DEIM Forum 2012 D5-3

Generating Slide Skeletons based on Expression Styles for Presentation Contents

Yuanyuan WANG † and Kazutoshi SUMIYA ‡

† Graduate School of Human Science and Environment, University of Hyogo
1-1-12 Shinzaike-honcho, Himeji, Hyogo, 670-0092 JAPAN
E-mail: † {ne11u001@st, sumiya}shse.u-hyogo.ac.jp,

Abstract By using PowerPoint or Keynote that can effectively create attractive presentation slides, a presentation-based communication environment can now be created in which people can use these presentation slides to exchange and discuss ideas together. However, because it is necessary to prepare many slides to enable audiences to understand the content, authors need to prepare the best possible slides. Our skeleton generation method is designed to help authors to prepare slides with ease by constructing slide layouts based on the expression styles of words from the text in the textbooks they use. To do this, we analyze the expression styles of the words presented in the slides; our method can then extract context-role of the words by the differences between the words in the texts and their slides. To generate skeletons for slides from target texts in a textbook, our method derives the expression styles of the words from pre-existing texts and their slides. Finally, it generates slide skeletons by using the derived expression styles of the corresponding words from the target texts arranged in slides, which are the same as the layouts of pre-existing slides. We also present the results of an evaluation of the method's effectiveness.

Keyword e-Learning, Skeleton generation, Expression styles, Slide making support

1. Introduction

Presentations now play a socially important role in many fields, including business and education, among others. Many university teachers have used Web services such as SlideShare [1] and CiteSeerX [2] to store the slides they use in lectures. However, because teachers prepare many slides to enable students to understand their content, the teachers should prepare the best possible slides. In fact, when authors plan their slides often refer to texts (e.g., lectures in a textbook) to determine the information should be conveyed. It is important to focus on how to express the information that will appear in slides from texts. We can generate skeletons serve as slide layouts that express typical words from the texts based on their context-role in slides by considering how to convey the words to create the layout of the slide. For example, a word “vegetable” appears in the body of text in one slide entitled “Agriculture Market,” which is a contextual summary of the information is referred from sections entitled “Agriculture Market Analysis” and “Vegetable Production” in a text of a textbook.

Our approach creates an editable slide skeleton for slide-making that is able to produce a slide layout based on specific words to help authors prepare slides easily and efficiently. In order to explore slide skeleton creation, we find that a word is expressed in different ways in slides. For instance, a word may be the title of one slide, or the same word appears in the body of text in another slide in a presentation content. We found that there are variety styles of presentation slides made by the same text based on the different expressions of the words. We therefore present a skeleton generation method for making new slides from a target text based on expression styles of words in the slides by analyzing the differences between pre-existing texts and their slides are referred (see Fig. 1). In this paper, we define the expression styles that the level positions of the words are arranged in slides for the expression of presentation, based on the context-role of the words in the slides by considering how each word represented differs from a slide and its text. We derived the document structure from texts by focusing on their logical units, and the document structure of slides by focusing the level of indentation in slide text that are often used to help users better organize their slide contents.

As depicted in Fig. 2, when a textbook contains a number of lectures, an author can take a target text as #2's text to prepare slides. When the logical units that constitute #2's text are the same as in the pre-existing #1's

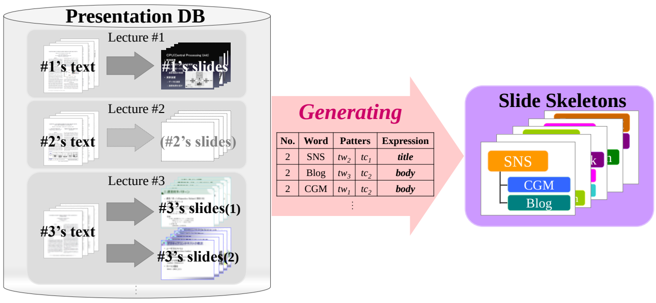

Fig. 1 Conceptual diagram of skeleton generation from textbook and its slides

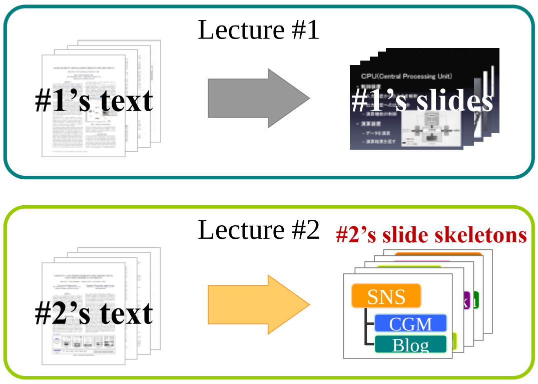

Fig. 2 Skeletons generation from target text using pre-existing text and their slides

text, we can generate skeletons for #2's slides from #2's text, based on the expression styles of the words in the #1's text and their slides that are referred.

