Automatically Extracting Important Sentences from Story based on Connection Patterns of Propositions in Propositional Network

HIDEJI ENOKIZU
Shibaura Institute of Technology
3-7-5 Toyosu, Koutou-ku, Tokyo
Japan
enokizu@shibaura-it.ac.jp

SATOSHI MURAKAMI
Oki Electric Industry Co., Ltd
4-10-16 Shibaura, Minato-ku, Tokyo
Japan
murakami724@oki.com

MORIAKI KUMASAKA
Shibaura Institute of Technology
Graduate School of Engineering
3-7-5 Toyosu, Koutou-ku, Tokyo
Japan
m105043@shibaura-it.ac.jp

KAZUHIRO UENOSONO
Shibaura Institute of Technology
Graduate School of Engineering
3-7-5 Toyosu, Koutou-ku, Tokyo
Japan
k-uenosono@komiya.ise.shibaura-it.ac.jp

SEIICHI KOMIYA
Shibaura Institute of Technology
Graduate School of Engineering
3-7-5 Toyosu, Koutou-ku, Tokyo
Japan
skomiya@shibaura-it.ac.jp

Abstract: In recent years, the world is filled with a large amount of information through the internet and so on. Such a situation increasingly enhances the worth of the automatic text summarization which can support a quick grasp of the text content. The automatic text summarization has so far been accomplished by extracting some important sentences from a text based on various surface cues. To be compared, we tried to devise a new method to extract the important sentences from the story according to the way in which the people comprehend it. In designing this new method, we took account of the text comprehension model, that is, how people comprehend the story text. Then we devised the procedure for transforming from a set of the propositions to the propositional network. In Experiment 1, the participants were asked to select the sentences regarded as important from five stories. Then we examined how the propositions drawn from each important sentence were connected in the propositional network of each story. As a result, we identified three distinctive connection patterns. In Experiment 2, it was examined whether those connection patterns are valid as the rules to extract the important sentences from five new stories. From the sentences extracted our system and the important sentences selected by the participants, we calculated the aggregation accuracy measures. As a result, it was found that they were clearly higher than the baselines. Moreover they were equal to or higher than ones obtained in the previous researches. This finding was replicated to the stories used in Experiment 1.

Key–Words: Automatic Text Summarization, Extracting Important Sentences, Extended Rules of Connecting Propositions, Propositional list, Propositional network, Connection Patterns

1 Introduction

The world is currently filled with a vast amount of information through the internet so on. It regretted, however, that the human's information processing capability can't improve so much even if the information which must be handled increases. Therefore, it is necessary to control the quantity of the text read by the human using the technology of the automatic text summarization. The automatic text summarization, which can help the human grasp the necessary information quickly, increased in importance as the fundamental technology to support the human intellectual activity [1].

There are several kinds of approaches in the researches on the automatic text summarization [2].

The classical approaches try to extract the important sentence from a text based on several surface cues: the word frequencies, the critical phrases (for example, "significant," "impossible," and etc.), the title and the heading words, and the sentence location. And the corpus-based approaches try to determine the importance of various text features based on the frequencies which such features occur in the text corpora (well-known measure is tf.idf). Although these two approaches focus on the surface cues, the discourse structure approaches exploit the discourse models while they are relatively domain-independent and knowledge-poor. The discourse models are likely to be useful because it appears to play an important role in the strategies used by human summarizers and in the structure of their summaries. In other words, a summary is not just a collection of the sentences, each of which represents some salient information in the text. In addition, the knowledge-rich approaches focus primarily on the structured information (for example, data and knowledge bases) as a starting point for summarization. Although the knowledge-rich approaches are interesting, they are just going to be developed theoretically. According to these approaches, a number of researchers have investigated the automatic text summarization until now. However, they have not obtained the fully satisfied results.

In the present research, we propose a new way of automatic text summarization that extract the important sentences from the story based on the text comprehension models. Our approach is most relevant to the discourse-structure approaches. However, its distinctive feature is to device some connection rules, which make the text coherent, on the basis of various text comprehension models. Especially Kintsh's text comprehension model is the core one given us some ideas that underlie the new way.

2 Text Comprehension Model

Kintsch and his colleagues indicated that the text comprehension is a series of transforming a representation into another representation: each sentence of a text to the surface code, the surface code to the text base, and the text base to the situation model. [3] [4] Thought the final process may be mainly top-down, the previous two processes are appeared to be essentially bottom-up. The surface code is the verbatim representation roughly corresponding to the product of the morphological and the syntactic processing. The text base is also roughly corresponding the semantic, then, the propositional representation. Therefore, it is suggested that the core process of the text comprehending is the process that transforms the text base into the situation model.

They assumed that there are two types of representation between them: the microstructure and the macrostructure (sometimes mentioned superstructure is seemed to refer to the situation model). The microstructure is constructed by connecting propositions which are derived from each sentence and clause to make the representation coherent. The local coherence is achieved by connecting the propositions successively. In the other hand, the macrostructure is the compilation of the microstructure. Specifically, it is constructed by applying the macro-rules such as the deletion rule, the generalization rule, the construction rule, and the composite deletion rule to the microstructure.

