Structural Analysis of Cooking Preparation Steps in Japanese

Reiko HAMADA†, Ichiro IDE‡, Shuichi SAKAI†, Hidehiko TANAKA†

†Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan
‡National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan
Email: { reiko | ide | sakai | tanaka }@mtl.t.u-tokyo.ac.jp

Abstract

We propose a method to create process flow graphs automatically from textbooks for cooking programs. This is realized by understanding context by narrowing down the domain to cooking, and making use of domain specific constraints and knowledge. Since it is relatively easy to extract significant keywords from cooking procedures, we create a domain specific dictionary by statistical methods, and propose a structural analysis method using the dictionary. In order to evaluate the ability of the proposed method, we applied the method to actual procedures as an experiment, which showed effective results. The same experiment was also performed on a different program, which showed lower accuracy but also showed realistic results.

Keywords: Cookbooks; Preparation steps; Structural analysis; Domain specific dictionary

1 Introduction

Reflecting the increasing importance of handling multimedia data, many studies are made on indexing to TV broadcast video [2]. Multimedia data consist of image, audio and text, and various researches on analysis of each individual medium have been made. Especially, image analysis has been the main theme for handling multimedia. But recently, it has started to be considered that image analysis alone is insufficient for the semantic analysis of multimedia data. In the 1990s, integrated analysis that supplements the incompleteness of information from each medium has become a trend [1].

Following this trend, we are trying to integrate TV programs with related documents, taking advantage of the relative easiness of extracting semantic structures from text media [3]. Among various programs, educational programs are considered as appropriate sources, since (1) supplementary documents are available, and (2) the video contains plentiful implicit information that integration could be helpful to thorough understanding of supplementary documents. We are currently examining and constructing a practical system using relatively simple elemental technologies by reflecting the result of one medium's analysis to the other. Among educational programs, we will focus on cooking programs, so that we can take advantage of domain specific constraints and knowledge.

In this paper, we will propose the text analysis part of this system and structural analysis of texts from cookbooks will be presented. Although our aim is to integrate information from various sources, the method described in this paper should also be applicable to texts without supplementary video.

Analyzing document structure has been an important subject in the field of natural language processing for a long time. Although many attempts have been made to understand context of documents through structural analysis, it is still a difficult task [4, 5]. In this paper, we propose a method to understand context by narrowing down the domain and making use of domain specific constraints and knowledge. The targets of the method are documents that explain the processes of assembling, producing, or cooking. There are many documents that explain such processes in the form of supplementary textbook, CD-ROM or WWW documents. The amount of independent information lying in these

Proceedings of the 5th International Workshop Information Retrieval with Asian Languages

documents is enormous, but since each document exist independent from others in various forms, it is difficult to make full use of the information synthetically. We try to make a data flow graph from such documents through structural analysis. The graph makes the implicit structure visible and understandable. From this graph, we can normalize the procedures from documents in various forms, so that in the future they could be stored uniformly in a database. Furthermore, the flow graph can be used for many applications, such as optimization of the assembly process or automated cooking system in a kitchen.

In our method, a domain specific dictionary is used for making a practical system for structural analysis. We created this dictionary from text corpora first statistically and manually corrected afterwards.

2 Overview of the Method

In cooking programs, the order of steps often differs slightly between video and textbook. Nevertheless, there are still some restrictions, such as the temporal order of processing materials (a material once processed never returns as it originally was). Therefore, extracting such restrictions from documents is essential for structuring and association. We propose a method to create a process flow graph (Fig. 2.2) from a text (Fig. 2.1) to make such restrictions clear. By this


GRATIN

Ingredients

50mL flour
400mL milk
500g asparagus
8 slices ham

Preparation Steps

1.Melt butter. Blend in flour and Seasonings, stirring constantly. Gradually add milk, simmer until smooth.
2.Boil asparagus until tender-crisp. Wrap the ham around the asparagus.
3. Place noodles, ham rolls in baking dish; cover with sauce. Bake 20 min at 180℃.

