Doctoral Dissertation

A Study on Structural and Semantic Analysis for Presentation Content Management

Supervisor Professor Kazutoshi SUMIYA

Graduate School of Human Science and Environment
University of Hyogo

Yuanyuan WANG

2014

A Study on Structural and Semantic Analysis for Presentation Content Management

Yuanyuan WANG

Abstract

Due to the advent of usable presentation tools to create attractive presentation contents, such as Microsoft PowerPoint, Apple Keynote, and OpenOffice Impress, presentations now play a socially important role in promoting understanding in many fields, including business and education. With the development of Web services such as presentation sharing websites (e.g., SlideShare, Prezi) and MOOCs (e.g., Coursera, iTunes U), they provide the archives of oral presentations and materials in public speeches and lectures owing to the features for searching, browsing, reusing, and sharing presentation contents. For example, famous Coursera and SlideShare provide an online educational presentation archive for self-learning and later review with presentation slides or video recordings of lectures. Thereby, people can be readers of presentation contents even if they are not in the same place as presenters. Although these tools and Web services make it easy for creating and sharing presentation contents are widely used, criticisms have pointed out their problems from the viewpoint of understandability about relevant information, structural information, and contextual information of presentation contents. The goal of this research is to solve the problems of traditional presentation tools focusing on searching, generating, and grasping presentation contents in e-learning, and then to develop efficient and useful applications for improving understanding of presentation contents, enhancing user interactions through presentation contents.

In this doctoral dissertation, in order to achieve these goals, the fundamental approach of this research is to explore explicit and implicit semantics of presentation contents. This approach enables to determine semantic relationships, extract expression styles, and present presentation context in presentation contents. Specifically, we propose our approaches not only for a structural and semantic analysis of presentation contents, but also focus on support for retrieval, generation, grasping overviews of presentation contents.

This research consists of five themes: support for retrieval of presentation

contents to readers or searchers, (1) scene combination for slides with recorded videos, (2) semantic slide ranking and snippet generation; support for generation of presentation contents to presenters or authors, (3) outline generation for presentation slides; support for grasping overviews of presentation contents to readers, (4) dynamic word clouds of presentations, (5) iPoster: a collaborative browsing platform for slides.

A Study on Structural and Semantic Analysis for Presentation Content Management

Contents

List of Tables

List of Figures

List of Publications

Journal and International Conference Papers

1. Automatic Generation of Learning Channels by Using Semantic Relations among Lecture Slides and Recorded Videos for Self-Learning Systems
   Yuanyuan Wang, Daisuke Kitayama, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of the 11th IEEE International Symposium on Multimedia (ISM 2009), pp. 275–280, San Diego, California, USA, December 2009
2. Semantic Ranking of Lecture Slides based on Conceptual Relationship and Presentational Structure
   Yuanyuan Wang and Kazutoshi Sumiya
   In Proc. of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010), pp. 2801–2810, Barcelona, Spain, September 2010
3. Slide KWIC: Snippet Generation for Browsing Slides based on Conceptual Relationship and Presentational Structure
   Yuanyuan Wang and Kazutoshi Sumiya
   In Proc. of the 9th International Conference on Creating, Connecting and Collaborating through Computing (C5 2011), pp. 40–47, Kyoto, Japan, January 2011
4. A Browsing Method for Presentation Slides Based on Semantic Relations and Document Structure for e-Learning
   Yuanyuan Wang and Kazutoshi Sumiya
   Journal of Information Processing, Vol. 20, No. 1, pp. 11–25, December 2011
5. Skeleton Generation for Presentation Slides based on Expression Styles
   Yuanyuan Wang and Kazutoshi Sumiya
   In Proc. of the 5th International Conference on Intelligent Interactive Multimedia Systems and Services (KES IIMSS 2012), SIST 14, pp. 551–560, Gifu, Japan, May 2012
6. A Method for Generating Presentation Slides based on Expression Styles using Document Structure

Yuanyuan Wang and Kazutoshi Sumiya
International Journal of Knowledge and Web Intelligence, Vol. 4, No.1, pp. 93–112, March 2013

7. Dynamic Word Clouds: Context-based Word Clouds of Presentation Slides for Quick Browsing
   Yuanyuan Wang and Kazutoshi Sumiya
   In Proc. of the 6th International Conference on Intelligent Interactive Multimedia Systems and Services (KES IIMSS 2013), pp. 108–117, Sesimbra, Portugal, June 2013
8. iPoster: A Collaborative Browsing Platform for Presentation Slides based on Semantic Structure
   Yuanyuan Wang, Kota Tomoyasu, and Kazutoshi Sumiya
   In Proc. of the 1st Workshop of Quality, Motivation and Coordination of Open Collaboration (QMC 2013), Kyoto, Japan, November 2013 (to appear)

Technical Reports

1. Presentation Management System based on Semantic Relations using Metadata of Slides and Videos
   Yuanyuan Wang, Daisuke Kitayama, and Kazutoshi Sumiya
   In Proc. of the 1st Forum on Data Engineering and Information Management (DEIM Forum 2009), E9-4, March 2009 (in Japanese)
2. Slide Relation Extraction for Presentation Contents based on Conceptual Structure and Visible Structure
   Yuanyuan Wang, Daisuke Kitayama, and Kazutoshi Sumiya
   In Proc. of the 2nd Forum on Data Engineering and Information Management (DEIM Forum 2010), C1-1, March 2010 (in Japanese)
3. Slide Relation Extraction based on Conceptual Structure for Presentation Contents
   Yuanyuan Wang, Daisuke Kitayama, and Kazutoshi Sumiya
   In Proc. of Information Processing Society of Japan (IPSJ) 50th Anniversary, 72nd National Convention of IPSJ, pp. 869–870, March 2010 (in

Japanese)

4. Semantic Ranking of Lecture Slides based on Conceptual Relationship and Presentational Structure

   Yuanyuan Wang and Kazutoshi Sumiya

   In Proc. of the 2nd International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB Workshop 2010), pp. 65–71, August 2010
5. Slide-KWIC Browser: Snippet Generation Based on Semantic Relations and Document Structure for Presentation Slides

   Yuanyuan Wang and Kazutoshi Sumiya

   In Proc. of the 3rd Forum on Data Engineering and Information Management (DEIM Forum 2011), B7-3, February 2011
6. A Snippet Generation Method for Retrieving Slides in Presentation Content Sharing Sites

   Yuanyuan Wang and Kazutoshi Sumiya

   In Proc. of the 2nd Symposium on Social Computing (SoC 2011), p. 26, June 2011
7. A Browsing Method for Presentation Slides Based on Semantic Relations and Document Structure

   Yuanyuan Wang and Kazutoshi Sumiya

   In Proc. of the 3rd International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB Workshop 2011), pp. 86–97, August 2011
8. A Browsing Method for Digital Posters based on Semantic Relationships between Slides

   Kota Tomoyasu, Yuanyuan Wang, and Kazutoshi Sumiya

   In Proc. of the 5th Ubiquitous Wearable Workshop (UWW 2011), p. 12, December 2011 (in Japanese)
9. Generating Slide Skeletons based on Expression Styles for Presentation Contents

   Yuanyuan Wang and Kazutoshi Sumiya

   In Proc. of the 4th Forum on Data Engineering and Information Management (DEIM Forum 2012), D5-3, March 2012

10. A Browsing Method for Presentation Contents based on Semantic Relationships between Slides
    Kota Tomoyasu, Yuanyuan Wang, and Kazutoshi Sumiya
    In Proc. of the Institute of Electronics, Information and Communication Engineers (ISS), p. 225, March 2012 (in Japanese)
11. A Generation Method of Presentation Slides based on Expression Styles using Slide Structure
    Yuanyuan Wang and Kazutoshi Sumiya
    In Proc. of the 4th International Workshop with Mentors on Databases, Web and Information Management for Young Researchers (iDB Workshop 2012), August 2012
12. A Browsing Method for Posters using Zooming Transitions based on Posters and Slide Structures
    Kota Tomoyasu, Yuanyuan Wang, and Kazutoshi Sumiya
    In Proc. of IPSJ SIG Technical Reports, Vol. 2012-DBS-155, No. 1, November 2012 (in Japanese)
    item Context-based Word Clouds of Presentation Slides for Quick Browsing
    Yuanyuan Wang and Kazutoshi Sumiya
    In Proc. of the 5th Forum on Data Engineering and Information Management (DEIM Forum 2013), A10-4, March 2013 **[Student Presentation Award], [Interactive Award]**
13. Poster Generation System based on Silde Elements and Semantic Relationships
    Kota Tomoyasu, Yuanyuan Wang, and Kazutoshi Sumiya
    In Proc. of the 5th Forum on Data Engineering and Information Management (DEIM Forum 2013), C8-4, March 2013 (in Japanese)
14. A Word Cloud Visualization for Comparing Presentation Contents
    Yuanyuan Wang and Kazutoshi Sumiya
    In Proc. of the 4th Symposium on Social Computing (SoC 2013), pp. 47–52, June 2013

Technical Magazine

1. Fusion of Image and Heterogeneous Multimedia (1)
   Daisuke Kitayama, Shoko Wakamiya, Yuanyuan Wang, and Kazutoshi Sumiya
   Image Laboratory, Japan Industrial Publishing Co., Ltd, pp. 41–45, January 2014 (in Japanese)
2. Fusion of Image and Heterogeneous Multimedia (2)
   Daisuke Kitayama, Shoko Wakamiya, Yuanyuan Wang, and Kazutoshi Sumiya
   Image Laboratory, Japan Industrial Publishing Co., Ltd, pp. 59–67, February 2014 (in Japanese)

University Memoirs

1. Development of Web-based Calligraphy Contents and Design of Database System
   Yuanyuan Wang, Wuyi Yue, and Jun Liu
   Memoirs of Konan University, No. 6, CD-ROM, 28 pages, March 2008 (in Japanese)

Chapter 1 Introduction

1.1 Motivation and Our Approaches

With the advent of usable presentation tools to create attractive presentation contents, such as Microsoft PowerPoint [52], Apple Keynote [2], and OpenOffice Impress [1], presentations now play a socially important role in promoting understanding in many fields, including business and education. Many people have used Web services such as presentation sharing websites (e.g., SlideShare [83], Prezi [76]) and MOOCs (e.g., Coursera [13], iTunes U [27]) to store presentation contents that they use in speeches and lectures owing to the features for editing, browsing, sharing, and reusing presentation data. For example, most presentation contents with rich graphics and animations are prepared by using presentation tools such as Microsoft PowerPoint, Apple keynote, and recently online editor Prezi. In addition, presentation sharing websites such as SlideShare and Coursera provide an online presentation archive for later review with presentation slides or video recordings of speeches. Although useful and powerful support tools make it easy for creating presentation contents and Web services for sharing presentation contents are widely used, criticisms have pointed out their problems from the viewpoint of understandability of presentation contents [71, 91, 66, 18]. They are still a lack of support for users (1) acquiring relevant information implicit among presentation contents; (2) utilizing structural information explicit in presentation contents; and (3) grasping overviews of contextual information existing in presentation contents. This dissertation addresses these problems for improving understanding of presentation contents, enhancing user interactions through presentation contents by analyzing structural semantics and lexical semantics in presentation contents. In this doctoral dissertation, we are challenging to support for retrieval, generation, grasping overviews of presentation contents by focusing three critical issues, that is, a) determining semantic relationships, b) extracting expression styles, and c) presenting presentation context as shown in Figure 1.

