

Relationship Between Features of Reading Behaviors and Dynamic Abstract of Novel

Haruna Mori, Ryosuke Yamanishi, Yoko Nishihara, and Junichi Fukumoto

Abstract—After the interrupted the reading, the reader sometimes reads again the past story in a short time to remember the contents. As grasping the abstract before resuming the reading, the reader can smoothly restart his/her reading. This paper conducted the experiment to create human-made abstract to investigate the abstract required before resuming the reading. Analyzing the features by focusing on the positions of the sentences selected for abstracts in the novels, we discussed the variation of abstracts depending on the reading progress. Through the discussion, we obtained the following seven findings: (1) the position where the bookmark is inserted is biased to a boundary of the story, (2) there are three patterns to create abstracts, (3) whether the sentence just before the bookmark is included in the abstract or not depends on the attribute of each reader, (4) the important information for the abstracts is changeable depending on each reading progress, (5) there is no relation between the reading intervals and which sentences are selected for the abstract, (6) the information of character name and place may work effectively to generate the abstracts, and (7) it seems that the words that have a climax on the part just before the bookmark is important for the abstract.

Index Terms—features of abstract, reading behaviors, dynamic abstract

I. INTRODUCTION

CURRENT e-books are simply digitalized books of paper media, so we benefit a little from e-books. As the part of advantages of e-books, that is not bulky and easy to purchase, and the texts can be expanded. These features might be not enough as the good appeal point of e-books. As another advantage of e-books, the information in the e-books can be processed in real time on the terminal. Though the search function has been already realized, more utilized advantage should be expected. The purpose of the study is to create the additional value of e-books by enhancing the entertainment of novels. We tackle generating the dynamic abstract depending on the reading progress, that supports the user to resume the reading novel.

In dramas, animations, and weekly novels and comics in Japan, abstracts are used to remind viewers/readers of the contents until last time. These contents generally present the same abstracts to all viewers/readers because the same contents should be shown to all viewers/readers for each time. However, in non-series contents, it seems that the reading pace differs from each viewer/reader and the different abstracts for each viewer/reader should be required. While reading a long novel, the reader sometimes forgets the past story and reads again when the reading is interrupted and the reader resumes. In such a case, it seems that the abstracts work effectively; it is expected that the abstract shortens the

time to read again and let the readers resume the reading smoothly. In the previous study, we proposed a method to generate the abstracts based on the increasing rate of the word importance before and after the bookmark [1]. However, what kind of information is required for the abstracts is not clear. In this paper, we conduct the experiments to create human-made abstract in order to investigate the abstract for smoothly resuming reading. We analyze the features of human-made abstract focusing on positions of sentences selected for the abstract corresponding to reading progress. Also, we discuss the relationships between the abstracts and reading activity, such as the amount of text and time in each reading.

II. RELATED WORK

The abstracts are constituted by extracting information related to specific words, such as character names and events, and by summarizing them. From this point of view, generating the abstract of novels can be regarded as a part of general Query-Focused Summarization (QFS). On the other hand, the abstracts are changeable depending on the reading progress: that is, the target sentences of summarization is updated. From this point of view, generating the abstract of novels can be regarded as a type of Update Summarization (US) [2]. Therefore, it seems that generating abstract depending on the reading progress of the novel combines both properties of QFS and US. As the research related to generating the abstracts, Bamman *et al.* related the sentences in the abstract with the sentences in the novel [3]. They classified the sentences in the novels into two types: those which can be used in the abstract or not. Based on the classification, they also proposed a method to generate the abstract for the novel. It is considered that their method is capable of generating the abstract for the whole part of the novel. This study focuses on the abstract in each reading progress, so the important information for the abstract is sequentially varied depending on the reading progress. The method proposed by Bamman *et al.* does not satisfy the requirement to generate such specific abstract.

Green *et al.* showed the pleasure of reading novel would be improved as absorbed in the story [4]. According to this study, the trigger to be absorbed in the story, which reminds the reader of the read part, would improve the pleasure of the reading of novels. The abstracts have such roles and are important for improving the experiences in the reading of novels. To improve the experiences in the reading novels, web review for the novel has been studied. The spoiler is known as the description that reduces the fun of novels, comics, and movies. Once the spoiler is shown, the enjoyment and interest towards the entertainment can be dramatically reduced [5]. In the novels, the most

Manuscript received December 26, 2017; revised January 17, 2018. This work was supported in part by JSPS Grant-in-Aid for Young Scientists B #16K21482.

