Finding Replaceable Materials in Cooking Recipe Texts Considering Characteristic Cooking Actions

Yuka Shidochi
Graduate School of I.S.,
Nagoya University
shidochi@murase.m.is.
nagoya-u.ac.jp

Tomokazu Takahashi
Gifu Shotoku Gakuen
University
ttakahashi@gifu.
shotoku.ac.jp

Ichiro Ide
Graduate School of I.S.,
Nagoya University / NII
ide@is.nagoya-u.ac.jp

Hiroshi Murase
Graduate School of I.S.,
Nagoya University
murase@is.
nagoya-u.ac.jp

ABSTRACT

The number of cooking recipe texts published on the Web is increasing in recent years. However, in general, cooking recipe texts have little flexibility. So, it is not always easy to retrieve cooking recipe texts that satisfy users' various demands. Therefore, it is necessary to create and offer recipes that suit the user's requirements. In this paper, we propose a method for finding replaceable materials considering characteristic cooking actions from a large amount of cooking recipe texts. The proposed method finds the replaceable materials by first extracting the cooking actions that correspond to each material than measuring the similarity of the extracted cooking actions. Through an evaluation of recipe texts created by replacing some materials that were found by the proposed method, we verified the effectiveness of the proposed method.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

text mining, cooking recipe, recipe retrieval

1. INTRODUCTION

Cooking is a complex creative activity based on the advanced knowledge acquired through daily cooking experience. Therefore, there have been works that tried to assist cooking activities by various methods [2, 10]. For instance, there is a work that recognizes cooking operations with a camera, and informs the upcoming procedures in appropriate timing [3]. Also, there is a work that creates a database on cooking operations by detecting cooking actions and uses it to teach users the cooking operations [6]. There is also a work that developed a projector system that adds attractive colors on a dining table, based on the assumption that color is an important factor when eating [7].

Recently, cooking recipe texts are available in various forms including cookbooks and mobile game machines. Especially, the number of cooking recipe texts published on the Web is increasing in recent years rapidly[1], because people can publish them easily and freely. It is possible to decide a menu by searching through these recipes, especially for a person who decides and cooks a daily meal, which is sometimes considered as a troublesome daily task.

However, the cooking recipe texts lack flexibility on the originally decided materials and the procedure. When people actually cook, they might need to change an existing cooking recipe according to the situation and the feelings. For example, there are demands to select a recipe that satisfies various requirements, such as follows:

- A recipe that uses materials in the refrigerator.
- A recipe that uses materials that is on sale and/or could be obtained easily.
- A recipe considering health condition.
- A recipe considering preferences.

However, it is difficult to discover cooking recipes that satisfy one's demands as mentioned above. So, we aim to produce and to provide cooking recipes that meet the user's requirements by appropriately replacing materials in existing cooking recipe texts. There is a previous work that discovers replaceable materials by using a data mining technique [5]. However, the obtained replaceable materials are few, and also the method is impractical because it does not consider the relation between a material and its corresponding cooking action.

[1]"COOKPAD", http://cookpad.com/.

CEA'09, October 23, 2009, Beijing, China.

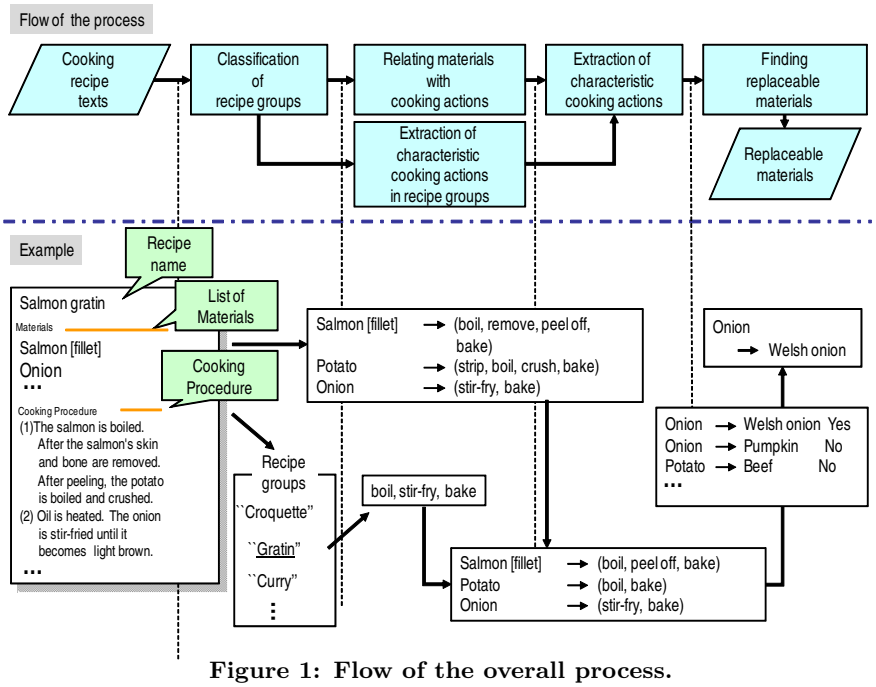

Figure 1: Flow of the overall process.