Support for Retrieval of Presentation Contents :

Scene Combination for Slides with Recorded Videos We propose

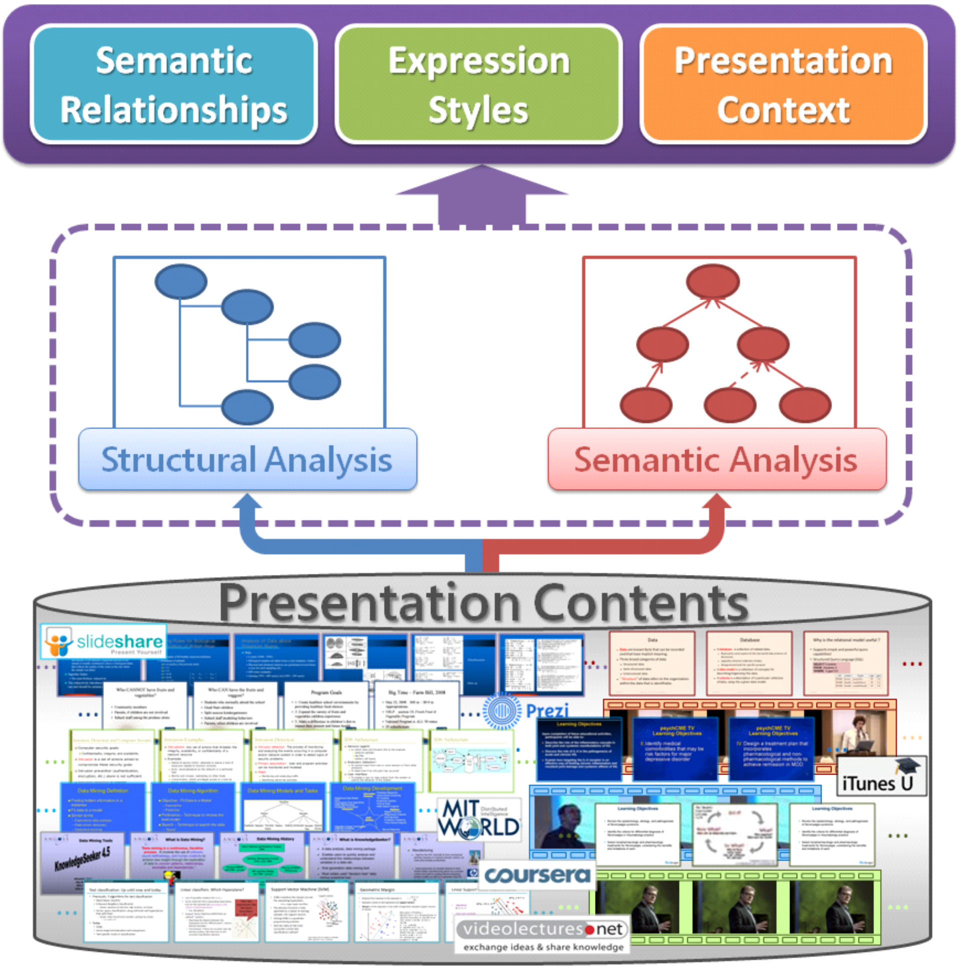


Figure 1: Concept of a Structural and Semantic Analysis of Presentation Contents

a method of automatically generating learning channels to extract scenes and combined scenes from slides with their recorded video based on semantic relations. The system analyzes the type of semantic relation on the basis of the metadata of structural information, such as indents and texts in slides, and the set of keywords in the text of the speech in the video. In this way, our newly generated learning channels let users easily focus on either highly detailed slides or introductory slides without needing to examine all of the data.

Semantic Slide Ranking and Snippet Generation We develop a slide retrieval system involving semantic ranking and snippet generation, and we discuss how to present the retrieval results to users by considering what rank orders of slides and what portions of slides are relevant to a query, on the basis of relationships between slides that enable the browsing of slide retrieval at the conceptual level. With our novel slide ranking method and snippet-generation method, not only precise retrieve target slides but also the semantic ranking of them to help users easily learn through slides; and the relevant portions of them giving their surrounding context to help the users easily decide which slides to learn are useful or not.

Support for Generation of Presentation Contents :

Outline generation for presentation slides We attempt to generate outlines for lecture slides from textbook chapters. We aim to organize slide layouts from target chapters based on the expression styles of referred slides. Therefore, we analyze level positions of words in the referred slides and arranged words from target chapters to generate slide outlines based on difference in document structure (i.e. text structure within a chapter, slide structure within a slide). Finally, our method generates slide outlines by reusing the expression styles of existing slides to help presenters or authors make slides in their desired styles.

Support for Grasping Overviews of Presentation Contents :

Dynamic word clouds of presentation We challenge to develop a quick browsing tool to help users effectively compare presentations for their specific needs. For the purpose, we will provide a word cloud visualization that summarizes information to help the users visually understand the context of each presentation. Words important to the "presentation context," is first extracted based on components of the presentation (i.e., intra-slide and inter-slide structures). Finally, our word cloud visualization shows the words are interactively presented with visual effects in presentations.

iPoster: a collaborative browsing platform for slides We attempt to build a collaborative browsing platform for presentation slides based on interactive poster generation, called "iPoster," for presenting textual and graphic elements in a meaningfully structured layout with automatic transitions, such as zooms and pans, to promote user interaction. Through this, users can interactively browse an iPoster on their tablets. The navigation information maps each user's specific needs by considering the user's operations, and detects other users who have similar requirements to help them share their interests with each other.

1.2 Outline of the Doctoral Thesis

In this dissertation, we introduce our approaches on a structural and semantic analysis of relevant information, structural information, and contextual information from presentation contents. In order to achieve the support for retrieval, generation, grasping overviews of presentation contents, this dissertation addresses five themes and organized into eight chapters, including this chapter as the introduction.

In Chapter 2, in order to position our research comparing with others and show the value of our research, we overview related work.

In Chapter 3 and Chapter 4, we describe methods to support for slide retrieval to readers or searchers by utilizing semantic relationships between slides. Concretely, we measure semantic importance and relationships with semantic relations between keywords and document structure in presentation contents; i) scene combination for slides with recorded videos and ii) semantic ranking and context summarization of presentation slides. As for i), we proposed a method to automatically generate learning channels by using the semantic relations among scenes, which lets users easily focus on either highly detailed scenes or introductory scenes without needing to examine all of the data. In the case of ii), we attempt to rank slides by using the semantics of relationships without relying on the existence of any specific structure in a slide or relevant information between slides. In addition, we consider that retrieved slides also contain irrelevant information to a query. For this, we challenge to generate snippets that capture relevant portions of the retrieved slides as their surrounding context, which help users understand them in presentation contents easily.

In Chapter 5, we show a method to support for slide generation from textbook chapters to presenters or authors by reusing expression styles of existing slides. Although most slides can be automatically generated by conventional methods follow structured document summaries (e.g. academic papers), we aim to organize slide layouts from target chapters by reusing various styles of referred slides based on level positions of words in the referred slides. To achieve this, we extract differences between tendency of word appearance in chapters and their associated slides (referred slides).

Support for grasping overviews of presentation contents to readers. In Chapter 6, we explain a method to visualize presentation contents by extracting words important to the context of presentations. Here, we focus on how to decide which files are worth learning, because most of presentation contents in search results are similar; it can be difficult to identify differences in them. Therefore, we develop a quick browsing tool provides a word cloud visualization shows the words are interactively presented with visual effects. In Chapter 7, we present a collaborative browsing platform that generates a meaningfully structured presentation by transporting slides. It promotes user interaction and communication and is called the "iPoster." A collaborative browsing platform based on the iPoster, which can share and navigate information, matches each user's specific requirements by analyzing the users' operations. Further, it detects other users who have similar requirements by mapping the similarity in their operations and conveys their interests to each other.

Finally, we conclude this dissertation and discuss the future direction of the research in Chapter 8.

Chapter 2 Related Work

2.1 Presentation Content Retrieval

Most of the research related to presentation contents has been focused on the retrieval of slides. Yokota et al. [100] proposed a system named Unified Presentation Slide Retrieval by Impression Search Engine (UPRISE) for retrieving a sequence of lecture slides from archives containing a combination of slides and recorded videos. Okamoto et al. [65] proposed UPRISE to retrieve scenes matched with given keywords from a video recording a lecture using presentation slides based on the adaptability of keywords and synchronous information about slides and videos. Kobayashi et al. [34] proposed a method based on the use of laser pointer information for retrieving lecture slides by UPRISE. Le et al. [38] proposed a method for extracting important slides by automatically generating digests from recorded presentation videos. However, we considered that retrieving only the important slides decreases the relevance of the results of a user query to the given context, and their method cannot be used to browse important slides containing information related to a query.

In our previous works, Kitayama et al. [32, 33] proposed a method for extracting slides with corresponding video scenes based on relationships between slides and their roles, and Wang et al. [95] described a process for automatically generating learning channels by using the semantic relationships that implicitly exist in the slides of a lecture with an accompanying recorded video. These studies were similar to our study, in that a method for understanding the desired slides using relationships between, and relevant information about, these slides was proposed. In general, context is useful for understanding, and several studies, which we now briefly explain, have exploited context in different ways. Pattanasri et al. [67] utilized information in textbooks to construct an entailment ontology, finding that two types of entailment relations were helpful for identifying context when trying to understand search results inside e-learning material. Thus, users can browse a collection of items or documents to understand them easily. Similarly, we analyze contextual information in presentation contents to give the surrounding context for the focused slides.