H. Mori, R. Yamanishi, Y. Nishihara and J. Fukumoto are with Ritsumeikan University. e-mail: ryama@media.ritsumei.ac.jp

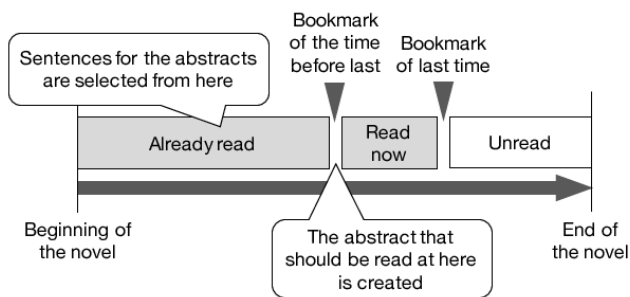


Fig. 1. The Concept of How to Create Abstract in the Experiment.

of the spoilers are related to the detail and epilog of the story, thus the spoilers are considered to be a kind of the most important information in the novel. Boyd-Graber *et al.* classified whether the sentence includes the spoiler word or not by using a supervised learning [6]. As the learning data, the words in the review posted on the abstract-sharing website and the sentence structure were used. Their method might be useful before starting the reading of novels but should be not suitable for the resuming the reading. Guo *et al.* also studied the spoilers [7]. Their target was to detect the reviews including the spoilers and automatically tag the spoilers by their proposed method. Maeda *et al.* also studied the spoilers [8]. In their research, it is reported that the spoiler would appear more in the last parts of the story. This paper focuses on the abstract depending on the reading progress, that is dynamic abstract and investigates what information should be used for the dynamic abstract.

III. THE EXPERIMENTS TO CREATE HUMAN-MADE ABSTRACT

Novels which have over 20,000 characters were used in the experiment. It is considered that readers may interrupt their reading many times when they read a novel which has many words, and then it is expected that abstract should be useful to remember the story of the novel. It seems that a novel with a lot of words has many characters, then organizing information with abstract should be more important. Text data of “Kokoro,” “Night on the Galactic Railroad (Shincho Bunko ver.),” and “Hell Screen” were used in the experiment to create abstract. These three novels have been publicly available on Aozora Bunko ¹. Novel IDs and metadata are shown in TABLE I; the data was calculated from each novel without metadata such as title, author, footnote, information of copytext, and headline. In the case of ID N_1 , the first part was extracted and used to let the number of characters be almost the same as other novels. These three novels were ranked the second, 11th, and 58th respectively in the access ranking of Aozora Bunko ². Authors of each novel are top three persons of the total number of accesses to novels written by each author. Thus, these novels should be common and popular novels.

The experiment to create abstract was conducted by 22 participants in their late teens to twenties who did not know the content of the assigned novel. The number of participants

¹<http://www.aozora.gr.jp>

²The ranking of XHTML ver. in the whole year of 2016. There are all 13,969 novels, as of 2016.12.31.

in the experiments for each novel is shown in TABLE I. The participants were instructed not to obtain information that could become the spoiler such as an outline of the novel. The experiment was conducted by repeating the following three procedures.

- 1) Participants read the novel.
 Participants record the time of starting and finishing the reading, and the position corresponds where bookmark was inserted.
- 2) Participants answer the questions to verify whether participants understand the contents of the novel or not.
 We prepared 10 questions for each novel. The questions are based on the contents and not difficult. They can be answered by the participants if they just read the novel. We asked the participants the questions after finishing each reading.
- 3) Participants create the abstract.
 Eight sentences were selected from all read part. Those sorted in order of appearance in the novel were treated as the abstract.

The number of sentences selected for abstract was decided by 8.17647, that is the average number of sentences in the abstract of young-adult novels with different titles and author names.

The following information was given to the participants.