Although the knowledge rich approaches are trying to construct the macrostructure, they have a long way to provide the good practical outcomes. Therefore, we determined to extend the rules for connecting the propositions in order to elaborate the microstructure, that is to say the propositional network, to extract the important sentences for the text summarization based on the connection patterns of the proposition in such a network.

3 Extended Connection Rules

The microstructure, which is referred to as the propositional network, is constructed by means of connecting propositions derived from the text. [5]. The researchers have proposed various methods of connecting the propositions based on the different cues.

Kintsch suggested that the propositions with a common argument are connected [6]. On the other hand, some researchers indicated that the reader comprehends the text by linking the events described in it on the basis of their causality [7] [8]. It was suggested that the propositions are connected by their cause and effect. Furthermore, some researchers decomposed the proposition derived from one sentence to the main proposition and the quasi-proposition. The main proposition includes the predicate, the agent, and the object. On the other hand, the quasi-proposition includes the other concept qualifying the main proposition. Then, they have found that the sentence composed of the main proposition with more quasi-propositions is easier to be remembered [9]. And other researchers found that the story is comprehended by linking the events described in each sentence so that the rhetorical structure appeared in the story grammar is reflected on [10] [11]. Their findings suggested that the propositions are connected so as to reflect the change of the time or the space.

After considering the connection rules suggested in the previous researches, we added three new rules to make seven (then, extended) connection rules altogether. Two of new rules are relevant to the conversation. One connection rule is made to distinguish between the expository sentence and the conversational sentence. Another rule is made on the basis of the conversational reciprocality. And other new rule is made so as to reflect the association in the working memory. Those connection rules are specified in Table 1.

Table 1
Extended rules for connecting propositions

Common argument connection (Ar) If two propositions share a common argument, they are connected.
Development connection (D) If a proposition represents an event occurring in a new stage, it is connected with the first proposition.
Causal connection (Ca) If two propositions have the causal relationship, they are connected.
Grammatical connection (G) If a proposition represents a part of a sentence (the time, the space, the property of a character and an object, and so forth), it is connected with another proposition that represents an event described in the sentence.
Mere adjacency connection (Ad) If a propositions is not connected with other proposition by any other rules, it is forced to be connected with the proposition that represents an event described in an adjacent sentence.
Conversation connection (Co) If a conversational sentence follows a descriptive sentence, the propositions derived from each sentence are connected.
Reciprocality connection (R) If a conversational sentence follows another conversational sentence, the propositions derived from each sentence are connected.

Note. The short name of each connection rule is given in a parenthesis.

4 General Description of Our System

Inputs of our system are two kinds of the list generated from outcomes of the semantic analysis that has performed to each sentence contained in the text. Then, output is a series of the sentences that are judged as important and extracted by our system. Therefore, our system must perform several tasks between them; transforming inputs to propositional lists, connecting them to construct a propositional network represented as a matrix using some connection rules, detecting distinctive connection patterns of propositions that are involved in the important sentences by means of pattern matching through the propositional network, and extracting sentences that involve propositions forming such connection patterns. Figure 1 presented the flow of the tasks in our system.

In this section, we specify the processing procedures of our system from the input to the construction of the propositional network. And in the next section, we introduce two experiments performed to find out the distinct connection patterns of the propositions derived from the important sentences in the propositional network and assess the output of our system in which we implemented them as the extraction rules. In our research, we consider the construction of the propositional network as the construction of the microstructure. We designed our system on the ground that the important sentences may be extracted from the text by the reader on the basis of the distinctive connection pattern of the propositions in the propositional network.

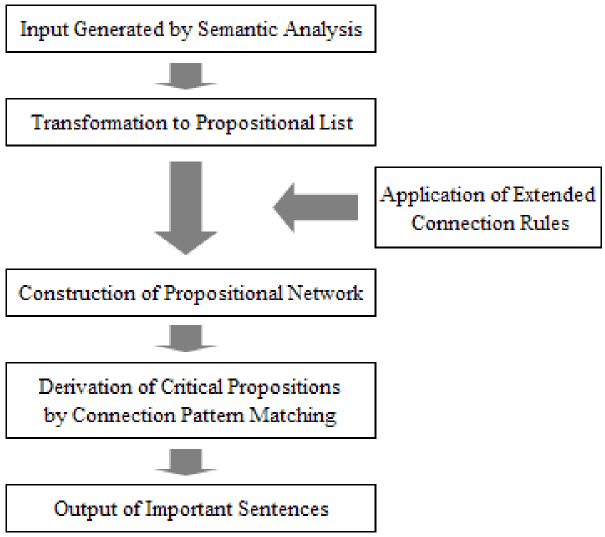

Figure 1. Processing flow of our system.

In this section, we specify the processing procedures of our system from the input process to the construction process of the propositional network. And in the next section, we introduce two experiments performed to find out the distinct connection patterns of the propositions extracted from the important sentences in the propositional network and assess outcome of our system in which we implemented them as the selection rules. In our research, we consider the construction of the propositional network as the construction of the microstructure.

We designed our system on the ground that the important sentences may be selected by people on the basis of the distinctive connection pattern of the propositions in the propositional network.

