Research and Practice in Technology Enhanced Learning
Vol. 8, No. 1 (2013) 153－169

A MINING TECHNIQUE FOR EXTRACTION OF PRESENTATION SCHEMA FROM PRESENTATION DOCUMENTS ACCUMULATED IN LABORATORY

Shinobu HASEGAWA

Center for Graduate Education Initiative, JAIST
1-1, Asahidai, Nomi, Ishikawa, 923-1292, Japan
hasegawa@jaist.ac.jp
http://dlc.jaist.ac.jp/shinobu/

Akihiro KASHIHARA

Graduate School of Informatics and Engineering, The University of Electro-Communications
1-5-1, Chofugaoka, Chofu, Tokyo, 182-8585, Japan
akihiro.kashihara@inf.uec.ac.jp

The main topic addressed in this paper is how to automatically extract the presentation schema, which represents typical heuristics to be shared by laboratory members for composing presentation structure of presentation documents, from a number of the presentation documents the laboratory members have composed. The key idea is to propose a model of the presentation structure, which represents roles of and sequences among presentation slides included in the presentation documents with metadata. Following this model, the presentation schema is defined as a common presentation structure for the members who have the same methodology and/or domain of research in a research group in the laboratory. This paper accordingly introduces a machine learning technique based on association rule mining for automatically extracting the presentation schema from the repository of the documents accumulated in the research group. In addition, we report experiments with the comparison of the presentation schemas extracted from three different research groups to investigate the validity of the proposed extracting technique.

Keywords: Presentation schema; presentation structure; association rule mining; laboratory education.

1. Introduction

In our daily research activities, presentation is one of the most important activities for researchers and students to report research findings with well-organized presentation documents consisting of presentation slides. There are many tasks for a presenter to make a good presentation such as preparing mentally and physically, brainstorming and selecting the topics, developing the outline, structuring the topics, finding examples and images, preparing handouts, making rehearsals, and so on (McCarthy & Hatcher, 2002). We divide presentation skills in conducting these tasks into two main skills which are oral communication skill and presentation composition skill (Saito, Tanida, Kashihara, & Hasegawa, 2010). A skill in making oral communication is to orally deliver the presentation to audiences clearly. Such skill can be viewed as one of the most important transferable skills (Greenan, Humphreys, & McIlveen, 1997) and be trained as presentation courses in many institutions. The latter skill, on the other hand, involves not only designing the presentation slides but also constructing presentation structure so as to specify what to present and what order to present in the presentation documents (Hasegawa & Kashihara, 2011). Such presentation structure is particularly important to represent research contents logically with suitable segmentations and organization. However, it is quite difficult for novice graduate/undergraduate students newly assigned to a laboratory to be aware of the presentation structure since they have few experiences and knowledge about composing the presentation structure. They usually need to learn the presentation composition skills from the presentation documents described by their laboratory members or from discussions with the expert researchers in the research meetings, which can be viewed as a part of laboratory education (Kirschner, 1988). In general, a laboratory includes a few research groups, each of which would have its own methodology and/or domain of research. The presentation documents accumulated in the research group tend to have a common presentation structure, which represents its own style of composing the presentation documents. This could be a basis for the novices to specify what to present and what order to present in the laboratory style for composing their presentation documents (Hasegawa & Kashihara, 2011). In this paper, such common presentation structure is called presentation schema.

The final goal of our research is accordingly to help the novices develop their skill in composing the presentation documents with the presentation schema. However, it is not enough for them to learn the presentation schema by referring to good presentation documents since it is often embedded in the presentation documents. Textbooks for presentation are not also suitable for learning it since they do not always cover the laboratory own style for composing the presentation documents. On the other hand, it is not so easy for the expert researchers to prepare such schema for the novices as tangible standards in advance because it is strongly dependent on the diversity of the presentation contexts such as presentation time limitation, audiences, research domain, research methodology, and so on (Hasegawa & Kashihara, 2012). Our approach to the final goal is consequently to reuse the presentation documents accumulated in the research group whose members have the same methodology and/or domain of research in the laboratory to extract the presentation schema. The novices could use the schema as a scaffold for composing the presentation documents and for developing their skill.

In order to achieve the goal, it is necessary to define the presentation schema. The main issue addressed in this paper is how to extract the presentation schema automatically from a number of the presentation documents the research group members have composed. We first discuss the presentation composition skill and introduce a framework for representing the presentation structure and the presentation schema with metadata representing the roles each presentation slide plays and the sequence of the slides. Following this framework, we then utilize a machine learning technique to analyze and extract a typical presentation structure as the presentation schema from the repository of the documents attached with the metadata in advance. The extracted schema could become a scaffold for the novices to develop the presentation composition skill because they can become aware of the typical presentation structure to compose their documents. In addition, the presentation schema of the research group's own is dependent on the presentation context. It is accordingly hard for the expert researchers to consider each context to define the corresponding schema. The machine learning technique aims to automatically extract the presentation schema suitable for the presentation context, which could reduce their burdens of defining the schema. Finally, we report experiments with the comparison of the presentation schemas extracted from three research groups with different presentation contexts, which intended to investigate the validity of the proposed extracting technique and the practical possibilities whether the technique could extract the different schemas reflecting on their own presentation styles.

2. Presentation Composition Skill and Presentation Schema

2.1. *Presentation composition skill*

In this paper, a presentation composition skill is the capability to determine how to divide the research contents into the slides and then how to sequence the slides in an understandable way as shown in Figure 1. In fact, it is difficult for the novices to compose the presentation structure appropriately since they have little knowledge about the structure to be followed. In order to learn how to make the division and determine the sequence, in addition, it is desirable to refer to good presentation documents or textbooks such as Atkinson (2011) and Holsinger (2003).

However, these resources do not always imply the presentation style in the laboratory.

