Structural Analysis of Preparation Steps on Supplementary Documents of Cultural TV Programs

Reiko HAMADA, Ichiro IDE, Shuichi SAKAI, and Hidehiko TANAKA
Graduate School of Electrical Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
TEL: +81-3-5841-7413, FAX: +81-3-5800-6922
{ide,reiko,sakai,tanaka}@mtl.t.u-tokyo.ac.jp

Abstract

We propose a method to create process flow graphs automatically from documents which explain the process of assembling something. This is realized by understanding context by limiting the subject and making use of domain specific characteristics. As an example, we restrict the domain to cooking procedures, which is comparatively easy to retrieve significant keywords for the domain. We will also introduce a future plan to make the dictionary automatically from large corpora.

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 has 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 semantic analysis of multimedia data. In the 1990's, 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. Among various programs, cultural programs are considered as appropriate sources since (1)supplementary documents are available and (2)the video contains a lot of implicit information that integration could be helpful to thorough understanding of supplementary texts. 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. 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. Although our aim is to integrate information from various sources, the proposal in this paper should be applicable to texts without supplementary video.

Analyzing document structure has been an important subject in the field of artificial intelligence for a long time. Although many researches have been done to understand context of documents through structure analysis, it is still a difficult problem[3, 4].

In this paper, we propose a method to understand context by limiting the subject and making use of domain specific characteristics. Our targets are documents which explain the process of assembling, producing, or cooking something. There are many documents which explain such processes in the form of supplementary textbook, CD-ROM or WWW. Information lying in these documents is enormous, but since each document exists independent from each other in different forms, it is difficult to make full use of the information systematically. We try to make a process flow graph from these documents through structure analysis. The graph makes the implicitly structured information visible and understandable. From this graph, we can normalize the procedures from differently formatted documents, so that in the future they could be used to make a database. Furthermore, the flow graph can be used for many applications, such as optimization of the assembly process.

In our method, a domain specific dictionary is used for making a practical system for structure analysis. We are planning to create this dictionary automatically from text corpora by statistical analysis in the future.

2 Structural Analysis of Preparation Steps

In this section, we propose a structural analysis method of preparation steps in cooking textbooks.

An example of a document that explains a cooking process is shown in Fig.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" and create a process flow graph.

In cooking programs, the order of steps often differ between video and textbook. Nevertheless, there are still some restrictions, such as the time flow of processing materials (A material once processed never returns as it was). Therefore, extracting such restrictions from documents is essential for structuring and association. We propose a method to create a data flow graph to make such restrictions clear, as shown in Fig.2. By this 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 the 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.

2.1 Categorization of Cooking Terms

In our method, we only use nouns and verbs as keywords to make a data flow graph. We categorize nouns and verbs according to Tables 1 and 2. Words which could not be categorized to any of the categories in the Tables are omitted from the analysis.

As shown in Table 1, we categorize nouns to Ingredients([Ing]), Seasonings([Sea]), Receptacles, and Tools. In this experiment, we only use ingredients and seasonings for analysis. In future, we are planning to use receptacles to know which ingredients are put together in a receptacle.

Figure 1: Example of supplementary document for a cooking program.

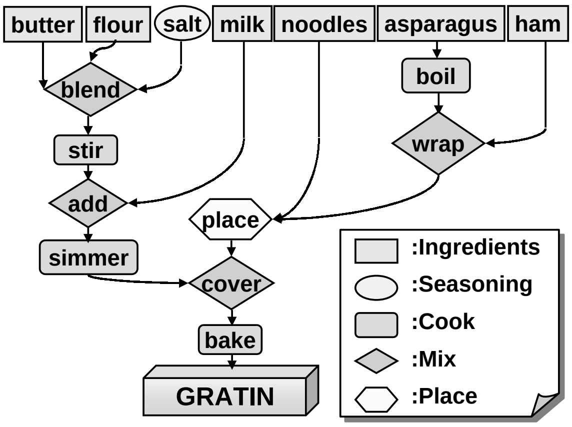

Figure 2: Example of a data flow graph.

