step 1- Step 5) using the concrete examples in Figure1. First, the user enters his/her health problem in natural language, e.g., 'I want to cure my acne.' In Step 2, the system breaks the sentence input down into its morpheme units by MeCabMeCab(the software of morphological analyzer)[14] and extracts the noun. The noun in this example is acne. In Step 3, the system searches the co-occurrence dictionary for the nutrient that is most closely related to the noun extracted in Step 2. We explain the co-occurrence dictionary in the next section (Section 3.2.) In Step 4, the dishes that are most closely related to the nutrient that was identified in the previous step are searched for in the food database. In this example, the nutrient is pantothenic acid, and the search in the food database revealed that dishes that contain a lot of liver and mushrooms are rich in pantothenic acid. The web server that received the search request retrieves the results in the XML file format. The software in the iPhone presents the results to the user in a user-friendly manner.

3.2 The Co-occurrence Dictionary

As mentioned above, the system performs morphological analysis of an input sentence such as 'I want to cure my acne,' identifies the noun (that is, acne) and searches for it in the co-occurrence database to locate the nutrient that co-occurs most with it. The dictionary contains co-occurrence data for 45 nutrients

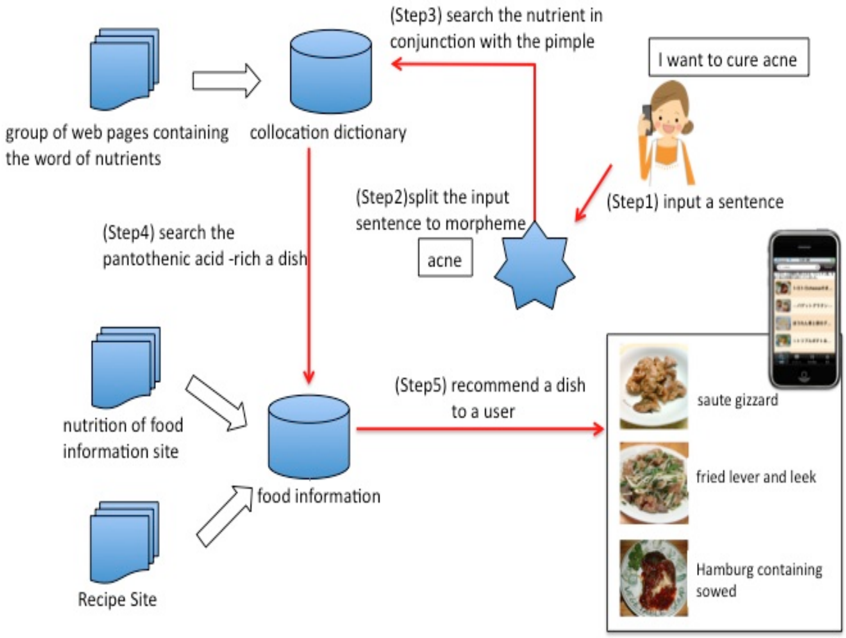

Fig. 1. System Configuration

such as vitamin A (Retinol), Vitamin B, and Vitamin C. The dictionary was created by using up to 500 ranked web pages from a Google search result for each nutrient. The web pages found from the search result are morphologically analyzed to identify the nouns in them. For example, a search for calcium usually returns pages about the effects of calcium. In addition, these pages often contain the words 'tooth' and 'bone.' In this way, a row in the co-occurrence database contains a nutrient, a noun, and a number showing how many times the nutrient and the noun co-occur. Table 1 shows part of the co-occurrence data rows for pantothenic acid. As can be seen, the acid usually co-occurs with immunity and metabolism. This indicates that the acid improves immunity and metabolism. Table 2 shows the search result from the dictionary when acne is used as a key word. As can be seen, pantothenic acid co-occurs with acne many more times than does any other nutrient. This data is used to conclude that the acid can aid in treating acne. Currently the database contains 350,000 entries. This amounts to roughly 8,000 co-occurrences for each of the 45 nutrients.