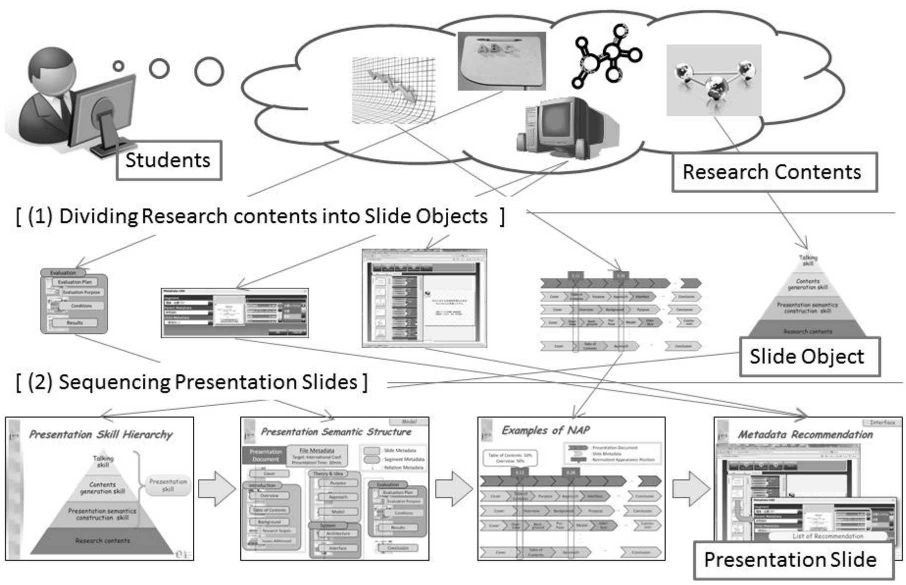

Figure 1. Overview of presentation composition skill.

The resources do not also demonstrate how to compose the presentation structure in accordance with the presentation contexts. In order to make a report for an internal meeting, for instance, the presenter would often omit explaining background knowledge of the research since all audiences belonging to the laboratory have preliminary knowledge of his/her research. In order to make a presentation for an external conference, on the other hand, he/she should compose the presentation structure including the background as introduction to the audiences who do not have preliminary knowledge of the research.

2.2. *Presentation schema framework*

In this paper, the presentation schema is represented as a typical presentation structure, which implies some heuristics for composing the presentation documents in the research group own style. Here, the research group consists of the researchers and the students who have the same research methodology such as system development and system application. A certain laboratory in the institute may include a few research groups. The presentation schema could provide the novices with how to divide the research contents into a number of the slides and how to construct the presentation structure that expresses roles of and sequences among the slides. However, it would be difficult to consider such presentation schema since it is often embedded in the presentation documents accumulated in the research group. In order to deal with the presentation structure and the presentation schema explicitly, we have provided a framework for representing them (Hasegawa & Kashihara, 2011) with three types of metadata as shown in Figure 2. This

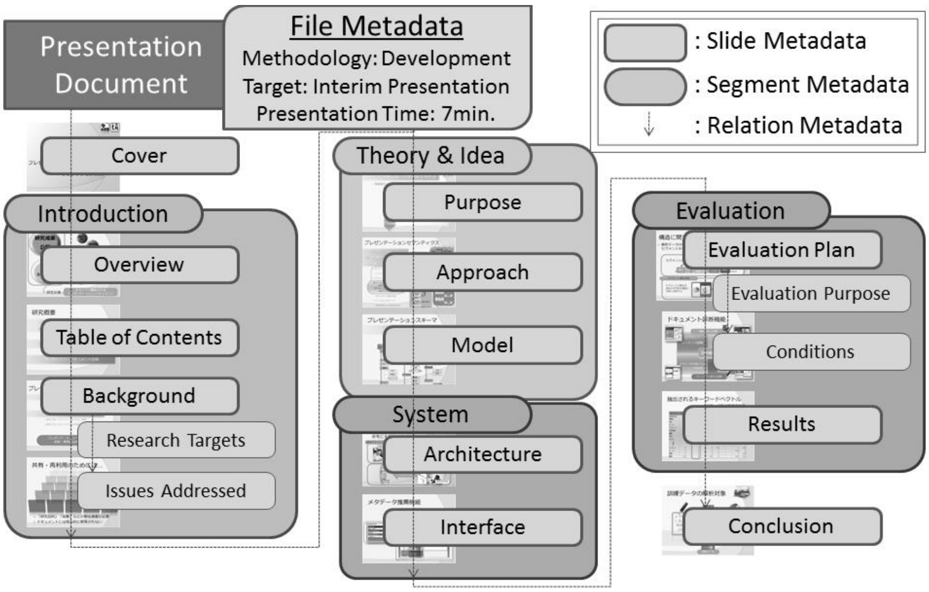

Figure 2. Examples of presentation schema.

framework allows the novices to be aware of the presentation structure in composing the presentation document.

Slide metadata represent the role that each slide included in the presentation document plays in presenting the research contents. Such metadata do not necessarily correspond to the slide title. Examples of the slide metadata are "*Cover*", "*Overview*", "*Background*", "*Issues addressed*", "*Purpose*", "*Architecture*", "*Evaluation plan*", "*Conclusion*", etc. Some of them are nested such as "*Research Target*" and "*Issue Addressed*" since these slide metadata often appear as compound metadata in one slide. Some of them also vary according to the presentation context. It means each research group has its own slide metadata set.

Segment metadata also represent the section of the presentation document that several slide metadata compose for presenting the research contents. We have defined four kinds of main segment metadata. Each kind of segment metadata is associated with several slide metadata in advance. For example, the segment "*Introduction*" includes the sequence of "*Overview*", "*Table of Contents*", and "*Background*". The segment "*Theory & Idea*" also includes the sequence of "*Purpose*", "*Approach*", and "*Model*". This type of metadata shows a rough sequence of the presentation structure.

File metadata represent some attributes of the presentation context, which includes the research methodology, presenter information and presentation contextual information such as "*System Development*", "*Target*", "*Presentation Time*", and so on.