We found that there were two main features particularly helpful for deriving expression styles, based upon the differences between context-role of words by analyzing the document structure of texts and their slides: (1) A word appears dispersed or centered in a text; and the word may appear in the slide title or in lines that are less indented, or the body of the slide in lines that are more indented from one slide to another slides [3]. (2) When a word in the title or in the body of one slide, which has the referential context, indicates from what kind of content in sections in a text that the word in slide is referred [4]. We consider a slide is made that it has its referential context always refer to one or multiple sections in a text. In this paper, we supposed that when an author prepares presentation slides for each lecture from texts in a textbook, he/she can take the texts to arrange the words in slides according with the same context-role of the words in the pre-existing slides made from the pre-existing texts. Therefore, we can generate skeletons for slides from a target text based on the expression styles of words by extracting the differences between the role of the words in pre-existing texts and their slides that are in a textbook.

The remainder of this paper is organized as follows. The next section provides a brief survey of related works. Section 3 describes how to determine the role and context of words in texts and slides. Section 4 presents the skeleton generation for slides. Experimental results and conclusions are given in Sections 5 and 6, respectively.

2. Related Work

Most of the research related to slide-making support has focused on slide generation. Mathivanan et al. [5], Beamer et al. [6] and Yasumura et al. [7] proposed a system for generating slides from academic papers. Their method extracts information from a paper by the TF-IDF method, and assigns the sentences, figures and tables in slides by identifying important phrases for bullets. Shibata et al. [8] converted Japanese documents to slides representation by parsing their discourse structure and representing the resulting tree in an outline format. However, conventional approaches that focus only on the consistency of the document structure in the text and slides, both ignore the context-role played by how to express words from the text to the slides. Our method focuses on the differences between the role of words in texts and their slides with the context information, and it generates slide skeletons based on the expression styles of words.

Kan [9] proposed a system for the discovery, alignment, and presentation of such document and slide pairs. Hayama et al. [10] aligned academic papers and slides based on Jing's method, which uses a hidden Markov model. These studies are similar to ours for analyzing information that is common to texts and their slides. Our approach focuses not only on the information that is in common, but also on information that differs between texts and slides. Zettsu et al. [4] discovered aspects for characterizing Web pages based on their contexts. They considered that a Web page is referred to by other pages in various contexts through links, these contexts indicate the reputation of the page. Then, we consider that a word in slides refer to some sections in a text in various contexts. Therefore, our goal is to generate skeletons for slides by analyzing the differences between the role of words in texts and their slides.

3. Determination of Context-Role of Words Using Document Structures

In our proposed method, we determine context-role of words based on the document structure in the text and then taking the document structure in slides. A chapter in a textbook is referred to as a text. We define the document structure of a text in terms of its logical units, which consist of sections, which in turn consist of a section head and paragraphs. The content of a presentation includes a number of slides that have structured text information. We define the document structure from slides, based on the indentations in the slide text. We define the slide title as the 1st level. The first item of text is considered to be on the 2nd level, and the depth of the sub-items increases with the level of indentation (3rd level, 4th level, etc.).

3.1. Determination of Role of Words in a Text

When the location in which a word appears dispersed in the text, the role of the word is deemed dispersion; this word is called W_d. In contrast, when the location in which a word appears centered in the text, the role of the word is deemed concentration; this word is called W_c. We explain the determination of W_d and W_c using the word *b*, and we calculate the degree of dispersion and concentration of *b* in the text. When *b* is dispersed to a high degree, *b* is determined to be W_d; and when *b* is centered to a high degree, *b* is determined to be W_c.

$$W_d = \{b \mid \min(\frac{\sum_{v=1}^{n} dist(c_1, b_v)}{n}, \cdots, \frac{\sum_{v=1}^{n} dist(c_j, b_v)}{n}) > \alpha\} \quad (1)$$

$$W_c = \{b \mid \min(\frac{n}{\sum_{v=1}^{n} dist(c_1, b_v)}, \cdots, \frac{n}{\sum_{v=1}^{n} dist(c_j, b_v)}) > \alpha\} \quad (2)$$