- Abstract means the information that is required by the reader when resuming the reading.
- Abstract created by the participants should be read just before reading the text for this time.
- The abstract was created by extracting sentences from the text between the beginning of the novel and the bookmark for last time.

Fig. 1 shows the image of the range where sentences for abstract are selected. When the participant read the novel until the position where he/she interrupted his/her reading that is the position of the bookmark at this time, the sentences for abstract should be selected from the beginning of the novel to the position of the bookmark at last time. For example, after the second reading, the sentences for the abstract should be selected from the beginning of the novel to the position of the first bookmark. After the third reading, the sentences for the abstract should be selected from the beginning of the novel to the position of the second bookmark. There is no previous bookmark after the first reading, so no abstract is created only at this time.

All 22 participants showed a correct answer rate of 80% or more in the questions asking contents of each novel. According to this result, it was considered that all participants understood the contents of each novel. Therefore, the abstracts created by all 22 participants were assumed as reasonable reference information.

IV. ANALYSIS OF HUMAN-MADE ABSTRACT

The relationships between the features of abstracts created by the participants and the reading behaviors of each participant will be analyzed and the discussions will be described below. The results were discussed from five kinds of viewpoint: the trends for positions of a bookmark, the positions from which sentences for abstract were extracted, the

TABLE I
 METADATA OF NOVELS USED IN THE EXPERIMENT AND THE NUMBER OF PARTICIPANTS IN THE EXPERIMENT FOR EACH NOVEL.

Novel ID	Author	Title	The amount (in Japanese)				# of participants
			Characters	Words	Sentences	Paragraphs	
N_1	Natsume, Soseki	Kokoro (the first part)	49,103	32,415	1,779	282	7
N_2	Miyazawa, Kenji	Night on the Galactic Railroad (Shincho Bunko ver.)	38,212	22,963	1,100	158	8
N_3	Akutagawa, Ryunosuke	Hell Screen	26,103	18,313	480	98	7

relationships between the reading interval and the abstract, the semantic information in the abstract, and the relationships between the appearance rate of each word and the abstract.

A. Positions of Bookmark

TABLE II shows the positions that the participants interrupted the reading and insert the bookmark. Duplication was not allowed in TABLE II. Where the bookmark was inserted is checked through in the order of boundary of sections, blank lines, between paragraphs, and between sentences. When the corresponding position was confirmed, the checking was finished. For example, the bookmark is inserted between paragraphs and also blank lines, then we assume the bookmark is inserted at the blank lines.

More than 50% of the bookmarks was inserted at the positions where the boundary of sections in every three novels. It was considered that 83% of the bookmarks was inserted between paragraphs, because the boundary of sections and the blank line were included in between paragraphs. In generic texts or documents, a paragraph is a group of sentences concerning the same one topic. The point that the reader interrupts reading is the boundary of the story; a boundary of paragraphs is a boundary of the story.

B. Positions from which Sentences for Abstract were Extracted

We analyzed the relationships between each reading text of each participant and the positions of sentences for the abstract in the novel. Regarding positions of sentences selected for the abstracts, there were three patterns of the trend: the sentences selected from only the part just before the bookmark, from the beginning part of the novel and the part just before the bookmark, and exhaustively from all of the read parts. The beginning part of the novel means the part that each participant read first. TABLE III shows these three patterns and trend of selecting sentences for an abstract by each participant. In TABLE III, the participants are represented by using participant ID: S_i . And, S_1 to S_7 , S_8 to S_{15} , and S_{16} to S_{22} are the participants of experiments with N_1 , N_2 , and N_3 , respectively.

All of the participants required the information included in the part just before the bookmark for the abstract. It is considered that the contents just before interrupting the reading should be required for the abstract. Though most participants considered it was enough if there is the information of the read part just before bookmark, there were some participants required information from the beginning of the novel or other read parts. From these results, it was suggested that it was necessary to prepare the abstract appropriate to each participant. Also, it is inferred that the influence of the

differences of novels is smaller in the patterns of positions to select sentences for abstract.

