Slide Retrieval Technique Using Features of Figures

Seitaro TANAKA, Taro TEZUKA, Atsushi AOYAMA, Fuminori KIMURA, Akira MAEDA.

Abstract— **Existing systems for presentation slide retrieval target only text information. However, such systems cannot always attain the user's intension precisely. Non-textual elements, such as figures, illustrations, and layouts, also have important information. However, existing techniques do not utilize them for slide retrieval. In this paper, we propose a slide retrieval method that considers shapes and arrangements of objects in a figure, and the area ratio of the figure in a slide, by calculating similarities between a query and each figure in presentation slides. By using our proposed slide retrieval system, users are able to retrieve presentation slides more intuitively and efficiently.**

***Index Terms*—Presentation slide, OpenXML, retrieval, image**

I. Introduction

PRESENTATION slides are used in many organizations for various purposes, and the number of these slides is rapidly increasing. Past presentation slides are sometimes reused as-is or with slight modifications for improving work efficiency. For this purpose, slide retrieval techniques are coming to be considered as important.

Conventional slide retrieval techniques consider only text information. However, presentation slides usually consist of more figures than documents in other formats such as Word. When a user searches for a slide, figures are an important factor for understanding the content of a slide. Moreover, meaningful object groups contained in a figure (hereinafter called “figure groups”), such as flow charts and block diagrams, are frequently reused.

However, conventional slide retrieval techniques depend on text information in slides. In this case, the user can retrieve a figure group only by its surrounding text.

Company employees regularly give presentations that consist of several hundred slides. When a user searches for a slide from a large number of past presentation slides, text-based search might result in many different slides that contain text matching the query. In such cases, it is difficult for the user to find the desired slide. Also, similar text might be contained in multiple slides in a presentation. In such a case, it is difficult for a user to find keywords to search effectively. However, when searching for a known slide, the user might remember the figure group contained in the figures rather than the text contents.

We propose a slide retrieval method that considers the figure information in the slides. Our method uses the user's image information as the query. For the features of the figures, we use shapes and their arrangement, and we calculate their similarities with the query on the basis of these features. Using these features, our system can find slides that meet the user’s needs.

II. Related Work

Hayama et al. proposed a technique for extracting the information about a search demand in accordance with arbitrary display domains [1]. This research is related to the information extraction from a slide. In this technique, only the information on the portion relevant to a search demand is extracted and shown.

Moreover, they developed a structured method for slides [2]. This technique systemizes a layout, a figure, etc. in a slide. In this research, the information included in a figure is extracted from the slide.

Min et al. [3] performed slide searches by regarding a slide as a picture. Their research is in the field of image processing. However, this method needs the background and insert of a slide to be distinguished. It also separates a background and object. Our research can specify object information. Therefore, our system can identify objects, such as a background. Thus, the problem of Min et al. does not occur.

Moreover, Kimoto [4] performed picture search by using a sensitivity word. This method calculates similarity of the sensitivity word to words in the figures. Furthermore, the value of the similarity is compared with the color scheme pattern of a figure. Then relationships among the sensitivity word, hue, chroma saturation, and brightness are clarified. This sensitivity search is used when searching based on the impression received from a figure.

Kitayama et al. [5] developed an information retrieval system for multimedia contents including presentation slides. They presented the surrounding context of the part to match the search query. Then scenes were extracted using the layered structure of the text information in a slide. The aim of our research is also to utilize the data structure of a presentation slide. However, we are researching not scene extraction but slide retrieval. Moreover, similarity is computed from the arrangement of the objects in a slide. In our research, the figure feature in a slide is extracted and used for slide retrieval.

