
Construction of a Cooking Ontology from Cooking Recipes and Patents

Hidetsugu Nanba

Hiroshima City University
3-4-1 Ozukahigashi, Asaminamiku
Hiroshima, 731-3194 JAPAN
nanba@hiroshima-cu.ac.jp

Yoko Doi

Hiroshima City University
3-4-1 Ozukahigashi, Asaminamiku
Hiroshima, 731-3194 JAPAN
doi@ls.info.hiroshima-cu.ac.jp

Miho Tsujita

Hiroshima City University
3-4-1 Ozukahigashi, Asaminamiku
Hiroshima, 731-3194 JAPAN
tsujita@ls.info.hiroshima-cu.ac.jp

Toshiyuki Takezawa

Hiroshima City University
3-4-1 Ozukahigashi,
Asaminamiku
Hiroshima, 731-3194 JAPAN
takezawa@hiroshima-cu.ac.jp

Kazutoshi Sumiya

University of Hyogo
1-1-12 Shinzaike-honcho,
Himeji,
Hyogo 670-0092 JAPAN
sumiya@shse.u-hyogo.ac.jp

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

UbiComp '14, Sep 13-17 2014, Seattle, WA, USA
ACM 978-1-4503-3047-3/14/09.
<http://dx.doi.org/10.1145/2638728.2641328>

Abstract

A cooking ontology is an indispensable language resource for the language processing of cooking recipes. We have constructed a cooking ontology by means of pattern matching, statistical natural language processing techniques, and manual steps to identify hyponymy, synonymy, attributes, and meronymy.

Author Keywords

Cooking Ontology; Cooking Recipe; Hyponymy; Synonymy; Attribute; Meronymy

Recipe; Ontology; Predicate Argument Structure Analysis

ACM Classification Keywords

H.3.1 Content Analysis and Indexing

Introduction

There has been a recent increase in research work focusing on cooking recipes, including recommendation [7], summarization [8], and predicate-argument structure analysis [5], has been increasing. However, different terms are used in different recipes, particularly in user-generated recipe-sharing sites, even though these terms often refer to the same thing. Moreover, anaphora and ellipsis resolution is often required, which can make it difficult to process recipes correctly. We have therefore constructed a cooking ontology that can be used in a variety of language processing tasks as a linguistic resource.

In general, an ontology is organized as a hierarchy of concepts involving relationships such as synonymy and meronymy. Various methods for constructing ontologies from text databases have been proposed for a variety of natural language processing fields [1, 2, 3, 4]. By applying these techniques to a cooking recipe database, we can construct a cooking ontology. In addition to a recipe database, we have used a patent database. Patent applicants tend to describe various matters explicitly to avoid patent-infringement lawsuits, even though they may refer to common-sense concepts. This indicates that we can expect to find common knowledge about cooking matters in patent documents.

The remainder of this paper is organized as follows. Section 2 describes related work. Section 3 explains the method for constructing a cooking ontology. Section 4 reports on an experiment, and discusses the results. We present some conclusions in Section 5.

Related Work

In this section, we describe some related studies on "construction of a cooking thesaurus," "natural language processing for cooking recipes," and "extraction of various relations between terms."

Cooking Ontology

Several cooking ontologies, such as "Cook's Thesaurus¹" and "Taaable²," have been developed. "Cook's Thesaurus" is an English cooking encyclopedia that covers thousands of ingredients and kitchen tools. This thesaurus comprises 17 categories, each of which

contains an ingredient image, synonyms, a pronunciation guide, notes, and substitutes. Our cooking ontology covers not only synonyms and hyponymy but also meronymy and attributes of the various terms. "Taaable" is another widely used cooking ontology, which contains meronymy and hyponymy. It also covers various languages, such as English, French, German, and Spanish. However, this ontology does not focus on Japanese, and also does not cover some ingredients, which are commonly used in Japanese food. Our cooking ontology focuses on Japanese cooking recipes, and covers ingredients for Japanese food.

Natural Language Processing Focusing on Cooking Recipes

Yamakata et al. [8] proposed a method that creates a typical cooking procedure from multiple recipes by converting each recipe into recipe trees and by integrating them. Then, they extract features of each recipe by comparing with the typical one.

Tachibana et al. [6] focused on various modifiers in the titles of recipes that point to the characteristics of the recipes, such as "kid-friendly" and "simple." They analyzed the reasons for these modifiers being used in the recipe titles. For example, soymilk, which is preferred by many children, can be used instead of garlic and milk as an ingredient in carbonara, which is why the modifier "kid-friendly" was used in the title. To identify this as the reason, it would be necessary to compare this recipe with other carbonara recipes, extracting the similarities and differences between them. For such purposes, our cooking ontology can contribute to improving the Tachibana's method, because different terms are often used in different recipes, even though they refer to the same thing.