4.1 Input Based On Semantic Analysis

The input of our system is a series of the predicate-centered lists and the modifier lists that have been transformed from the results of the semantic analysis. We used the modifier as the term to refer to the sate, the degree, the time, and the location concept qualifying the noun and the verb as well as the index indicating the transformation of declarative sentence. The fundamental analyses, that is to say, the morphological and the syntactic analysis, are performed using the free software.

4.1.1 Semantic Analysis

The semantic analysis performed to the sentence has three purposes as follows:
1) Judging whether the sentence meaning is appropriate (Syntax structure is right, but is a meaning right?).
2) Extracting a meaning such as the depths case and the conceptual dependency structure of a sentence etc. (Is the meaning of the sentences the same though their expressions are different?).
3) Converting the sentences into the inferable expression (The inference rule is described between concepts.). In other words, the semantic analysis to the sentence is equivalent to the Conversion of the vague and/or various linguistic expressions into the concept representation described systematically. A word corresponds to a concept in the conceptual system. So it is possible to say that the sentence expresses the relations between the concepts. Therefore, an expression form about the concept and the conceptual relation is necessary for the semantic analysis.

In the present research, the results of the semantic analysis were transformed to the frame representation. According to the case grammar advocated by Fillmore, the role, which each noun concepts play to the predicate, are written as the case in the frame representation [12]. That is to say, it is suggested that the case frame represents the things described in the sentence focusing on the predicate. Table 2 presented the deep cases used in the present investigation. Then Figure 2 presented the case frame that represents an example sentence "Taro hangs a picture on the wall".

Table 2
Deep cases used in our investigation

Agent	The instigator of an action *Hanako* gave me a book.
Experiencer	The person who experiences a psychological event *Taro* suspects that Rie is an enemy.
Instrument	The means for doing an action or the psychological stimulus of an experience or action elicitor Yasuo goes to school by *bus*.
Object	The object of an action that undergoes change Jiro hit *the desk* with his fist.
Source	The source of an action in time or space or the initial state of things in changing Ichiro left *Kamakura* for Kyoto at eight.
Goal	The goal or result of an action or the final state of things in changing Ichiro left Kamakura for *Kyoto* at eight.
Location	The place or the position where an event occurs Taro met Hanako at *the party* one week ago.
Time	The point of time when an event occurs Taro met Hanako at the party *one week ago*.

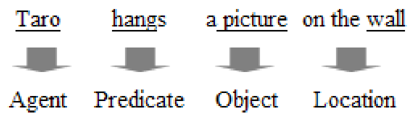

Figure 2. Predicate and case of each noun in an example sentence.

4.1.2 Predicate-Centered List

The results of semantic analysis were transformed into two kinds of list, the predicate-centered list and the modifier list. The predicate-centered list is the list in which the noun concept of each case is written with a focus on the predicate. In the present research, the predicate is limited to the verb concept. Table 3 presents the predicate-centered list of an example sentence "Once Upon a time, a poor grandpa lived in a certain place." The item stored in each column is as follows:
Column 1: The row number of the list is stored.
Column 2: The predicate of sentence is stored. If there are more than 2 predicate, it is stored in next Row.
Column 3 to 10: They express the agent case, the object case, the experiencer case, the instrument case, the goal case, the source case, the location case, and the time case respectively.
Column 11: The sentence number corresponding to the sentence order in the original text.
Column 12: The scene segmentation point. It is used as a clue when the development connection rule is applied to construct the propositional network.

4.1.3 Modifier List

The modifier list is the list in which the modifier class, the number referring to the corresponding row or column number in the predicate-centered list, and the state or the degree concept are stored. The modifier classes consist of the adverb, the adjective, and two indices of the transformation of the sentence. The state or the degree concept is primarily represented by the adjective or the adverb. And the index of the transformation is either the negative or the interrogative. Table 4 presented details of the list of words to an example sentence "Once upon a time, a poor grandpa lived in a certain place." The item stored in each col-

Table 3
Predicate-centered list of a sentence "Once upon a time, a poor grandpa lived in a certain place."

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7
No.	Predicate	Agent	Object	Experiencer	Instrument	Goal
1	LIVE	GRANDPA				

Column 8	Column 9	Column 10	Column 11	Column 12
Source	Location	Time	Sentence No.	Scene
	CERTAIN PLACE	ONCE UPON A TIME		

umn is as follows:

Column 1: The corresponding number of the modifier class is stored; 1 adverb, 2 adjective, 3 negative, 4 interrogatives.

Column 2: The corresponding row or column number in the predicate-centered list is stored.

Column 3: The corresponding column number in the predicate-centered list is stored.

Column 4: The state concept or the degree concept is stored.