There is a possibility that the sentence just before the bookmark contains information strongly related to the contents of the story after the bookmark. It seemed that the sentence just before the bookmark was considered as important to remind contents just before interrupting reading. We investigated the probability p_i that the sentence just before the bookmark was selected for the abstract by each participant S_i . The equation (1) calculates p_i ;

$$p_i = \frac{Abst_{last}}{N}, \quad (1)$$

where, $Abst_{last}$ and N each shows the number of abstracts including sentence just before the bookmark and the number of creating the abstract by each participant, respectively. In TABLE IV, the results are shown as $(S_i : p_i)$: the participant ID S_i and the value of p_i .

The majority of participants did not select the sentence just before the bookmark for the abstract; the sentence just before the bookmark should be not important for the abstract. However, there were some participants selecting the sentence just before the bookmark; two participants selected the sentence just before the bookmark every time. It was suggested that whether the sentence just before the bookmark is important or not depends on each participant.

C. Relationships between Reading Interval and Abstract

As mentioned in section I, the reader may forget the past story if the time interval between the reading is longer. From this assumption, it seems that the longer interval between the readings the more information included in the beginning of the novel should be required. We investigated the relationships between the reading interval and the sentences selected from the part that each participant read first. The results in the abstracts that should be presented before the third and fourth reading are shown in TABLE V and TABLE VI, respectively. In TABLE V and TABLE VI, the results are shown as $(S_i : f_i)$ with participant ID S_i and the ratio f_i that sentences for abstract selected from the first part that each participant read. The ratio f_i is calculated by the equation (2);

$$f_i = \frac{Sent_{first}}{TS}, \quad (2)$$

where, $Sent_{first}$ and TS each shows the number of sentences selected from the first read part and the total number of sentences for the abstract, respectively. As mentioned in section III, the total number of abstract is always eight in this paper. In TABLE VI, the participants who completed the experiment with three times of reading are not listed.

As shown in TABLE V and TABLE VI, the sentences in the first read part were selected also in the case that interval

TABLE II
 POSITIONS WHERE THE BOOKMARK IS INSERTED BY THE PARTICIPANTS FOR EACH NOVEL.

		Boundary of sections	Blank line	Between paragraphs	Between sentences	Total number of bookmark
# of bookmark	N_1	20	0	0	0	20
	N_2	14	1	5	4	24
	N_3	13	0	1	0	14

TABLE III
 POSITIONS FROM WHICH SENTENCES FOR ABSTRACT WERE EXTRACTED.

Novel ID	Positions from which sentences was selected		
	Only from the part just before the bookmark	From the beginning part of the novel and the part just before the bookmark	Exhaustively selected from all of the read parts
N_1	S_2, S_3, S_4, S_6, S_7	S_1	S_5
N_2	$S_9, S_{10}, S_{11}, S_{12}, S_{14}, S_{15}$	S_{13}	S_8
N_3	$S_{17}, S_{18}, S_{19}, S_{20}$	S_{16}, S_{21}, S_{22}	

TABLE IV
 SELECTION PROBABILITY OF SENTENCE JUST BEFORE THE BOOKMARK FOR THE ABSTRACT FOR EACH PARTICIPANT.

Novel ID	$p_i = 0.00$	$0.00 < p_i \leq 0.50$	$0.50 < p_i \leq 1.00$
N_1	$(S_1 : 0.00), (S_3 : 0.00), (S_4 : 0.00), (S_5 : 0.00), (S_6 : 0.00)$	$(S_7 : 0.50)$	$(S_2 : 1.00)$
N_2	$(S_{11} : 0.00), (S_{12} : 0.00), (S_{15} : 0.00)$	$(S_{10} : 0.50), (S_{13} : 0.33)$	$(S_8 : 0.75), (S_9 : 0.60), (S_{14} : 0.67)$
N_3	$(S_{16} : 0.00), (S_{17} : 0.00), (S_{18} : 0.00), (S_{19} : 0.00), (S_{20} : 0.00)$	$(S_{21} : 0.50)$	$(S_{22} : 1.00)$

between readings is “less than one day” and “three days or more and less than seven days.” In the abstract which should be presented just before third reading, the sentences in the first read part were not selected in the case that interval between readings is “more than 30 days.” From these results, the interval between reading is not affecting the abstract, and the assumption that the longer interval between readings requires the information in the beginning of the novel was denied.