Table 1. Search results for 'pantothenic acid' in the co-occurrence dictionary

Nutrient name	Word	Number of co-occurrences
Pantothenic acid	health	248
Pantothenic acid	folic acid	198
Pantothenic acid	Vitamin C	193
Pantothenic acid	stress	184
Pantothenic acid	skin	151
Pantothenic acid	metabolism	127
Pantothenic acid	acne	125
Pantothenic acid	effective	115
Pantothenic acid	lipid	107
Pantothenic acid	body	106

Table 2. Search results for 'acne' in the co-occurrence dictionary

Nutrient name	Word	Number of co-occurrences
Pantothenic acid	acne	125
Vitamin B2	acne	47
Vitamin B6	acne	29
Vitamin C	acne	19
Retinol	acne	17
alpha-Carotene	acne	16
Vitamin B1	acne	15
alpha-Tocopherol	acne	13
Polyunsaturated	acne	9
Saturated fatty acid	acne	7

The food databases mentioned in Section 3.1 consist of two databases– the ingredient nutrient database and a nutritional information database for recipes.

3.3 Ingredient Nutrient Database

Here ingredients refer to the food materials such as meat, dairy products, spices, etc., that make up a dish. This database contains information on the amount of nutrients in an ingredient. We were able to collect such data for 1,861 ingredients using a crawler that we developed. Each ingredient contains each nutrient in different proportions and amounts. Table 3 shows the amount of pantothenic acid in 100 grams of certain ingredients. It can be seen that chicken liver and shiitake mushrooms contain lots of the acid. Therefore, dishes that contain these ingredients are given priority for selection when a user queries 'I want to cure my acne'.

A recipe must show not only the amount of a nutrient in an ingredient, but also the percentage of the daily intake requirement for the nutrient that this

Table 3. Pantothenic acid in 100 grams of certain ingredients

Food name	Content	Unit
Chicken, Liver - Raw	10.1	mg
Shiitake, mushrooms - Dried	7.93	mg
Pork, Smoked Liver	7.28	mg
Pork, Liver - Raw	7.19	mg
Meat, Liver - Raw	6.4	mg
Cheese, whey powder	5.95	mg
Lamprey	5.76	mg
Spices and seasonings / Yeast / Brewer's yeast	5.73	mg
Fish roe	5.17	mg
Sparrow, Meat with bone and skin - Raw	4.56	mg
Egg yolk - Raw	4.33	mg

amount includes. The amount of a nutrient in a recipe should not be over or under the proper daily limit. For this, we created a database based on the 2005 Japanese nutrient intake requirements [10]. The requirements vary according to age and sex.

3.4 Nutritional Information Database

This database contains recipes with the type and amount of nutrient that each recipe contains. The ingredients in a recipe were identified, and then nutrient amounts were calculated using the database on the amount of nutrients in ingredients. We were able to collect 800,000 dish recipes from cookpad.com[11].

The ingredient nutrient database also contains information of how cooked ingredient or raw. We analyzed the cooking procedures and extracted verbs such as boil or roast. Then we referred to the database and identify the ingredient. For example, the cooking procedure contains 'I boiled a bean sprouts'. The database contains 'bean sprouts -raw' and 'bean sprouts -boil'. Then we extract 'boiled' from the cooking procedure and identify 'bean sprouts -boil'.

We evaluated the effectiveness of our system in identifying ingredients from a recipe by using it to analyze ten sample dishes. Table 4 shows the result of the evaluation. The ten dishes contained 79 ingredients. Our system was able to identify 70 of them correctly. Even we could not identify one of the ingredients, so it was probably a typo. The remaining five ingredients were simply not registered in our ingredient nutrient database.

Although the recipes describe quantities of the ingredients in different units, we changed all of them into grams. Table 5 shows the result of changing grams by using it to analyze 70 ingredients we showed above. The word "half" converted it as half grams of a gram (per none of the foods) registered with a gram conversion dictionary. We did not calculate the unjust word such as not the unit and a number. We also calculated the number including "2 or 3" and "1-2" to the average between numbers. We assume that a tablespoon equals 15 grams, a teaspoon is equal to 5 grams, and the words ”a little” and an ”appropriate amount” correspond to 10 grams of the ingredient material.

Table 4. Classification results of applying the system to ten sample dishes

Category	Number
Ingredients correctly identified	70
Mistaken name for an ingredient	1
Ingredients not registered in our database	5
Ingredient correctly identified but misunderstanding by mean	3
total	79

Table 5. Changing to gram results of applying the system

Category	Number
the word ”a little”	9
the word ”appropriate amount”	9
unjust word	4
the word ”half”	5
the unit ”g” or ”cc”	26
tablespoon / teaspoon	17
total	70

4 System Overview

4.1 System Flow

We chose the iPhone because it is convenient to use since it is mobile, and it does not require that much IT literacy. We introduced a guide to using the system with figure 2. Part (1) of the figure shows the top screen of the system. The system has interfaces that enable a user do the following:

(2) search using his/her own words (3) search with the nutrient name. (4) search with predefined search keywords (5) view a list of dishes recommended by the system and a detailed recipe of a dish (6) view nutrition information on the dish. (7) view the day’s menu (8) view favorite dishes (9) edit user setting information and other information.