Smith et al. [85] proposed a method of extracting video segments from a video by detecting the features of scenes by analyzing their CC (Closed Caption), color, and speech features. Pradhan et al. [74] proposed a gluing operation that generated a new video segment from a set of video segments. Tanaka et al. [89] focused on the manipulation of complex (database) objects and introduced the concept of "element-based" generalization relationships between complex objects as generalization hierarchy and two new abstraction operators, namely, reduction and unification operators. However, we focused on semantic relations including generalization relationships between keywords in the slide text and utilized indents in the slide containing the keywords. Kushki et al. [36] proposed a novel XML-based system for the retrieval of presentation slides. This system analyzing contextual information, such as structural and formatting features, is extracted from the open format XML representation of slides. Our complementary method considers both the structure of the slide, and the semantic relations between different keywords contained therein, and it analyzes these two features to determine the relationships between slides.

2.2 Presentation Content Generation

There are several studies related to slide-making support has focused on reusing documents (i.e., academic papers, textbooks). Mathivanan et al. [46], Beamer et al. [4], Miyamoto et al. [56], and Yoshiaki et al. [101] proposed a system for generating slides from academic papers. Their method summarizes and extracts information from an academic paper by means of tf-idf term weighting, and assigns sentences, figures, and tables in slides by identifying important phrases for bullets in order to generate slides. Shibata et al. [79] converted documents to slides by parsing their discourse structure and representing the resulting tree in an outline format. Kan [30] proposed a system called SlideSeer for the discovery, alignment, and presentation of documents and slide pairs. This system modifies the maximum similarity in alignment to favor monotonic alignments, and it incorporates a classifier to handle slides that should not be aligned. Hayama et al. [24] proposed a method for aligning academic papers and slides based on Jing's method, which uses a hidden Markov model (HMM). From the viewpoint of reusing slides, Sharmin et al. [78] and Mejova et al. [51] proposed a system for composing presentation slides from existing ones and modifying them for specific events, such as lectures and conferences. We share a common point in reusing expressions of words in existing slides to create new ones.

Kurohashi et al. [35] detected the importance of a word in a document based on its location with the highest density. This study is similar to ours in terms of the retrieval of characteristic information in documents and slides. However, they focus on important information, while our method considers the differences between text structures within documents and slide structures within slides. Because of the problem associated with providing adequate support for the organization of slide components has been addressed, Watanabe et al. [97] and Hanaue et al. [22] have focused on semantic relationships among slide components for a presentation strategy. Other approaches to presentation composition have included outline matching [7], topic clustering [87], and hierarchical organization [5]. These approaches are similar to ours in that they help presenters better organize their slide contents. Our goal, however, is to support presenters in their slide preparation with semi-automatically generated slide outlines based on reusing expressions of words in existing slides.

2.3 Presentation Content Visualization

In the area of text analytics, researchers have developed a number of approaches for text summarization, of which there are two main techniques: sentence-based and word-based text summarization. Sentenced-based approaches identify the most salient sentences in a document [9, 94]. For example, Murai and Ushiama [59] proposed a browsing method that presents users with a review-based recommendation of attractive sentences in a novel. However, it may be time consuming for users to read several sentences per document, especially when handling a large number of documents. Alternatively, word-based methods summarize documents by topic, each of which is characterized by a set of words [16, 47, 31]. Our quick browsing method is built on the latter method, but its focus is on enhancing the summarization results through word clouds.

In the area of information visualization, researchers have developed various visualization approaches to text analysis. These systems can be classified into two categories: metadata-based and content-based text visualization. In email analysis for instance, metadata-based text visualization can use a time-based visualization to explain text summarization results derived by a text analytic engine [43], or a create a relationship-based visualization of email senders and receivers [69]. For content-based text visualization, Viegas et al. [93] used Themail to visualize keywords based on TF-IDF scores in an email collection. Similarly, Strobelt et al. [88] used a mixture of images and TF-IDF-based keywords to create a compact visualization of a document. More recently, Chen et al. [10] and Iwata et al. [28] focused on visualizing document clustering results. In contrast, others have concentrated on representing text content at the word or phrase level, including TextArc (www.textarc.org), WordTree [98], Phrase Net [21], and FeatureLens [15]. Our work focuses on visualizing presentation contents used word clouds with transitions between the word clouds.

NextSlidePlease [86] is a novel application for authoring and delivering slideware presentations. This tool addresses issues of content integration, presentation structuring, time-management, and flexible presentation delivery. Our generated interactive poster, called "iPoster," is similar to this work, as we utilize a structured layout, rather than one or more slide lists, to allow interactive and collaborative browsing by users for each other. Good and Bederson [19] proposed replacing the card stack or film strip metaphor with a ZUI [6] in their CounterPoint application, borrowing insights from the domain of mind-maps or concept maps [63] and visual storytelling [48]. The Fly application addresses graph-based presentation authoring [42]. It provides a set of tools for authoring presentations from scratch in a two-dimensional canvas with defined paths. Our iPoster uses a ZUI to navigate to the elements from slides. On the other hand, Laufer et al. [37] argue about the use of avatars in ZUIs are providing a uniquely efficient environment for collaboration in productivity applications. Multiple users can develop a presentation together; create a mindmap, and a storyline or do brainstorming. In our method, we create an iPoster by transporting slides, which aims to utilize the strengths of ZUIs for providing an enhanced collaborative browsing tool.

Chapter 3 Scene Combination using Semantic Relations among Presentation Slides with Recorded Videos

3.1 Introduction

These days, a lot of lecture materials made from many actual classes in universities or other education organizations are shared on websites such as Coursera [13] and VideoLectures.NET [92]. Free online educational contents often consist of presentation slides and recorded video. Thus, not only students who missed the class but also any other people interested in the topic can review the class and study the content by themselves later. However, compared with actual participation, learning throughout such achieved material is more passive and tedious because of the lack of interactivity and intensity; the dominant learning style for online presentation contents is just viewing sequentially arranged slides and video. Such unidirectional injective learning cannot easily attract a self-learner's interest and requires the user's effective attention. On the other hand, making the static contents much more dynamic and interactive would require a lot of effort by lecturer. To fill the gap between the lecturers' limitations and students' diverse requirements in practice, we propose a dynamic reorganization of the almost raw contents, which are easily available on the Web but unable to meet the needs of students having various levels of understanding.

Reorganizing presentation contents to suit users' interests or capabilities could be achieved mainly by (1) summarizing long contents into short intensive highlights [11] that include what users need to know and (2) constructing a hierarchical or graph-like structure like HTML documents [50] without keeping the sequential ordering, but focusing on relationships among slides or video segments. However, the first approach often fails to cope with dynamic changes in users' interests because once the highlights have been created, they are hard to reorganize according to given summarization criteria. Unlike the relatively static approach, the second approach is very flexible at supporting dynamic changes in users' interests during the learning processes, but there are generally no special linkages explicitly represented in either slides or videos. Thus, it would be necessary to translate presentation contents into dynamic and semantic learning channels, where each student is supported throughout dynamic changes in his or her interests according to his/her learning level by using the semantic relationships among slides that has accompanying video segments from archives. The concept of our approach is shown in Figure 2.

We propose a method of automatically generating learning channels with a differential base by analyzing specific relationships. We extract structural information such as indents and the logical set of text in slides and the keyword set for text in the speech in a video and provide this information so that our method can use it (see Figure 3). The content is divided on the basis of the speaker's slide changes. Each slice produced by the division is called a scene, which consists of *one* slide and *one* video segment containing a recording of the speaker's explanation of that slide.

We defined the structural information on the basis of indents in slide text. The slide title (1st level indent) is the upper level. The first item of text is on the 2nd level and subitems deepen with the level of indentation (3rd level, 4th level, and so on). Indents outside text such as figures or tables are on the average level for the slide. It is usual that the lower-level indented keywords are supplementary and explain the upper-level keywords. We define semantic relationships between scenes by using the metadata of slides and videos, for example, a certain scene is a much more detailed explanation than other scenes.

Our basic method is to extract corresponding scene pairs by using the relationships between the two scenes of the selected pair. Our approach involves two types of input: (1) selecting one scene and one relationship for which the user wants to get corresponding scene pairs for some purpose and (2) selecting a scene pair and the relationship between the two scenes for which the user wants to get other scene pairs related to the selected scene pair. First, we determine semantic relationships by examining how a keyword's indent position varies in different slides and how frequently the keyword appears the video. As a result, users can get semantic scenes as learning channels from inputs that may or may not be contiguous.

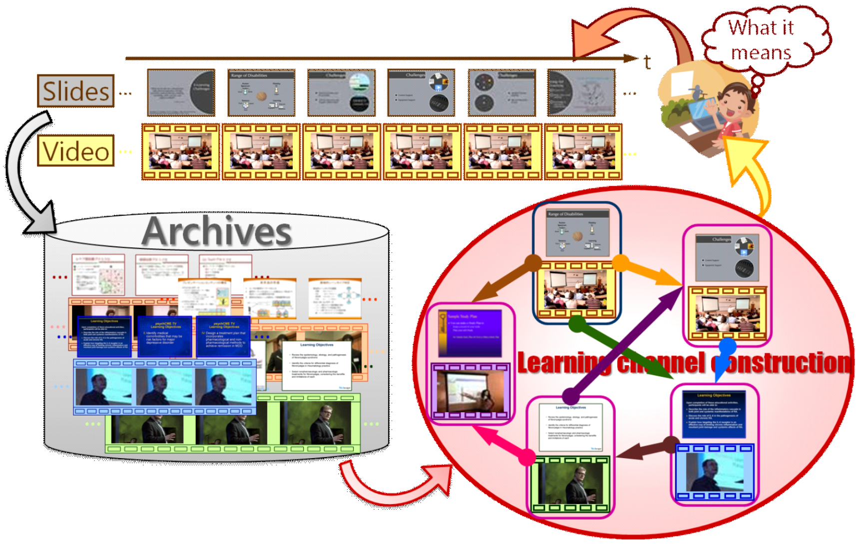


Figure 2: Concept of Our Approach

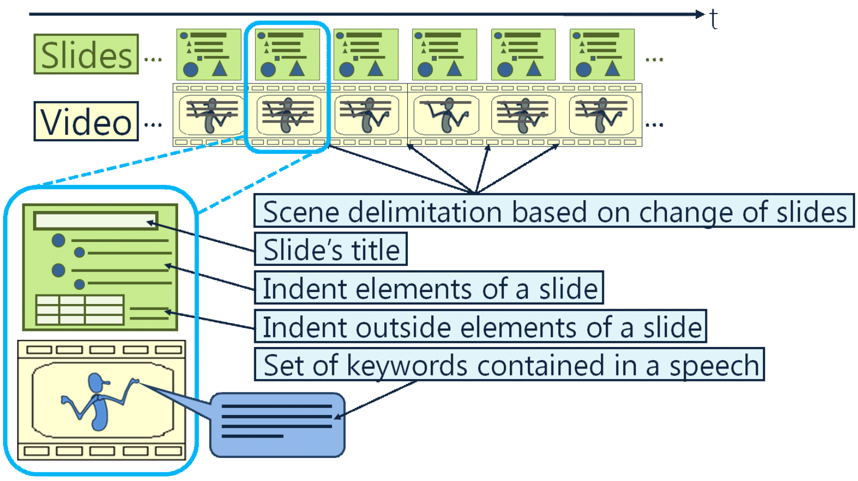


Figure 3: Metadata of an Unified Presentation Content

3.2 Determination of Semantic Relationships

3.2.1 Semantic Relationship Types

We define the selected scene of interest as the basic scene and define other scenes that have specific relationships as being semantically related to the basic scene through one of four semantic relationship types: *detailed*, *generalized*, *similar*, and *additional* (see Figure 4). Scenes that have semantic relationships are called semantic scenes, i.e., if a scene has a detailed relationship, we call this scene a detailed scene.