Figure 2.1: Example of a supplementary document of a cooking program.

graph, restricted and un-restricted orders could be distinguished clearly (directly linked orders can not be changed), and the structure of a cooking process becomes very clear. Once a graph is created, it is possible to optimize the process, or to gather the video segments corresponding to each node and restructure a new video associated with the document.

Many attempts have been made on structural analysis of Japanese documents. In most cases, it is necessary to use background knowledge on the target document to obtain accuracy. So, in many cases, the application domain is narrowed down to take advantage of domain specific constraints and knowledge. For example, there are works specified to editorials or manuals [6, 7, 8], or on connection of scenes in a story using its grammatical characteristic [5, 9].

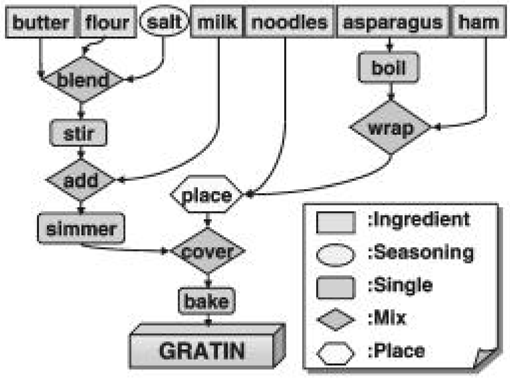

Figure 2.2: Example of a data flow graph.

The number of works on preparation steps from cookbook is very small. One of them is a discussion about the role of propositional sentences including a Japanese particle *'ha'* [14]. Another is on semantic analysis of the name of a dish, using several recipes that have partially the same procedure in common (*Ex.* Extract semantics of 'Salad' from a 'Chicken Salad' recipe and a 'Bacon Salad' recipe) [13].

In the proposed method, we aim at practical structural analysis by making use of constraints and knowledge peculiar to cooking preparation steps. An example of a document that explains a cooking process is shown in Fig. 2.1. The document consists of an "Ingredients" part and a "Preparation Steps" part. The "Ingredients" part can be used to extract nouns such as ingredients and seasonings. The "Preparation Steps" part gives explanation on how to cook the "Ingredients". We aim at analyzing the "Preparation Steps" with the help of "Ingredients" to create a process flow graph. To realize this, first, we create domain specific dictionaries by statistically gathering keywords from many cookbooks and later categorizing and manually correcting them. Next, structural analysis on cooking preparation steps is done using the category of each word in the dictionary.

3 Creation of Domain Specific Dictionary

3.1 Overview

When we create a dictionary, it is necessary to make clear the characteristic of the dictionary, such as size, type and target.

Our aim is to analyze cooking steps, so we need professional knowledge on the target domain (cooking), and general knowledge necessary for basic analysis. We create noun and verb dictionaries with cooking terms as special knowledge, and a general keyword dictionary as general knowledge.

Since cookbooks are explanatory documents, the number of sentence patterns and ambiguity are relatively small compared to general documents. So, nouns, verbs and general keywords are only categorized into several semantic categories. If more information is needed during the analysis, we can get it from other sources referring to the part of speech and the category of the word.

3.2 Creation of Dictionaries

3.2.1 Noun Dictionary

Nouns were categorized into 6 categories as shown in Tab. 3.1. Words that could not be categorized to any of them were omitted from the analysis. Usage of each category is also shown in the Table.

Table 3.1: Categorization of nouns.

Category	Example	Usage
Ingredient	*carrot, chicken*	Targets of structure analysis
Seasonings	*salt, pepper*	
Receptacle	*dish, plate*	Supplementary information to show the situation
Tool	*oven, knife*	
Pronominal noun	*this, it*	All nouns that may indicate ingredients
Action	*cutting in slices*	Nouns that express cooking actions

First, "Ingredients" parts in cookbooks were used to extract nouns that denote ingredients and seasonings. Cookbooks always have a "Ingredients" part, and other textbooks on assembly and production also have lists of materials. The nouns extracted from "Ingredients" of about 880 actual cooking documents is shown in Tab. 3.2, ordered by frequency. The documents were gathered from a single WWW cooking program site.