2.3. *Related work*

The followings describe related work on presentation support from semantic or structural information and technical points of view. Kohlhase (2007) developed CPoint as a semantic PowerPoint extension that allows the authors to enrich PowerPoint documents by means of semantic annotation. The key feature was to deal with domain knowledge included in the presentation document as semantic information and to visualize it by means of concept mapping. Ihsan, Rehman, Ahmed, and Qadir (2008) and Verbert, Jovanovic, Grasevic, Duval, and Meire (2005) proposed e-learning content development tools used for PowerPoint/OpenOffice.org. These tools managed not only the domain knowledge but also the information for education/learning using the metadata standards such as IEEE Metadata Standards or SCORM.

Hayama, Nanba, and Kunifuji (2005) proposed an automatic approach for generating presentation slides from a technical paper. Following a machine learning technique, they could obtain a set of generating rules from the relationships between technical papers and presentation slides collected from the Web. Seta and Ikeda (2006) developed a support environment with which the novices can produce persuasive presentation documents and develop their presentation skills. This support environment was designed to encourage the presenter to perform meta-cognitive activities in presentation document design and import expertise of other experienced learners through presentation rehearsal. Li and Chang (2009) developed the management model and tools that enable the users to better exploit and transfer presentational knowledge assets for representing the domain knowledge.

There is also diverse related work on schema extraction from semi-structured documents automatically. Ji, Zeng, Zhang, and Wu (2010) proposed a Web information extraction method from tag tree template. It could extract not only text contents in pages but also data schema of the texts by generating the template automatically. Yang, Dai, and Chen (2010) also proposed an automatic ontology extraction method from XML documents so that they can get common domain knowledge as XML schema. These approaches mainly dealt with the schema not as the general representation of the semantic structure but as the local rules for information extraction.

In spite of the significance of the presentation structure and the presentation schema, each of these researches has not dealt well with such information embedded in the presentation documents. We accordingly address how to deal with the presentation structure and the presentation schema in composing the presentation document. We also proposed a metadata estimation framework for the presentation documents (Hasegawa & Kashihara, 2011) as the previous work. The function of the metadata estimation is to estimate the slide metadata corresponding to the target slides without any metadata attached. The presentation slide often includes typical keywords for identifying the slide metadata. The function accordingly calculates a keyword vector representing relationship between each slide metadata and keywords described in the slide. However, it needs a certain number of training data, which means the presentation documents annotated with the metadata, to execute the machine learning technique in advance.

In order to help the novices compose the presentation structure appropriately, on the other hand, we have proposed a scaffolding system as shown in Figure 3 (Shibata, Kashihara, & Hasegawa, 2012). Most novices do not have a full understanding of the presentation structure and enough chances of creating the presentation document with attention to the presentation schema to be followed. Therefore, the proposed system provides them with views for the presentation schema in composing the presentation structure to create the presentation documents as add-in of Microsoft PowerPoint.

In providing the presentation schema, we have two options as follows. The first one is to define the schema by the expert researcher, which provides the novices with ideal presentation structure based on his/her experience. Such schema tends to represent the presentation structure which covers diverse presentation contexts. It is not easy for him/her to prepare the presentation schema suitable for a specific presentation context.

The second option is to automatically extract the presentation schema from the repository of the presentation documents that were composed for the same presentation context. The presentation schema extracted is represented as the typical presentation structure suitable for the context. Although the automatic extraction needs to in advance attach metadata to the presentation documents accumulated in the research group, it enables the novices to look at the presentation schema suitable for the context they want to present.

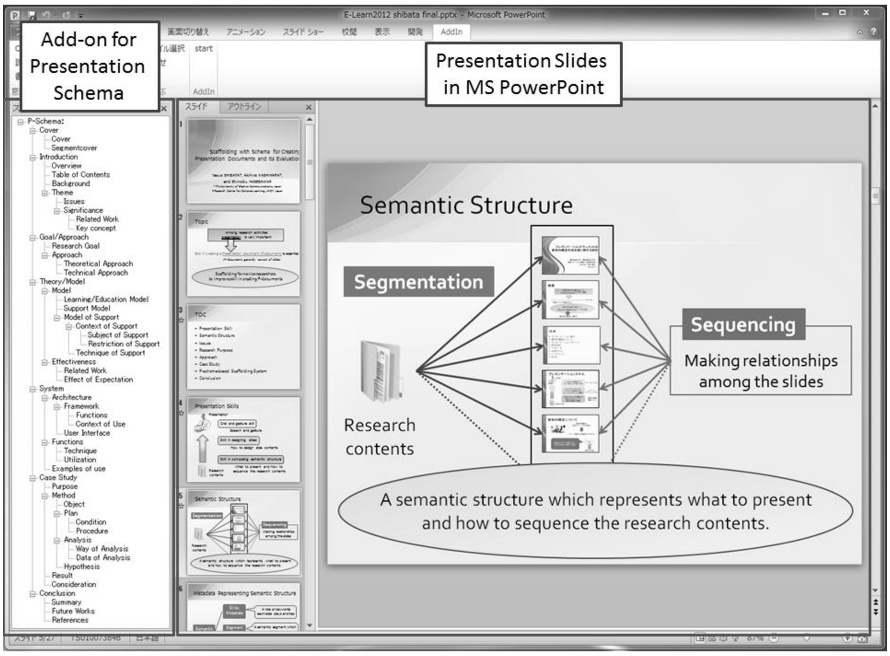

Figure 3. User interface for iomposing presentation structure of presentation document.

In this paper, therefore, we propose how to extract the presentation schema automatically from a certain number of the presentation documents annotated with the metadata by hand in advance that have been created for the same presentation context.

3. Extraction Technique for Presentation Schema

3.1. *Requirement*

In order to extract the presentation schema, we have made a requirement to consider the extraction technique. Among the presentation documents, some local orders of the slides often appear repeatedly. This means the presentation schema could be represented by means of the combination of the local orders. With the requirement, we decide to use association rule mining for extracting frequently appeared metadata and their sequence in this research.