Table 4
Modifier list of a sentence "Onece upon a time, a poor grandpa lived in a certain place"

Column 1	Column 2	Column 3	Column 4
No.	Row No.	Column No.	Modifier
2	1	3	POOR

4.2 Propositional list

The propositional list is a kind of the notation of the sentence meaning, that is to say the proposition, with the list format. It involved the main proposition and the semi-proposition. The main proposition preserves the essential information of a sentence while disregarding such details as word order, voice, and tense. It is composed of one predicate that refers to the verb concept primarily and one or more arguments that refer to the noun concepts. We treated separately each concept that represents the location, the time, the state, or the degree. The semi-proposition is composed of these concepts and the index that indicates some transformation rules applied to the sentence. The former two concepts of the semi-propositional list are also critical in the story structure. In short, the main proposition is central to the sentence meaning. Then, the semi-proposition elaborates the main proposition

4.2.1 Generation of Propositional list

The propositional list was transformed from the predicate-centered and the modifier lists according to the following procedure.

1) A list is drawn from a set of the preprocessed lists.

2) If it is the predicate-centered list, the concept is drawn in order Predicate, Agent, Object, and so on from it.

3) If the number referring to the transformation of the sentence is in the column 1 of the list of words, the list is connected with the propositional list.

4) If the concept representing the time or the location is in the appropriate column of the predicate-centered list, it is connected with the propositional list.

5) If the number referring to the state or the degree is in the column 1 of the list of words, the list is connected with the propositional list that contains the concept modified by it.

6) If the preprocessed list is conversational, No. of the proposition representing the content of beginning and ending conversation is identified.

7) If the period indicated the sentence end is drawn, the period symbol is outputted as a list.

8) If other types of the preprocessing list are drawn, they are discarded.

Table 5 presented the propositional list formed to an example sentence "Once upon a time, a poor grandpa lived in a certain place." The item stored in each column is as follows:

Column 1: The proposition number is automatically counted and stored.

Column 2: The predicate is stored. If two or more predicates are provided, they are stored in next row successively.

Column 3 to 8: The concept representing each case is stored in the appropriate column.

Column 9: The concept expressing the time, the location, the state, or degree is stored.

Column 10: The number, which refers to the kind of the concept stored in the column 9, is stored. The number to the kind is as follows: 1 adverb, 2 adjective,

Table 5
Propositional list of a sentence "Onece upon a time,
a poor grandpa lived in a certain place"

No.	Predicate	Agent	Object	Experiencer	Instrument	Goal	Source	Location/Time/Modifier	Type	No.	Final No.	Scene
1	LIVE	GRANDPA										
2								CERTAIN PLACE	7	1		
3								ONCE UPON A TIME	8	1		
4								POOR	2	1		

3 negative, 4 interrogative, 5 conversation, 6 conjunction, 7 location, 8 time, 9 period, 0 conversational.
Column 11: No. of the proposition, which contains the concept modified by the concept stored in the column 9, is stored. It is used when the propositions are connected to construct the propositional network.
Column 12: It is used only when the propositional list represents the conversation. No. of the proposition which indicates the end of the conversation is stored in this column. But No. of the proposition that indicates the beginning of the conversation is stored in the column 11.
Column 13: No. of the scene developed in the story is stored. It is used when the propositions are connected using the rule of the development connection to construct the propositional network.

4.3 Propositional Network

The relation between the propositions can be clarified by constructing the propositional network. Therefore the construction of the propositional network is needed as the preliminary step to identify the semantic relation between the sentences in the story. Specifically, the propositional network is constructed by applying the extended connection rules to the propositional lists to connect the propositions each other. It is noticed that these rules are applied to the propositional lists in a consistent order. Figure 3 shows the flow of applying the extended connection rules to the propositional lists.

Table 6 shows the matrix representation of the propositional network that is constructed from the propositional lists generated from two consecutive sentences in the story ”Bamboo Hats for Six Ksitigarbhas” as an example. The second row of this table refers to the order of the sentence in the story. Therefore it is showed that the first sentence contains the propositional list 1, 2, 3, and 4 referred in the first row. In a same way, the second sentence contains the propositional list 5, 6, 7, and 8. The symbols in the matrix refer to the extended connection rule applied in order to connect the propositional lists. Specifically, Table 6 presented as following things. The

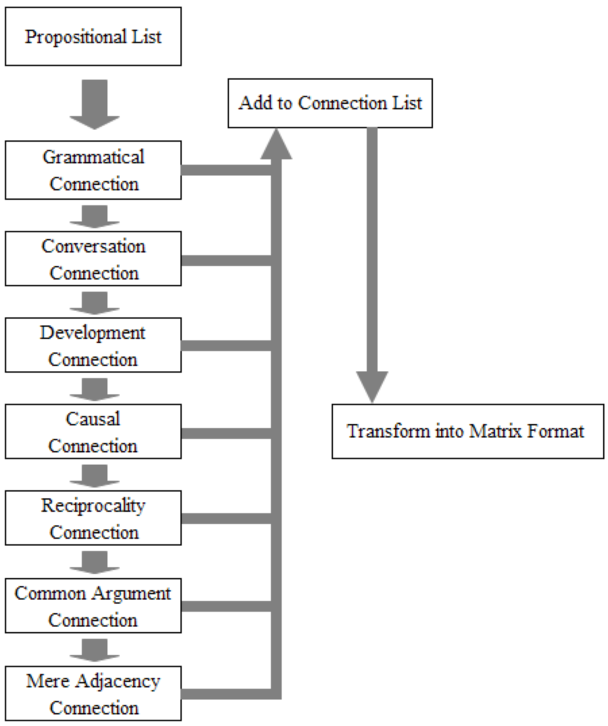

Figure 3. The flow of application of connection rules in constructing propositional network