Table 1: Example of recipe groups.

Recipe group name	Number of recipes	Examples
salad [サラダ]	496	bonito sashimi salad [かつおの刺身サラダ], makaroni salad [マカロニサラダ]
simmered dish [煮]	447	simmered radish [大根煮], simmered pork in sukiyaki sauce [豚すき煮]
soup [汁]	272	bamboo sprout soup [たけのこ汁], potato soup [芋煮汁]
soup [スープ]	259	cabbage soup [キャベツのスープ], celery soup [セロリのスープ]
fry [焼き]	244	fried ram with miso sauce [ラム肉のみそくわ焼き], fried egg [卵焼き]
stir-fry [炒め]	239	stir-fried egg in sweet sauce [うずら卵の甘酢炒め], stir-fried spicy oyster [かきの辛味炒め]
poured dish [かけ]	132	taro poured on mozuku seaweed [もずく山かけ], oyster oil poured on rape [油菜のカキ油かけ]
fritter [揚げ]	131	cod fritter [たらの白扇揚げ], pork fritter [豚肉の南蛮揚げ]
garnish [添え]	130	egg garnished with vegetables [卵の野菜添え], egg roll garnished with mint [春巻のミント添え]
pickles [漬け]	128	pickled radish [大根の黒酢漬け], pickled turnip [かぶの即席漬け]

On the other hand, there are works that propose cooking recipes that a user needs [4, 9, 8]. For instance, literature [4] considers the preference of a user referring to the preference and the frequency of the use of a material in daily life. However, the cooking recipe texts provided by this method can not completely meet the user's demands because the method searches existing recipe texts only.

In this paper, we propose a method for discovering replaceable materials from a large number of cooking recipe texts on the Web as the knowledge necessary to produce and provide new recipes flexibly according to the user's situations.

2. FINDING REPLACEABLE MATERIALS FROM COOKING RECIPE TEXTS

This section describes the method of extracting the materials and actions in the cooking recipe texts, and finding replaceable materials. The flow of the overall process is shown in 2.1, and details of each process are described in 2.2, 2.3, 2.4, and 2.5, respectively.

2.1 Flow of the process

The flow of the overall process is shown in Figure 1. We aim to obtain materials that could replace a material by first considering the relation between each material and its corresponding cooking actions, and later measuring the similarity between the materials.

Cooking recipe texts are input as a HTML file obtained from a specific recipe site[2]. Then, they are analyzed to extract the "List of Materials" and the "Cooking Procedure".

[2]AjinomotoCo., Inc., Ajinomoto recipe site, http://www.ajinomoto.co.jp/recipe/.

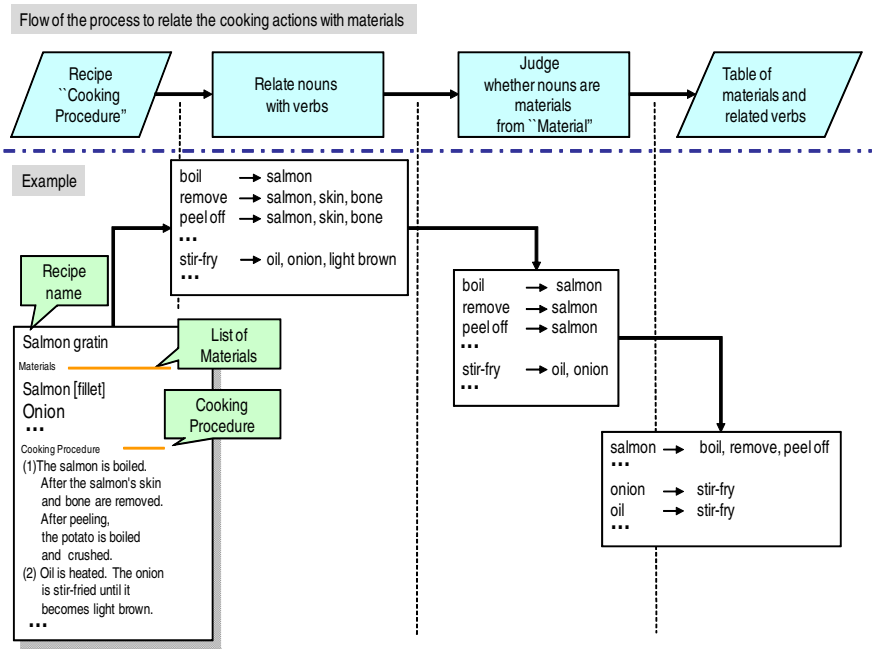

Figure 2: Flow of the process to relate the cooking actions with materials.