Manuscript received January 8, 2013;
S.TANAKA is with Graduate School of Science and Engineering, Ritsumeikan University, Shiga, JAPAN, e-mail:cm007073@ed.ritsumei.ac.jp.
A. TEZUKA is with Graduate School of Library, Information and Media Studies, University of Tsukuba, Ibaraki, JAPAN, e-mail:tezuka@slis.tsukuba.ac.jp
A.AOYAMA is with Graduate School of Technology Management, Ritsumeikan University, Shiga, JAPAN, e-mail:aoyama@mot.ritsumei.ac.jp
F.KIMURA is with College of Information Science and Engineering, Ritsumeikan University, Shiga, JAPAN, e-mail: fkimura@is.ritsumei.ac.jp
A.MAEDA is with College of Information Science and Engineering, Ritsumeikan University, Shiga, JAPAN, e-mail:amaeda@media.ritsumei.ac.jp

Wang et al. [6] also analyzed presentation slides and investigated the relationships among slides. Between the keywords in the slide exists a notional relation. Therefore, the conceptual structure created by a keyword is used. The relationship within a slide is judged in accordance with the conceptual structure. Because our research is not concerned with the relationship within a slide, a figure is not necessarily used in each slide.

III. Proposed Method

Figs. 1 and 2 show the processing flows of our proposed method. The method has two main processes: data storage and retrieval. In the data storage process (Fig. 1), the system extracts data (i.e. features of figures) from presentation slides such as PowerPoint documents and stores them in the database. The retrieval process (Fig. 2) shows the processing flow in which the system returns retrieval results for a user when the user retrieves images. Both processes have a common process module: the figure information extraction module.

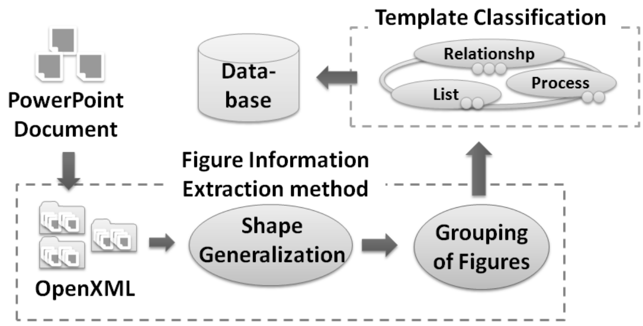

Fig. 1. Outline of data store process.

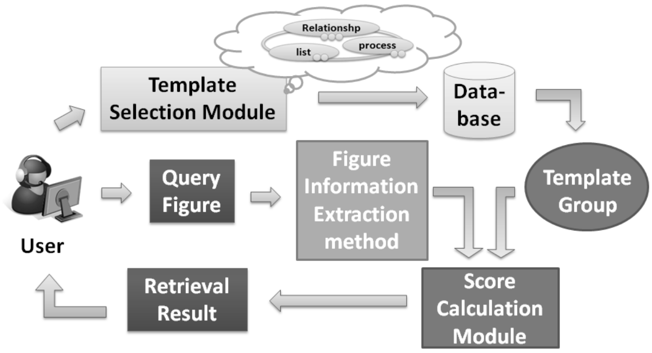

Fig. 2. Outline of retrieval process.

IV. Data Storage Process

Our system supports PowerPoint documents as presentation slides. First, the system extracts data from PowerPoint slides by using the figure information extraction module. Second, the system indexes these extracted data for each slide. Third, each slide is classified into one of the template types, which will be explained in Section B. Fourth, the system stores these indexes in the database.

Next, we explain the processing in the figure information extraction module. First, the module extracts figure information from OpenXML. Next, the system determines the form of the figure by shape generalization, which will be explained in detail in Section A.2. Finally, we group individual figures that make up one composite figure.

A. Figure Information Extraction Module

First, the system converts slides (or input queries) into Open XML. Open XML represents the information of slides or input queries in the form of an XML document that is written in text format. Next, the system groups figures in each slide. In the case of slides, figures in one slide are grouped into some groups (or one group in some cases). In the case of queries, all figures in a query are considered as one group. Lastly, the system extracts data for each group. The data to be extracted are the shape style as the figures, the coordinates of the figures, occupied area of figures, and so on.

1) OpenXML

The pptx extension is the PowerPoint document format adopted from Microsoft Office 2007. The document converted in pptx extension format can be converted into "OpenXML". "OpenXML" is written as an XML document. OpenXML also contains binary data, such as a picture. The XML document is the text data arranged with the tag. These tags append diverse metadata to each text. Therefore, it can acquire the information about each element. An example description in OpenXML format is shown in Fig. 5. We extract figure information by way of shape, arrangement coordinates, and size from XML document.