Next, ingredient nouns those were not in the "Ingredients" part but were in the "Preparation Steps" part needs to be collected. They are gathered from nouns that partially match with ingredients nouns collected from the "Ingredients" part (*Ex*. 'cornflakes' is extracted from 'corn', and vice versa). "Preparation Steps" parts were extracted from 880 recipes by the help of the HTML structure. Since word boundaries are not explicit in Japanese, a morphological analysis system JUMAN [16] was employed for this task. JUMAN outputs morpheme and their parts-of-speech. We extracted nouns and *Katakana* (Japanese phonetic alphabet mostly used to represent non-Japanese words) words, and matched them with ingredients already extracted from the "Ingredients" parts.

Table 3.2: Nouns extracted from "Ingredients".

Noun	Frequency	Noun	Frequency
Salt	864	Soup	147
Oil	437	Leek	144
Pepper	361	Egg	144
Soy sauce	345	Sweet *sake*	139
Sugar	339	Vinegar	132
Sake	287	Sesame oil	105
Ginger	212	White wine	97
Butter	199	Water	92
Garlic	164	Olive oil	88
Onion	164	Bouillon	85
Carrot	155	…	…
Starch	148	Total	11,700

The result was corrected manually, and words that can be ingredients were collected. The words except ingredients were also categorized manually as "Tool" or "Receptacle" or "Action", and words not categorized anywhere were omitted.

Ingredients that appear very frequently might be seasonings, so we categorized the top 8 frequent ones except "Ginger" as seasonings.

Words categorized to "Action" were originally verbs or names of dishes. Many of them indicate specific cooking actions co-occurring with specific verbs, such as "do". So, noun and verb sets co-occurring in a sentence were extracted from the 880 recipes. The result is shown in Tab. 3.4. From Tab. 3.4, we extracted nouns that co-occur with "do" more than twice, excluding tools and receptacles. Among the rest, we categorized the words suitable as "Action" and got 45 words. Adding 35 words that were categorized when we categorized tools and receptacles, 80 "Action" verbs were collected.

Table 3.3: Final vocabulary of the noun dictionary.

Category	Vocabulary	Category	Vocabulary
Ingredients	1,636	Tools	132
Seasonings	7	Pronominal	10
Receptacles	56	Action	89
Total			1,930

After manual addition and correction, the final vocabulary of the dictionary was as shown in Tab. 3.3.

Table 3.4: Result of extracted noun-verb sets by co-occurrence frequency.

Noun	Verb	Frequency
器(dish)	盛る(serve)	480
皮(peel)	剥く(peel)	377
水気(moisture)	きる(remove)	315
千切り (cutting into strips)	する(do)	298
大さじ(tablespoon)	熱する(heat)	278
塩(salt)	ふる(sprinkle)	274
塩(salt)	する(do)	274
薄切り(slice)	する(do)	265
微塵切り(cut fine)	する(do)	262
油(oil)	熱する(heat)	251
長さ(length)	切る(cut)	242
塩(salt)	加える(add)	226
鍋(pot)	入れる(put)	221
…	…	…
TOTAL		61,350

3.2.2 Verb Dictionary

Verbs were categorized as shown in Tab. 3.5. Explanation of each category is also shown in the Table. "Causative verb" is a Japanese auxiliary verb, but they co-occur with other words and imply many cooking terms (*Ex.* Make water boiled), so they could be considered as actions, and were categorized as verbs.

As noted in the previous Section, we applied the morphological analysis tool JUMAN to "Preparation Steps" of 880 recipes. We extracted verbs and categorized them manually. The result is shown in Tab. 3.6. "Error" and "General" in Tab. 3.6 were excluded, but some intransitive verbs can become cooking actions with a causative verb, so some of the intransitive verbs were registered to the dictionary.

Table 3.5: Categories of verbs.