As shown in Table 2, we categorize the verbs into "Cook", "Blend", "Separate" and "Place". But in this experiment, we categorize these types into 2 big titles, [Sol] and [Mix]. [Sol] includes "Cook" verbs and [Mix] includes "Blend", "Separate" and "Place" verbs. [Sol] cooks single ingredients, such as 'Bake' and 'Cut'. On the other hand, "Blend" mixes several ingredients, such as 'Add', 'Mix' and 'Sprinkle', "Place" puts ingredients in receptacles, such as a 'baking dish'. If several ingredients are put in a dish, these ingredients are mixed in the dish. So we categorized "Place" verbs in the [Mix] category. In the case of a "Separate" verb, it is clear that originally, these ingredients were one ingredient. So, "Separate" verbs are also categorized in the [Mix] category.

We are planning to make a domain specific noun and

Table 1: Categorization of Noun.

Noun	Ingredients	Seasonings	Receptacles	Tools
Usage	[Ing]	[Sea]	(Currently not used.)	
Example	*Carrot* *Chicken*	*Salt* *Pepper*	*Dish* *Plate*	*Oven* *Knife*

Table 2: Categorization of Verb.

Verb	Cook	Blend	Separate	Place
Usage	[Sol]	[Mix]		
Example	*Bake* *Cut*	*Add* *Mix*	*Separate* *Divide*	*Place* *Put*

verb dictionary, in which words are categorized according to Tables 1 and 2. In the rest of this Section, we will discuss our method assuming that an ideal dictionary already exists.

2.2 Structural Analysis Experiments

We will introduce the process of structural analysis for creating a data flow graph. The steps to create the flow graph is as follows:

Step1. **Extract nouns and verbs that the document has in common with the dictionary.**
Extract nouns and verbs in the document, then make "Noun - Verb" sets. Nouns and verbs are extracted by matching with words in the dictionary. Words that do not exist in the dictionary are ignored. We consider that a verb modifies the nearest noun, satisfying the no-cross condition[5].

Step2. **Create intermediate states referring to verbs.**
The sets with [Mix] verbs are connected with any previous set. New numbers are assigned to each of the new states in this process.

Step3. **Connect intermediate states referring to nouns.**
The intermediate states created in the previous step are connected and data stream is completed in this step. First, if there is a "process number" in the process, all the states in that process are connected with the precess with the "process number". Next, other states are connected with the nearest state which has the same "ingredients". And finally, if there are some states left, these are connected with the nearest states which has the same "seasonings". If the categorization of "ingredients" and "seasoning" in the dictionary is correct, we may not need the last rule. But, there may be mistakes in the dictionary, such as miscategorization of an ingredient as a seasoning or vice versa, since it is difficult to statically categorize some materials as an ingredient or a seasoning.

An example of the creation process is shown in Fig.3. We actually applied the method to Japanese documents, so noun and verb in the example is in reverse order to that in English.

2.3 Preliminary Experiment

We applied the previous steps to an actual supplementary document for a cooking program. The text book and its dictionary is showed in Fig.4. The actual experiment was performed to a Japanese document, and the English sentences in the following experiment are translations from a Japanese textbook. Note that the domain specific dictionary used in the experiment was manually created in this case.

The data flow graph of the result is shown in Fig.5. The flow of the cooking procedure is visible and easily understandable from the graph. It is shown that the proposed method works at least for an easy cooking textbook.