2.2 Classification of recipes by recipe name

It is difficult to extract a common replaceable material from all the collected cooking recipe texts because the variation is wide. Therefore, we decided to first classify similar recipes based on the assumption that the term at the end of a recipe name represents a similar recipe. Here, the group of the classified cooking recipe texts is named a "recipe group". Table 1 shows the ten largest recipe groups obtained by classifying 6,789 recipes.

2.3 Relating cooking actions with materials

There exists similar steps in recipes with common cooking actions in the same recipe group. So, we focus on the common cooking actions in a recipe group. At the same time, it is necessary to handle the cooking actions together with the materials involved, to find replaceable materials that could replace a material in a certain context.

First, materials and cooking actions are extracted from the cooking recipe texts. To extract them, morphological analysis is applied to the cooking recipe texts. Here, we define the nouns that appear commonly in the "List of Materials" and in the "Cooking Procedure" as materials.

Moreover, we define the verbs that appear in the "Cooking Procedure" are cooking actions. In addition, we handle some cases as exception in this process. Because they would affect the process negatively, we use the following special dictionaries that were manually made to cover them.

Dictionary1 : Words improperly segmented into more than two morphemes by the morphological analysis

Dictionary2 : Pairs of a transitive verb and an intransitive verb

Dictionary3 : Same words with different notations

Figure 2 shows the detailed process of relating the cooking actions with materials. The "Cooking Procedure" part has the following features.

- Each step has a number assigned to it ((1), (2), $\cdots$), and referred from other steps; When the result of the previous procedure is used in a procedure, it is represented by the procedure number for simplification.
- When many materials are processed together, they are expressed as a group (A, B, $\cdots$) for simplification.

The following rules were defined to associate a material and a cooking action.

1. All materials following a verb in a sentence are associated to the verb.
2. When no material following a verb is found in a sentence, all the materials that appear in the previous sentence are associated to the verb.
3. When a procedure number is used before a verb in a sentence, the materials that appear in the procedure of the procedure number are associated with the verb.

In Japanese, a subject generally appears before a predicate, where a subject is generally a noun. This is reflected in Rule 1. We define Rule 2 because the previous subject can be omitted.

Note that seasonings are considered not as materials.

2.4 Extraction of characteristic cooking actions

We consider it is possible to substitute a material if the corresponding cooking actions match. However, even the same cooking actions do not always match because of the difference in the writing style of each recipe or omission of short preparation procedures. Therefore, we need to evaluate the similarity of cooking actions related to a material considering the difference in the description of recipe texts for finding replaceable materials. We consider two kinds of cooking actions to find the similarity: 1) actions characterizing a recipe group and 2) actions associated to each material

Table 2: Recipe groups used in the experiment.

Recipe groups	Number of recipes
(1) croquette [コロッケ]	26
(2) marinade [マリネ]	67
(3) omelet [オムレツ]	32
(4) toast [トースト]	32
(5) curry [カレー]	51
(6) gratin [グラタン]	70
(7) roll [巻き]	72
(8) stir-fried burdock & carrot [きんぴら]	27
(9) pork & potato in broth [肉じゃが]	14

Table 3: Replaceable materials in the "croquette" group.

	Material	Replaceable materials
Main material in a recipe	onion	Welsh onion
	Welsh onion	onion, ground beef
	ground chicken	ground beef
Material used for decoration	endive	chervil
	paprika	chervil
	green perilla	endive, chervil
	radish	endive, chervil

in the recipe group. For instance, a characteristic action of the "croquette" recipe group may be "deep-fry".

However, the characteristic action associated with "potato" that appears in the cooking recipe texts in the recipe group may be "crush". Therefore, characteristic actions to a "potato" should be both "crush" and "deep-fry".