Category	Example	Explanation
Single	*roast, cut*	Cook or process single ingredient
Mix	*add, sprinkle*	Possible to mix several ingredients
Separate	*divide, peel*	Possible to divide or separate an ingredient
Put	*put, place*	Place an ingredient on receptacles or other places
Polysemy	*spread, return*	Have several meanings over categories
Employment	*make*	Co-occur with various words and represent cooking action

Table 3.6: Verbs extracted from "Preparation Steps"

Category	Example	Frequency
Action	*cut, roast, mix*	379
Intransitive verb	*be torn off , crumble*	125
General	*decide, know, like*	35
Error	*misprinted words*	78
Total		617

Since word matching is used to detect words in documents, all conjugations of verbs were registered in the dictionary. After manual addition and correction, the final vocabulary of the dictionary was as shown in Tab. 3.7.

Co-occurring expressions gathered as shown in Tab. 3.4 were also registered to the dictionary. Noun and verb sets those the meaning of the verb can be determined by the noun were collected.

Table 3.7: Final vocabulary of the verb dictionary.

Category	Vocabulary	Category	Vocabulary
Single	231	Place	36
Mix	70	Polysemy	6
Separate	44	Employment	2
Total			389

3.2.3 General Keywords

In addition to domain specific nouns and verbs, there are many useful keywords for structural analysis. We categorized these words according to Tab. 3.8 considering grammatical rules [15].

Table 3.8: Categories of general keywords.

Category	Example
Addition	still more, addition to
Condition	if, in the case of
Time	just before, after, when
Negation	no, not
Particle	*ha, ga, no, wo* (Japanese particles)
Conjunction	*ta, da, te, de* (Japanese conjunctions)

Since automatic creation of dictionaries is difficult because they should be precise to be reliable for analysis, final check must be done manually. But even if the target of the dictionary is reduced to cooking terms, they have a wide variation and sometimes there are many types of representation for one term (*Ex.* sometimes a fish has several names in Japanese). To cover all of these terms, statistical collection is realistic compared to manual creation, and we are planning to use statistic methods such as TF-IDF or make use of existing dictionaries to reduce the burden.

4 Structural Analysis of Preparation Steps

4.1 Process of Structural analysis

The outline of the structural analysis is shown in Fig. 4.1. The dictionaries created in the previous Chapter were used to analyze cooking recipes. Our final goal is to create a flow graph as shown in Fig. 2.2. In this paper, visualization itself is not the main topic, so we will focus on the structural analysis.

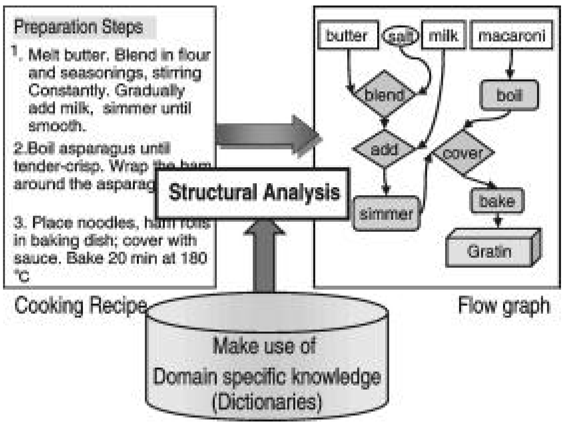

Figure 4.1: Outline of structural analysis.

Here, the result of the structural analysis is shown as an operation flow graph as shown in Fig. 4.3. At the center of Fig. 4.3, cooking actions (operations) are arranged vertically in order. In the left and the right side of them, there are ingredients and seasonings as data. An ingredient or seasoning with an arrow pointing toward the center indicates an input data, and an opposite arrow indicates an output data. Since only ingredients and seasonings in the graph is the target of structural analysis, extra information such as tools, receptacles or conditional sentences are just added in the graph as supplementary information.

The next 4 processes are performed to analyze the structure.