Where b_v is the v^{th} word *b*, and c_j is the j^{th} section in the text. The function *dist* calculates the distance between sections. The distance between sections is a number that indicates how many sections there are between two words. *n* is the number of times that *b* appears in a text. When the words appear in the same section, the distance between them in the section is 1. The minimum value of the word is extracted using the function *min* because there are unknown expectations. The highest degree of expectation is obtained for a position in a section with the lowest degrees of dispersion and concentration. W_d or W_c is a bag of words in the text, if the formula is greater than a threshold α in Eq. (1), and the role of *b* is determined to be the dispersion in W_d; if the formula is greater than a threshold α in Eq. (2), and the role of *b* is determined to be the concentration in W_d.

3.2. Determination of Role of Words in Slides

If a slide has more information in terms of a given word than is contained in a prior slide in the presentation file, the role of the word is upper, and this word is called W_u. In contrast, if a slide has generalized information in terms of a given word than is contained in a prior slide in the presentation file, the role of the word is lower, and this word is called W_l. We explain the determination of W_u and W_l using the word *g*, which is present in both slide *x* and slide *y*. When *g* and the other words in slides *x* and *y* satisfy certain conditions, *g* is determined to be W_u or W_l.

$$K(x,g) = \{k_i \mid k_i \in x, l(x,g) < l(x,k_i)\} \quad (3)$$

Here, $K(x,g)$ is a bag of words that can be considered to provide an explanation in terms of *g* in slide *x*. $l(x,\mathrm{g})$ is a function that returns the level of indentation of *g* in slide *x*. When *g* appears frequently in slide *x*, $l(x,\mathrm{g})$ will return the lowest possible value; that is, the uppermost level at which *g* occurs in slide *x*. This because we consider that when *g* appears in an upper level, all the other levels in which *g* appears in the body of that slide are explanatory points related to a deeper occurrence of *g*. The word k_i is included in the levels that have a hierarchical relationship with the level of *g*, and k_i belongs to the bag of words $K(x,g)$ in slide *x*. $l(x,k_i)$ is greater than $l(x,\mathrm{g})$, in that k_i is a child of *g* in the document structure. When k_i is not present in slide *x*, $K(x,g)$ will be empty. Based on the above criteria, we compute the number of words in detailed information related to *g* for slides *x* and *y*, and compare their numbers using the following formulas:

$$W_u = \{g \mid |K(x,g)| < |K(y,g)|\} \quad (4)$$

$$W_l = \{g \mid |K(x,g)| > |K(y,g)|\} \quad (5)$$

where the function $|K(x,g)|$ extracts the total number of k_i in $K(x,g)$ in slide *x*. $K(y,g)$ are also bags of words in slide *y*, and they satisfy the same conditions as $K(x,g)$ in Eqs. (4) and (5). Thus, Eqs. (4) and (5) can be used to count the number of words in $|K_l(x,g)|$ for slide *x* and the number of words in $|K(y,g)|$ for slide *y*. W_u or W_l is a bag of words in the slides, if the number count for slide *x* is lower than that for slide *y* in Eq. (4), the role of *g* is determined to be upper in W_u; and if the number count for slide *x* is greater than that for slide *y* in Eq. (5), the role of *g* is then determined to be lower in W_s.

3.3. Determination of Referential Context of Word

We define the referential context for words in slides from the text. Here, a word is described in Subsections 3.1 and 3.2. The referential context consists of sections or subsections that the word in slides is referred, and the titles of other slides containing the word that are referred. The referential context indicates from what kind of content the important word is referred. For a given word *m*,