¹ <http://www.foodsubs.com/>

² http://wikitaaable.loria.fr/index.php/Main_Page

Automatic Construction of an Ontology from Text Databases

Several methods for constructing an ontology from text databases have been proposed. Hearst [2] proposed a method for extracting hyponymy from text databases using a set of patterns. For example, "cabbage," "radish," "eggplant," and "cucumber" are extracted as hyponyms of "vegetable" from the following sentence, using the pattern "NP₀ such as {NP₁, NP₂, (and/or)}* NP_n."

Furthermore, the method of this specific embodiment can be used to dry *vegetables such as cabbage, radish, eggplants and cucumber*, which were difficult to dry with prior art methods

Here, NP_x stands for a noun phrase. We apply this method to Japanese patents to extract hyponymy.

Chung [1] proposed a method for extracting synonyms based on a recipe data structure. From the observation that the main ingredient is usually written first in the ingredient list of a recipe, he assumed that this first ingredient is strongly related to the category to which the recipe belongs. They confirmed experimentally that his method for calculating relation scores between ingredients and category names using the ingredient position was effective for collecting synonyms from a recipe database. We use this method to extract synonyms.

Although, we confirmed the effectiveness of Chung's method via the experiment described in Section 4, we found that Chung's method was insufficient for collecting all synonyms. Furthermore, the method could not collect verb synonyms. To overcome this weakness,

we examined another approach. Lin [4] and Lee [3] proposed distributional similarity methods for calculating the similarity between terms. They focused on the contexts in which terms are used, defining the similarity between two terms as the amount of information contained in the commonality of the terms divided by the amount of information in their contexts. We apply this method to a cooking recipe database, collecting related terms as candidates for synonyms of a given term.

Construction of a Cooking Ontology

In this section, we describe our cooking ontology, which comprises synonyms, hyponymy, meronymy, and attribute relations of terms in a cooking domain. We give an overview of our ontology, with examples, and explain the procedure used for its construction.

Overview of Our Cooking Ontology

The structure of our cooking ontology is shown in Figure 1. Our ontology employs a two-level hierarchy. The top level comprises the following seven categories.

- Ingredient - seafood
- Ingredient - meat
- Ingredient - vegetable
- Ingredient - other
- Condiment
- Kitchen tool
- Movement

Among these categories, "Movement" is a category involving verbs, while the others involve nouns. Each category contains several entry words. For example, in Figure 1, "Ingredient-seafood category" comprises several entry words such as "squid" and "shrimp." For

each entry, several related terms are classified into three categories: "synonymy," "meronymy," and "attribute."

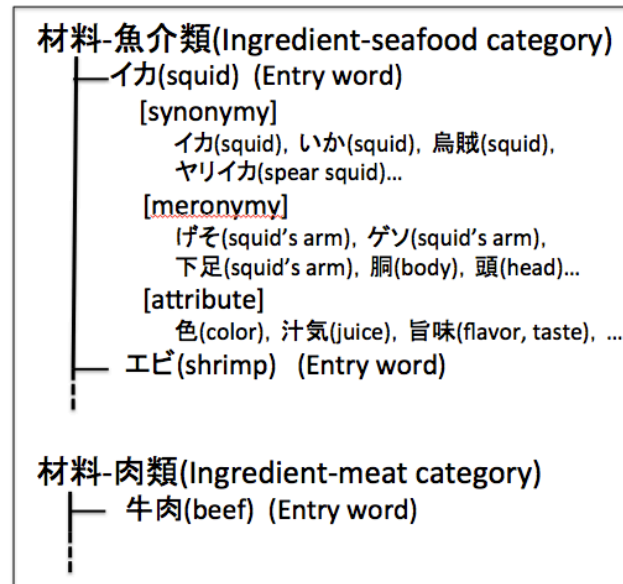


Figure 1. Structure of our cooking ontology.

Procedure for Constructing the Cooking Ontology

The procedure for constructing the cooking ontology comprises the following three steps.

- (Step 1) Determination of entry words
- (Step 2) Collection of synonyms for each entry word
- (Step 3) Collection of meronymous words and attributes for each entry word

We now describe these steps in detail.

Step 1. Determination of Entry Words

We determined entry words via the following two substeps.