1. Extract keywords and their categories by matching with the dictionaries
2. Make noun-verb sets
3. Make blocks by connecting sets
4. Connect blocks

Details of each process are explained in the following Sections.

4.2 Process 1: Extract Keywords and Their Categories by Matching with the Dictionaries

First, as shown in Fig. 4.2, words in the dictionaries and step numbers are extracted and the categories in the dictionaries are tagged. Uncategorized words are labeled as "Others".

If there are noun and verb sets in one sentence registered in the co-occurrence dictionary, the category of the verb is rewritten to a new one defined in the co-occurrence dictionary. By using the co-occurrence with a noun, we can disambiguate the semantics of verbs with various meanings.

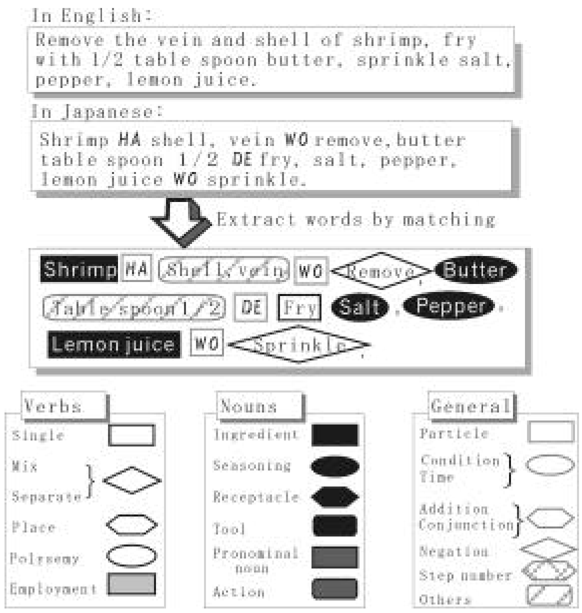

Figure 4.2: Process 1. Extract words by matching with the dictionaries.

4.3 Process 2: Make Noun-Verb Sets

In the next process, we make noun-verb sets considering that a verb modifies the nearest noun, satisfying the no-cross condition [12]. A set consists of noun(s) and nearest verb(s) and words between them. We deduced the role of the sentences whether they indicate condition, time relation or declaration of a dish name, referring to Tab. 3.8, and excluded them from the sets.

Whether each ingredient is added or removed is determined by the particle and the category of the verb. This corresponds