3.2. *Basic concepts of association rule mining*

Association rule mining is one of well-used techniques of data mining in various areas, which was first proposed by Agrawal, Imielinski, and Swami (1993). It aims to extract frequent patterns and casual structures among sets of items in the transaction databases. For example, Pannurat, Kerdprasop, and Kerdprasop (2010) proposed a method to discover conceptual database schema from database instances and relations, and

Malarvizhi and Mary (2012) also extracted frequent access patterns of educational institution's web log data by applying the association rule mining. In this section, we describe general definitions of the association rule mining to prepare our schema extraction algorithm.

Let $I = \{i_1, i_2, \cdots, i_n\}$ be a set of n distinct items, $T = \{t_1, t_2, \cdots, t_m\}$ be a set of m different transaction records, where each $t_m \subseteq I$. An association rule is indicated by the form of $X \Rightarrow Y$, where $X, Y \subset I$ are sets of items called item sets, and $X \cap Y = \phi$. X is called antecedent while Y is called consequent, the rule means X implies Y.

There are two important basic measures for association rules, a support described in $sup(X \Rightarrow Y)$ and a confidence described in $conf(X \Rightarrow Y)$. $sup(X \Rightarrow Y)$ is defined as the proportion of the transactions that contain $X \cup Y$ to all transactions in *T*. $conf(X \Rightarrow Y)$ is also defined as the proportion of the transactions that contain $X \cup Y$ to all transactions that contain *X*.

In regard to such $X \Rightarrow Y$, thresholds of support and confidence are usually predefined by analyzers to extract important rules. These thresholds are called a minimal support described in *min_sup* and a minimal confidence described in *min_conf* respectively. The association rules are finally extracted as the item sets that satisfy both *min_sup* and *min_conf*. However, there are several well-known problems in setting the thresholds (Kotsiantis & Kanellopoulos, 2006). The lower the thresholds are set, the larger the numbers of rules are extracted, which are difficult to recognize the relationships among the rules. The higher the thresholds are set, the smaller the numbers of just known rules are extracted.

3.3. *Schema extraction algorithm*

Our schema extraction algorithm aims to extract the typical presentation structure as the presentation schema based on sequence of the slide metadata of the presentation documents accumulated by the research group members, which are attached in advance with the metadata. Therefore, we adopt the association rule mining as shown in Figure 4, where *I* is a set of all kinds of slide metadata, *T* is a set of all presentation documents that have the same file metadata in the research group's repository, *X* is an arbitrary slide metadata appeared in a certain presentation document as an antecedent, and *Y* is a slide metadata appeared next to *X* in the presentation document as a consequent. Suppose *conf("Overview"* $\Rightarrow$ *"Background") = 50%* and *sup("Overview"* $\Rightarrow$ *"Background") = 33%*, it means that *50%* of *"Background"* slides are next to *"Overview"* slides and *33%* documents include such order relation.

In usual association rule mining, antecedent *X* and consequent *Y* are able to include multiple items but are not able to specify the order relation among these items. Therefore, our algorithm restricts the number of items to one per each *X* and *Y*, which means that *X* and *Y* only contain one slide metadata. This makes it possible to extract partial order relations, and to represent whole sequence of presentation schema by accumulating such partial relations as shown in lower right of Figure 4.

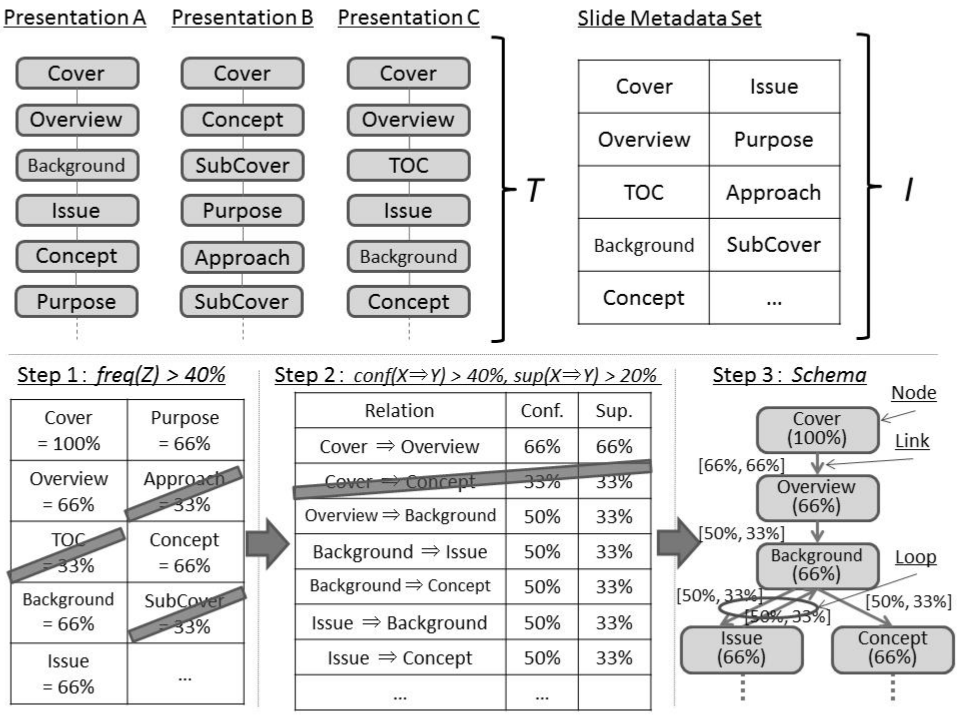

Figure 4. Overview of schema extraction algorithm.

In addition, the infrequently appeared metadata are discarded in order to reduce amount of calculation. *freq(Z),* where $Z \subseteq I$, is defined as the proportion of the number of transactions that contain *Z* to the total number of transactions in *T*, and the threshold of frequency described in *min_freq* is also predefined. Based on the above assumptions, the algorithm contains the following steps:

Step 1. The algorithm extracts a set of frequently-appeared metadata that have *freq(Z)* larger than or equal to *min_freq*. Suppose *min_freq* is *40%* in Figure 4, then metadata *TOC*, *Approach*, and *SubCover* are discarded.