We calculate TFIDF for each verb in the cooking recipe texts to determine the characteristic action. Verbs with a TFIDF value larger than 0.8 was determined as characteristic cooking actions in the recipe group, and other verbs were discarded. The TFIDF is defined as follows:

$$TFIDF_{d,t} = TF_{d,t} \cdot IDF_t \tag{1}$$

$$TF_{d,t} = \frac{F_{d,t}}{N_d} \tag{2}$$

$$IDF_t = \left(1 + \log \frac{N}{F_t}\right) \tag{3}$$

where N is the number of all cooking recipe texts collected from the Web, N_d the number of cooking recipe texts in a recipe group d, $F_{d,t}$ the frequency of appearance of cooking action t in recipe group d, and F_t the frequency of appearance of the cooking action t in all the cooking recipe texts.

Next, cooking actions that appear more than 75% with a material in all the cooking recipe texts in a recipe group were extracted as characteristic cooking actions to the material. We measure the similarity of materials by comparing the cooking action vectors. The vector is created for each material. The cosine distance between two vectors is used to evaluate the degree of similarity. In addition, the distances are adjusted by weights because we considered that the possibility of substitution changes depending whether the materials are homogeneous (ex.[beef, pork], [potato, taro]). The decision of homogeneity is based on the "Food Composition Table" [1]. When a material is not in the Table, we judge it manually.

2.5 Finding replaceable materials

We consider that two materials are cooked in a similar context in a recipe group if the distance of the two vectors is small. In that case, the two materials are extracted as replaceable materials. The replaceable materials are sometimes obtained from the same cooking recipe text. In this case, both materials are characteristic to the recipe and not necessarily replaceable. So, when a pair of materials that appears in more than 3% of all the recipes in a recipe group, we do not consider them to be replaceable.

3. EXPERIMENT OF FINDING REPLACEABLE MATERIALS

This section introduces the experiment on extracting replaceable materials by using the method described in Section 2.

3.1 Experiment

3.1.1 Experimental conditions and procedure

We extracted replaceable materials from 391 recipe texts classified into nine groups shown in Table 2. There were 168 kinds of materials included in these recipe groups.

3.1.2 Result

92 replaceable material candidates were obtained from the recipe groups shown in Table 2. For example, "Onion (Vegetables) ⇒ Welsh onion (Vegetables)", "Ground chicken (Meat and poultry) ⇒ Ground pork (Meat and poultry)", and "Potato (Potato and starchs) ⇒ Pumpkin (Vegetables)", where words in parentheses are material classes in the Food Composition Table. For instance, the result of replaceable materials for the "croquette" group is shown in Table 3. Materials with different properties appear in cooking recipe

Table 4: Evaluation of the created recipes.

Evaluation	1	2	3	N/A
Number of recipes	15	15	1	9

Table 5: Materials evaluated as appropriate replacement in the evaluation.

Recipe groups	material	Replaceable materials
gratin	bacon	cauliflower, ham
	mushroom	shimeji mushroom
	onion	cabbage, Welsh onion, cauliflower
	beaten egg	beaten egg yolk
	potato	tofu, cauliflower
	spinach	bacon, broccoli
curry	potato	pumpkin
	green pepper	shimeji mushroom
marinade	whitebait	salmon
	paprika	tomato
croquette	Welsh onion	onion
	ground chicken	ground beef
	ground beef	ground chicken, Welsh onion
	endive	radish, green perilla, chervil
	onion	Welsh onion
toast	salami	carrot
omelet	bacon	carrot
stir-fried burdock	carrot	udo
roll	smoked salmon	salami, tuna
	green asparagus	tomato, tuna

Table 6: Replaceable materials in the "pizza" group.

Material	Replaceable materials
(a) salami	carrot
(b) green pepper	orange

texts, such as those materials used as a main material in a recipe or supplementary materials used only for decoration, and so on. In Table 3, we manually divided materials into two group that have different properties.

3.2 Evaluation

3.2.1 Procedure and result

A housewife with a dietitian's qualification evaluated the results. We prepared 13+40 cooking recipes for the evaluation, where the 13 recipes were the original ones obtained from the Web, and the other 40 recipes were created by replacing the materials with replaceable materials based on the original ones. All of the recipes were presented to the examinee, and were examined whether they were appropriate or not from the view point of a housewife and dietitian.

The examinee was asked to select one of the followings for each recipe.

1: The materials are appropriate. The cooking procedure is also appropriate.

2: The materials are appropriate. However, it is necessary to change the description of the "Cooking Procedure" (cooking actions).

3: The replacement of the material is impossible.

N/A: It is difficult to judge.

Table 4 shows the evaluation results concerning the 40 created recipes. Needless to say, the 13 original recipe texts were all judge as 1. As a result, the ratio of appropriate recipes (Evaluations 1 or 2) was 75%. We consider that this result was relatively good. Table 5 shows the replacements of the materials evaluated as 1 or 2.