Detailed and generalized scenes are functionally interchangeable, while a basic scene is a generalized scene from the viewpoint of a detailed scene.

This section explains how the semantic relationship type is determined. Let a_i be the slide number of a basic scene and a_j be the slide number of a scene we want to detect. The semantic relationship types are determined for all scenes.

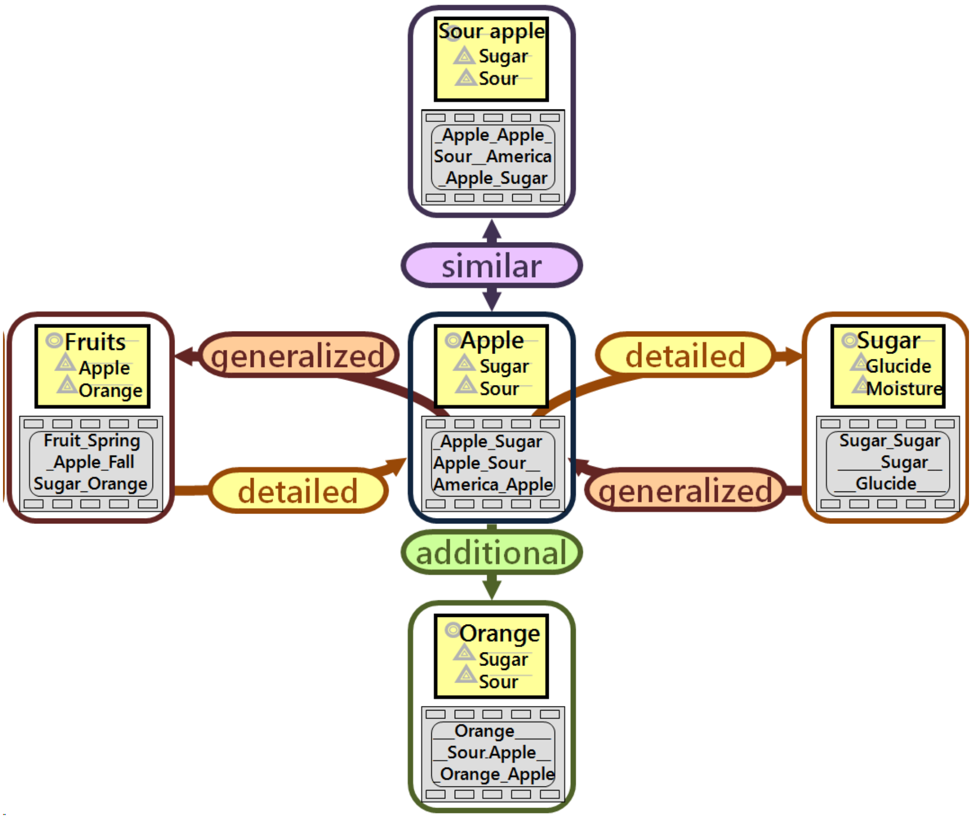


Figure 4: Semantic Relationship Types

3.2.2 Judgment of Detailed or Generalized Relationships

If a scene has more information than the basic scene, its relationship to the basic scene is *detailed*. The slide of a scene that contains specialized content is talked about at greater length by the speaker. If a scene contains content in the outline given in a *generalized* scene, it is described in relationship to the basic scene. The slide of a scene that contains generalized content is talked about less by the speaker. Because *detailed* and *generalized* scenes are equivalent, we explain only the determination of *detailed* scenes by using keywords present in basic scene a_i and the scene to detect a_j here.

$$\begin{aligned} |U(a_i, a_j)| &> |S(a_i, a_j)| \quad (1) \\ |U(a_i, a_j)| &> |D(a_i, a_j)| \quad (2) \\ \frac{vn(U(a_i, a_j), a_i)}{vc(a_i)} &< \frac{vn(U(a_i, a_j), a_j)}{vc(a_j)} \quad (3) \end{aligned}$$

If the levels conform to Eqs. (1) and (2), and the ratio of keywords in the video conforms to Eq. (3), then a_j is determined to be detailed. This is because the keywords in the a_jth slide appear more frequency than in the a_ith slide, and the a_jth video segment contains many keywords. $U(a_i, a_j)$ is the set of keywords in levels that ascend from the a_ith slide to a_jth slide. $S(a_i, a_j)$ is the set of keywords in the same level in both the a_ith and a_jth slides. $D(a_i, a_j)$ is the set of keywords in levels that descend from the a_ith slide to a_jth slide. If the number of keywords in $U(a_i, a_j)$ is extracted more frequently than the number in $S(a_i, a_j)$ and $D(a_i, a_j)$ in Eqs. (1) and (2), then the level of the slide is judged to be higher. In Eq. (3), $vn(U(a_i, a_j), a_i)$ is the number of keywords in $U(a_i, a_j)$ in the a_ith video segment and $vc(a_i)$ is the total number of keywords in the a_ith video segment. If a scene has more information than the basic scene, its relationship to the basic scene is *detailed*.

3.2.3 Judgment of Similar Relationships

If the slide of a scene contains similar content to the basic scene and its video segment has a similar quantity of speech to that of the basic scene, then the relationship between these scenes is *similar*.

$$\begin{aligned} |S(a_i, a_j)| &> |U(a_i, a_j)| \quad (4) \\ |S(a_i, a_j)| &> |D(a_i, a_j)| \quad (5) \\ \left| \frac{vn(S(a_i, a_j), a_i)}{vc(a_i)} \right. &- \left. \frac{vn(S(a_i, a_j), a_j)}{vc(a_j)} \right| < \alpha \quad (6) \end{aligned}$$

If the levels conform to Eqs. (4) and (5), and the keyword ratio in the video conforms to Eq. (6), then a_j is determined to be similar. This is because the keywords in the a_jth slide appear with a similar frequency in the a_ith slide, and the keyword ratio is similar in both the a_ith and a_jth video segments. If the number of keywords in $S(a_i, a_j)$ is extracted more frequently than the number in $U(a_i, a_j)$ and $D(a_i, a_j)$ in Eqs. (4) and (5), then the hierarchical structure of the slide is determined to be the same. In Eq. (6), $vn(S(a_i, a_j), a_i)$ is the number of keywords in $S(a_i, a_j)$ in the a_ith video segment and α is a threshold.

3.2.4 Judgment of Additional Relationships

If a scene contains another topic related to the basic scene, its relation to the basic scene is *additional*. The speaker's additional comments can also be used to describe the content in other scenes. The speaker descriptions also include the keywords contained in the slide of the basic scene. This *additional* scene helps users to understand the basic scene by providing extra information.

$$\begin{aligned} |inter(a_j, a_i)| &< |differ(a_j, a_i)| && (7) \\ \frac{sn(inter(a_j, a_i), a_i)}{sc(a_i)} &> \frac{sn(inter(a_j, a_i), a_j)}{sc(a_j)} && (8) \\ vn(inter(a_j, a_i), a_i) &> 0 && (9) \\ vn(inter(a_j, a_i), a_j) &> 0 && (10) \\ l(k_x, a_i) < \frac{l_l(a_i)}{2} &\quad \text{or} \quad l(k_x, a_j) > \frac{l_l(a_j)}{2} && (11) \end{aligned}$$

If the levels conform to Eqs. (7) and (11), the keyword ratio in the slide conforms to Eq. (8), and the keywords in the video conform to Eqs. (9) and (10), then a_j is determined to be additional. This is because there is an explanation of the a_ith scene in the a_jth scene, and there is more explanation in the a_ith slide. Let $inter(a_j, a_i)$ be the set of keywords that appear in both the a_ith and a_jth scenes and let $differ(a_j, a_i)$ be the set of keywords that do not appear in a_ith scene. If the number of keywords in $differ(a_j, a_i)$ is more than the number in $inter(a_j, a_i)$ in Eq. (7), then the keywords in $inter(a_j, a_i)$ are common keywords in both the a_ith and a_jth scenes. In Eq. (8), $sn(inter(a_j, a_i), a_i)$ is the number of common keywords in the a_ith slide and $sc(a_i)$ is the total number of keywords in the a_ith slide. In Eq. (9), $vn(inter(a_j, a_i), a_i)$ is the number of common keywords in the a_ith video segment. In Eq. (11), $k_x \in inter(a_j, a_i)$, function l_l is the lowest level of the slide, so we can estimate whether common keywords ascend from a lower level to an upper level.

3.3 Automatic Generation of Learning Channels

Our learning-channel construction method involves two types of input: (1) by selecting one scene and one relationship, the user wants to get corresponding scene pairs for some purpose and (2) by selecting a scene pair, the user wants to get corresponding scene pairs that are related to the selected scene pair. When one scene is selected, our system extracts the corresponding scene pairs that have the selected type of semantic relationship for the user's interest by detecting scenes that correspond to the selected scene. When the user selects a scene pair of interest, our system determines its relationships with all other scene pairs and extract ones judged to be corresponding ones.

3.3.1 Detection of Corresponding Scenes

The contents of three lectures are shown as examples in Figure 5. When a user studying geography (A) is interested in scenes a_4 and a_6 about "pumpkin as a vegetable", the scene pair [a_4, a_6] is selected, and the relationship from a_4 to a_6 to be *detailed*. In this case, we consider that it would be useful for the user if the system presented him/her with other detailed scene pairs having relevant points from other lectures. [b_3, b_4] in B explains about "pumpkin as a vegetable", and b_4 explains in more detail than b_3. Thus, [b_3, b_4] also corresponds to the selected scene pair. On the other hand, although c_7 provides more details about "pumpkin" than c_6 in C, these scenes are not related to the selected scenes because they explain about the pumpkin as a symbol of Halloween. Therefore, [b_3, b_4] is extracted as a corresponding scene, where b_3 is a scene in other contents that corresponds to the basic scene a_4 of the selected scene pair. Thus, in extracting corresponding pairs, it is necessary to refer to the corresponding scene.