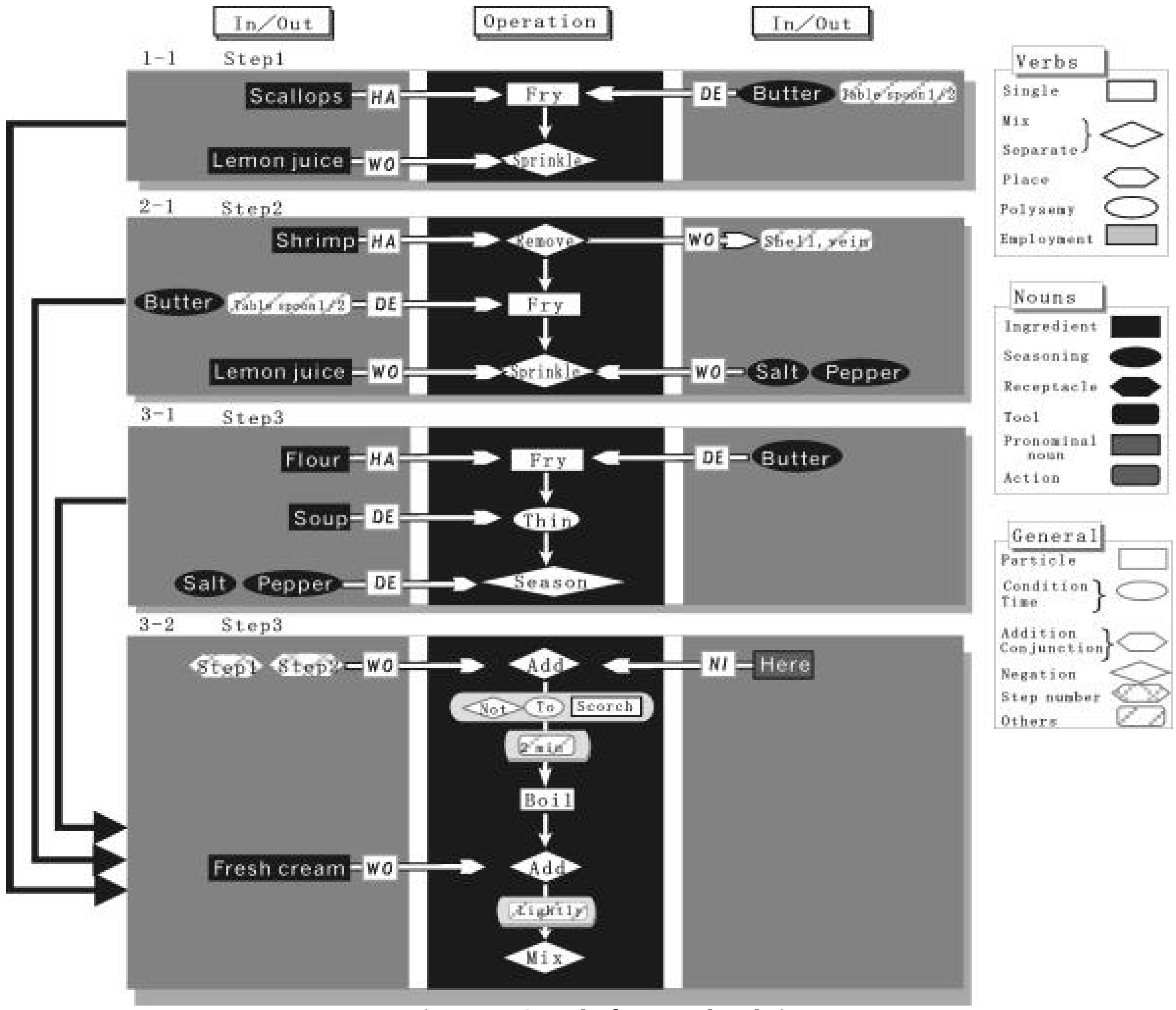

Figure 4.3: A result of structural analysis.

to the input and output of data in the operation flow. Here, data is either ingredient, seasoning, pronominal noun or step number. Operation is only verbs including co-occurred expressions. Other words that belong to tools or receptacles or sentences that explain the situation are placed at the side of the corresponding positions in the graph.

4.4 Process 3: Make Blocks by Connecting Sets

Sets that are considered to be continued from the previous set in the same step are connected and form a block with a new number.

If the first verb of a set does not satisfy the case frame, it is considered to be continuous to the previous set. "Mix" or "Put" verbs need more than 2 nouns and "Single" verbs need more than 1 noun to satisfy the case frame. At the same time, we consider not only the first verb in the set, but also those in the middle to connect sets.

In blocks formed in this step, several ingredients are mixed and cooked, and each block becomes a new intermediate state.

4.5 Process 4: Connect Blocks

In the last process, each block is connected to the nearest block which has common ingredients or step numbers. If the first set of the next block contains an "Addition" word, a pronominal noun, or has no ingredient, this block and the previous one is connected.

When searching common ingredients in 2 blocks, we refer to the hierarchical semantic structure of a conceptual dictionary[1] and consider the matching with upper concepts (*Ex.* "carrot" will also be referred as "vegetable").

With this last process, the structural analysis is finished. An example of the result of structural analysis is shown in Fig. 4.3.

5 Evaluation of Structural Analysis

We applied the structural analysis method introduced in the previous Chapter to 32 recipes (135 steps in total) chosen at random from a single WWW site and performed an evaluation experiment. We used the dictionaries created in Chapter 3.