Table 6
An example of propositional network

No. of propositional list		1	2	3	4	5	6	7	8
No. of sentence		1	1	1	1	2	2	2	2
No. of propositional list	1		G	G	G				
	2								
	3								
	4					Ad			
	5						Ca		
	6							G	Ar
	7								
	8								

Note. A part of story "The Runaway Riceball" is as follows: Once upon a time, the poor grandpa and gramdma lived in a certain place. Althogh tomorrow is the New Year, there is no money for buying rice.

propositional list 1 is connected with the propositional list 2, 3, and 4 by means of the grammatical connection. And the propositional list 4 is connected with the propositional list 5 by means of the mere adjacency connection. The propositional list 5 is connected with the propositional list 6 by means of the causality connection. The propositional list 6 is connected with the propositional list 7 by means of the grammatical connection and 8 by means of the common argument connection.

5 Extraction of Important Sentences

5.1 Purposes of Experiment 1 and 2

Experiment 1 was performed to find out how the propositions derived from the important sentences are connected with other propositions in the propositional network. Here the important sentences refer to the sentences that a number of the human participants determine to be important in a story. Subsequently, Experiment 2 was performed to confirm that the connection patterns obtained in Experiment 1 are valid as the selection rules to extract the important sentences from the story. If it is confirmed, the connection patterns are likely to be usable as the selection rules of our system.

We calculated three aggregate accuracy measures to evaluate the validity of the selection rules for extracting the important sentences [13]. Table 7 shows the contingency table for the participant's and our system's extraction of important sentences. Then Table 8 indicates the definitions of three aggregate accuracy measures. The recall rate is the percentage of the sentences extracted by our system to the important sentences. Therefore, it means that our system can extract more important sentences as the rate increases. However, it must be noticed that many unimportant sentences may be falsely extracted together. The precision rate is the percentage of the important sentences contained in the sentences that our system extracts from the story. It also means that our system can extract more important sentences as the rate increases. Specifically, our system can't extract so many sentences, but extracted ones are likely to be the important sentence. F-measure is the index that represents a comprehensive assessment of our system performance. Hence, it is interpreted as a weighted average of the recall rate and the precision rate, where F-measure reaches its best value at 1 and worst value at 0.

These three measures were calculated in both experiments. However, there was a noteworthy difference between Experiment 1 and 2. That is to say, the input in Experiment 1 were five stories that we used to find out the connection patterns, but the inputs in Experiment 2 were five new stories that we used to confirm whether the connection patterns are valid as the selection rules for extracting the important sentences. In order to confirm such validity, it is sufficient to compare the sentences automatically extracted by our system implementing such selection rules with the important sentences selected by the human. In Experiment 1, we also examined how correctly the selection rules can extract the important sentences from each story originally used to derive them. Through these two experiments, we tried to confirm that the important sentences can be extracted from any stories using the selection rules based on the connection patterns found out in Experiment 1.

5.2 Experiment 1

5.2.1 Results of Experiment 1

We defined that the important sentence is the sentence which more than half of the participants selected. Table 9 presented the number of the important sentences

Table 7
Contingency table for participant's and our system's selection of important sentences

Ground Truth (Participant's Selection)	Judgment (Our System's Selection)	
	True (Impotrant)	False (Unimportant)
True (Important)	True Positive (TP)	False Negative (FN)
False (Unimportant)	False Positive (FP)	True Negative (TN)

Table 8
Aggregate accuracy measures used in evaluating our system

Measure	Definition
Recall	TP/(TP+FN)
Precision	TP/(TP+FP)
F-Measure	2 × Recall × Precision/(Recall+Precision)

that were selected from each story. Through five stories, we examined how the propositions included in each important sentence connect with each other and the propositions included in the adjacent sentence.

Using an example story "Click-Clack Mountain", we try to explain the procedure for deriving connection patterns used in order to extract the important sentences. Table 10 presents all sentences included in this story. Furthermore, in this table, the sentences written in the bold letters are selected as important by more than half of the participants. Table 11 shows a part of the propositional network and the connection patters of the propositional list. And several circled signs indicate the connection types of the propositional lists that are derived from each important sentence.

As a result, we found three common connection patterns as fallows:

Connection pattern 1: A proposition connects with two or more propositions based on either the grammatical or the causal connection rule.

Connection pattern 2: A proposition connects with another proposition based on the development connection rule and two or more propositions based on any other connection rules.

Connection pattern 3: A proposition connects with another proposition based on the common reference connection rule and one or more propositions based on any other connections.

We implemented our system with the selection rules based on these connection patterns and applied it to five stories used in Experiment 1. For example, the sentences written in italic letter in Table 10 are ones extracted from the story "Click-Clack Mountain" by our system. Using these outputs and the important sentences selected by the participants, we calculated three indices. Table 12 presented three aggregation accuracy measures and the number of sentences extracted by our system.