Let a_i be a basic scene in content A, and let b_n be a candidate scene in content B. If the keywords in a_i and b_n satisfy the following conditions, then the scene

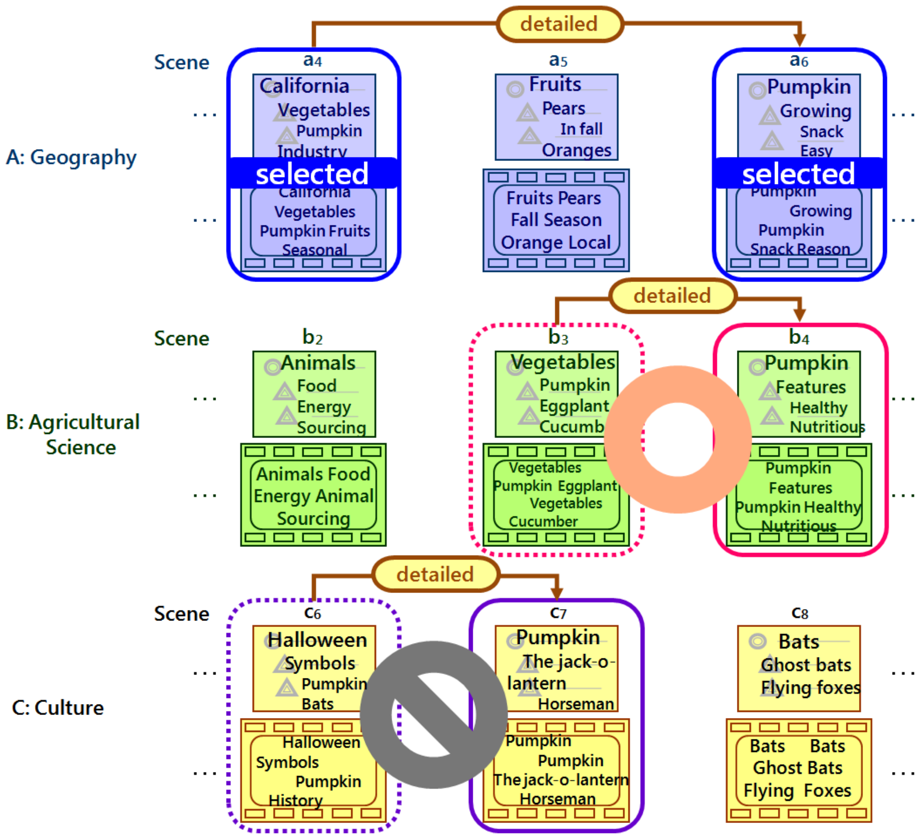


Figure 5: Scene Pair Extraction by Detecting Corresponding Scenes

b_n is judged to be a corresponding scene in other contents (see Figure 6).

$$K(a_i) = \{\{k_x, k_y\} \mid l(k_x, a_i) < l(k_y, a_i)\} \quad (12)$$

$$\frac{|K(a_i) \cap K(b_n)|}{min(|K(a_i)|, |K(b_n)|)} > \beta \quad (13)$$

$$SVratio(a_i, b_n) = \frac{sn(K(a_i) \cap K(b_n), a_i)}{vn(K(a_i) \cap K(b_n), a_i)} \quad (14)$$

$$|SVratio(a_i, b_n) - SVratio(b_n, a_i)| < \gamma \quad (15)$$

In Eq. (12), $K(a_i)$ is the set of keyword pairs, where the level of keyword k_x is higher than the level of keyword k_y in the a_ith slide. In Eq. (13), $|K(a_i) \cap K(b_n)|/min(|K(a_i)|, |K(b_n)|)$ calculates the degree of the hierarchical relation between the keywords in the a_ith and b_nth slides, and β is a threshold. In Eq. (14), function $SVratio(a_i, b_n)$ calculates the degree of the number of $K(a_i) \cap K(b_n)$ keeping the hierarchical relation in the a_ith slide and video segment. $s_n(K(a_i) \cap K(b_n), a_i)$ is the number of $K(a_i) \cap K(b_n)$ in the a_ith slide, and

$v_n(K(a_i) \cap K(b_n), a_i)$ is the number of $K(a_i) \cap K(b_n)$ in the a_ith video segment. In Eq. (15), the function calculates the similarity of $K(a_i) \cap K(b_n)$ in a_i and b_n, and γ is a threshold.

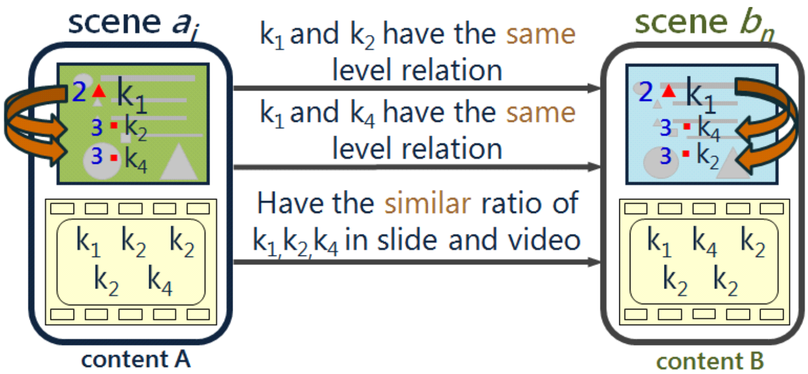


Figure 6: Determination of Corresponding Scene

3.3.2 Generation of Learning Channels

Learning channels extract corresponding scene pairs determined by selecting a scene pair that has the same semantic relationship between its scenes. They produce different outputs depending on the type of input (there are two types of input).

Input One Scene

When a scene is selected, the system searches for scenes in other contents using the selected relationship type. We think that the user understood the selected content but wants to gain a better understanding of the topic.

In Figure 5, after the user understood A, he selected a_4 and *detailed* in order to study a_4 and gain a more *detailed* understanding of it. a_6 explains a_4 in detail, but it is also useful to present the user with relevant detailed scenes in other lectures. b_3 corresponds to a_4. The corresponding scene pair [b_3, b_4] in B explains about “pumpkin as a vegetable”, and b_4 explains in more detail than b_3. So, b_4 is a detailed explanation of the content in a_4 that could help the user to understand “pumpkin as a vegetable” in detail by utilizing content from other lectures. Therefore, we think that the user understood a_6 in A so it is not presented, extracting only b_4 can satisfy the user’s demand.

Input One Scene Pair

When a scene pair is selected, our method estimates all relationships between it and corresponding scene pairs in other lectures. We think that if the user looks at only the content of a single lecture, he/she can understand it well with a little supplementary explanation from other lectures.

In Figure 5, when the user selects [a_4, a_6] out of interest in "pumpkin as a vegetable", a_6 explains a_4 by providing more detail, so our method estimates the relationship from a_4 to a_6 to be *detailed*. In this case as well, we think that it is useful for the user to be presented with relevant detailed scenes from other lectures. b_3 corresponds to a_4. The corresponded scene pair [b_3, b_4] in B explains about "pumpkin as a vegetable", and b_4 explains in more detail than b_3. So, b_4 is treated as a supplementary explanation of the selected scenes and can help the user to understand "pumpkin as a vegetable". Therefore, extracting b_4 combined with [a_4, a_6] can help the user to understand his/her topic of interest.

3.4 Evaluation

3.4.1 Prototype System

We have developed a prototype system to support the learning-channel construction engine (see Figure 7) in Microsoft Visual Studio 2008 C#. This prototype implements the determination part and the output part. In the determination part, all semantic relationship types are determined using the video and slide metadata, and scenes corresponding to the basic scene are detected. The corresponded scene pairs are determined by using the semantic relationships and the corresponding scene. The terms in the slides and video are extracted using the morphological analyzer *Mecab* [49], which is in SlothLib [64, 84].

A list of slides and scene numbers is displayed in the input window. The user can select scenes of interest by inputting the scene numbers in the textbox and by checking the semantic relationship type in the list. If the user selects either (1) a scene and a relationship or (2) a scene pair, then slides of semantically related scenes are presented in the output window. The code for controlling the output in other windows is described in the Synchronized Multimedia Integration Language (SMIL).

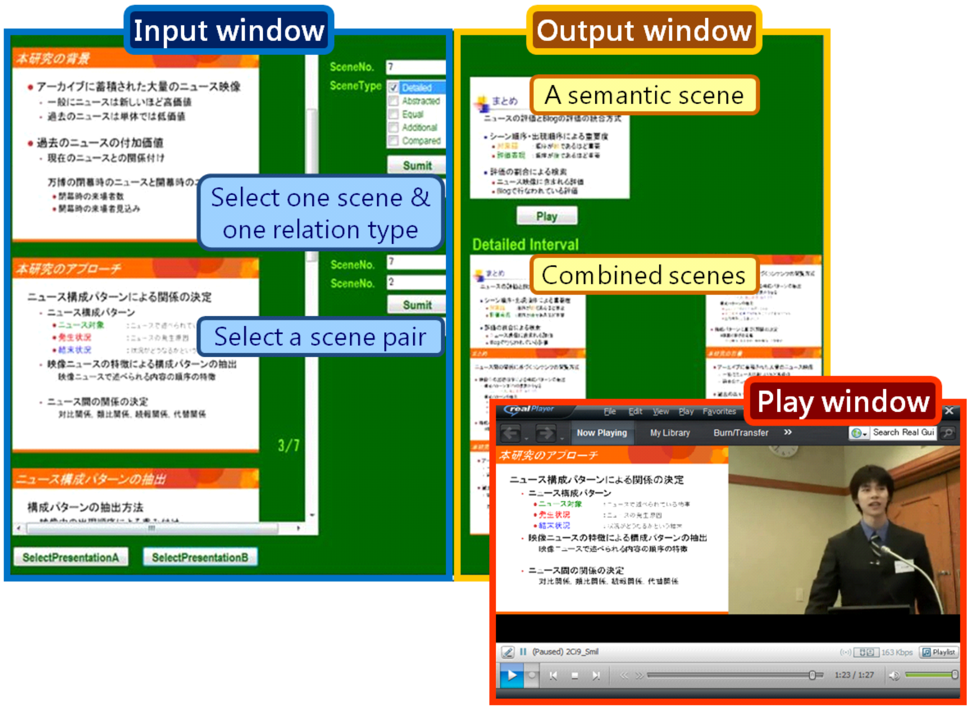


Figure 7: Screenshot of Prototype System

Table 1: Experimental Results

		Results of system				
		detailed	*generalized*	*similar*	*additional*	*others*
Correct answers	*detailed*	**<u>52</u>**	12	5	3	22
	generalized	23	**<u>81</u>**	14	10	42
	similar	13	9	**<u>25</u>**	3	16
	additional	6	7	13	**<u>17</u>**	22
	others	27	24	12	8	48

3.4.2 Experiment 1: Validity of Semantic Relationship Types

There were five participants freely described the relationships between scenes that were extracted by our system and they assessed 115 sets of two scenes sampled at random from 8 real presentation contents [14]. Table 1 shows the results of determine the classification. The vertical shows the results of system, the horizontally shows correct answers were defined by participants. We evaluated the coverage calculated by using the scene set, which was determined to be any scene type identified by our system. The *others* is not able to determine by our system.