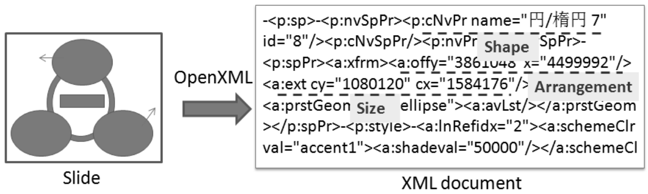

Fig. 5. Example description in OpenXML format.

2) Shape Generalization

There are many kinds of shape types of figures, but they have similar shapes and meanings. However, users often do not remember which shape type of the figure was used in the past.

In the proposed method, figures with similar shapes and meanings are recognized to be the same shape type. The similar shape name is generalized as in the following examples.

Example

[Straight Line Arrow, Block arrow => Arrow]

[Square, Round rect => Rectangle]

Three shapes (arrow, quadrangle, and triangle) are generalized in this research, because these are frequently used in figures. The shapes that belong to these three generalized figures are shown in Figs. 11 and 12. Text box is contained in the generalization figure group of a quadrangle, because it tends to be used as a quadrangle.

Fig. 11. Figure group accepted as "arrow".

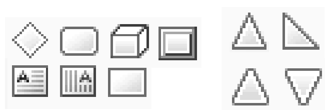

Fig. 12. Figure group accepted as “quadrangle" and "triangle".

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

3) Grouping of Figures

The figure group expression included in the slide is shown in Fig. 6. In general, the aggregate of multiple figures constitutes one figure group. Therefore, the system needs to perceive which figures should be consisted one figure group for visual expression. A figure group is judged from the distance between figures. If the figures are nearer to each other than the threshold, they will be considered as the same group. The processing flow of grouping is shown in the grouping rule. The example of the grouping rule is shown in Figs. 7, 8, and 9. In addition, the coordinates of figures are acquired from the OpenXML document of each slide.

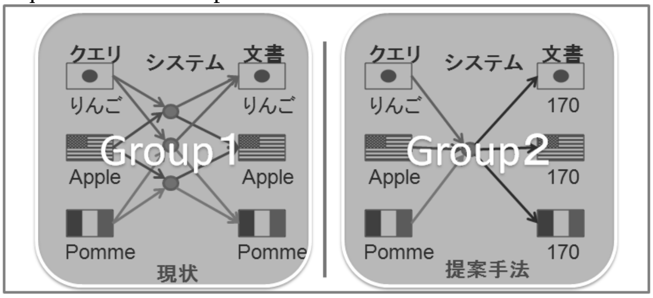

Fig. 6. The example of figure group.

[The grouping rule]

1. A group is formed when other figures are within the limits of the figure used as a reference point.
2. The next reference point is the furthest figure out of the range of a reference point. The figure is checked to see whether it should be in the group.
3. Grouping is completed when there are only figures confirmed to be within the limits. When unchecked figures remain, they are considered as another group and start from 1.

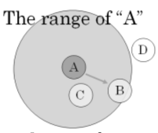

Fig. 7. Flow 1 of grouping.

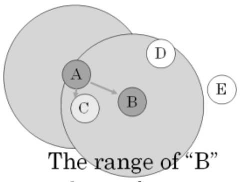

Fig. 8. Flow 2 of grouping.

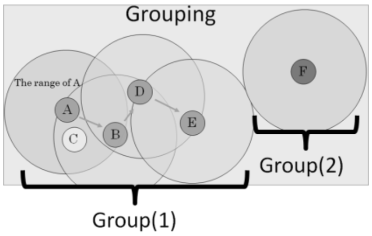

Fig. 9. Flow 3 of grouping.

The judgment range of a group uses two lengths. The formula to calculate the judgment range of group is shown in Fig. 10. The length of the straight-line arrow expresses the length of half the diagonal line of the figure. The dotted line arrow shows the length of the Compensation. Compensation is the distance of the diagonal line for a square with 50*50 pixels. The threshold is a value of adding these two lengths. Therefore, the judgment range indicates the range to threshold radius from the center of a figure.