4.2 System Output

Figure 3 is a screen shot of an output of the system. The user can tap on any of the recommended dishes to view detailed recipe information. Figure 4 is a

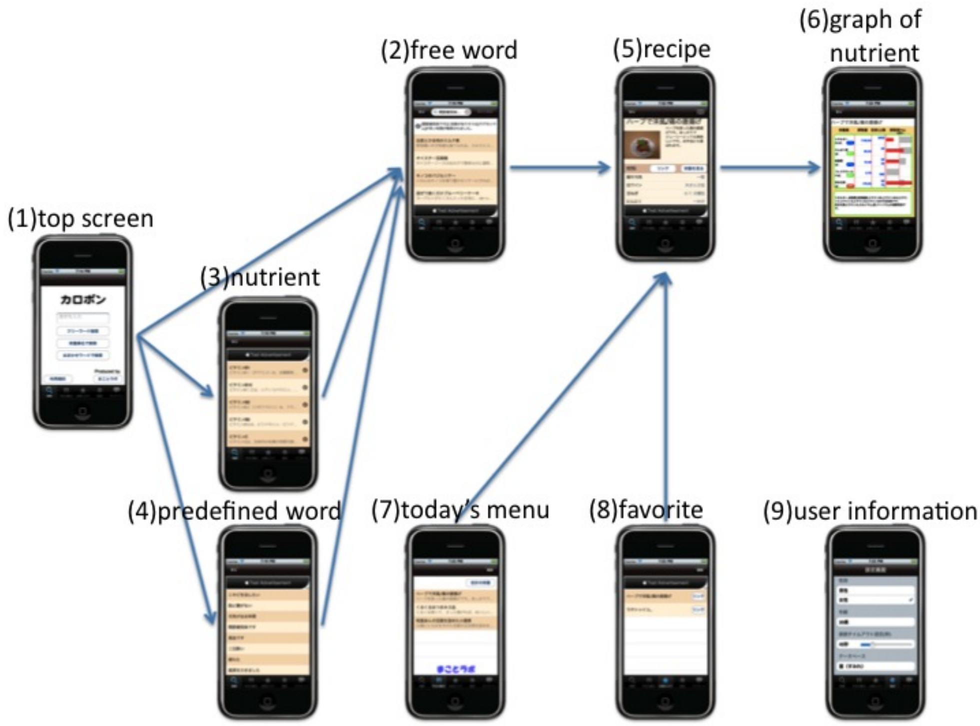

Fig. 2. system flow

screen shot of a graph about nutrients. Table 6 shows the result for a search with 'want to prevent a cold'. Because the nutrient which usually co-occurs with 'cold' is Vitamin C, a Vitamin C rich dish should be recommended. It is common knowledge that green vegetables are rich in vitamin C, and as can be seen, dishes that contain green vegetables like green peppers are recommended. Table 7 shows the recommendation results when the user inputs "want to cure my acne". The nutrient that usually co-occurs with acne is pantothenic acid. Therefore, dishes that contain liver, mushrooms, dairy products, or eggs should be recommended. As can be seen, such dishes are recommended.

5 Performance Evaluation of the System

5.1 Recommendation Accuracy

We evaluated the effectiveness of our system when it calculated the amount of nutrients in a dish (recipe). The system first had to identify the ingredients, next calculate the amount of each ingredient, and then use the ingredient nutrient

Table 6. The top 10 recommendations when the user inputs "want to prevent a cold"

Summer vegetable nachos
Colorful paella in a fry pan
Plum-flavored pasta with rich mushrooms and green peppers
Ratatouille
Spicy fried clams
Warm tofu with crab sauce
Richly nutritious vegetable juice
Anchovy-flavored red peppers
Easy nasi goreng
Stir-fried chicken and green peppers

Table 7. The top 10 recommendations when the user inputs "want to cure my acne"

Hamburger steak with maitake mushrooms and cheese topping
Baguette gratin
Spinach and egg gratin
Triple potato oyaki (Japanese baked bun)
Rich mushrooms sauteed in butter
Pork fillet with black pepper
Three-berry tarts
Rich cheese hamburger steak
Deep-fried koya tofu sandwiches
Sandwiches using French bread

table to calculate the amount of nutrient in each ingredient. Moreover, when recommending a dish, it had to consider the allowable upper and lower intake limits for the nutrient based on the intake requirements database described in section 3.3. Evaluation was done on 1000 dishes. We compared the results of our system to the results we obtained by calculating the nutrient information manually. We measured the effectiveness of the system using three indices– for Precision, Recall, and F-Measure.