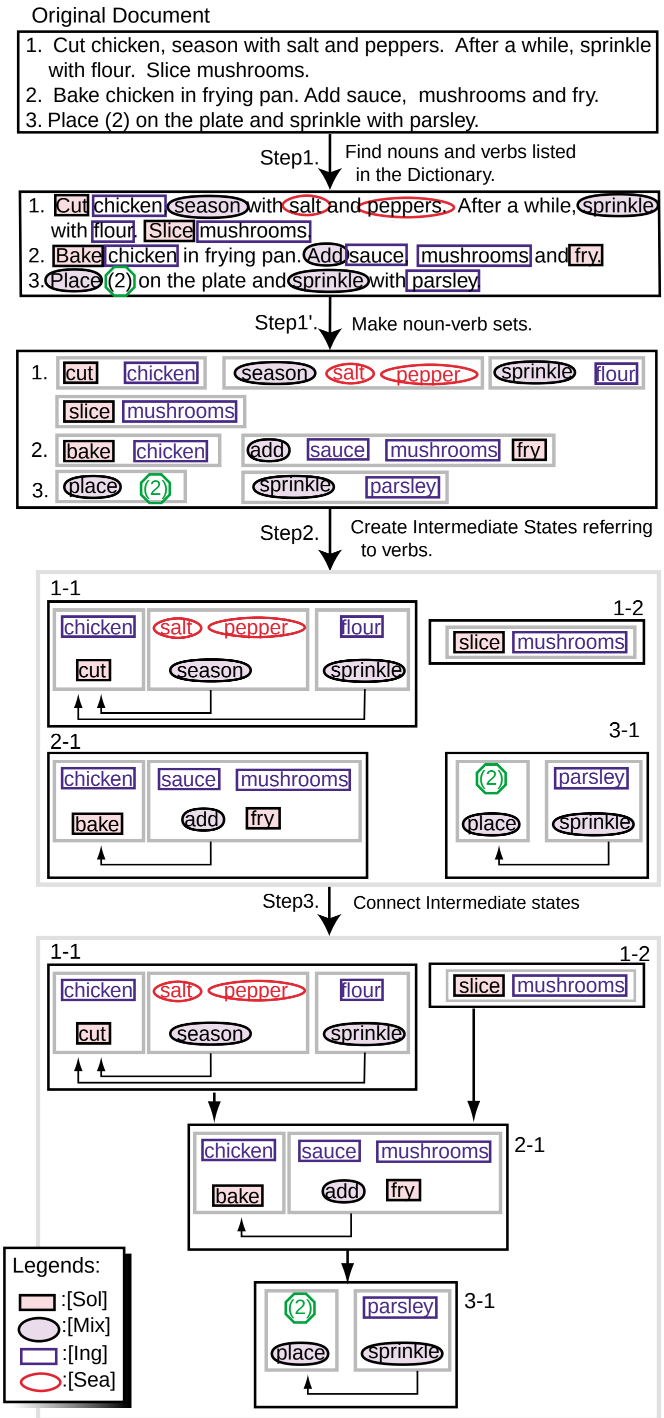

Figure 3: Structuring Preparation Steps in Cooking Textbook

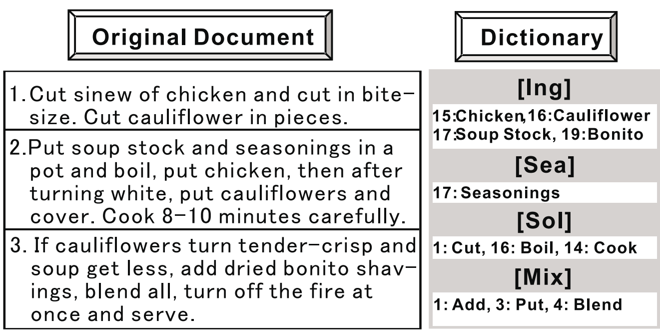

Figure 4: The Document and Dictionary for Preliminary Experiment.

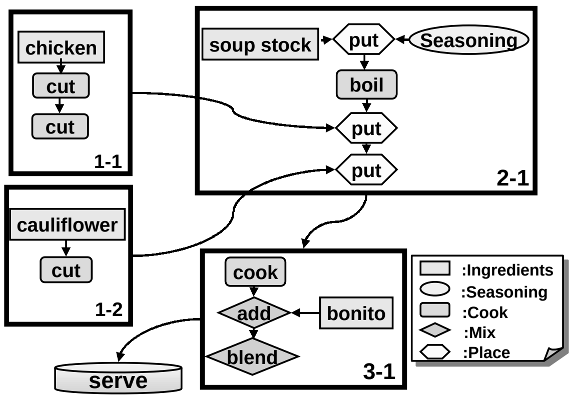

Figure 5: Data flow graph from Result of a Preliminary Experiment.