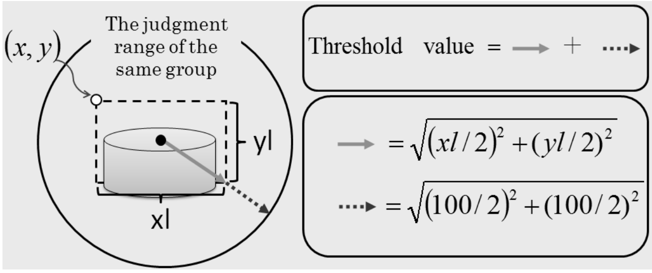

Fig. 10. Calculation of judgment range.

B. Template Classification

When retrieving the figure group from past presentation slides, the user can limit the retrieval object by the template style of his/her intended figure group (the figure used for the system outline, list, etc.). Our system can eliminate different types of similar figure groups with the query from the search results. The template is used to classify figure groups. The hierarchy of templates used in the system is shown in Fig. 3.

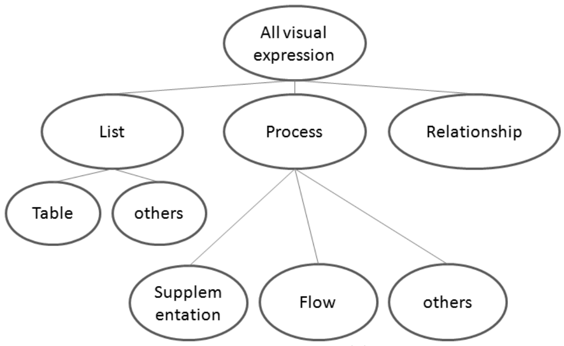

Fig. 3. Template types of figure group.

We use Support Vector Machine (SVM) [7] for template classification. First, the feature vector for classification is created. Each figure contained in a figure group is classified into one of three sizes. The feature vector consists of the figure shapes as dimensions and the frequency of each shape as its elements. This module classifies figure groups into one of three template types: list, process, and relationship. In the experiments, we classified 1,127 figure groups into these three templates. The accuracy of classification using SVM is shown in Table 1. The types of figure shape used as the elements of a feature vector in the experiments are shown in Fig. 4. In the results of the exploratory experiment, each classification is appropriately performed with precision.

TABLE I.
ACCURACY OF TEMPLATE CLASSIFICATION
BY SVM (20-FOLD CROSS-VALIDATION)

	List	Process	Relationship
accuracy	88.7342%	86.1456%	88.9876%

1:Rectangle1	2:Text1	3:Arrow3	4:Arrow1
5:Text2	6:Column1	7:InteractiveArrow1	8:Arrow2
9:Oval1	10:Oval3	11:Chord1	12:Oval2
13:Text3	14:Table1	15:Parenthesis1	16:Line3
17:Line1	18:Arc3	19:Line2	20:Figure1
21:Object1	22:Figure2	23:Rectangle2	24:Rectangle3
25:Balloon1	26:Column2	27:Document3	28:File3
29:Object2	30:InteractiveArrow2	31:Document2	32:Balloon3
33:Balloon2	34:Parenthesis3	35:Object3	36:Document1
37:Arch3	38:Organization1	39:Parallelogram1	40:Rhomboid1
41:Triangle3	42:tower1	43:computr31	44:Arch1
45:Display3	46:Display2	47:InteractiveArrow3	48:Figure3
49:Parenthesis2	50:Column3	51:Display1	52:Triangle1
53:Homebase1	54:Homebase2	55:Triangle2	

Fig. 4. Components of feature vector.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

C. Data Reduction

The coordinates and size of each figure are acquired from the OpenXML document. Furthermore, a data set is created. The contents of the data set are the five features: central coordinates of figure group detected by "grouping of the figure", shape type of figure, size of figure, the conservative field place of a PowerPoint document, and slide page number. Coordinates and area make the upper left of the slide the starting point.