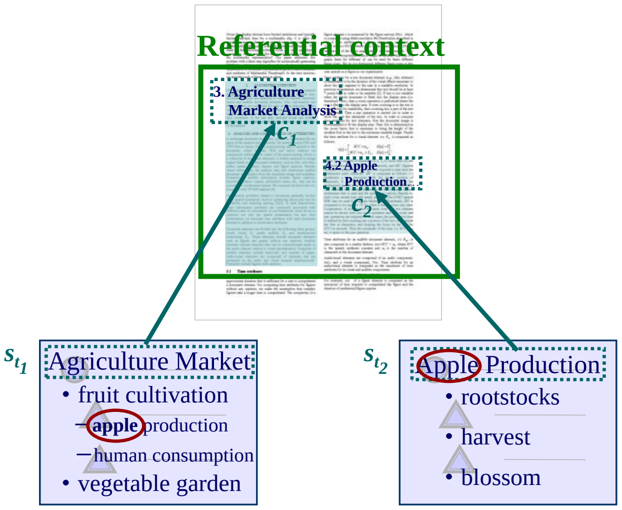

Fig. 3 Example of referential context

words in the titles of sections or subsections and slides are extracted: …, c_2, c_1, s_{t_1}, s_{t_2}, m. Here, c_i is the section title or subsection title, S_{t_i} is the slide title. A slide is titled s_{t_i} corresponds to the section or subsection is titled c_i in the referential context $C(s_{t_i},m)$, while the words in s_{t_i} is similar to that in c_i. This is done using the Simpson similarity coefficient [11] as $Sim(c_i,s_{t_i})=|c_i \cap s_{t_i}|/\min(|c_i|,|s_{t_i}|)$. When $Sim(c_i,s_{t_i})$ exceeds a predefined threshold, the words in c_i and s_{t_i} are similar. The slide s_{t_1} contains the word m, and the section or subsection c_1 is referred in the referential context $C(s_{t_1},m)$.

Fig. 3 illustrates the referential context of the word in a slide. The referential context of the word “apple” in a slide “Agriculture Market” indicates that a section on the “Agriculture Market Analysis,” which describes the information regarding to “apple” that is contained in the slide “Agriculture Market” refers to it. On the other hand, a slide “Apple Production” refers to “apple” in the slide “Agriculture Market” and it also refers to a section on the “Apple Production” in the referential context of “apple.”

4. Skeleton Generation

4.1. Detecting Expression Styles

To generate skeletons for slides, a slide layout is used, which consists of words based upon expression styles using the differences between the context-role of words in the pre-existing text and their slides.

For the differences between the role of word q in the slides and the text, we distinguish the following 3 categories:

- tw_1: $q \in W_d \cap W_u$, the role of q in the text is dispersion and the role of q in the slides is upper.
- tw_2: $q \in W_d \cap W_l$, the role of q in the text is dispersion and the role of q in the slides is lower.
- tw_3: $q \in W_c \cap W_u$, the role of q in the text is concentration and the role of q in the slides is upper.

P	tc_1	tc_2
tw_1	*title/body*	*body*
tw_2	*title*	*body*
tw_3	*title/body*	*body*

Table 1 Patterns in expression styles

For the differences between a word that appears in slides and their referential context by considering what sections or subsections in the referential context that are corresponded. For this, we compute the degree of the similarity of the bag of words S_{t_i} in the title of a slide that contains the word q to the bag of words C_i in the title of a section or subsection. We consider that q in one slide may correspond to one or multiple sections, based upon the similarity of the slide title and the section title, and there are 2 categories:

- tc_1: q in one slide whose referential context is one section or subsection in the text.
- tc_2: q in one slide whose referential context is multiple sections or subsections in the text.

From the differences between the role of words in the text and slides, we can find which of these changes both in the text and slides. In addition, from the differences of the words in slides and their referential contexts in the text, we can find how the words should be summarized from the text to slides, and whether multiple sections are referred, or one section is corresponded in referential context of slides. In the example shown in Fig. 4, the word “document” is dispersed in all sections in Chapter 5, and “document” also is a title of slide a6 of a Presentation 5. When the role of “document” is dispersion in the text and upper in slides as tw_1, “document” in the slide a6 refers to in some sections in referential context from the text as tc_2. Slide a6 is concentrated the topic of “document” when it summarizes multiple sections in referential context in terms of “document” in Chapter 5. On the other hand, when the word “summary” repeatedly appears in a certain text segment, slide a3 is titled “summary” of Presentation 5. When the role of “summary” is concentration in the text and upper in slides as tw_3, and one section is corresponded in referential context of slide a3 as tc_1. We consider that slide a3 offers specialized information regarding “summary” refers to a specific context from Chapter 5.