Step 2. It extracts partial order relations $X \Rightarrow Y\ (X, Y \subseteq Z)$ that satisfy both *min_conf* and *min_sup*. Suppose *min_conf = 40%* and *min_sup=20%*, the relation between *Cover* and *Concept* is discard.

Step 3. It composes a presentation schema diagram by combining the extracted metadata and relations. In the diagram as shown in Figure 4, the nodes are the slide metadata left in Step 1 and the links show the order relations left in Step 2. The loops mean dual-ordered relations such as *Background* and *Issue,* which have the links from node *Background* to *Issue* and from node *Issue* to *Background* at the same time.

Table 1. Presentation context and basic statistical data of each condition.

	Group 1	*Group 2*	*Group 3*
Research Methodology	System Development	System Development	System Application
Research Domain	Self-directed Learning, etc.	Web-based Learning	Distance Learning System
Presentation Time	7 min	15-20 min	15-20 min
Number of Presentation	30	10	10
Average of Slides	20.1	21.9	22.2
Standard Deviation of Slides	3.68	4.01	3.95

4. Experiments

This section describes experiments which investigated how to configure the thresholds of frequency, support and confidence in our association rule mining for extracting presentation schema, and compared the presentation schemas between different research groups, laboratories, audiences, and presentation time limitations since the presentation schema would vary according to such factors. The followings are detailed information on sets of the presentation documents used in these experiments.

The presentation documents accumulated in *Laboratory A* were final versions of the ones for graduation research of *30* undergraduate students who belonged to one research group where they focused on the development of support systems for self-directed learning, research activity, and experiential learning. We called this group *Group 1*. The audiences of the presentations were faculties and students of their affiliation of the university, and the presentation time was *7* minutes.

The presentation documents accumulated in *Laboratory B* at the university different from *Laboratory A* were also final versions of the ones for domestic conferences of *20* graduate students or researchers. They belonged to two research groups called *Group 2* and *Group 3*, which focused on the development of web-based learning support systems and the practice of application for distance learning systems respectively. Ten documents were generated in *Group 2*, and the remains were also generated in *Group 3*. The audiences of those presentations were related researchers, and presentation time was *15-20* minutes. Table 1 shows the average and standard deviation for the number of the slides of the documents in the experiments.

Every document did not refer to any presentation schema, and was brushed up through research meetings and the expert researchers' comments in each research group. Although slide metadata should be defined for each research group, in these experiments, we had prepared 34 kinds of the slide metadata and annotated to all the documents in advance as shown in Table 2. The reason was to compare the results extracted from each research group.

Table 2. Lists of metadata used in the experiments.

Metadata	Description	Metadata	Description
Cover	Cover page of presentation	Advantage	Desired effects or advantages
TOC	Table of contents	Architecture	Block diagrams for system
Sub Cover	Section breaking	Interface	Explanation of system interface
Overview	Outline of research	System	Whole picture of system
Background	Research contexts	Features	Explanation of system features
Concept	Keywords explanation	Function	Explanation of each function
Situation	Target and Situation of research	Examples	How to use system
Issue	Issues addressed in research	Demo	System demonstration
Related work	Related work done by others	Practices	Practice of educational activity
Purpose	Purpose of research	Achievement	Achievement of practice
Approach	Theoretical research methods	Conditions	Experimental conditions
Model	Model of issues and/or contexts	Procedure	Experimental Procedure
Technology	Technical solution	Results	Results of experiments
Previous work	Previous work done by own self	Discussions	Discussions of experiment results
Instances	Specific examples of issues	Conclusions	Conclusion of presentation
Target	Coverage of research contexts	Future work	Future challenge
Elements	Theoretical building blocks	References	References of presentation

4.1. *Experiment 1: Analysis for thresholds by each group*

In order to consider the suitable thresholds, we first investigated how the numbers of nodes, links, and loops included in a presentation schema diagram changed by values of *min_freq*, *min_sup* and *min_conf*. Figure 5 compares the numbers of them extracted by the thresholds on the abscissa in the proposed schema extraction algorithm. From the definitions of *min_sup* and *min_conf* described in Section 3.2, the value of *min_sup* should be equal to or lower than the one of *min_conf*. However, it is difficult to justify these values from theoretical point of view. In our previous work (Ota & Kashihara, 2010), which had been conducted in different research domain, *min_sup* had been set as the half value of *min_conf* empirically and the results seemed to be extracted as important rules with the association rule mining. Therefore, we have ascertained that *min_sup* could be set as the half value of *min_conf*. On the other hand, *min_freq* could be set as the same value as *min_conf* so that the algorithm could not extract isolated nodes or complicated links in the extracted rules. Although these relations among the thresholds should be changed on actual conditions, we followed the assumption to set these thresholds in this experiment. In case that the thresholds were sufficiently small, Figure 5 shows the numbers of links are larger than the number of nodes in all groups. In such case, there are many alternative links as the typical presentation schema as shown in the left of Figure 6. This may not be so good for the novices to learn how to compose the presentation document clearly since they get confused with too many links and loops. The larger the thresholds were, the smaller the numbers of nodes, links and loops were. In case that *min_freq* and *min_conf* are *35%* as shown in the right of Figure 6, the sequence of slide