The purpose of our research is retrieval considering the figure group in a slide. Therefore, the system needs to consider the placement relations of figures in the figure group. Thus, each figure is corrected to the coordinate value on the basis of the central point of figure group. Moreover, there are figure groups of various sizes in slides. The size of a figure group is standardized into 720x540 pixels.

Five data features explained above and the copy of a PowerPoint document are stored in a database.

V. Retrieval Process

In the retrieval process, first the user makes a query figure. The user inputs a query figure and the template type of his/her demanded slides. In template selection, the user chooses the template type of the query figure. Second, the system extracts figure features from the query by using the figure information extraction module, except for in figure grouping. Third, the system calculates the similarity between the query and each figure group in the slides using their indexes in the database. We call this the score calculation module in this process. Fourth, the system ranks the slides using their similarities to the query and the figure groups in the slide. Lastly, the system returns higher ranked slides as the retrieval results.

A. Score Calculation Module

The score calculation module calculates the high resemblance degree of query figure and each slide in the database. First, the system searches the slides including the same figure in the query. Second, when the size of each figure is different from those of the database and query, the system removes the slide from the search object by using the Figure Size Comparison, which will be explained in Section A.2. Third, the Coordinate Similarity Measurement calculates the similarity between a query and each slide in the database. Fourth, the system calculates the score for each slide.

1) Figure Shape Comparison

When similarity retrieval of query figure is performed, the figures of the same shape are compared. Therefore, the system compares each form of the figure group in a query figure and the figure group in a database. This process extracts figures with corresponding shapes from the database and query figure. The extracted data are used to calculate similarity by the next processing. An example of shape comparison is shown in Fig. 13.

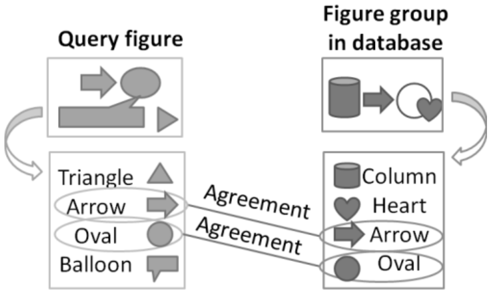

Fig. 13. Example of shape comparison.

2) Figure Size Comparison

In this process, the system compares the sizes between each figure in the query and those in extracted slides. When the sizes of all figures are different, the impression of the figure group will change. The system removes figures that are too large or too small for the target figure in the query. In this paper, we set the minimum bounce of the size threshold at 0.5 times the size of the target figure in the query and its maximum at 1.5 times. The judgment range of size is shown in Fig. 14.

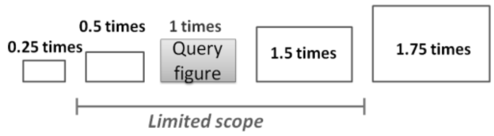

Fig. 14. Size threshold.

3) Coordinate Similarity Measurement

This process calculates the position similarity between figures in the query and in each figure group. It is useful when figures are positioned for the purpose of visual impression.

First, the system extracts figures that are limited by shape and size threshold from the database. Next, each coordinate vector is created from the center of the figure group. Cosine similarity is calculated using these coordinate vectors. This cosine similarity is denoted by "CosSim." (formula (1)).

The coordinates of each figure in a query figure are (x1, y1). Moreover, the coordinates of each figure of the figure group stored in the database are (x2, y2).

$$\mathrm{CosSim} = \frac{x_1 x_2 + y_1 y_2}{\sqrt{{x_1}^2 + {y_1}^2} \times \sqrt{{x_2}^2 + {y_2}^2}} \quad --(1)$$

4) Score Calculation

We calculate the final score between the query and each figure group in slides. This score is calculated by the total amount of CosSim that is calculated in the similarity measure of coordinates. We select a target figure with a higher CosSim for the final score if there are some candidates for comparison figures for the one in the queries. Furthermore, in the score calculation, we ignore figures that have CosSim lower than 0.9.

An example of score calculation is shown in Fig. 15.
In Fig. 15, the system detects two arrows in the figure query (hereinafter, the query arrows.).
The target slide has two arrows that are the same shape type as detected ones in the query (hereinafter, the database arrows.).