Generally, Microsoft's AutoSummarize summarizer is assumed to provide the baseline of each aggregation accuracy measure. Through five stories, Microsoft's AutoSummarize summarizer provided about 40% for each of three aggregation accuracy measures under same compression rates. Specifically, to the story "Click-Clack Mountain", the recall, the precision, and the F-measure are 36%, 44%, and 40% respectively. Therefore, at least, our system could clearly outperform Microsoft Summarize summarizer.

5.3 Experiment 2

5.3.1 Method of Experiment 2

We performed another experiment to confirm whether our system can extract the important sentences from the other stories using the selection rules found in Experiment 1. Several days after Experiment 1, same eighteen participants were returned to our laboratory and asked to read five new stories as follows: "Grandfather Cherry Blossom", "Bamboo Hats for Six Ksitigarbhas", "Kintaro, the Golden Boy", "The moon Princess", and "A Spouse for a Mouse". The subsequent procedure was same as one in Experiment 1.

Table 9
Number of important sentances selected
by participants in Experiment 1

Titles of Nursery Tales	Number of sentences	Number of Selected Sentences
Sumou Wrestling of the Rats	31	17
The Monkey and the Crabs	49	23
Click-Clack Mountain	47	20
Momotaro, the Peach Boy	47	18
The Grateful Crane	32	14

Table 11
Propositional network generated from a part of an example story "Click-Clack Mountain"

No. of Propositional List	1	53	54	55	56	57	58	59	60	61	62	63	64	65	66
No. of Sentence		23	***24***	***24***	***24***	***24***	25	25	***26***	***26***	***26***	***26***	***28***	***29***	***29***
No. of Propositional List															
52		Ad													
53			Ar												
54	D			G	Ca	Ca	Ad								
55															
56															
57															
58								Ca	Co						
59															
60										Ad					
61											G	G	Ar		
62															
63															
64														Ad	
65															Ca

Table 12
Number of important sentences selected by our system and
three values about its validity to results of Experiment 1

Titles of Nursery Tales	Number of Sentences Selected by Our System	Recall	Precision	F-measure
Sumou Wrestling of the Rats	18	71%	67%	69%
The Monkey and the Crabs	19	70%	84%	76%
Click-Clack Mountain	15	60%	80%	69%
Momotaro, the Peach Boy	15	67%	80%	73%
The Grateful Crane	13	64%	69%	67%

Needless to say, we have already applied our system to these stories to extract the important sentences from each new story.

5.3.2 Results of Experiment 2

According to the definition of the important sentence, the important sentences of each of five new stories were determined. Table 13 presented the number of the sentences involved in each story and the number of the important sentences selected by the participants. We also calculated the aggregation accuracy measures to examine the validity of the selection rules. Table 14 presented these measures and the number of the important sentences that our system extracted from each new story. For example, Table 15 presents the sentences that our system extracted as important ones from the story ”Bamboo Hats for Six Ksitigarbha”.

Through five new stories, all aggregation accuracy measures provided by Microsoft’s AutoSummarize summarizer are less than 40% under same compression rates. Specifically, to the story ”Bamboo Hats for Six Ksitigarbha”, Rcall, Precision, and the F-measure are 75%, 71%, and 73% respectively. Therefore, our system could clearly outperform Microsoft Summarize summarizer again.

Table 10
All sentences included in a simple story "Click-Clack Mountain"

1)	**When a grandpa was working on a farm, a raccoon dog appeared.**
2)	***The grandpa got angry with the raccoon dog getting up mischief always, bound it with the rope, hung in the door of the house, and said.***
3)	"Grandma, please make the raccoon dog soup."
4)	"Well, I will make rice dumplings to put them in soup."
5)	**When the grandma was hulling rice,**
6)	The raccoon dog call to her, "Untie a rope, since I help you."
7)	"No! Grandpa will get angry with me."
8)	**"Don't worry since I am hung again after finishing hulling rice."**
9)	Then, the grandma untied the rope.
10)	**The raccoon dog pretended to hull rice and bludgeoned the grandma to death.**
11)	***Then, the raccoon dog changed himself into the grandma and waited for the return of a grandpa.***
12)	***In the evening, the grandpa came back.***
13)	The grandpa said "Did you make the raccoon dog soup?"
14)	The grandma replied, "Yes, it was fully getting warm."
15)	The grandpa said, "It's a delicacy.
16)	It's very tasty."
17)	**The grandpa ate his belly.**
18)	***In a moment, the raccoon dog returned to the original figure and said while escaping, "Hey, the grandpa ate the grandma soup!"***
19)	The grandpa was so surprised as to cry violently.
20)	**Then, a rabbit come rushing over and asked, "Grandpa, what happened? I was deceived by the raccoon dog."**
21)	*When the grandpa said what the raccoon dog did, the rabbit replied.*
22)	"Don't cry, Grandpa.
23)	**I take revenge for you."**
24)	**When the rabbit went to Mt. Kaya, he cut the thatch while singing in a loud voice.**
25)	Then, the raccoon dog appeared immediately and began to cut the thatch together.
26)	***"Dear Raccoon, please carry the thatch because you are stronger then me."***
27)	"Sure!"
28)	***The raccoon dog carried the thatch on his shoulder.***
29)	***The rabbit set fire to the thatch with the flint while following him. "Click-Clack"***
30)	"What sound is it?" "Probably, I thought that a bird sang Click-Clack."
31)	***After a while, the thatch caught fire and began to burn. "Bow-Bow"***
32)	"What is that strange sound?"
33)	"Probably, I thought that another bird sang Bow-Bow."
34)	**"Yipe! I am burning!" The raccoon dog took the burn injury on his back severely.**
35)	***When the rabbit was making the woods of Mt. Sugi into a ship, the raccoon dog came.***
36)	"Hey! You look at me.
37)	You do bad things to me."
38)	"I am a rabbit living in Mt. Sugi.
39)	*I don't know the rabbit living in other mountain."*
40)	The raccoon replied, "Oh, Yeah.
41)	***Then, dear Rabbit of Mt. Sugi, for what purpose are you making the ship?" "I want to go fishing."***
42)	"Please take me together."
43)	The ship of the raccoon dog was making with the mud.
44)	***The rabbit went to the middle of the river and struck his ship with the oar and said.***
45)	"Use this method to gather the fish."
46)	***"OK! Come in, many fish!" As the raccoon dog struck his ship with oar many times, it was cracked to careen gradually.***
47)	**The raccoon dog sink down and died by drowning.**