3 Automatic Retrieval of Domain Specific Dictionary

3.1 Overview

In the previous section, we assumed that an ideal dictionary already exists. But, creating a dictionary in each domain manually is burdensome. So, we are attempting to automatically create a domain specific dictionary.

The text which explain how to make something uses characteristic keywords compared to general sentences. For example, there are few words such as 'cook', 'boil', 'bake' and 'chicken' in newspapers. Then, it is relatively easy to extract such characteristic keywords from these documents statistically. We are now investigating several ideas introduced in this Section.

3.2 Automatic Retrieval of Cooking Terms

1. **Extraction of Nouns from "Ingredients"** Cooking textbooks always have a "Ingredients" part, and other textbooks on assembly and production also have lists of materials. We are planning to extract nouns from these lists. The result of nouns extracted from "Ingredients" of actual cooking documents is shown in Table 3. The list is ordered by frequency. We gathered the documents from WWW sites.

Table 3: Noun Lists Extracted from "Ingredients"

Noun	Freq.	Noun	Freq.
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

2. **Extraction of Verbs from "Preparation Steps"** Verbs can be extracted from "Preparation Steps". Words are not separated in Japanese sentences, so we must detect the boundaries. We used the morphological analysis system "JUMAN"[6] for this task. "JU-

MAN" outputs morpheme and their parts-of-speech. In the "Preparation Steps", there are many general words that can not be used as keywords. So, we use TF-IDF to separate these general words. Using a general corpus S, and a specific domain's document C, we can determine the TF-IDF as:

$$f_{TF-IDF} = \frac{Freq_c / All_c}{Freq_s / All_s} \quad (1)$$

Where All_s and All_c stand for all the verb's number in the corpus and the document, and $Freq_s$ and $Freq_s$ the target verb's frequency in the corpus and the document. We are planning to remove the words which get lower marks in TF-IDF and extract the true keywords of each domain. As a general corpora, newspaper articles in RWC text database (RWC-DB-TEXT-95-2) [7] were used.

The verbs ordered by frequency in the "Preparation Steps" are shown in Table 4. Many cooking terms in the document are so domain specific, that they never appear in newspaper articles, thus get infinity score in f_{TF-IDF}, so it is more significant to show the words with lower score than higher score as shown in Table 5. The words listed in the Table does not seem to be suitable as cooking keywords. We will remove these terms from the dictionary.

Table 4: Result of Extraction of Verbs from Documents.

Verb		Freq.
する	(Do)	3,400
加える	(Add)	2,056
入れる	(Put)	1,976
切る	(Cut)	1,342
炒める	(Fry)	858
かける	(Put over)	819
混ぜる	(Mix)	733
とる	(Take)	696
つける	(Put on)	636
盛る	(Serve)	584
ふる	(Sprinkle)	525
熱する	(Heat)	505
むく	(Peel)	444
きる	(Remove)	438
ゆでる	(Boil)	413
煮る	(Boil)	395
まぶす	(Sprinkle)	325
焼く	(Roast)	320
洗う	(Wash)	319
合わせる	(Add)	306
...		...
TOTAL		30,161

3. **Extraction of Associated Noun-Verb Pairs.** There are some expressions that specific nouns and verbs associatively function. They can not be analyzed properly individually. Futhermore, from these associated pairs, sence of nouns or verbs can be guessed mutually. We extracted these associations by cooccurrence frequency. It could generally be considered that a verb modifies the nearest noun and we counted such noun and verb pairs' cooccurrence frequency. List of the result is shown in Table 6. As shown in Table 6, reasonable pairs were extracted.