Fig. 15 shows database arrows for the query arrows.

The similarities of each arrow between query and database are calculated. The arrow pair with the highest similarity is used for the score. In addition, the similarities of the oval figure are also calculated. Since all the values are over the threshold value (0.9), the total of three values turns into a score.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Similarly, we calculate final score for each figure group in the slides. Then, we rank slides using final score and show the output results to a user.

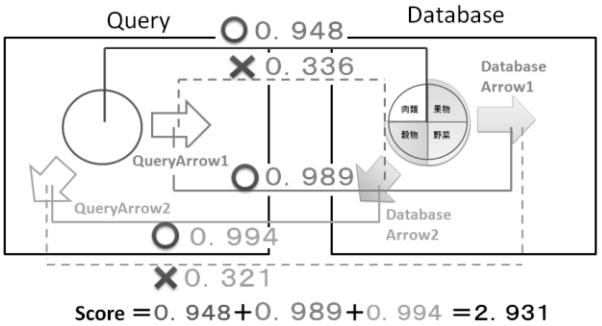

Fig. 15. Example of score calculation.

VI. Experiment

A. Experiment and Result

In this experiment, 1,127 slides were stored in a database. We manually made 15 figure groups in order to retrieve the answer slides. These 15 figure groups were used for the query. We retrieved proper slides for each query. We checked whether retrieved slides matched the query. We named these matching slides “answer slides”. Answer slides show slide(s) relevant to the query figure. In this experiment, we made answer slides manually. For the procedure, we show the query figure and the top 10 retrieved slides to nine evaluators as the experimental results for each query figure. We also call the slides with the support of the majority "answer slides". We evaluated the accuracy of the retrieval results by Mean Average Precision (MAP). The results of MAP are shown in Table 2.

TABLE II
RESULTS OF MAP

	MAP		MAP
query1	1.000	query9	1.000
query2	1.000	query10	1.000
query3	1.000	query11	0.500
query4	0.555	query12	1.000
query5	0.859	query13	1.000
query6	1.000	query14	1.000
query7	1.000	query15	1.000
query8	1.000	Average	0.960

We show five results for the example in order to do detailed analysis. Also, we show the top three retrieved slides as the experimental results for each query. The answer slides for each query are ranked in Table 3. Table 4 shows the number of retrieved slides for each query in the case of considering and not considering a template. Moreover, the pictures of five query figures and top three ranks of retrieval results are shown in Figs. 16-20.

TABLE III
RANK OF ANSWER SLIDES

	Rank
Query figure 1	1,2
Query figure 2	1
Query figure 3	1
Query figure 4	1,3
Query figure 5	1,2,3,4,5,6,8

TABLE IV
NUMBER OF OCCURRENCES

	Number of occurrence	
	No template	Template
Query figure 1	450	125
Query figure 2	610	308
Query figure 3	110	97
Query figure 4	145	111
Query figure 5	515	283

B. Discussion

In all the query figures, the answer slides were given a higher ranking. However, the proposed method tends to give a higher score to figure groups containing many figures. Therefore, we need to improve the score calculation so it is not influenced by the number of figures.

The proposed method retrieves figure groups on the basis of cosine similarity. Thus, we can consider the placement relations of each figure. However, we cannot consider the distance between the figures. Therefore, the proposed method must be improved to consider the distance between the figures.

Based upon the above, the following two methods are needed.

1. A system method that considers the ratio of object matching between group figures and a query figure by all objects included in group figures.
2. Similarity calculation that considers distance between the figures.

VII. Conclusion

This research aims at realizing the retrieval of figure groups included in presentation slides. We proposed the similarity retrieval of figure groups in slides as one such way to do this. In the results of the experiments, all the correct slides were retrieved with high ranking in the ranking lists. Therefore, we can conclude that our proposed technique is effective for the retrieval of figure groups.

One of our future works is to consider a better user interface. In our proposed method, the user has to make a query figure by him/herself, which imposes a considerable burden on the user. Alternatively, a hand-drawn image search technique using sketch may be usable. However, the current sketch retrieval methods have low precision of about 60%. Therefore, we must think about ways to reduce the burden on users.