3.2.2 Discussion

It seemed that it is insufficient to create a recipe by simply replacing a material. The replacement that was evaluated as 3 was "macaroni" and "steamed rice" in the recipe "Macaroni salad". The replacement was possible if it were "uncooked rice". Thus, we can see that it is necessary to take caution of pre-cooked materials.

(a) salami ⇒ carrot (b) green pepper ⇒ orange

Figure 3: Examples of actually cooked recipes. Both are "pizza toasts".

3.3 Actually cooking the produced recipes

We actually cooked recipes produced by replacing some replaceable materials. The recipe group was "pizza toast". We substituted the materials as shown in Table 6. the replaced "pizza toasts" are shown in Figure 3. Some people commented that they were delicious, but some others said they were unpalatable. Eventually, it is necessary to choose a replaceable material according to a user's favor.

4. CONCLUSIONS

In this paper, we proposed a method for finding replaceable materials from cooking recipe texts. In the experiment, we obtained relatively good results, such as "onion" ⇒ "Welsh onion", "potato" ⇒ "pumpkin" and so on. Evaluation by an examinee also proved the appropriateness of the results.

Problem of classification of recipes into "recipe group":
There are many recipe names which end with a material or do not specify a cooking method. Therefore, it is necessary to improve the classification accuracy of recipe groups by introducing other methods such as classification by referring to "List of Materials" and "Cooking Procedure".

Problem in finding replaceable materials:
The materials do not necessarily correspond to appropriate cooking actions when cooking recipes are created by simply replacing replaceable materials. Therefore, it is necessary to replace the cooking actions together. In the future, we should consider the order of the cooking actions and also replace the materials with the cooking actions as a set.

Problems as a whole:
The proposed method relies greatly on the text processing accuracy in clustering recipe groups, detecting a material, and extracting the cooking action, and so on. Recipe names, materials and cooking actions are spelled differently even in the same recipe. Moreover, cooking recipe texts contain special vocabulary. In this work, we made dictionaries when needed to cover these problems, but they should be done automatically in the future. We will also examine the appropriate granularity of recipe groups because the method used in this work for creating recipe groups depends on the variety of recipe texts in a recipe group. Therefore, common cooking actions corresponding to them could be few.

We need to further expand the data set to extract more replaceable materials. Finally, we will implement a system that provides new recipes created by replacing the materials.

5. ACKNOWLEDGMENTS

The authors would like to thank their colleagues for useful discussions. Parts of this research were supported by the Grant-In-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

6. REFERENCES

[1] *Food Composition Table, Fifth Edition [in Japanese].* Ishiyaku Publishers, Inc., 2001.

[2] K. J. Hammond. Chef: A model of case-based planning. In *Proc. 5th National Conf. on Artificial Intelligence*, volume 1, pages 267–277, August 1986.

[3] A. Hashimoto, N. Mori, T. Funatomi, Y. Yamakata, K. Kakusho, and M. Minoh. Smart kitchen: A user centric cooking support system. In *Proc. 2008 Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 848–854, June 2008.

[4] K. Ishihara, M. Ueda, Y. Hirano, S. Kajita, and K. Mase. An evaluation on the recommendation method for personal taste recipe based on the ff-irf. Technical Report MVE2007-77, IEICE, January 2008.

[5] T. Karasawa, R. Hamada, I. Ide, S. Sakai, and H. Tanaka. Extraction of knowledge on ingredients and cooking steps from cookbooks. In *Proc. IPS Japan 66th Bi-Annual Convention*, volume 2, pages 119–120, March 2004.

[6] S. Kuai, T. Takahashi, I. Ide, and H. Murase. Classification of cooking video segments based on sequences of image feature [in japanese]. In *Proc. 2009 IEICE General Conf.*, pages 291–301, March 2009.

[7] M. Mori, K. Kurihara, K. Tsukada, and I. Siio. A system to enrich food color [in japanese]. Technical Report MVE2007-80, IEICE, January 2008.

[8] M. Ohira, T. Ozono, and T. Shintani. Mining similarity assessment knowledge in cooking [in japanese]. In *Proc. 2000 JSAI Annual Conf.*, pages 396–399, July 2000.

[9] M. Ohira, T. Ozono, and T. Shintani. Implementing a recipe search system "minerecipe" using similarity-assessment knowledge[in japan]. In *Proc. IPS Japan 62nd Bi-Annual Convention*, volume 3, pages 129–130, March 2001.

[10] S. Russell and P. Norving. *Artificial Intelligence, A Modern Approach*. Prentice Hall, 1994.