The first index, for Precision, measures the usability of the system. It shows the percentage of dishes correctly recommended by the system. A dish was considered to be correctly recommended if it was also recommended by manual analysis.

$$Precision = \frac{dishes(manual) \cap dishes(system)}{dishes(system)} \tag{1}$$

The second index, Recall, measures the comprehensiveness of the search results. That is, it measures the number of the system's recommended dishes as a percentage of the number of dishes recommended by manual analysis.

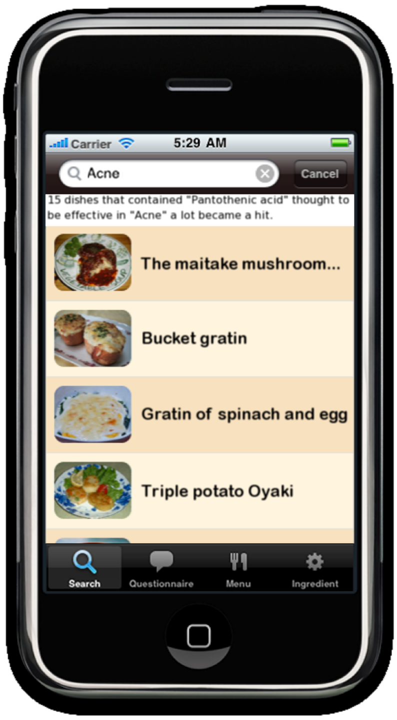

Fig. 3. Search Results

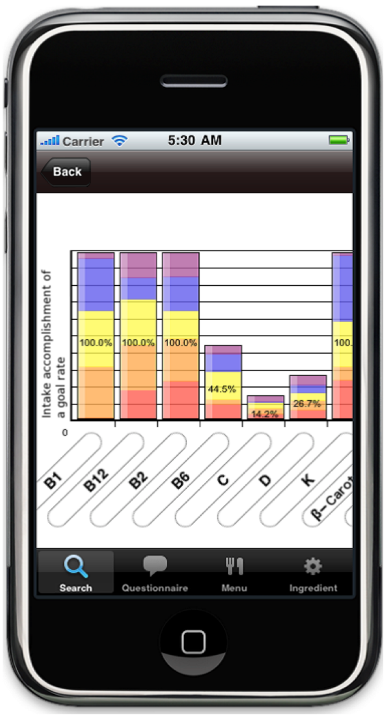

Fig. 4. Display of Nutrients

$$Recall = \frac{dishes(manual) \cap dishes(system)}{dishes(manual)} \quad (2)$$

F-measure is the harmonic average of Precision and Recall. It is used as the general evaluation index of the system.

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

Table 8 shows the evaluation results for some nutrients. The average Precision, Recall and F-measure were 62.6%, 65.6% and 0.64 respectively. When the system chose dishes randomly the average Precision and Recall were 13% and 18%. Hence, we can confidently say that using the system to select dishes is better than randomly selecting them.

5.2 Evaluation System Trial

We conducted a survey of subjects on the system that we have described in this report. In this questionnaire, we asked the subjects some questions about age,

Table 8. Recommendation Accuracy

Nutrient name	Unit	Precision	Recall	F-measure
Calcium	mg	60.3%	64.8%	0.63
Phosphorus	mg	62.9%	72.5%	0.67
Sodium	mg	53.3%	60.3%	0.57
Niacin	mg	68.9%	77.0%	0.72
Pantothenic acid	mg	76.2%	68.8%	0.72
Potassium	mg	69.6%	68.9%	0.69