We proposed a slide retrieval technique that focuses on the figures in presentation slides. However, the text information is nevertheless an extremely important element in searching presentation slides. It is possible to combine our technique with the conventional retrieval techniques that use text information. In this way, we can accommodate users' search

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

demands more widely, and the retrieval effectiveness of our system can further be improved.

REFERENCES

[1] Tessai Hayama, Susumu Kunifuji "Rtelevant Piece of Information Extraction from Presentation Slide Page for Slide Information Retrieval System" Knowledge, Information, and Creativity Support Systems Lecture Notes in Computer ScienceVolume 6746, 2011, pp 22-31, DOI: 10.1007/978-3-642-24788-0_3

[2] Tessai Hayama, Hidetsugu Nanba, Susumu Kunifuji "Structure Extraction from Presentation Slide Information" Proc. of 10th Pacific Rim International Conference on Artificial Intelligence (PRICAI2008), pp.678-687, 2008.

[3] Guo Min Liew and Min-Yen Kan, Slide image retrieval: a preliminary study, Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL2008), pp.359-362, 2008.

[4] Kimoto Haruo, "An Image Retrieval System Using Impressional Words and the Evaluation of the System" Transactions of Information Processing Society of Japan, Vol.40, No.3, pp.886-898, 1999 (in Japanese)

[5] Daisuke Kitayama, Akiko Otani and Kazutoshi Sumiya, "A Scene Extracting Method based on Structural and Semantic Analysis of Presentation Content Archives" Proc. of The Seventh International Conference on Creating, Connecting and Collaborating through Computing (C5 2009), pp. 128-135, 2009

[6] Yuanyuan Wang and Kazutoshi Sumiya, "A Browsing Method for Presentation Slides Based on Semantic Relations and Document Structure for e-Learning" Journal of Information Processing, Vol.20, No.1, pp.11-25, 2011

[7] Chih-Chung Chang and, Chih-Jen Lin "LIBSVM -- A Library for Support Vector Machines", http://www.csie.ntu.edu.tw/~cjlin/libsvm/ (last accessed 2012/12/18)

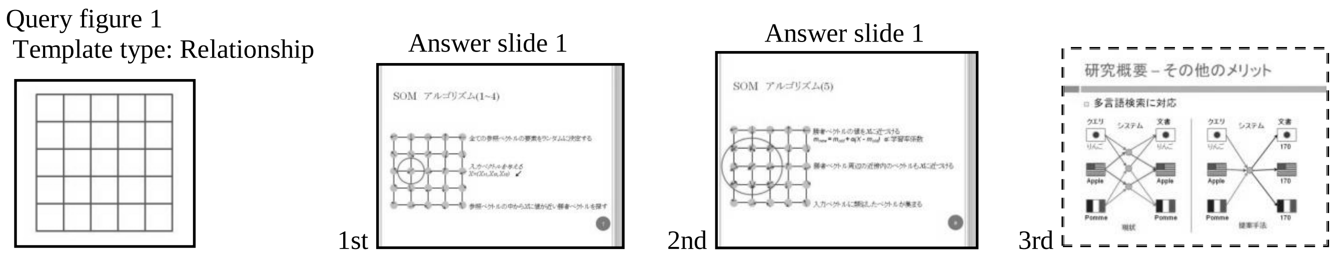

Fig. 16. Query figure 1 and 1st – 3rd slides.

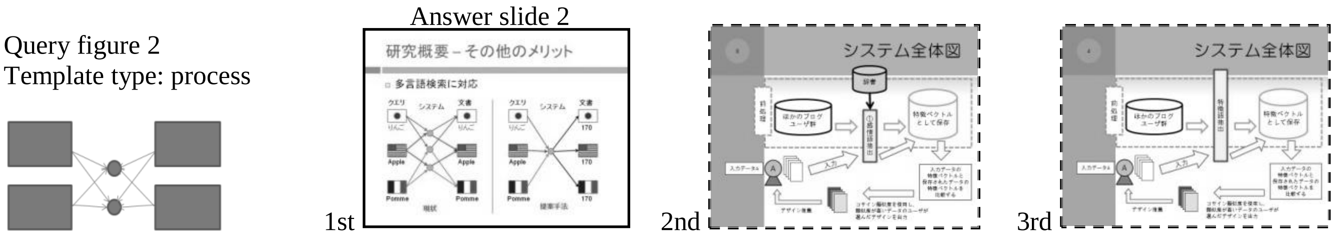

Fig. 17. Query figure 2 and 1st – 3rd slides.