Note: The sentences written in the bold letters were selected as the important ones by the participant in Experiment 1. And the sentences written in the italic letter were extracted as important ones by our system.

6 General Discussion

Aggregation accuracy measures were partially reported in the previous researches on the automatic text summarization. To be compared with our results, we cite the values of those measures. In particular, several researches, which used the discourse-level information like ours, reported the aggregation accuracy measures. For example, Barzilay and Elhadad have reported 67% Recall and 61% Precision (therefore 64% F-measure) for the lexcal chain method [14]. In their research, the baselines provided by Microsoft's AutoSummarize summarizer are 37% Recall and 33% Precision (therefore, 35% F-measure). In addition, using a summarization program based on rhetorical structure theory, Marcus has obtained the high 60% Recall and Precision [15]. And, as compared with the baselines (40% both Recall and Precision), his automated system was said to outperform Microsoft's AutoSummarize summarizer. Furthermore, Teufel and Moens

Table13
Number of important sentances selected
by participants in Experiment 2

Titles of Nursery Tales	Number of Sentences	Number of Selected Sentences
Grandfather Cherry Blossom	39	20
Bamboo Hats for Six Ksitigarbhas	55	20
Kintaro, the Golden Boy	39	16
The Moon Princess	40	17
A Spouse for a Mouse	28	15

Table 14
Number of important sentences selected by our system and
three valus about its validity to results of Experiment 2

Titles of Nursery Tales	Number of Sentences Selected by Our System	Recall	Precision	F-measure
Grandfather Cherry Blossom	13	55%	85%	67%
Bamboo Hats for Six Ksitigarbhas	21	75%	71%	73%
Kintaro, the Golden Boy	17	63%	59%	61%
The Moon Princess	13	59%	77%	67%
A Spouse for a Mouse	11	53%	73%	61%

Table 15
The summary of the story "Bamboo Hats
for Six Ksitigarbha" generated by our system

Long time ago, poor grandpa and grandma lived somewhere. For new year ceremony, they want to get some moneys. Grandpa went to the city to sell some straw hats.Nobody bought his straw hat at the city. Grandpa decided to go back home. On his way home, it started to snow. He found six Ksitigarbhas (made of stone) . He put the straw hat on each head of Ksitigarbas. Grandpa came home and told the affairs to grandma. At dawn, they heard the noise and the singing voice from the outside. They went outside and found many goods. They saw six Ksitigarbhas walking in the distance.

have reported 64.2% Precision in their research based on the document scheme (the genre-specific document structure and the document's theme) [16]. As compared with the baseline of 40.1% Precision, their Precision was clearly high. In summary, Aggregation accuracy measures provided by the previous researches were the high 67% Recall, the high 64.2% Precision, and the high 64% F-measure.

Aggregation accuracy measures provided by Microsoft's AutoSummarize summarizer were about 40% Recall and Precision in Experiment 1 and under 40% Recall and Precision in Experiment 2. Then it fallows that F-measures were less than 40% in both experiments. In comparison to these baselines, three aggregation accuracy measures obtained in our experiments are substantially high as follows: In Experiment 1, mean Recall was 66% (ranged from 60% to 71%), mean Precision was 76% (ranged 67% to 845), and mean F-measure was 71% (ranged from 67% to 765). In Experiment 2, mean Recall was 61% (ranged 53% to 75%), mean Precision was 73% (ranged from 59% to 85%), and mean F-measure was

66% (ranged from 61% to 73%). Moreover they are equal to or higher than those obtained in several previous researches.

In Experiment 1, we examined the connection patterns of the propositions, that is how the propositions included in the important sentence are connected with each other and the other proposition. And we turned up three distinct connection patterns. Then we made our system search the propositional network to find out these patterns. From the number of the important sentences selected by the participants and the number of the important sentences extracted by our system, we calculated three aggregation accuracy measures of each story used in Experiment 1 and 2. These measures obtained in Experiment1 were inevitably higher than ones obtained in Experiment 2 on the whole, because the connection patterns were found out in the propositional network of each of five stories used in Experiment1. Therefore, our system is appeared to be so usable one that can extract the important sentences from the simple story.