- (Step 1-1) Collecting candidate words from patents
- (Step 1-2) Selecting entry words manually

In Step 1-1, we applied Hearst's method [2] to patents³, and collected candidates for entry words. For the following five categories, we prepared seed words that were synonyms of each category name, and then collected hyponyms using the pattern "NP₀ ((, | や)NP_n)* (等|など)の[seed word]" ([seed word] such as NP₀ ((, |and|or) NP_n)*).

- Ingredient – seafood
魚類(fish), 魚介類(fish), 海産物(sea product), 水産物(fishery product)
- Ingredient – meat
肉類(meat), 食肉(edible meat), 食肉類(edible meat), 原料肉(ingredient meat)
- Ingredient – vegetable
野菜(vegetable), 果菜類(fruit vegetable), 野菜類(vegetable), 果菜物(fruit vegetable), 農産物(agricultural products)
- Condiment
調味料(condiment), 香辛料(spice), 薬味(condiment), スパイス類(spice)
- Kitchen tool

³ We used those unexamined Japanese patent applications over 19 years (1993-2011) to which any of the International Patent Classification codes A23L (foods, foodstuffs, or non-alcoholic beverages), A47J (kitchen equipment), or H05B (electric heating, electric lighting) were assigned.

調理器具(kitchen tool), 調理容器(cooking container), 調理器(cooking device), 調理具(cooking tool), 調理道具(cooking utensil)

For example, when we collected candidate words for the "Ingredient - seafood" category, we found sentences that contained patterns such as "(NP₀ ((、|や)NP_n*)などの魚類" (fish such as ((、|and|or) NP_n*)) or "(NP₀ ((、|や)NP_n*)等の水産物" (fishery product such as ((、|and|or) NP_n*)). We then extracted noun phrases (NP₀ and NP_n), such as "イカ" (squid) or "エビ" (shrimp), as candidates for entry words in the "Ingredient - seafood" category.

Although Hearst's pattern-based method is useful for collecting hyponyms from texts, there are several cases where inappropriate words are mistakenly extracted. In the following sentence, "食用" (edible use) and "鑑賞用" (ornamental purpose) are mistakenly extracted as candidates for the "Ingredient - seafood" category.

食用や観賞用等の魚介類をいう。

(This indicates fish for edible use and for ornamental purposes)

We therefore delete such inappropriate words manually from the candidate list in Step 1-2. From among the remaining candidates, we statistically determined one representative word for each group of synonyms. As an example, for the three candidates "サケ" (salmon), "鮭" (salmon), and "さけ" (salmon), we manually selected "サケ" as the representative word, because the frequency of the phrase "サケ(など|等)の魚介類" (fish such as a salmon) is greater than those of "鮭(など|等)の魚介類" (fish such as a salmon) and "さけ(など|等)の魚介類" (fish such as a salmon).

We selected ingredient words that do not belong to any of the categories "Ingredient - seafood", "Ingredient - meat", and "Ingredient - vegetable" as entry words in the "Ingredient - other" category. Most of the words in this category are processed foods, such as cheese and pasta.

For entry words in the "Movement" category, we manually selected verbs appear frequently in the Rakuten recipe database⁴.

Step 2. Collection of Synonyms for Each Entry Word

The procedure of collecting synonyms for each entry word comprises the following two substeps.

(Step 2-1) Collecting candidates for synonyms.

(Step 2-2) Identifying synonyms manually.

In Step 2-1, we used the following three methods.

(Method 1) Using words that were deleted in the process of determining representative words in Step 1-2

(Method 2) Chung's method [1]

(Method 3) Distributional similarity method [3, 4]

We have already described Methods 1 and 2. Here, we explain Method 3. As we explained in the section on related work, the basic idea of distributional similarity is to calculate the similarity between two words in terms of their context words. Our algorithm is as follows.

1. Analyze the dependency structures of all sentences in the Rakuten recipe database, which contains about

⁴ <http://recipe.rakuten.co.jp/>

440,000 recipes, using the Japanese parser CaboCha⁵

2. Extract <noun phrase><verb> pairs that have dependency relations from the dependency trees obtained in Step 1
3. Count the frequencies of each <noun phrase><verb> pair
4. Collect verbs and their tf*idf scores for each noun phrase, creating indices for each noun phrase
5. Calculate the similarities between two indices for noun phrases using the cosine distance
6. Obtain a list of synonymous noun phrases

In addition to collecting verbs for each noun phrase in Step 4 of the algorithm, we collected noun phrases for each verb similarly, obtaining a list of synonymous verbs.

In Step 2-2, we selected synonyms from the candidates obtained using the three methods. The characteristics of these methods are summarized in Table 1. We checked all candidates collected by Methods 1 and 2, because the numbers of candidates were small. The candidates collected by Method 3 were checked in order of similarity to each other as much as possible.