This experiment confirmed semantic relationships between scenes could be

Table 2: Results of Semantic Relationships

	detailed	*generalized*	*similar*	*additional*	all
Precision	43.0% (52/121)	60.9% (81/133)	36.2% (25/69)	41.5% (17/41)	48.1% (175/364)
Recall	68.4% (52/76)	63.3% (81/128)	50.0% (25/50)	37.8% (17/45)	58.5% (175/299)
F-measure	0.53	0.62	0.42	0.40	0.53

covered by using the concept of them. We should improve the definitions of *similar* and *additional*, because *similar* includes development relationship and *additional* includes practice relationship. But they were not enough that difficult to define. The results of semantic relationships are listed in Table 2, and they can be explained as follow about the features of the academic contents.

- In case of *detailed*, some of correct answers were no relationship. System extracted more *detailed* than correct answers. Figure 8 shows the example of adequate result. Figure 9 shows the example of inadequate result that the number of keywords are not enough in the video, *detailed* could not be determined by only higher level keywords. We considered that a set of keywords in the slide can define *detailed*. In addition, the precision was higher. We considered a lot of scenes as detailed explain in academic contents.

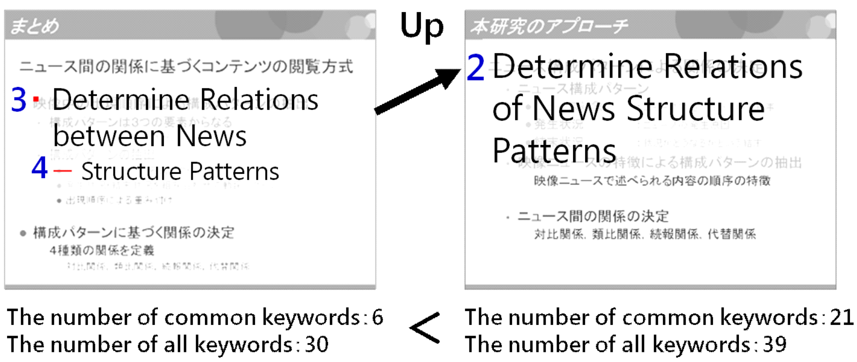


Figure 8: An Adequate Result of *detailed* Relationship

- In case of *generalized*, even if same scene set, participants' answers were difference as "generalize" and "additional". If participants answered "generalize", participants can well understand content. However, participants

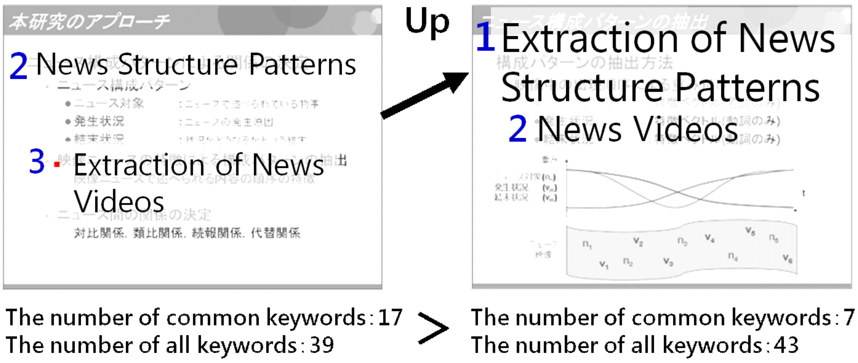


Figure 9: An Inadequate Result of *detailed* Relationship

answered "additional" if participants understudied that scenes have a relationship at a minimum. We considered the *generalized* is effective when the user can understand scenes have a relationship at a minimum, but cannot detect relationship type. Thus, our method can show semantic relationships to understand academic content.

- In case of *similar*, the definitions hard to narrow down decisions by subdividing slides when different levels of a set of keywords are described similar content. In addition, our method extracts fewer results that academic contents have some scenes contain similar content but not more.
- In case of *additional*, a lot of correct answers were not detected any semantic relationships by our method. We considered that the ratio of included keywords and the degree of explain are not in agreement as well as making *detailed*. For example, *Conventional Study(4)* is an additional explain to *Delivery Data*, and system is correct answer. Meanwhile, the same scene set, participants' answers are difference as "additional" and "similar". We considered some keywords have additional explain, but the content of scene is not additional.

Although these results were lower. However, this experiment confirmed that academic contents have some kinds of relationships between scenes. *detailed* and *generalized* might give appropriate definition of using a slide level, and we should enhance to choose method of using the number of keywords in the video. As *similar*, we have to improve the method of using a slide level that different levels of keywords are described similar content. We should relax definition of

Table 3: Results of Semantic Relationships among Combined Scenes

	detailed	*generalized*	*similar*	*additional*	all
Precision	43.8% (14/32)	34.8% (8/23)	55.6% (10/18)	25.0% (1/4)	42.9% (33/77)
Recall	73.7% (14/19)	88.9% (8/9)	83.3% (10/12)	11.1% (1/9)	67.3% (33/49)
F-measure	0.55	0.50	0.67	0.15	0.52

additional and enhance to choose method of keywords set in the slide.

3.4.3 Experiment 2: Validity of Learning Channel Generation

We evaluated the validity of the detected relationship among combined scenes. The participants assessed 50 set of combined scenes from the data set we used in Experiment 1. The results are listed in Table 3.

Although these results were lower. However, this experiment confirmed that users used generated learning channels effectively by semantic relationships among combined scenes. *detailed* and *generalized* among combined scenes might give appropriate definition of using a slide level. For example, [*Conclusion*, *Attached by Order*] is *detailed* in $2C$-$i9$, and *Decision of Relations between News* of $5C$-$i4$ into after two scenes of $2C$-$i9$ to be combined is not detailed explain, but system is detailed explain. We considered that the scene of $5C$-$i4$ into before two scenes of $2C$-$i9$ to be combined is detailed explain. We have to improve the entering position of scenes can be improved the precision. *similar* among combined scenes are useful for participants and we have to improve the determination algorithm of *similar*. As *additional* among combined scenes, most of correct answers are no meaning by our method. We should improve the corresponding scene algorithm that is not to the basic scene, is to the detecting scene.

3.5 Summary

We have proposed a learning-channel construction engine that uses semantic relationships. It automatically generates learning channels to extract scenes and combined scenes from unified contents based on semantic relationships. The type of semantic relationship is determined on the basis of the metadata of structural information, such as indents and texts in slides, and the set of keywords in the text of the speech in the video. Thus, users use the learning-channel engine to look for scenes that have relevant points to the scene of interest or appropriate combined scenes from unified contents. This approach is very effective. We have also developed a prototype system and evaluated it using actual presentation data. We confirmed an improvement in the coverage of semantic relationship types and their definition and in the detection of corresponding scenes for extracting corresponding scene pairs effectively by using the semantic relationships. In the future, our method could extend the range of available educational materials that would be useful if other related content, such as related papers, graphics, and Internet content, were also unified.

Chapter 4 Semantic Ranking and Snippet Generation based on Document Structure and Keyword Semantic Relations

4.1 Introduction

Presentation slides (e.g., PowerPoint, Keynote) are now one of the most frequently used tools for educational purposes. A considerable amount of slide-based lecture material, often prepared from teaching material used in actual classes at universities or other educational institutions, is freely shared on Web sites such as SlideShare [83] and edubase Stream [17]. In particular, students can view lectures on their iPhone or iPad by using MPMeister [57], which has hosted presentation slides and recorded lecture videos from Kyoto University since 2010 [62] (see Figure 10). Other online e-learning material archives include those of the Nara Institute of Science Technology [61], which has provided presentation content recorded from lectures for about seven years, and the Database Society of Japan (DBSJ) [14], which stores 1200 presentations from workshops (DEWS and DEIM) for members of the society. These presentations provide varying levels of knowledge, and are useful and valuable to students. Thus, content can be reviewed and studied alone and when convenient, not only by students who missed a lecture or presentation, but also by anyone interested in the topic.

Currently, self-learners (e.g., students) must formulate a query consisting of proper keywords in order to retrieve the required lecture slides. However, e-learning material provides varying levels of knowledge associated with the various levels of university courses or seminars, and so many presentation slides will require prior knowledge and expertise. Moreover, if the keywords in the query are common, the large number of search results returned will make it difficult for self-learners to find material appropriate to their level of understanding. This current method does not consider the relevance of the information contained in slides returned by the query, so it is impossible for students to easily determine

Figure 10: Captured Image of a Student Studying Lecture Material on an iPad

which of the slides retrieved by the query are appropriate for study.

We present a novel retrieval system for retrieving slides to meet user requirements for presentations containing different levels of knowledge. We discuss how to present the retrieval results to users by considering what rank orders of the slides related to a user query and what portions of the slides are relevant to the query, on the basis of the relationships between slides that enable the browsing of slide retrieval at the conceptual level. To achieve our goal, we analyzed the implicit semantic relations between keywords, and how the keywords at different indent levels of slides are related to a user query. We derived keyword conceptual structure focusing on 'is-a' and 'part-of' relations between keywords extracted from the slide text. However, the usage of keywords in slides varies depending on the author. We derived document structure by focusing on certain features of the slides, such as the levels of indents in the slide text, as these are often used to help users to better understand the content in slides. It was then necessary to use the semantic relations and document structure to determine the portions of slides related to the user query; furthermore, we detected the relationships between slides in terms of the query.