Therefore, we define the expression style ES that the

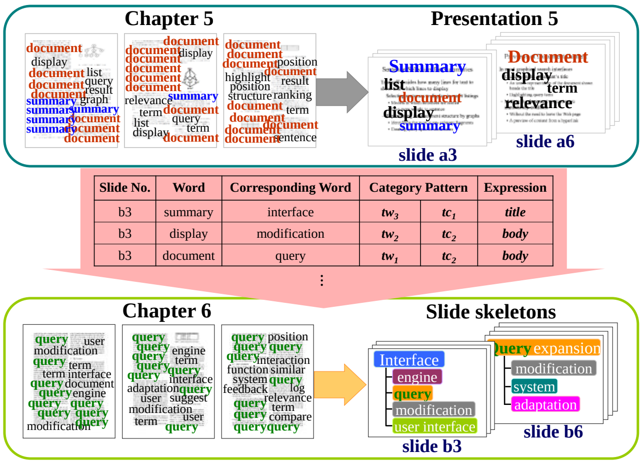

Fig. 4 Example of skeleton generation

context-role R of words with the expression E of presentation is represented by the level positions of the words in slides as follows:

$$ES = (R, E) \tag{6}$$

$$R = (w_i, p_{w_i})(w_i \in W, p_{w_i} \in P) \tag{7}$$

$$W = (W_d \cup W_c) \cup (W_u \cup W_l) \tag{8}$$

$$P = \{p_{w_1}(tw_1, tc_1), \ldots, p_{w_3}(tw_2, tc_1), p_{w_4}(tw_2, tc_2), \ldots, p_{w_6}(tw_3, tc_2)\} \tag{9}$$

here, W is a bag of words that belongs to W_d, W_c, W_u or W_l that can be considered as the words that play key context-role in the slides. E denotes the level positions of the words in slides by the context-role of the words in R, and P denotes the total of 6 patterns that intend the context-role of the words in R, and the words belong to W. These patterns combine 3 categories of differences in the role of words and 2 categories of differences of the words in slides and their referential contexts in text that are shown in Table 1.

4.2. Generating Skeletons for Slides

Presentations consist of slides that rely on a combination of words and images to drive home a point. The way can combine these elements creates the design that layout of the slide. Layouts are crucial to making a slide understandable and unforgettable. In this paper, we define skeletons for slides that different slide layouts to best communicate key points from texts focused on how to express key points in slides. We consider that key points as context-role of words from texts to their slides are extracted by our proposed method. Therefore, we create slide skeletons that construct different layouts to express the words following the specific role are determined as expression styles of the words.

Based upon the expression styles drawn from pre-existing texts and slides, we can generate skeletons for slides from a target text in the same textbook by extracting the word in the target text that corresponds to the words in pre-existing texts. Therefore, we can use the same expression styles of the words in the pre-existing texts applying to the corresponding words from the target text to generate skeletons for slides from the target text.

We consider texts in which the chapters in a textbook have the same document structure as the sections in each chapter. When a tree of a word z belongs a tree t_A of document structure of the pre-existing text A and a tree of a word z’ belongs a tree t_B of document structure of the target text B are consistency, we consider that z’ in B corresponds to z in A. Next, we extract z’ in B by matching the partial trees of z and z’ in t_A and t_B have employed a structure matching method [12]. We consider that the words in A and B that are not the common words, the structure matching method can help identify non-linguistic matches and disambiguate between seemingly identical structures in different contexts as A and B.

For each word, when there are many words in two partial trees of z and z’ in t_A and t_B to compare, the number of the partial trees are consistency to be larger. We define a set of the partial trees $SP(t_A(z))$ of z in t_A and a set of the partial trees $SP(t_B(z'))$ of z’ in t_B. We extract a pair C_p of z in A and z’ in B as the following formula:

$$C_p = \left\{ (z, z') \middle| \frac{1}{2}\left(\frac{sum(SP(t_A(z)), SP(t_B(z')))}{N_A} + \frac{sum(SP(t_A(z)), SP(t_B(z')))}{N_B} \right) > \theta, z \in W \right\}$$

where the function *sum* extracts the total number of $SP(t_A(z))$ of z in t_A and $SP(t_B(z'))$ of z’ in t_B are consistency. N_A is the number of $SP(t_A(z))$ of z in t_A, and N_B is the number of $SP(t_B(z'))$ of z’ in t_B. We calculate the similarity of the trees $t_A(z)$ of z in A and $t_B(z')$ of z’ in B by the above formula. If the formula is greater than a threshold θ that $SP(t_A(z))$ of z and $SP(t_B(z'))$ of z’ are similar, z’ is determined to be the corresponding word of z. Therefore, we are able to generate skeletons for layout slides by using the expression style of z’ in the same expression style as z, which is performed according to Eqs. (6), (7), (8) and (9), and the number of skeletons for slides is the same as the number of pre-existing slides.