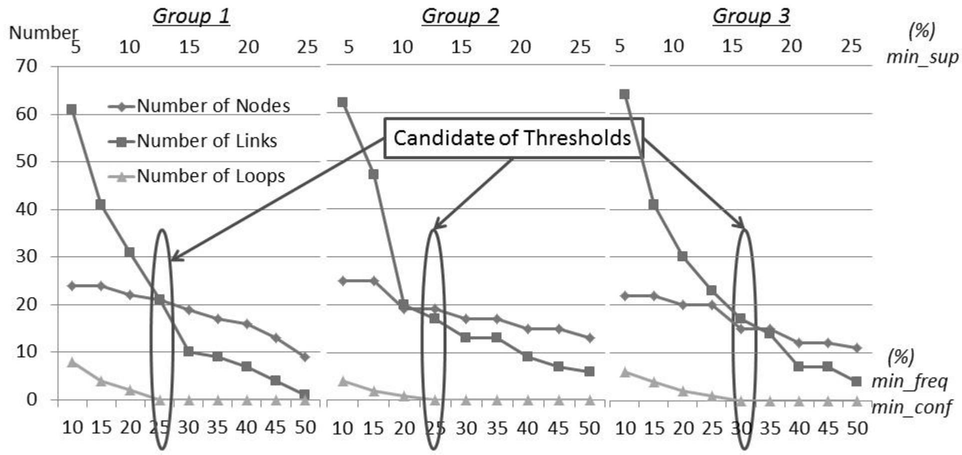

Figure 5. Numbers of extraction by changes in thresholds.

metadata is not clear from the extracted schema. We can see the points around *min_freq* = *min_conf = 25%* and *min_sup=12.5%* at where the numbers of the nodes were higher than the ones of the links and the numbers of loops were zero as shown in the center of Figure 6. These points can be important candidates for setting the thresholds. In other words, suitable presentation schema could be obtained by finding out such points to set the thresholds.

4.2. *Experiment 2: Assessment of validity for schema extracted from each group*

The purpose of this experiment was to assess the validity of the proposed mining technique by comparing the presentation schemas among *Group 1*, *2*, and *3*. Figure 7 illustrates three presentation schemas by setting *min_freq* = *min_conf* = *25%* and *min_sup* = *12.5%* from *Group 1* and *2*, and *min_freq = min_conf =30%* and *min_sup = 15%* from *Group 3* respectively. Values in round brackets are probabilities of appeared metadata *freq*(*Z*), and values in square brackets are probabilities of confidence $conf(X \Rightarrow Y)$.

First of all, each schema seems to show a main path to be followed by the novice students with some branches. The branches would depict the difference among the research group members. Furthermore, some slide metadata did not have any arrows to indicate transition. These shows there are no significant (over thresholds) transitions from the metadata because such metadata repeatedly appeared in some different positions in the presentation documents. Comparing these schemas, the one extracted from *Group 3* was quite different from the ones extracted from *Group 1* and *2*. The reason would be due to a difference in the research methodology among the groups. The members of *Group 1* and *2* mainly engaged in system development and both presentation schemas clearly included *“Issue”*, *“Purpose”*, and *“Technology”* of each research. This implies the expert researchers of *Group 1* and *2* believe that presentation for system development

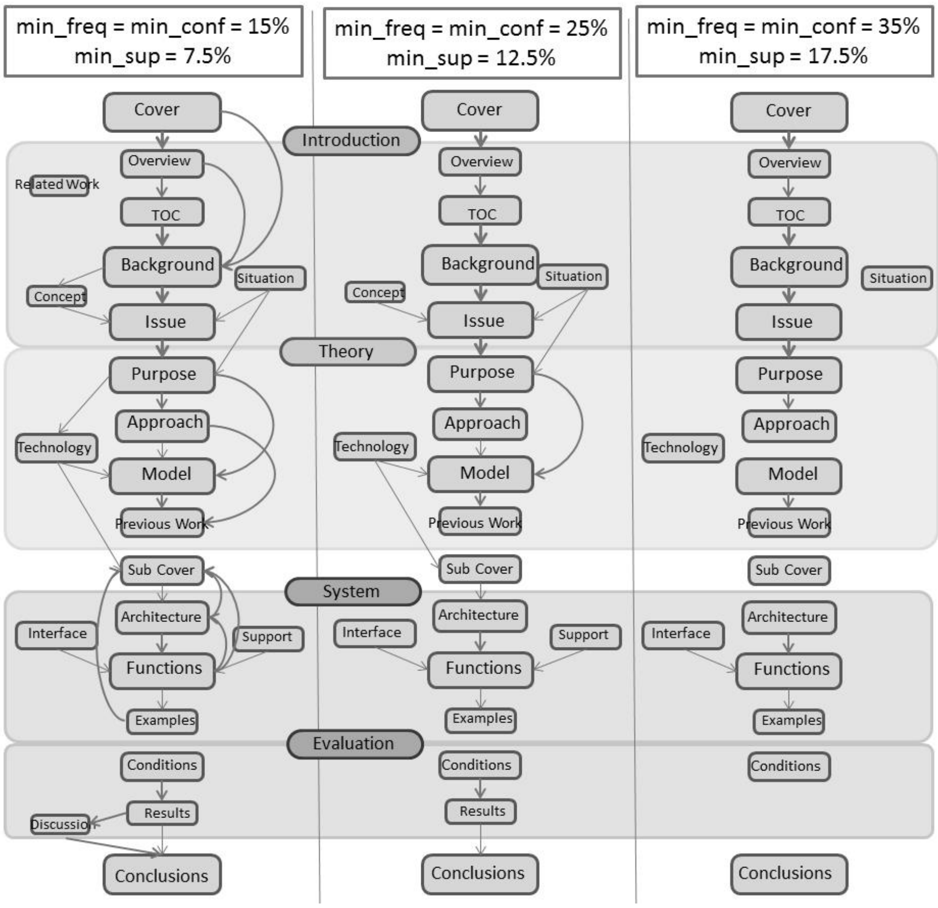

Figure 6. Presentation schema of *Group 1* extracted by changes in thresholds.

should emphasize issue-oriented and technological approach. In addition, *"Conditions"* of its experiments played important roles in both schemas from system evaluation point of view. On the other hand, *"Situation"* and *"Practice"* stood out in the presentation schema extracted from *Group 3* because these presentations focused on system application, which should describe a circumstance of the application clearly and the results derived objectively from the practice. Besides, the *"Evaluation"* segment in the schema from *Group 1* was different from the ones from *Group 2* and *3*. These shows the longer presentations tend to include more detailed slides related to *"Evaluation"* segment as a capacity of the schema representation.