	Reliability	Number of candidates	Target category
Method 1 (Deleted words in Step 1-2)	Fully reliable	Very small	All
Method 2 (Chung)	Highly reliable	Small	Except for "Movement"
Method 3 (distributional similarity)	Moderately reliable	Very large	All

Table 1. Characteristics of the three methods used for collecting synonyms.

Step 3. Collection of Meronymous Words and Attributes for Each Entry Word

In this step, we collected meronymous words and attributes for each entry word using the following two substeps.

(Step 3-1) Collecting candidates of meronymous words and attributes.

(Step 3-2) Identifying whether each candidate is meronymy or attribute.

In Step 3-1, we collected candidates from texts using the pattern "[entry word]のNP₀." For example, if we apply the pattern "サケの(NP₀)" ((NP₀) of salmon) to the patent database, we can obtain "色" (color), "頭" (head), or "フライ" (fry) as candidates for meronymous words and attributes.

We identified manually whether each candidate was a meronymous word, an attribute, or an inappropriate word. In the above example for "サケ" (salmon), "色" (color), "頭" (head), and "フライ" (fry) are identified as

⁵ <https://code.google.com/p/cabocho/>

an attribute, a meronymous word, and an inappropriate word, respectively.

Here, for more efficient identification in Step 3-2, we propose several methods to rerank the candidates for each entry word. In general, most attributes of an ingredient are also attribute of other ingredients. For example, “色” (color) and “鮮度” (freshness) are attributes of most meats, fishes, and vegetables. Therefore, if we were to collect pairs of NP₁ using the pattern “NP₀ の NP₁” (NP₁ of NP₀) from texts and rerank candidates for each entry word in terms of the frequencies of each NP₁, we would expect to find the correct attributes quickly. On the other hand, in the pattern “NP₀ の NP₁” (NP₁ of NP₀), the probabilities that attributes appear in NP₀ are considered smaller than in NP₁. Therefore, if we were to collect pairs of NP₀ using the pattern “NP₀ の NP₁” (NP₁ of NP₀) from texts and rerank candidates for each entry word in terms of the frequencies of each NP₀, we would expect that attributes would be ranked lower, which will be efficient for the identification of meronymous words.

Experiments

In this section, we report on some statistics and experimental results.

Determination of Entry Words

In Table 2, we show the numbers of entry words that were collected using the method mentioned in the previous section.

Category	Number of entry words
Ingredient - seafood	61
Ingredient - meat	6
Ingredient - vegetable	122
Ingredient - other	55
Condiment	51
Kitchen tool	48
Movement	131
Total	474

Table 2. The number of entry words for each category.

Collection of Synonyms for Each Entry word

In Table 3, we show the numbers of synonyms for each category together with the number of synonyms for each entry word.

Category	The number of synonyms (per each entry word)
Ingredient - seafood	453 (7.4)
Ingredient - meat	383 (63.8)
Ingredient - vegetable	947 (7.8)
Ingredient - other	732 (13.3)
Condiment	909 (17.8)
Kitchen tool	643 (13.4)
Movement	956 (7.3)
Total	5,023 (10.6)

Table 3. The number of synonyms for each category.

Collection of Meronymous Words and Attributes for Each Entry Word

CORRECT DATA SET

For some entry words, which were randomly selected in the “Ingredient - seafood” category, we collected 453 candidates of meronymous words and attributes using the pattern “[an entry word]の NP₀” from the patents. Then we identified manually whether each candidate was a meronymous word, an attribute, or an inappropriate word. The results are shown in Table 4. We use this data for our experiment.

Attribute	Meronymous word	Other	Total
146	144	163	453

Table 4. The number of manually identified words for the experiment of identifying meronymous word and attribute.

Alternatives

For confirming the effectiveness of the reranking approach mentioned in Step 3, we examined the following four methods.

- Patent NP₀: Reranking candidates in terms of the frequencies of each NP₀ in the pattern “NP₀ の NP₁” (NP₁ of NP₀) from patent applications.
- Patent NP₁: Reranking candidates in terms of the frequencies of each NP₁ in the pattern “NP₀ の NP₁” (NP₁ of NP₀) from patent applications.
- Recipe NP₀: Reranking candidates in terms of the frequencies of each NP₀ in the pattern “NP₀ の NP₁” (NP₁ of NP₀) from the Rakuten recipe database.

- Recipe NP₁: Reranking candidates in terms of the frequencies of each NP₁ in the pattern “NP₀ の NP₁” (NP₁ of NP₀) from the Rakuten recipe database.