In this theme, we discuss how to help users understand the context of slides so they can select appropriate ones for self-learning purposes from retrieval results. We aim to build a retrieval system in e-Learning for retrieving lecture slides using the relationships between slides (see Figure 11). This retrieval

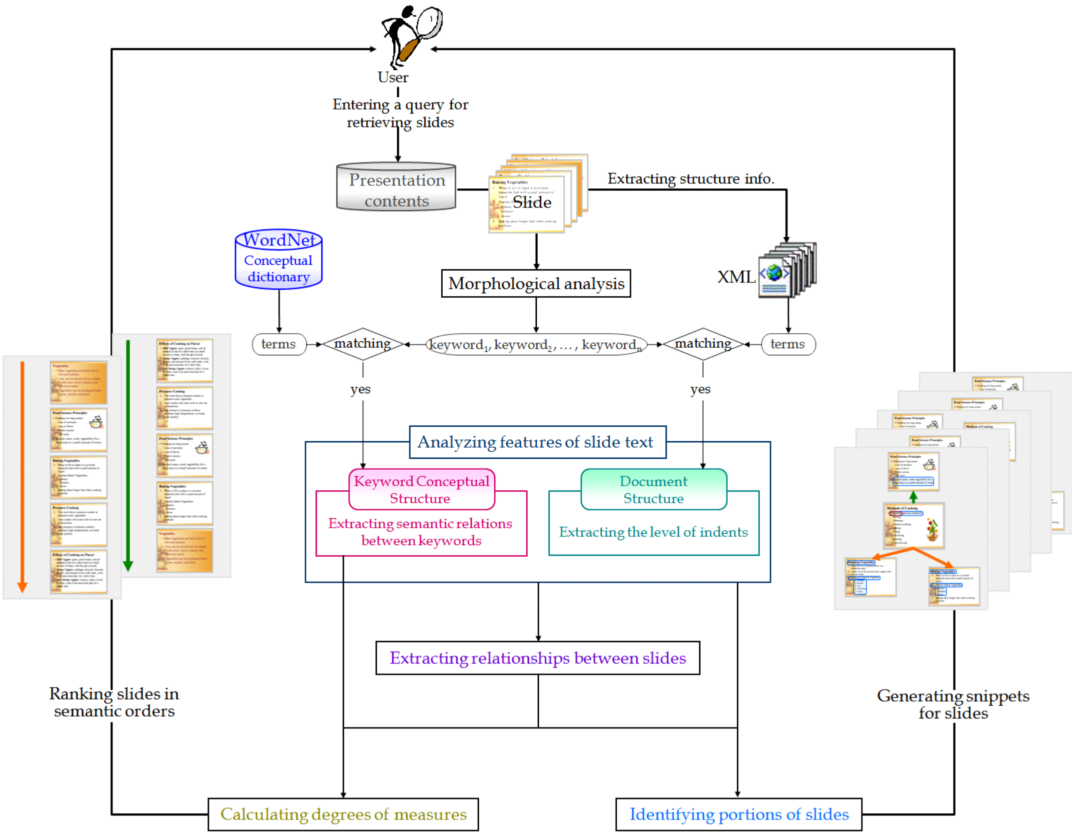


Figure 11: Concept of a Retrieval System for Presentation Slides

system can be implemented by

- Semantic ranking of target slides
- Snippet generation of target slides

For this, we have to consider a method for ranking slides and for generating snippets for slides, not only precise retrieve slides but also the semantic ranking of them, thus ranking either highly detailed slides or generalized slides in an order related to a user query to help users easily learn through slides; and the relevant portions of them in the presentation by focusing on portions from either detailed or generalized slides, thus giving their surrounding context to help users easily determine which slides to learn are useful or not. As the follows we describe our approach and present the concepts of semantic ranking and Slide KWIC.

4.1.1 Semantic Ranking

A conventional retrieval method ranks the retrieved slides in an order by a user query with a high frequency. Through this slide ranking, users cannot understand what the relevance of slides and the query, and how to browse this slide ranking. We propose a semantic ranking method of slides from presentation content with differential base by analyzing conceptual level of slides [96]. In this work, our goal is to use semantic relationships for ranking slides without relying on the existence of any specific structure in a slide or relevant information between slides. Instead, the relevance of slides is determined using relationships that are known to exist between slides by using populated semantic relations between the keywords as ontology and a specific document structure in a slide. We introduce two measures of relevance that are based on the semantics of relationships that link slides. Therefore, our slide ranking method calculates the degrees of measures for ranking slides on the basis of the relationships between slides related to the query.

4.1.2 Slide KWIC

A traditional snippet used to obtain a retrieval result consists of a portion of the retrieval result containing the user query with its surrounding text. We propose Slide KWIC (means keyword-in-context) that snippet generation for the focused slides with their surrounding context, which helps users understand them in presentation content easily. A snippet for the target slide, which we call the focused slide, is shown in Figure 12; There are three layers: the basic layer is the focused slide, the high layer is a generalized slide of the focused slide, and the low layer is a detailed slide of the focused slide. We then generate a snippet consisting of a captured portion of the focused slide with the relevant portions of the related slides in terms of the query, helping users to understand the presentation content in the focused slide. As mentioned above, the semantic relations and document structure can then be used to identify portions of sentences for a given indent level relevant to the query in the focused slide, along with the relevant portions from other slides.

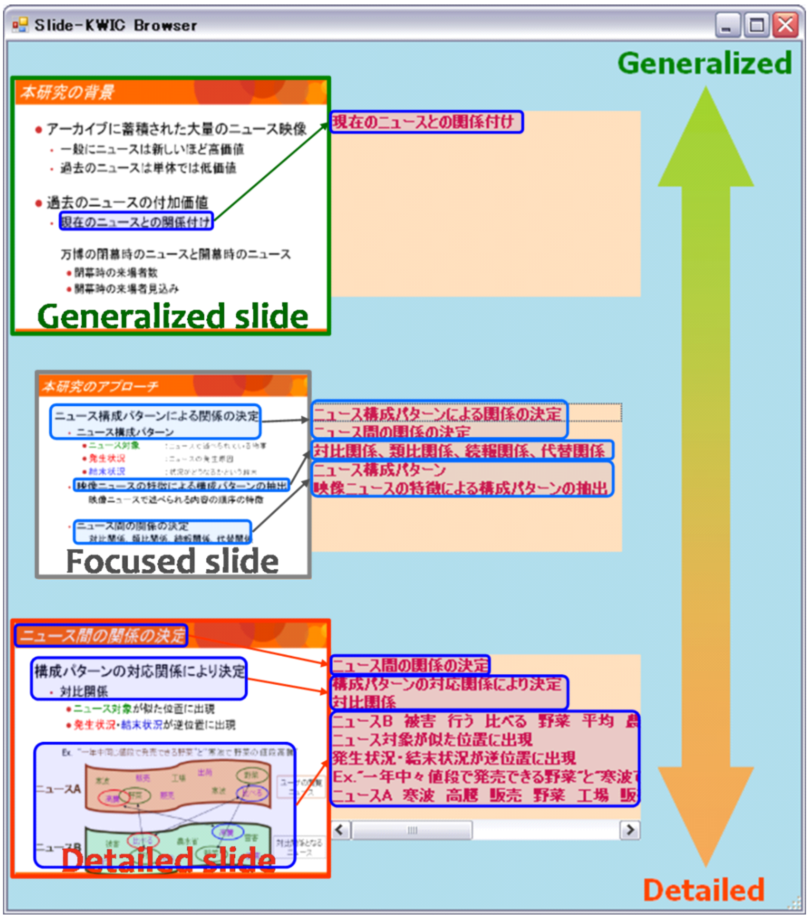


Figure 12: Screenshot of a Snippet Generated by Slide KWIC for Focused Slide

4.2 Determination of Relationship Types

To determine the relationship between two slides, we define one as a focused slide, and consider the other to be conceptually related to the focused slide through one of two types of relationships: *detailed* and *generalized*. In other words, the relationship has a direction. The focused slide is the starting point, and the other slides are end points, of the direction of the relationship. In the example shown in Figure 13, slide x is the focused slide, and the relationship from slide x to slide y is a *detailed* relationship. If a slide has a *detailed* relationship with the focused slide, it is called a *detailed* slide. If a slide has a *generalized* relationship with the focused slide, it is called a *generalized* slide. Let x be the slide number of a focused slide and y be the slide number of the slide that we want to retrieve. The relationship types are determined for all slides containing the keyword q from a user query.

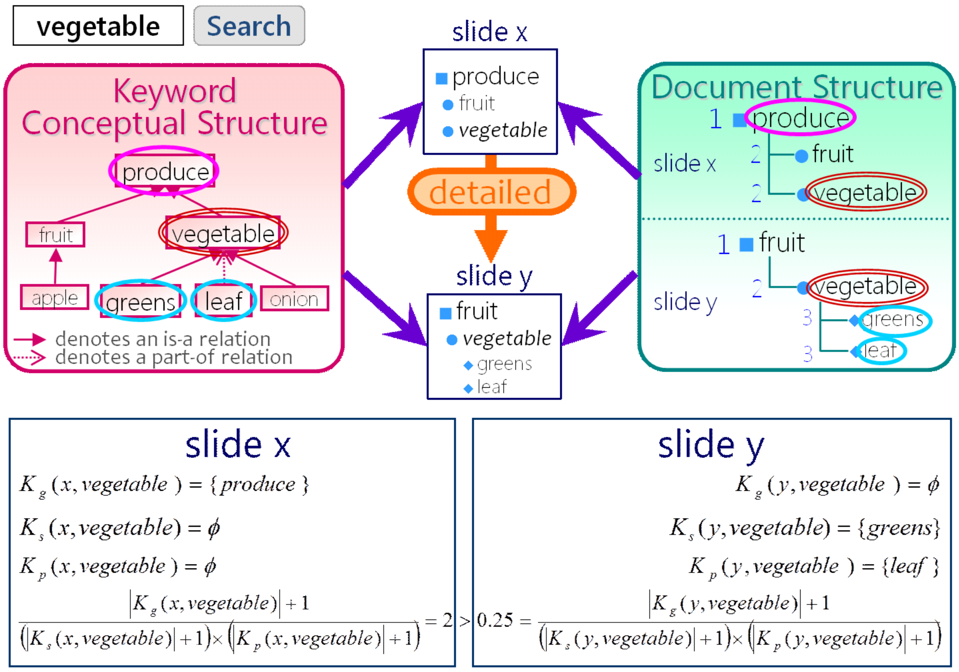


Figure 13: An Example of a *detailed* Relationship between Two Slides

4.2.1 Keyword Conceptual Structure and Document Structure

The content of one presentation contains thumbnails (images) of slides and their text information. We consider semantic relations exist implicitly between keywords extracted from the slide text. For example, when the keyword "fruit" is included in the user query "kinds of fruit," a semantic relationship is assumed to exist between the keyword "fruit" and other keywords in the slide text; for example, another keyword "apple" describes a specialized explanation of the keyword "fruit." Furthermore, other keywords such as "pulp" and "peel" also give explanations of the keyword "fruit." Therefore, various semantic relations such as is-a and part-of [54, 53] are used as a basis for the most common semantic relations between keywords. "X subsumes Y, or Y is-subsumed-by X" (Y is-a X) usually means that concept Y is a specialization of concept X and that concept X is a generalization of concept Y. Moreover, "Z is part of X, or X has Z as a part of itself" (Z is part-of X) usually means that Z is a meronym of X and that the whole X has Z as a part. For example, "fruit" is a generalization of an “apple,” “orange,” and many other fruits; in other words, an “apple” is-a “fruit.” Furthermore, “fruit” is a holonym of “pulp,” “peel,” and many other meronyms; in other words, “pulp” is a part-of “fruit.” Therefore, we define the keyword conceptual structure as consisting of an is-a or part-of relation between keywords extracted using WordNet [29, 99].