We focused on the sentences selected from the first read part for the abstract at the third and fourth reading. The participants who selected sentences from the first read part for the abstract at the third reading selected sentences in the first read part for the abstract at the fourth reading, too. On the other hand, the participants who did not select sentences from the first read part for the abstract at the third reading also did not select the sentences in the first read part for the abstract at the fourth reading. This tendency is not depended on novels. From these results, it was considered that whether the information in the beginning of the novel is important or not would be depended on individuals. In addition, the participants who select sentences from the first read part for the abstract at the third reading select the sentences in the first read part for the abstract at the fourth reading, but the ratio was decreased. Thus, the important information for the abstracts is change depending on each reading progress.

D. Semantic Information in the Abstract

TABLE VII shows the number of sentences including semantic information, such as character names, persons, places, and lines, that were selected for the abstracts. The percentages in TABLE VII are the ratio of sentences including semantic information in each total number of sentences: all sentences of the novel, the sentences for abstract selected by more than one participant, and the sentences for abstract selected by more than two participants.

The appearance rate of character names in the sentences selected by more than one or two participants for the abstract is higher than that in all of the sentences of the novel. From this, it was suggested that the information of character name contributed to extract the sentences for the abstracts that were required by many readers. Also regarding the information of place, the appearance rate of places in sentences selected by more than one or two participants for the abstract is higher than the appearance rate in all of the sentences of the novel. However, there is a difference of more than 10% in appearance rate in N_1 and N_2 , but there is only 2.5% difference in N_3 . Therefore, it was suggested that there might be differences in the effectiveness depending on the novels. Concerning the information of persons, in the results of N_1 and N_2 , the appearance rates in the sentences selected by more than one or two participants for the abstract are higher than the appearance rate in all sentences of the novel. However, in the result of N_3 , the appearance rate in all sentences of the novel is higher than the appearance rates in the sentences selected by more than one or two participants for the abstract. The ratios of the sentences including persons in all sentences of each novel were $N_1 : 42.5\%$, $N_2 : 12.5\%$, and $N_3 : 10.2\%$. It seemed that because of the narrator “I” frequently appeared, higher appearance rate was showed in sentences of abstracts in N_1 . In N_1 , “I” is one of the important people, and it seems that “I” was thought as a character. As for the lines, in all three novels, the appearance rates in sentences selected by more than one or two participants for the abstracts is 10% lower than the appearance rate in all of the sentences of the novel. Thus, information of lines may be effective to generate the abstracts by subtracting the importance of sentence including lines from the importance of sentence.

TABLE V
 RELATIONSHIPS BETWEEN READING INTERVAL AND ABSTRACT. (THE ABSTRACT SHOULD BE PRESENTED BEFORE 3RD READING.)

Novel ID	Less than 1 day	1 day or more and less than 3 days	3 days or more and less than 7 days	7 days or more and less than 14 days	14 days or more and less than 30 days	30 days or more
N_1		$(S_2 : 0.00)$, $(S_5 : \mathbf{0.63})$, $(S_6 : 0.00)$	$(S_1 : \mathbf{0.38})$, $(S_4 : 0.00)$, $(S_7 : 0.00)$	$(S_3 : 0.00)$		
N_2	$(S_{10} : 0.00)$, $(S_{11} : 0.00)$, $(S_{13} : \mathbf{0.63})$, $(S_{15} : 0.00)$			$(S_9 : 0.00)$	$(S_8 : \mathbf{0.25})$	$(S_{12} : 0.00)$, $(S_{14} : 0.00)$
N_3	$(S_{20} : 0.00)$	$(S_{18} : 0.00)$, $(S_{19} : 0.00)$, $(S_{21} : \mathbf{0.50})$	$(S_{22} : \mathbf{0.38})$	$(S_{19} : 0.00)$	$(S_{16} : \mathbf{0.25})$	

TABLE VI
 RELATIONSHIPS BETWEEN READING INTERVAL AND ABSTRACT. (THE ABSTRACT SHOULD BE PRESENTED BEFORE 4TH READING.)