EVALUATION METHOD

Collecting all meronymous terms and attributes are required for constructing a comprehensive ontology. Therefore, we compared the effectiveness of the above methods by precision at 100% recall (when all meronymous terms and attributes are collected). For the calculation of precision, we made use of trec_eval⁶, which is an evaluation tool developed for the Text Retrieval Conference (TREC).

Experimental Results

Experimental results for collecting meronymous words and attributes are given in Tables 5 and 6, respectively.

Methods	Precision
Patent NP ₀	0.452
Patent NP ₁	0.493
Recipe NP ₀	0.411
Recipe NP ₁	0.452

Table 5. Evaluation results of collecting attributes.

Methods	Precision
Patent NP ₀	0.222
Patent NP ₁	0.201
Recipe NP ₀	0.257
Recipe NP ₁	0.215

Table 6. Evaluation results of collecting meronymous words.

⁶ http://trec.nist.gov/trec_eval/

From the results, we can see that “Patent NP₀” obtained the highest precision in collecting attributes, while “Recipe NP₁” was the highest in collecting meronymous words.

Discussion

It is generally considered that the number of attributes is less than that of meronymous words. For example, “色” (color) and “鮮度” (freshness) are attributes of most vegetables, meat, and fish, while there are no meronymous words commonly used for vegetable, meat, and fish. Therefore, we expected that the frequencies of attributes are relatively larger than those of meronymous words, and that we could collect attributes efficiently by “Patent NP₁.” Actually, “Patent NP₁” was superior to others for collecting attributes. To confirm this, we checked top ten words collected by “Patent NP₁,” which are shown in Table 7. In this table, “○” and “×” indicate that a human judge identified each word as “correct” and “incorrect”, respectively. From the table, we can find that some meronymous words were still contained in top ten words. For improving precision of “Patent NP₁” for collecting attributes, using “Recipe NP₀” might be useful. As can be seen from Table 6, “Recipe NP₀” obtained the best performance, and this indicates that meronymous words frequently appear as NP₀ in a pattern “NP₀ の NP₁” (NP₁ of NP₀) rather than NP₁. Therefore, if we degrade the value of frequency of NP₁ words collected by “Patent NP₁” according to the frequency of NP₀ words collected by “Recipe NP₀.”

Candidates	Attribute	Meronymous words
場合 (case)	×	×
表面 (surface)	×	○
間 (while)	×	×
量 (amount)	○	×
状態 (condition)	○	×
水 (water)	×	×
上面 (upper surface)	×	○
種類 (kind)	○	×
部分 (part)	×	○
面 (side)	×	○

Table 7. Top ten words collected by “Patent NP₁.”

Conclusion

In this work, we constructed a cooking ontology that comprises 474 entry words, 5,023 synonyms, 1,512 attributes, and 2,429 meronymous words, using several statistical natural language techniques. We have now published this ontology on the Web (<http://www.ls.info.hiroshima-cu.ac.jp/cooking/ontology.html>).

Acknowledgement

The Rakuten recipe database was provided by Rakuten Data Release from the Rakuten, Inc.

References

- [1] Chung, Y. Finding food entity relationships using user-generated data in recipe service. *Proc. 21st ACM International Conference on Information and Knowledge Management (CIKM2012)* (2012) 2611-2614.

- [2] Hearst, M.A. Automatic acquisition of hyponyms from large text corpora. *Proc. 14th International Conference on Computational Linguistics* (1992) 539-545.
- [3] Lee, L. Measures of distributional similarity. *Proc. 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics* (1999) 25-32.
- [4] Lin, D. Automatic retrieval and clustering of similar words. *Proc. COLING/ACL 1998* (1998) 768-774.
- [5] Mori, S., Maeta, H., Yamakata, Y., and Sasada, T. Flow graph corpus from recipe texts. *Proc. 9th Language Resources and Evaluation Conference (LREC 2014)* (2014).
- [6] Tachibana, A., Wakamiya, S., Nanba, H., and Sumiya, K. Extraction of naming concepts based on

- modifiers in recipe titles. *Proc. International MultiConference of Engineers and Computer Scientists 2014 (IMECS 2014)* (2014)
- [7] Ueda, M., Takahara, M. and Nakajima, S. User's food preference extraction for cooking recipe recommendation. *Proc. 2nd Workshop on Semantic Personalized Information Management: Retrieval and Recommendation* (2011) 98-105.
- [8] Yamakata, Y., Imahori, S., Sugiyama, Y., Mori, S., and Tanaka, K. Feature extraction and summarization of recipes using flow graph. *Proc. 5th International Conference in Social Informatics, LNCS 8238* (2013) 241-254.