Following the above consideration, we can say that the presentation schemas extracted satisfy specific conditions of the presentation contexts including the research methodology each research group has, audiences, and presentation time limitations. The proposed technique accordingly seems to be valid.

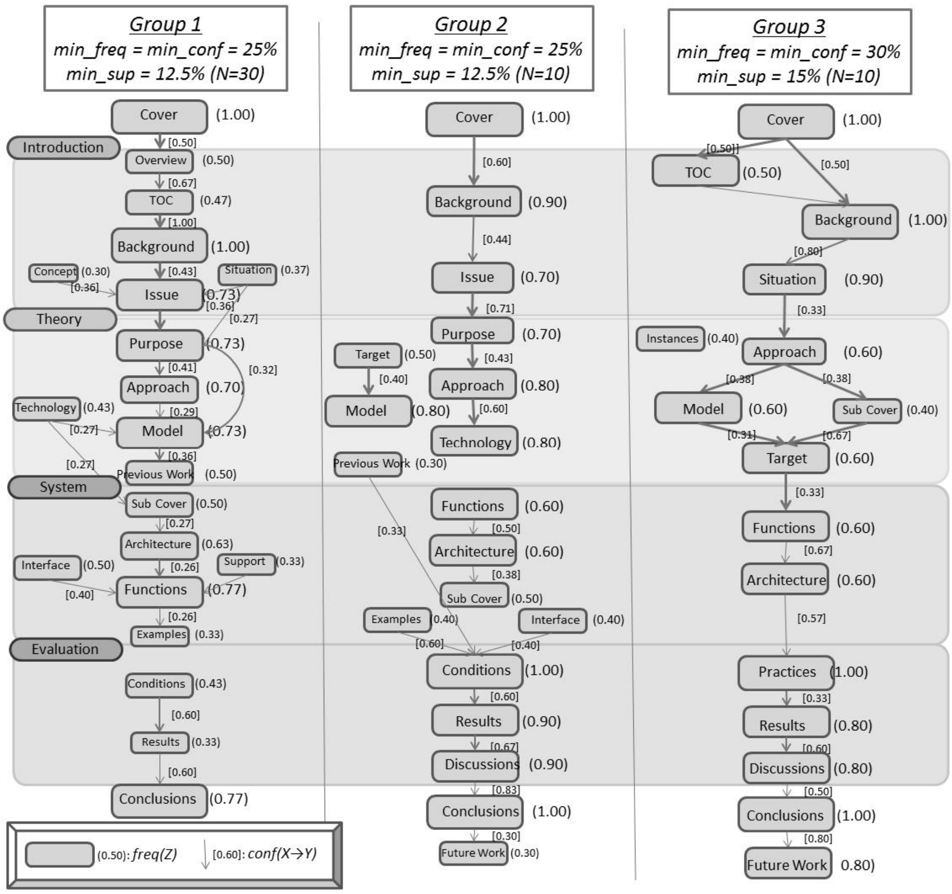

Figure 7. Results of extracted presentation schemas.

5. Applications

As described in Section 2.3, we have developed a scaffolding system, which provides novice students with the presentation schema as a scaffold for composing the presentation structure of presentation documents (Shibata et al., 2012). In considering an application of the proposed extraction method to the system, we can see the possibility to control the level of scaffolding with the schema in accordance with the novices' skill in presentation composition.

The schema to be extracted with the proposed thresholds would be suitable for the novices with low composition skill. On the other hand, the schema to be extracted with the higher thresholds would present fewer mediations and their relations as the core structure, and would allow more skillful students to compose their presentation structure with fewer restrictions. The presentation schema extracted with *min_freq = min_conf = 50%* and *min_sup = 25% or larger* as shown in 8, for instance, could allow the students to follow the core slide metadata to design the corresponding slides, and to select other slide metadata and make a sequence of slide metadata in their own style since the presentation schema would include only nine slide metadata and one link as the core structure. Such adaptive scaffolding would enable the novices to utilize the presentation schema dependent on their skill level.

6. Conclusions

This paper has described the framework for the presentation schema and proposed a fundamental technique to represent it automatically by combination of partial order relations extracted with the association rule mining. The diagram representing the roles of and sequence among the presentation slides would enable the novices to be aware of the presentation structure to create the presentation documents in the research group's style. Accordingly this is one of scaffolding ways for them to learn the presentation composition skill practically. We have also discussed the experiments with the presentation schema. The results indicated a reasonable setting approach for the