Novel ID	Less than 1 day	1 day or more and less than 3 days	3 days or more and less than 7 days	7 days or more and less than 14 days	14 days or more and less than 30 days	30 days or more
N_1	$(S_5 : \mathbf{0.50})$, $(S_6 : 0.00)$	$(S_4 : 0.00)$	$(S_3 : 0.00)$			
N_2	$(S_{13} : \mathbf{0.50})$, $(S_{14} : 0.00)$		$(S_8 : \mathbf{0.13})$, $(S_{12} : 0.00)$		$(S_9 : 0.00)$	
N_3						

TABLE VII
 THE NUMBER OF SENTENCES INCLUDING EACH SEMANTIC INFORMATION IN ALL SENTENCES OF NOVEL AND SENTENCES FOR ABSTRACT. THE RATIO OF THE NUMBER OF SENTENCES INCLUDING EACH INFORMATION IN THE NUMBER OF ALL SENTENCES IS SHOWN IN PARENTHESES.

		Characters	Persons	Places	Lines	# of sentences
N_1	All sentences of the novel	708 (39.8%)	756 (42.5%)	135 (7.6%)	715 (40.2%)	1179
	Sentences selected by more than 1 participant	61 (64.9%)	65 (69.1%)	23 (24.5%)	26 (27.7%)	94
	Sentences selected by more than 2 participants	13 (65.0%)	17 (85.0%)	6 (30.0%)	3 (15.0%)	20
N_2	All sentences of the novel	393 (35.7%)	138 (12.5%)	77 (7.0%)	579 (52.6%)	1100
	Sentences selected by more than 1 participant	62 (55.9%)	22 (19.8%)	22 (19.8%)	37 (33.3%)	111
	Sentences selected by more than 2 participants	22 (62.9%)	7 (20.0%)	8 (22.9%)	11 (31.4%)	35
N_3	All sentences of the novel	232 (48.3%)	49 (10.2%)	53 (11.0%)	116 (24.2%)	480
	Sentences selected by more than 1 participant	41 (78.8%)	1 (1.9%)	7 (13.5%)	2 (3.8%)	52
	Sentences selected by more than 2 participants	20 (80.0%)	1 (4.0%)	4 (16.0%)	0 (0.0%)	25

TABLE VIII
 THE RESULTS OF THE t -TEST. ($C_{q,q-1}$)

	# of data	Average	Standard deviation	t -value
N_1	$C_{q,q-1} > 1$	13	14.77	5.59
	$C_{q,q-1} < 1$	13	7.54	2.27
N_2	$C_{q,q-1} > 1$	16	20.50	5.47
	$C_{q,q-1} < 1$	16	11.25	3.63
N_3	$C_{q,q-1} > 1$	7	36.57	7.31
	$C_{q,q-1} < 1$	7	14.00	2.56

E. Relationships between Appearance Rate of each Word and Abstract

Basically, it has been known that the more the word appears the more important the word should be. If the information just before the bookmark is important, the appearance of the words in the read part just before the bookmark might be relatively higher than that after the bookmark. On the other hand, if the information related to the contents in the next reading would be required for the abstract, the words which would be more appeared in the next reading part should be included in the abstract. As mentioned in section I, we have already suggested the method based on this idea [1]. To verify this idea, we analyzed the relationships between the appearance rate of each word and the abstract.

The appearance rates of the word x selected by each

participant for the abstract is calculated by the equation (3);

$$\epsilon(x_i) = \frac{f(x_i)}{\sum_{i=0}^M f(x_i)}, \quad (3)$$

where, word x is only a noun, verb, and unknown word, and the word whose total number of appearance in each novel is over five times or more. M shows the total number of words x . Also, the words that do not have a meaning by themselves such as suffixes were excluded.

Let q be the current part that the reader starts to read, and then $q - 1$ means the part that the reader read last time. Each increasing rate of word x for part q and $q - 1$ is each calculated by the equation (4) and (5).

$$R_q(x_i) = \frac{\epsilon_q(x_i)}{\epsilon_{q-1}(x_i)}, \quad (4)$$

$$R_{q-1}(x_i) = \frac{\epsilon_{q-1}(x_i)}{\epsilon_{q-2}(x_i)}. \quad (5)$$

Then, the equation (6) calculates the contrast between $R_q(x_i)$ and $R_{q-1}(x_i)$.