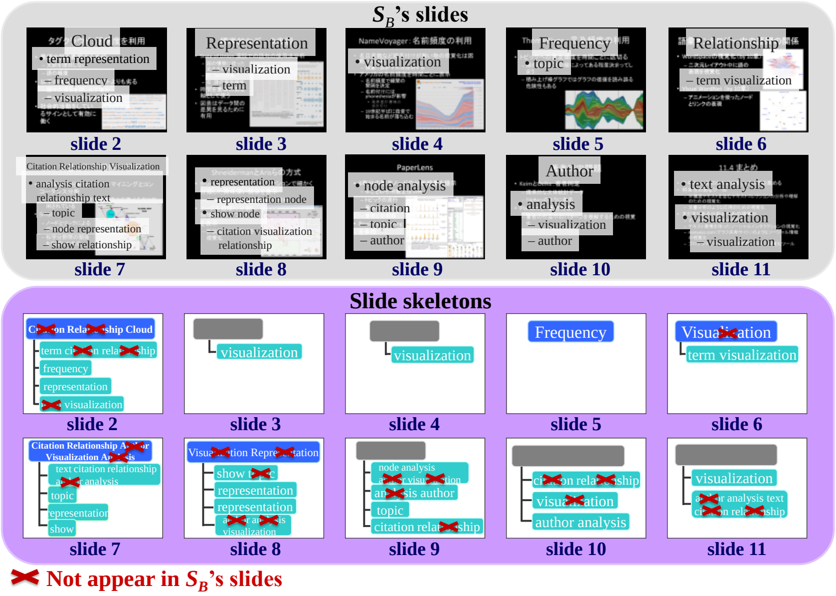

Fig. 5 Generated slide skeletons compared with slides in S_B

For example, an author wants to make presentation slides for lecture 6 regarding Chapter 6 about Query Reformulation in a textbook named Search User Interface. Our method generates skeletons for slides from Chapter 6, referring to slides in Presentation 5 from Chapter 5 about Presentation of Search Results (see Fig. 4). In Chapter 5 the word "document" appears in all sections. Meanwhile, if "document" is a title of slide a6 of a Presentation 5, then the expression of "document" that the level position of it in slide a6 is title. In Chapter 6 the word "query" appears in all sections that correspond to "document" in Chapter 5. The skeleton for slide b6 generated from Chapter 6 shows that "query" appears in the title of slide b6, which explains "query refinement suggestion" in referential context in terms of "query" from some sections of Chapter 6. Next, "query" in slide b6 has the same expression style as "document" in slide a6. When the author makes slides referring to the slide skeletons, such as slide b6, the information for "query" in slide b6 is constructed in the same way as it is for "document" in slide a6, based upon the same expression style by arranging the words to express "query" in the title of slide b6. The generated skeletons can be used to create slide layouts that construct words according to the same role the words play in pre-existing slides, and these skeletons then enable the author to make slides easily.

5. Evaluation

5.1. Dataset

The aim of this experiment was to verify whether our method is useful for generating skeletons for slides. We first prepared two presentation files: S_A from text A and S_B from text B were made by the same person, both from Chapter 11 on Information Visualization for Text Analysis in a textbook called Search User Interface [13]. Because of their single authorship, S_A and S_B both have the same expression styles, and A and B have the same document structure. Each presentation file contains 10 slides, not counting the cover slide. We used A and S_A to generate skeletons from B based on our method; the slides in S_B serve as correct answers regardless of whether the level positions of the words in the slides generated from skeletons from B are correct or not.

5.2. Experiment 1: Validity of Determining Referential Context of Word

For generating slide skeletons, we used A and S_A to completely extract the expression styles of 42 words appeared in S_A from A, which correspond to 19 sections as the referential context of words in A, based on our method. Then, we verified the sections are corresponded to the words in the referential context of these words are correct or not by human-based experimental results.

In the experimental results, the precision of referential context of words in slides by the corresponded sections based our method was 45.2%(19/42), and the recall of that was 40.4%(19/47). The result for the referential context of words was low. For example, our method determined the expression style of a word in S_A that corresponded to multiple sections in referential context from A; however, in the correct answer the word corresponded one section in referential context from A. In addition, we need to consider the figure captions for determining the referential context in the text. S_A contains a number of words in slides, and they appear in figure captions in the texts. However, the corresponding information of those words in figure captions in the text that cannot be determined in the referential context by our method.