sex, specialized knowledge about nutrition, recipe sites used frequently, and our system. The questions in the survey were a part of the system's software on the subjects' terminals. The subjects transmitted their answers to the survey from their terminals. We then analyzed the answers from the subjects. The following are the results from the questions about the system. Twenty persons, nine women and eleven men, replied to the questionnaire,. The breakdown of the subjects by age was as follows: Two persons were 15-19 years old, twelve persons were 20-24 years old, two persons were 25-29 years old, two persons were 30-34 years old, and two persons were 40-45 years old. Regarding "personal expertise on nutrition", one person answered, "I have a little detailed expertise," ten persons answered, "I have average knowledge", six persons answered, "I don't have much expertise," and two persons answered", I don't have any expertise." Regarding the degree of specialized knowledge of nutrition, most respondents turned out not to know much about nutrition. About the system, there were many positive evaluations. Many users said in the survey that they felt that the system provided effective recommendations; however, users tended to feel that a dish that had many vegetables was a healthy dish because actual users do not know much about nutrition. Thus, a dish with ingredients of high nutritional value such as green vegetables was generally recommended by this system. There were some evaluations that the user would like to continue to use the system. It is necessary to improve the system to make it easier to use.

6 Conclusion and Future Work

In this paper, we described a dish recommendation system that takes input like 'want to prevent a cold' or 'want to cure my acne'. The system outputs a list of dishes (with detailed recipes) that contain the right type of nutrient to treat the health condition. To do that, we first created a co-occurrence database that listed the co-occurrence of 45 common nutrients with nouns such as cold, acne, bone, teeth, metabolism etc. Common nutrients included vitamins, proteins, fats, etc. Then we created an ingredient nutrient database that listed 1,860 commonly known ingredients with the amount of nutrients available in them. Lastly, we created a recipe database by collecting 800,000 recipes from www.cookpad.com the system. Then we analyzed each recipe to identify the ingredients in it and the amount of each ingredient. This information was used to calculate the amount of a nutrient in a dish. This result was combined with the co-occurrence database to select the dishes that would best treat the health condition entered by the user. We developed a system to be used primarily with an iPhone. This is because users can use it conveniently. Even during cooking, a user can hold it in one hand to read recipe details and procedures. We evaluated the effectiveness of the system by comparing its result with results from manual analysis. The evaluation was done on the results of analyzing 1000 recipes. The similarity of the dish recommendations from the manual analysis and the system analysis was compared.

We have abundant sources of recipe data. Therefore, our system can usually find a recipe with ingredients that contain the necessary nutrient to address the health issue. Yet perhaps there should be a way for a professional dietician to check the validity of the results. Our current system does not consider the past search history of the user. Therefore, if the user inputs the same search phrase every day, he will get the same results. If the system were able to take past search history into account, a greater variety of dishes could be retrieved from the search, and repetition of the same dishes could be avoided. Another task of future research is to find ways of avoiding recipes with ingredients that the user is allergic to.

References

[1] Japan Health Promotion & Fitness Foundation 2000
http://www.kenkounippon21.gr.jp/index.html

[2] Increase of the importance of lifestyle-related disease measures, Annual Report on Health and Welfare, 2007
http://wwwhakusyo.mhlw.go.jp/wpdocs/hpax200701/b0039.html

[3] Bob and Angie recipe site
http://www.bob-an.com/recipe/daily/daily.asp

[4] Health and Nutrition Information Infrastructure Database
http://dbs.jst.go.jp/HandNDB.htm

[5] Fumiaki Minami, Mikito Kobayashi, Takayuki Ito, An Implementation of Goal-Oriented Book Recommendation System with Concept Dictionary, Japan Society for Software Science and Technology, 24th, 2007

[6] Mikito Kobayashi, Fumiaki Minami, Takayuki Ito, Satoshi Tojo, An Implementation of Goal-Oriented Fashion Recommendation System with Concept Dictionary, The 70th National Convention of IPSJ, 2007

[7] Kanwipa Lertsumruaypun, Chiemi Watanabe, Satoshi Nakamura, Onomatoperori : Recipe Recommendation System Using Onomatopoeic Words, SigDD2009, 2009

[8] Food Composition Database, http://fooddb.jp/

[9] Japanese meal intake standard, 2005

[10] COOCPAD, http://cookpad.com/

[11] 2001 national survey on nutrition, Food number list, The study group on standardization of diet surveys
http://www.nih.go.jp/eiken/nns/system/bangohyo.pdf

[12] allrecipes.com, http://allrecipes.com/, 2009.

[13] MOBILE MARKETING DATA LABO., 'Site Usage Survey housewife recipes'

[14] Mecab, http://mecab.sourceforge.net/