$$C_{q,q-1} = \frac{R_{q-1}(x_i)}{R_q(x_i)}. \quad (6)$$

We verified that the number of words which $C_{q,q-1}$ is higher and lower than 1.0 with t -test under the condition

that is the one-sided test and significant difference of 5%: the significant t -value should be 0.0167 after Bonferroni correction. TABLE VIII shows the result. The t -values of each novel were $N_1 : 0.0015$, $N_2 : 0.0000$, and $N_3 : 0.0001$. All of those values are smaller than the 5% significance level, moreover, N_2 and N_3 showed the 1% significance level. From this results, it was suggested that the abstract should include the words which appearance rate on $q - 1$ is higher than that on $q - 2$ and q . From these, it seemed that the words with a climax on the part just before the bookmark should be important for the abstract.

V. CONCLUSION

The purpose of this study is to generate the abstract depending on reading progress. In this paper, the experience to create human-made abstracts was conducted and we analyzed the features of the abstract. Through the discussion, we have obtained the following findings;

- The position where the bookmark is inserted is biased to a boundary of the story.
- There are three patterns to create abstracts:
 - Sentences are extracted from only the part just before the bookmark.
 - Sentences are extracted from the beginning part of the novel and the part just before the bookmark.
 - Sentences are extracted exhaustively from all of the read parts.
- Whether the sentence just before the bookmark is included in the abstract or not depends on the attribute of each reader.
- The important information for the abstracts is changeable depending on each reading progress.
- There is no relation between the reading intervals and which sentences are selected for the abstract.
- The information of character name and place may work effectively to generate the abstracts.
- It seems that the words that have a climax on the part just before the bookmark is important for the abstract.

In the future, we will discuss the abstracts for different novels created by the same reader. We will analyze the features of the abstracts that created by free description, though the participants create the abstracts by selecting sentences from the novel in this paper. Also, based on the findings effective for the abstract generation clarified in this paper, we will develop a method to automatically generate the abstracts suitable for each reader. In particular, as mentioned in section IV-E, we have proposed the method to generate the abstract focusing on the climax point of the words in the novels. We are considering the enhancing the entertainment of novel reading using information from other sources such as images and maps.

REFERENCES

- [1] H. Mori, R. Yamanishi, Y. Nishihara, and J. Fukumoto, "The Difference of Word Importance before and after Bookmark for Novel Abstract in Each Reading Progress," in *Proc. of the 21th International Conference on Knowledge Based and Intelligent Information and Engineering Systems*, 2017, pp. 1246–1253.
- [2] D. Wang and T. Li, "Document Update Summarization Using Incremental Hierarchical Clustering," in *Proc. of the 19th ACM Intl' Conf. on Information and Knowledge Management*, 2010, pp. 279–288.
- [3] D. Bamman and N. A. Smith, "New Alignment Methods for Discriminative Book Summarization," *CoRR*, vol. abs/1305.1, 2013. [Online]. Available: <http://arxiv.org/abs/1305.1319>
- [4] M. C. Green, T. C. Brock, and G. F. Kaufman, "Understanding Media Enjoyment: The Role of Transportation Into Narrative Worlds," *Communication Theory*, vol. 14, no. 4, pp. 311–327, nov 2004.
- [5] A. S. L. Tsang and D. Yan, "Reducing the Spoiler Effect in Experiential Consumption," *Advances in Consumer Research*, vol. 36, pp. 708–709, 2009.
- [6] J. Boyd-Graber, K. Glasgow, and J. S. Zajac, "Spoiler alert: Machine learning approaches to detect social media posts with revelatory information," *Proceedings of the American Society for Information Science and Technology*, vol. 50, no. 1, pp. 1–9, 2013. [Online]. Available: <http://dx.doi.org/10.1002/meet.14505001073>
- [7] S. Guo and N. Ramakrishnan, "Finding the Storyteller: Automatic Spoiler Tagging Using Linguistic Cues," in *Proceedings of the 23rd International Conference on Computational Linguistics*, ser. COLING '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 412–420. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1873781.1873828>
- [8] K. Maeda, Y. Hijikata, and S. Nakamura, "A Basic Study on Spoiler Detection from Review Comments Using Story Documents," in *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2016, pp. 572–577.