We evaluated the extraction accuracy of words (Process 1), noun-verb sets (Process 2), block structure of each step (Process 3), connection of blocks (Process 4), and total evaluation of all the steps. Only the structure of ingredients, seasonings, pronominal nouns and verbs were evaluated and other words such as tools were not considered.

Suppose that the result of human analysis is Ans_H, the result of automatic analysis is Ans_M and the answer in common is Ans_C, then recall is Ans_C/Ans_H, and precision is Ans_C/Ans_M.

Table 5.1: Result of evaluation experiment (Process 1, 2 and 4).

Contents	Recall	Precision
Process 1: Extract words	99%	98%
Process 2: Extract noun-verb sets	98%	97%
Process 4: Connect blocks	91%	92%

Table 5.2: Result of evaluation experiment (Process 3 and Total evaluation).

Contents	Accuracy
Process 3: Extract structure of blocks	87%
Total result of all steps	82%

Tab. 5.1 shows that the extraction rate of words and noun-verb sets are nearly 100%, and connection of blocks is more than 90%. Next, the success rate of block extraction and the total result of 135 steps is shown in Tab. 5.2. The result shows about 90% of the structure of all the steps are extracted, and more than 80% of the structure and connection of all the steps are extracted. This result shows that the proposed structural analysis method works in good accuracy on cooking recipes.

The previous experiment was performed to recipes from a single TV program. The dictionaries were created and the analysis method was developed from it. So, we performed the same evaluation experiment to recipes from another TV program to evaluate its generality. These recipes were also selected at random from a supplementary textbook. Other conditions were same as the previous experiment.

Table 5.3: Result of evaluation experiment from a different program.

Processes 1, 2 and 4	Recall	Precision
Process 1: Extract words	92%	93%
Process 2: Extract noun-verb sets	80%	84%
Process 4: Connect block	80%	82%
Process 3 and Total Evaluation	**Accuracy**	
Process 3: Extract structure of blocks	52%	
Total result of all steps	46%	

The result is shown in Tab. 5.3. According to Tab. 5.3, accuracy is much lower than in Tab. 5.1 and 5.2. This is mainly because there are many slight syntactic differences in the vocabulary.

Next, we added the lacking vocabulary to the dictionaries, and performed the same experiment to the same recipes. The result is shown in Tab. 5.4. The result has obviously improved than in Tab. 5.3, but still worse than in Tab. 5.1 and 5.2.

Table 5.4: Result of evaluation experiment on recipes from a different program with revision of dictionaries.

Processes 1, 2 and 4	Recall	Precision
Process 1: Extract words	98%	97%
Process 2: Extract noun-verb sets	96%	94%
Process 4: Connect block	82%	80%
Process 3 and Total Evaluation	**Accuracy**	
Process 3: Extract structure of blocks	78%	
Total result of all steps	60%	

These results show that vocabulary is so important that we must add new vocabulary to the dictionaries every time we analyze texts in a different series of recipes. We are expecting to overcome this problem by gathering new words statistically, or by creating a dictionary beforehand by making

[1] EDR electronic dictionary [17]

use of large-scale dictionaries. In addition to that, recipes from a different program failed more than the original recipes, although the reasons were mostly the same.

The reasons of the failures were as follows:

- Failure of word matching
- Sentences were too complicated to make noun-verb sets
- Lack of hierarchical semantic knowledge
- Complicated indications such as multiple indication
- Lack of considering continuity of actions
- Lack of situation analysis considering times and places
- Complicated branch of flow graph

In the future, we aim to improve the accuracy reffering to these reasons.

6 Conclusion

We proposed a structural analysis method for documents that explain cooking process, to create its data flow graph automatically. This is realized by making use of domain specific dictionaries. The creation processes of these dictionaries were described and the effectiveness of the proposed analysis method was shown through experiments.

The data flow graph of documents will be useful for integration with video and other applications such as optimizing the process or automated cooking system in the kitchen.