4. **Categorization of Words by cooccurrence Frequency.** We can estimate the sense of verb or noun from associated pairs if the category of the other side of the

Table 5: Result of TF-IDF (Lowest Score).

Verb		f_{TF-IDF}
いる	(Be)	0.0031
言う	(Say)	0.0049
受ける	(Receive)	0.0058
知る	(Know)	0.0071
いう	(Tell)	0.0099
見る	(See)	0.010
続ける	(Continue)	0.011
やる	(Do)	0.013
かかる	(Hang)	0.014
決まる	(be Decided)	0.018
生きる	(Live)	0.021
できる	(Can)	0.022
買う	(Buy)	0.023
よる	(Approach)	0.025
求める	(Seek)	0.027
待つ	(Wait)	0.027
くる	(Come)	0.028
当たる	(Hit)	0.029
集める	(Gather)	0.029
感じる	(Feel)	0.033
...		...

verb or noun is known. For example, seasonings almost always cooccur with a "Mix" category verb, or receptacles cooccur with a "Place" category verb.

First, we are planning to build a core dictionary which contains few correctly categorized basic terms. Using the previously introduced statistical methods, the vocabulary of each categories will be extended from the core dictionary.

4 Conclusion

We proposed a structure analysis method for documents which explain product process to create a data flow graph automatically. This is realized by making use of a domain specific dictionary. We showed the effectiveness of the proposed method by a preliminary experiment. Furthermore, we are planning to create the domain specific dictionary automatically in the future.

The data flow graph of documents are useful for integrating with video and other applications such as optimizing the process.

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.

Acknowledgment

RWC text database is a product of Real World Computing Partnership (RWCP), and was used under a licensed agreement.

References

[1] D. Ponceleon, S. Sriniivasan, A. Amir, D. Petkovic, D. Diklic; "Key to Effective Video Retrieval: Effective Cataloging and Browsing", *ACM Multimedia'98*, pp.99-107, Aug., 1998.

Table 6: Result of Extracted Association Sets by Cooccurrence Frequency

Noun		Verb		Freq.
器	(dish)	盛る	(serve)	480
皮	(peel)	むく	(peel)	377
水気	(moisture)	きる	(remove)	315
せん切り	(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
みじん切り	(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
半分	(half)	切る	(cut)	200
ふた	(cover)	する	(do)	179
電子レンジ	(microsave oven)	かける	(bake in)	178
しょうが	(ginger)	する	(do)	170
大さじ	(tablespoon)	入れる	(put)	170
ボール	(bowl)	入れる	(put)	163
こしょう	(pepper)	する	(do)	163
...		...		...
TOTAL				61,350

[2] Wactler, H. D., Hauptmann, A. G., Witbrock, M. J.; "Informedia News-on-Demand: Using Speech Recognition to Create a Digital Video Ligrary", *CMU Tech. Rep.*, CMU-CS-98-109, Mar, 1998.

[3] H. Fukumoto, J. Fukumoto, Y. Suzuki: "An Automatic Extraction of Key Paragraphs Based on the Degree of Context Dpendency(in Japanese)", *Journal of Natural Language Processing*, Vol.4, No.2, Apr., 1997.

[4] Y. Ogisawa, N. Inui, Y. Kotani, H.Nisimura: "Analyzing Stories based on Relation between Two Sentences(in Japanese)", *IPSJ SIG Notes*, 96-NL-112, Mar., 1996.

[5] S. Kurohashi, M. Nagao: "A Syntactic Analysis Method of Long Japanese Sentences based on Coordinate Structures' Detection", *Journal of Natural Language Processing*, Vol.1, No.1, Mar., 1994.

[6] Language Media Lab:, Grad. School of Informaatics, Kyoto Univ.; "Japanese Morphological Analysis System JUMAN version 3.6", Nov 1998.

[7] Real World Computing Partnership (RWCP); "RWC Text Database", Mar, 1996.