However, there are some problems in our research. We examined fully the connection patterns of the propositions included in the important sentences that can't be extracted by our system. As the results, most problems are attributed to the modifier list and the subsequent procedure relative to this list. Actually, we conveniently designed the modifier list that can include the mixed concept. We must redesign the modifier list to be able to use the critical information for connecting the propositions.

7 Conclusion

According to Kintsch's model of the text comprehension, we designed our system that can automatically extract the important sentences reflecting the gist of the simple story. While the reader is reading a story, he connects the propositions derived from each sentence, constructs the microstructure, and compresses it into the macrostructure as a gist of the story. In his model, the microstructure is represented as the propositional network in which the propositions derived from each sentence are connected with each other.

Although the rule for connecting the propositions was originally based on the common argument, we proposed the extended connection rules to construct more elaborative propositional network. Then we tried to design the system that can extract the sentences required to summarize the story based on the ways in which the propositions derived from the important sentences are connected with other propositions. The input of our system is the outcome that the semantic analysis produces to each sentence composing the story. Then the input is transformed into the predicate-centered list and the modifier list. Next, these lists are transformed into some propositional lists. And the propositional lists are connected using the extended connection rules to generate the propositional network. Finally the important sentences are extracted from the story using the selection rules that reflect on the connection patters of the propositions derived from important sentences.

Two experiments were performed to identify the connection patterns and to assess the validity of them as selection rules. In Experiment 1, we found out the three connection patterns of the propositions extracted from the important sentences that the participants selected from the story. In Experiment 2, we confirmed that the validity of these connection patterns is high enough as the selection rules. Therefore, these results indicated that our system, which was implemented these connection patterns as the selection rules, can do extract the important sentences required to summarize the story. Then three aggregation accuracy measures obtained in our experiments are clearly higher than the baselines provided by Microsoft's AutoSummarize summarizer. Moreover they are equal to or higher than ones obtained in the previous researches.

We found, however, some problems in our research. Specifically, the modifier list is so ambiguous and unorganized that the usability of our system is lowered. We need to redesign the modifier list in order to improve our system.

References:

[1] Okumura, M. and Nanba, H., *Automated Text Summarization:* Survey, Natural Language Processing, 6(6), 1999, pp. 1-26.

[2] Mani, I. and Maybury, M. T., *Advances in automatic text summarization*, Cambridge, Massachusetts: MIT press, 1999.

[3] Kintsch, W., *Comprehension: A paradigm for cognition,* Cambridge University Press, 1998.

[4] van Dijk, T. A. and Kintsch, W., *Strategies of discourse comprehension*, New York: Academic Press, 1983.

[5] Kintsch, W. and van Dijk, T, 1978, *Toward a Model of Text Comprehension and production*, Psychological Review, 85, pp. 363-394.

[6] Kintsch, W. and van Dijk, T. A., *Reading rate and retention as a function of the number of propositions in the base structure of sentences*, Cognitive psychology, 5(3), 1973, pp. 257-274.

[7] Trabasso, T. and Sperry, L. L., *Causal relatedness and importance of story events. Journal of Memory and Language*, 24, 1985, pp. 595-611.

[8] van den Broek, P. and Trabasso, T., *Causal networks versus goal hierarchies in summarizing text*. Discourse Processes, 9(1), 1986, pp. 1-15.

[9] Taylor, S. E. and Thompson. S. C, Stalking the elusive 'vividness' effect Psychological Review, 89(2), 1982, pp. 155-181.

[10] Rumelhart, D. E., *Note on a schema for Stories*. In D. G. Bobrow and A. M. Collins (Eds.), Representation and understanding: studies in cognitive science. New York: Academic press, 1975.

[11] Thorndyke, P. W., *Cognitive structures in comprehension and memory of narrative discourse*, Cognitive Psychology, 9(1), 1977, pp. 77-110.

[12] Fillmore, C. J., *The Case for case*, In E. Bach & R. Harms (Eds.), Universals in Linguistic theory, New York: Rinehart and Winston, 1968.

[13] van Rijsbergen, C. J., *Information retrieval*, Woburn, Massachusetts: Butterworths, 1979.

[14] Barzillay, R. and Elhadad, M., *Using lexical chains for text summarization. In I. Mani and M. T. Maybury (Eds.), Advances in automatic text summarization*, Cambridge, Massachusetts: MIT press, 1999, pp.111-121.

[15] Marcus, D., *Discourse trees are good indicators of importance in text. In I. Mani and M. T. Maybury (Eds.), Advances in automatic text summarization*, Cambridge, Massachusetts: MIT press, 1999, pp.123-136.

[16] Teufel, S. and Moens, M., *Argumentative classification of extracted sentences as a step towards flexible abstracting*. In I. Mani and M. T. Maybury (Eds.), Advances in automatic text summarization, Cambridge, Massachusetts: MIT press, 1999, pp.155-171.