Furthermore, through the usage of a supplementary document and its analysis, we aim for proposing a novel advanced multimedia integration method. Using the result of this method, we will also propose an integrative restructuring method of the multimedia data provided both from the video and the supplementary document in the future.

1. D. Ponceleon, S. Sriniivasan, A. Amir, D. Petkovic, D. Diklic: "Key to Effective Video Retrieval: Effective Cataloging and Browsing", *Proc. ACM Multimedia'98*, pp.99-107, Aug. 1998.
2. H. D. Wactlar, A. G. Hauptmann, M. G. Christel, R. A. Houghton, A. M. Olligschlaeger: "Complementary Video and Audio Analysis for Broadcast News Archives", *Comm. ACM,* Vol.45, No.2, pp.42-47, Feb. 2000.
3. R. Hamada, I. Ide, S. Sakai, H. Tanaka: "Associating Video with Related Documents", *Proc. ACM-Multimedia'99, Vol. 2*, pp.17-20, Oct. 1999.
4. H. Fukumoto, J. Fukumoto, Y. Suzuki: "An Automatic Extraction of Key Paragraphs Based on the Degree of Context Dependency (in Japanese)", *Journal of Natural Language Processing*, Vol.4, No.2, Apr. 1997.
5. Y. Ogisawa, N. Inui, Y. Kotani, H. Nishimura: "Analyzing Stories Based on Relation between Two Sentences (in Japanese)", *IPSJapan SIG Notes* NL-113-15, May 1996.
6. J. Fukumoto, H. Yasuhara: "Text Structure Analysis Based on Sentence Cohesion (in Japanese)", *IPSJapan SIG Notes* NL-88-2, Mar. 1992.
7. T. Tanaka, Y. Hayashi: "A Japanese Text Analysis Method Based on Sentence Adjust Patterns (in Japanese)", *IPSJapan SIG Notes* NL-89-1, May 1992.
8. N. Tamura, K. Wada: "Text Structuring by Composition and Decomposition of Segments (in Japanese)", *Journal of Natural Language Processing*, Vol.5, No.1, pp.59-78, Jan. 1994.
9. T. Yamamoto, A. Yamada, T. Nishida, S. Doshita: "Spatial Concept Analysis in Japanese Active Description ~Plot Recognization of Japanese Story~ (in Japanese)", *IPSJapan SIG Notes* NL-87-7, Jan. 1992.
10. K. Yoshikawa, N. Funato, K. Hanada, M. Miyamoto: "Context Analysis Based on Aspect Information (in Japanese)", *IPSJapan SIG Notes*, NL-82-11, Mar. 1991.
11. H. Ueda, N. Takiguchi, Y. Kotani: "Automatic Generation of Semantic Structure from a Fairy Tale 'Momotarou' (in Japanese)", *IPSJapan SIG Notes*, NL-84-4, July 1991.
12. S. Kurohashi, M. Nagao: "A Syntactic Analysis Method of Long Japanese Sentences Based on Coordinate Structures Detection (in Japanese)", *Journal of Natural Language Processing*, Vol.1, No.1, pp.3-20, Mar. 1994.
13. H. Adachi: "How to Acquire Cooking Definitions Based on Analogical Reasoning (in Japanese)", *IPSJapan SIG Notes*, NL-112-9, Mar. 1996.
14. F. Aoyama: "The Role of Propositions in Cooking Sentence (in Japanese)", *KeiryouKokugogaku-to-NihongoShori ~Riron to Ouyo~*, Prof. S. Mizutani Memorial Assos. Ed., pp.285-303, Akiyama Pub, 1987.
15. H. Furuya, H. Nakai, S. Yamazaki Ed.(in Japanese): "The Synthetic Japanese Text for Junior High School -revised and enlarged edition-", *Sho-shin-sha*.
16. Language Media Lab, Grad. School of Informatics, Kyoto Univ.: "Japanese Morphological Analysis System JUMAN version 3.6", Nov. 1998.
17. Japan Electronic Dictionary Research Institute (EDR): "EDR Electronic Dictionary version 1.5"