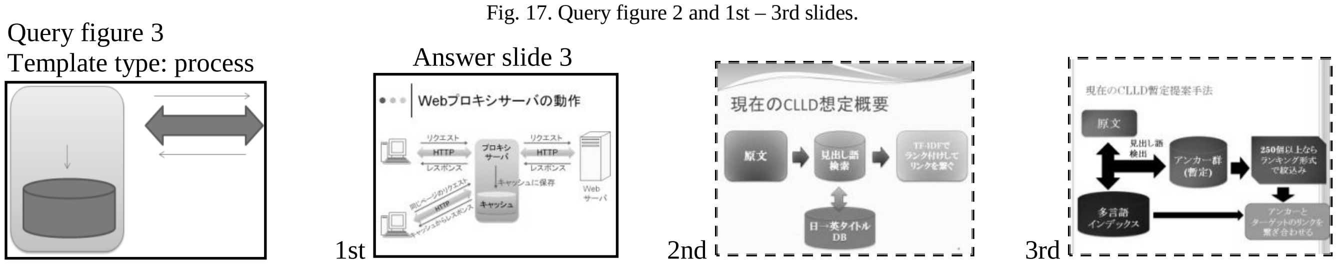

Fig. 18. Query figure 3 and 1st – 3rd slides.

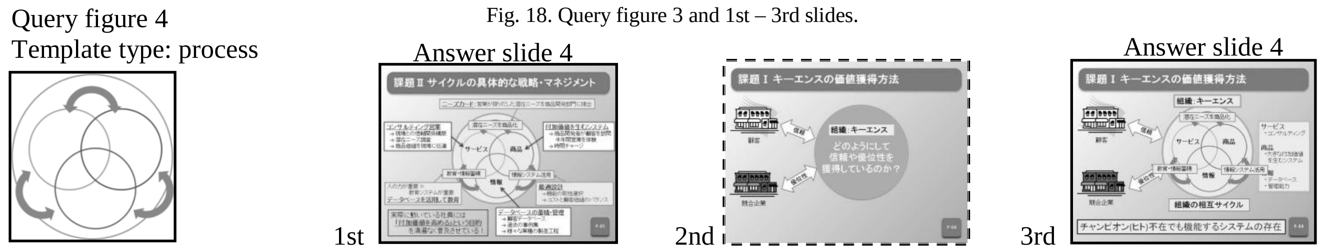

Fig. 19. Query figure 4 and 1st – 3rd slides.

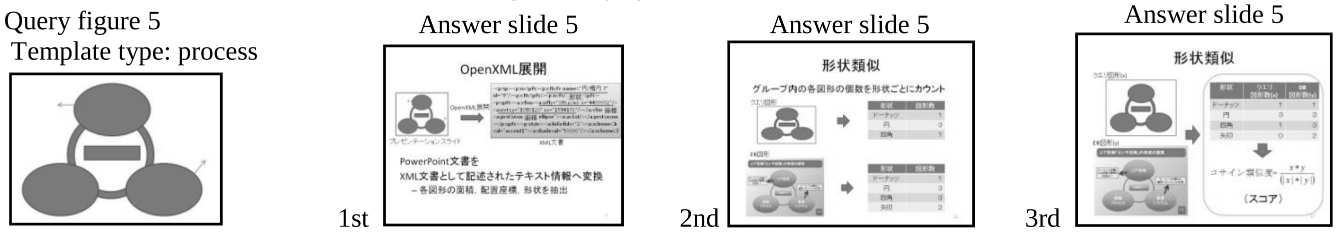

Fig. 20. Query figure 5 and 1st – 3rd slides.