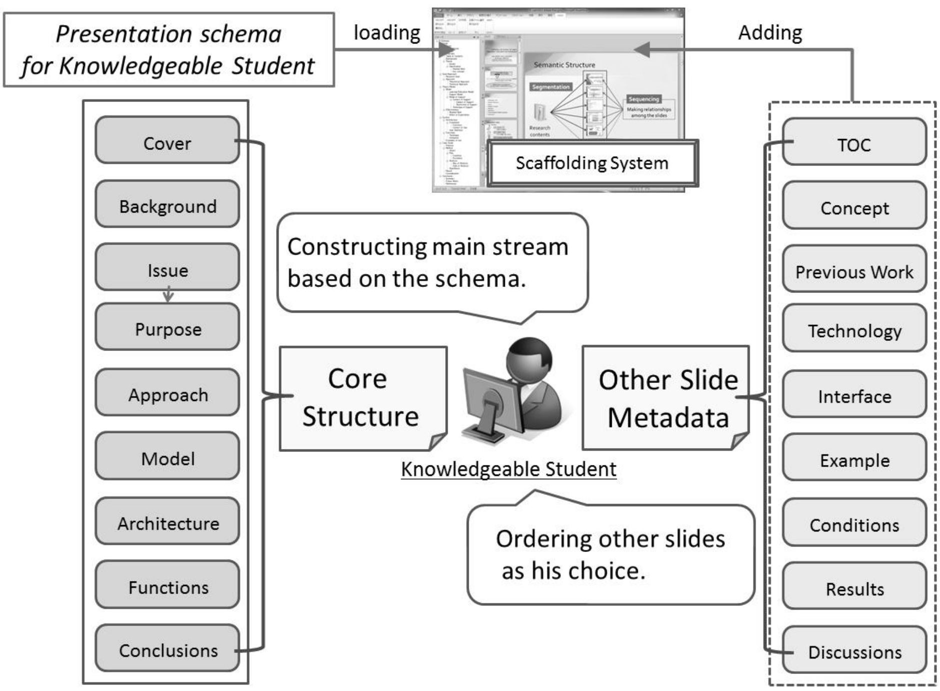

Figure 8. An example of presentation schema for knowledgeable student.

thresholds of the mining, and description capability of the schema which depends on the presentation context. In traditional laboratory education, presentation composition skill could be heuristically acquired through daily research activities as cognitive apprenticeship (Collins, 2006). Our challenge will make a contribution to develop such skill systematically.

In the near future, it will be necessary to try out the proposed technique to different domain of research groups. Especially, there are many presentation documents uploaded on the web like slideshare (http://www.slideshare.net) in these days. We would like first to analyze such documents and investigate common features as ontology and different ones among the research domain or other factors. In addition, we have to evaluate effectiveness of our related work by applying the concept of the presentation schema in a more detail.

Acknowledgments

This work is supported in part by Grant-in-Aid for Scientific Research (B) (No. 23300297), Grant-in-Aid for Scientific Research (C) (No. 23501141) from the Ministry of Education, Science, and Culture of Japan.

References

Agrawal, R., Imielinski, T., & Swami, A. N. (1993). Mining association rules between sets of items in large databases. *Proc. of the 1993 ACM SIGMOD International Conference on Management of Data* (pp. 207–216).

Atkinson, C. (2011). *Beyond bullet points: Using Microsoft PowerPoint to create presentations that inform, motivate, and inspire*. Microsoft Press; Third Edition.

Collins, A. (2006). Cognitive apprenticeship. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 47–60). Cambridge: Cambridge University Press.

Greenan, K., Humphreys, P., & McIlveen, H. (1997). Developing transferable personal skills: Part of the graduate toolkit. *Education + Training, 39*(2), 71–78.

Hasegawa, S., & Kashihara, A. (2011). Recommendation and diagnosis services with structure analysis of presentation documents. *Proc. of KES2011, Part I, LNAI 6881* (pp. 484–494).

Hasegawa, S., & Kashihara, A. (2012). An extraction technique for presentation schema embedded in presentation document. *Proc. of the 20th International Conference on Computers in Education (ICCE2012)* (pp. 214–221).

Hayama, T., Nanba, H., & Kunifuji, S. (2005). Alignment between a technical paper and presentation sheets using a hidden Markov model. *Proc. Active Media Technology 2005* (pp. 102–106).

Holsinger, E. (2003). *Presenting keynote: The insider's guide to creating great presentations*. San Francisco, CA; London: SYBEX.

Ihsan, I., Rehman, M., Ahmed, M. U., & Qadir, M. A. (2008). S-Point a semantic based e-learning content development tool. *Journal of Applied Sciences*, *8*(1), 127–133.

Ji, X., Zeng, J., Zhang, S., & Wu, C. (2010). Tag tree template for Web information and schema extraction. *Expert Systems with Applications*, *37*, 8492–8498.

Kirschner, P. A. (1988). The laboratory in higher science education, problems, premises, and objectives. *Higher Education*, *17*(1), 81–90.

Kohlhase, A. (2007). Semantic PowerPoint: Content and semantic technology for educational added-value services in MS PowerPoint. *Proc. of World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA2007)* (pp. 3576–3583).

Kotsiantis, S., & Kanellopoulos, D. (2006). Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering*, *32*(1), 71–82.

Li, S. T., & Chang, W. C. (2009). Exploiting and transferring presentational knowledge assets in R&D organizations. *Expert Systems with Applications*, *36*, 766–777.

Malarvizhi, M., & Mary, S. A. (2012). Preprocessing of educational institution web log data for finding frequent patterns using weighted association rule mining technique. *European Journal of Scientific Research*, *74*(4), 617–633.

McCarthy, P., & Hatcher, C. (2002). *Presentation skills: The essential guide for students.* London: Sage Publications Ltd.

Ota, K., & Kashihara, A. (2010). Mining collective knowledge for reconstructing learning resource. *Proc. of International Conference on Computer and Education (ICCE2010)* (pp. 104–106).

Pannurat, N., Kerdprasop, N., & Kerdprasop, K. (2010). Database reverse engineering based on association rule mining. *International Journal of Computer Science Issues*, *7*(2), 10–15.

Saito, K., Tanida, A., Kashihara, A., & Hasegawa, S. (2010). An interactive learning environment for developing presentation skill with presentation schema. *Proc. of E-Learn2010* (pp. 2696–2703).

Seta, K., & Ikeda, M. (2006). Design of an environment for developing presentation skills. *Proc. of International Conference on Computer and Education (ICCE2006)* (pp. 29–36).

Shibata, Y., Kashihara, A., & Hasegawa, S. (2012). Scaffolding with schema for creating presentation documents and its evaluation. *Proc. of E-Learn2012* (pp. 2059–2066).

Verbert, K., Jovanovic, J., Grasevic, D., Duval, E., & Meire, M. (2005). Towards a global component architecture for learning objects: A slide presentation framework. *Proc. of World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA2005)* (pp. 1429–1436).

Yang, Y., Dai, M., & Chen, H. (2010). An automatic semantic extraction algorithm for XML document. *Proc. of 2010 International Conference on Machine Vision and Human-machine Interface* (pp. 42–44).