5.3. Experiment 2: Validity of Generating Skeletons

We generated 10 slide skeletons from B with the same number of slides as in S_A, and 59 level positions of words from B were arranged in slide skeletons based on the expression styles of the corresponding words in A. Finally, we compared them with the correct answer as answers as S_B's slides (see Fig. 5).

In the experimental results, the precision of the level positions of words in slides by the generated skeletons based our method was 57.6%(34/59), and the recall of that was 75.6%(34/45). This experiment showed that our method can arrange the words in slides using generated skeletons based on their expression styles. However, when our method determine the expression style of some words that were the body of a slide in S_A; we used the same expression style for the corresponding words that were 2nd and 3rd level in the body of the corresponding slide in the correct answer S_B. Then, we need to consider hierarchical relationships between words in the body of text in slides. And the results of this experiment suggest that we need to improve the skeleton-generation algorithm not the expression styles of the words in slides, but also consider the phrases in slides.

6. Concluding Remarks

In this paper, we proposed a method of skeleton-generation that provides support for making slides based on the expression styles of words. We described in detail how to expression styles are determined by extracting the patterns that combine the differences between the role of words and their referential contexts in texts for slides, respectively. To generate skeletons for slides from a target text, we extracted the words in the target text that correspond to the words in pre-existing text, and we then used the same expression styles of the words in the target text.

In the future, we plan to improve our algorithm for the generation of skeletons of slides from texts and to evaluate it using a large set of actual texts and slides pairs, and we should evaluate our method can support to make slides or not by users make presentation slides using our generated slide skeletons.

References

[1] SlideShare, http://www.slideshare.net/.

[2] CiteSeerX, http://citeseer.ist.psu.edu/index.

[3] Y. Wang and K. Sumiya, "Skeleton Generation for Presentation Slides based on Expression Styles," in *Proc. of the 5th International Conference on Intelligent Interactive Multimedia Systems and Services* (*KES IIMSS 2012*), 2012.(to appear)

[4] K. Zettsu, Y. Kidawara, and K. Tanaka, "Discovering Aspects of Web Pages from Their Referential Context in the Web," in *Proc. of the 9th International Conference on Database System for Advanced Applications* (*DASFAA 2004*), pp. 618–629, 2004.

[5] H. Mathivanan, M. Jayaprakasam, K. G. Prasad, and T. V. Geetha, "Document summarization and information extraction for generation of presentation slides," in *Proc. of International Conference on Advances in Recent Technologies in Communication and Computing* (*ARTCOM 2009*), pp. 126–128, 2009.

[6] B. Beamer and R. Girju, "Investigating automatic alignment methods for slide generation from academic papers," in *Proc. of the 13th Conference on Computational Natural Language Learning* (*CoNLL 2009*), pp. 111–119, 2009.

[7] Y. Yoshiaki, T. Masashi, and N. Katsumi, "A support system for making presentation slides," in *Transactions of the Japanese Society for Arti cial Intelligence*, pp. 212–220, 2003 (in Japanese).

[8] T. Shibata and S. Kurohashi, "Automatic slide generation based on discourse structure analysis," in *Proc. of the 2nd International Joint Conference on Natural Language Processing* (*IJCNLP 2005*), pp. 754–766, 2005.

[9] M. Kan, "Slideseer: A digital library of aligned document and presentation pairs," in *Proc. of the 7th ACM/IEEE-CS joint conference on Digital libraries* (*JCDL 2007*), pp. 81–90, 2007.

[10] T. Hayama, H. Nanba, and S. Kunifuji, "Alignment between a technical paper and presentation sheets using a hidden markov model," in *Proc. of the 2005 International Conference on Active Media Technology* (*AMT 2005*), pp. 102–106, 2005.

[11] E. H. Simpson, "Measurement of diversity," in *Nature*, vol. 163, p. 688, April 1949.

[12] J. Madhavan, P. A. Bernstein and E. Rahm, "Generic Schema Matching with Cupid," in *Proc. of the 27th International Conference on Very Large Data Bases* (*VLDB 2001*), pp. 49-58, 2001.

[13] M. A. Hearst, "Search user interfaces," in *Cambridge University Press*, November 2009, pp.281-296.