We define a document structure as a slide that appears in the outline pane, on the basis of indents in the slide text extracted from the Office Open XML in Microsoft Office 2007 (see Figure 14). We defined the slide title (first level indent) as the upper level. The first item of text is considered to be on the second level, and the depth of the sub-items increases with the level of indentation (third level, fourth level, etc). Objects outside of the text, such as figures or tables, are considered to be at the same indent level as the text in which they are placed. If a given keyword appears in the title of the slide or in lines with smaller indents, we implicitly assume that the lower-level indented keywords are supplementary and that they explain the upper-level keywords.

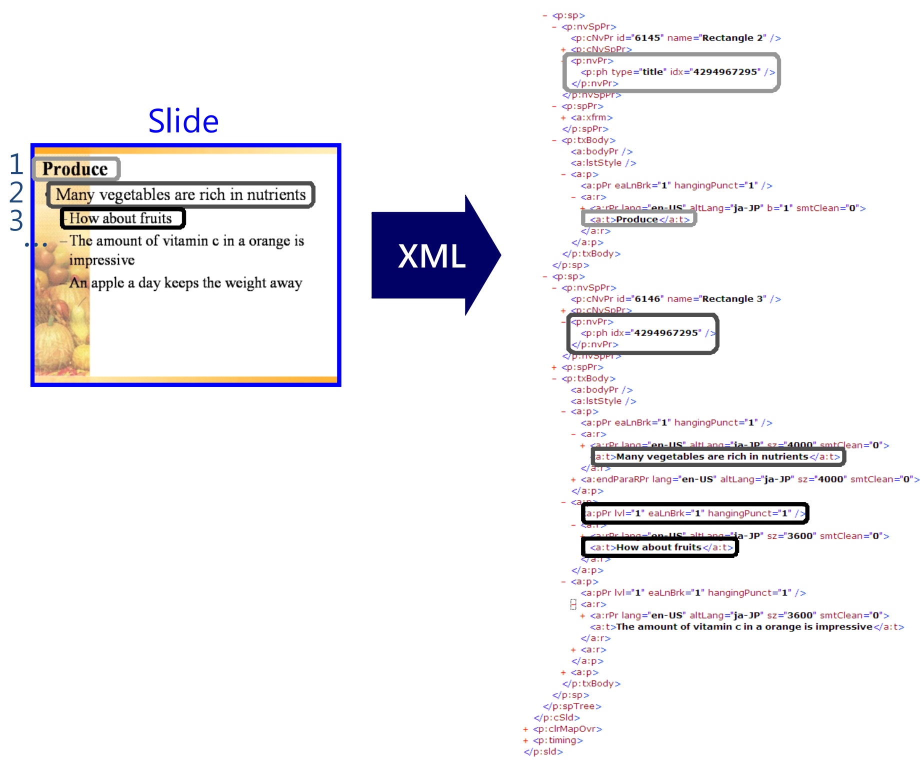


Figure 14: An Example of Practical Analysis of Document Structure

4.2.2 Preliminary Experiment: Usage Tendency of Document Structure

We conducted a preliminary experiment to confirm the usage tendency of this document structure, using presentation slides from the DBSJ Archives [14], HandsOut [23], and SlideShare [83] as experimental data. We extracted 50 slides from each site, giving a total of 150 slides, and analyzed the document structure and presentation category. The results are summarized as follows:

- Slides often used levels of indentation in the categories of academic, educational, and business content.
- This structure was used in 94.0%, 91.7%, and 100% of academic, educational, and business presentations, respectively.
- In 25% of the experimental presentations, no text indent levels were used, as the presentations contained only visual elements, such as pictures or videos.

In this experiment, we confirmed that document structure is used in academic, educational, and business slide presentations. Therefore, we considered semantic relationship types to adapt to the presentation content based on a method involving the document structure.

4.2.3 Determination of Detailed Relationships

If a slide has more information about a user query than the focused slide, its relationship with the focused slide is *detailed*. We explain the determination of *detailed* slides using the query keyword q, present in both the focused slide x and slide y (slide y as the *detailed* slide needs to be retrieved). As an example, Figure 14 shows the determination of the *detailed* relationship between slides x and y for a query on the word "vegetable."

When the query keyword q and other keywords in slides x and y satisfy certain conditions, slide y is determined to be the *detailed* slide of slide x. This is because q has more specific content in slide y than it does in slide x.

$$K_g(x,q) = \{k_i \mid k_i \in x, l(x,q) \geq l(x,k_i), q \text{ is-a } k_i\} \tag{16}$$

$$K_s(x,q) = \{k_j \mid k_j \in x, l(x,q) < l(x,k_j), k_j \text{ is-a } q\} \tag{17}$$

$$K_p(x,q) = \{k_m \mid k_m \in x, l(x,q) < l(x,k_m), k_m \text{ part-of } q\} \tag{18}$$

Here, $K_g(x,q)$ is a set of keywords that can be considered as general information in terms of q in slide x. In Eq. (16), $l(x,q)$ is a function that returns the level of indentation of q in slide x, and will thus return a value greater than 1. When q appears frequently in slide x, $l(x,q)$ will return the lowest possible value; that is, the uppermost level at which q occurs in slide x. This is because we consider that when q appears in an upper level, all of the other levels in which q appears in the body of that slide are explanatory points related to the upper level occurrence of q. Keyword k_i is included in the levels that have a hierarchical relationship with the level of q, and k_i belongs to the set of keywords $K_g(x,q)$ in slide x. $l(x,k_i)$ is less than or equal to $l(x,q)$ in the document structure, and q (e.g., "vegetable") has an is-a relation with k_i (e.g., "produce") in the keyword conceptual structure (see Figure 14). When k_i does not exist in slide x, $K_g(x,q)$ will be empty. In our method, the keyword conceptual structure is extracted as a tree-shaped structure. In general, an is-a or part-of relation between keywords is equivalent to a parent-child relation, and so our method may classify is-a or part-of as a descendant relation. $K_s(x,q)$ is a set of keywords that can be considered as specific information in terms of q in slide x.

In Eq. (17), keyword k_j is included in the levels that have a hierarchical relationship with the level of q, and k_j belongs to the set of keywords $K_s(x,q)$ in slide x. $l(x,k_j)$ is greater than $l(x,q)$ in the document structure, and k_j (e.g., "greens") has an is-a relation with q (e.g., "vegetable") in the keyword conceptual structure (see Figure 14). When k_j does not exist in slide x, $K_s(x,q)$ will be empty. $K_p(x,q)$ is the set of keywords that can be considered as additional information in terms of q in slide x.

In Eq. (18), keyword k_m is included in the levels that have a hierarchical relationship with the level of q, and k_m belongs to the set of keywords $K_p(x,q)$ in slide x. $l(x,k_m)$ is greater than $l(x,q)$ in the document structure, and k_m (e.g., "leaf") has a part-of relation with q (e.g., "vegetable") in the keyword conceptual structure (see Figure 14). When k_m does not exist in slide x, $K_p(x,q)$ will be empty. For the conditions mentioned above, when k_i, k_j, or k_m does not exist in slide x, then $K_g(x,q)$, $K_s(x,q)$, or $K_p(x,q)$, respectively, will be empty.

In general, detailed information means a more specific explanation of a term; a *detailed* relationship seems to be a mixture of is-a and part-of relations.

Based on the above criteria, we compute the ratio of general information to detailed information related to q for slides x and y, and compare their ratios using the following formula:

$$\frac{|K_g(x,q)|+1}{(|K_s(x,q)|+1)\times(|K_p(x,q)|+1)} > \frac{|K_g(y,q)|+1}{(|K_s(y,q)|+1)\times(|K_p(y,q)|+1)} \tag{19}$$

where the function $|K_g(x,q)|$ extracts the total number of k_i in $K_g(x,q)$, $|K_s(x,q)|$ extracts the total number of k_j in $K_s(x,q)$, and $|K_p(x,q)|$ extracts the total number of k_m in $K_p(x,q)$ in slide x. $K_g(y,q)$, $K_s(y,q)$, and $K_p(y,q)$ are also sets of keywords in slide y, satisfying the same conditions as $K_g(x,q)$ in Eq. (16), $K_s(x,q)$ in Eq. (17), and $K_p(x,q)$ in Eq. (18). Thus, Eq. (19) can be used to calculate the ratio of $|K_g(x,q)|$ to $|K_s(x,q)|$ and $|K_p(x,q)|$ for slide x and the ratio of $|K_g(y,q)|$ to $|K_s(y,q)|$ and $|K_p(y,q)|$ for slide y.

If the ratio calculated for slide x is higher than that calculated for slide y using Eq. (19), slide y is determined to be the *detailed* slide of slide x with regard to q.

4.2.4 Determination of Generalized Relationships

If a slide contains content about the query in the outline given in a *generalized* slide, it is described in relation to the focused slide. We explain the determination of *generalized* slides using the query keyword q present in the focused slide x and slide y; this keyword needs to be retrieved.

$$\frac{|K_g(x,q)|+1}{(|K_s(x,q)|+1)\times(|K_p(x,q)|+1)} < \frac{|K_g(y,q)|+1}{(|K_s(y,q)|+1)\times(|K_p(y,q)|+1)} \tag{20}$$

When the query keyword q and other keywords in slides x and y satisfy Eqs. (16), (17), (18), and (20), then slide y is determined to be a *generalized* slide of slide x with regard to q. This is because slide y has more general content on q than does slide x. Eq. (20) can be used to calculate the ratio of $|K_g(x,q)|$ to $|K_s(x,q)|$ and $|K_p(x,q)|$ for slide x and the ratio of $|K_g(y,q)|$ to $|K_s(y,q)|$ and $|K_p(y,q)|$ for slide y.

Thus, *detailed* and *generalized* slides are functionally interchangeable, whereas a focused slide is a *generalized* slide from the viewpoint of a *detailed* slide.

4.3 Slide Ranking and Snippet Generation

Our proposed method retrieves slides by determining the relationships between slides about a user query. It is also difficult for users to understand relevant information between the retrieved slides in terms of the user query. Moreover, we consider the users in the different levels of understanding that they have different desires on the retrieved slides in terms of the user query. Our method provides two types of semantic rankings that focus on two measures, namely, $DETAIL$ and $GENERALITY$.

To generate snippets, Slide KWIC takes the portions of the focused slides relevant to a user query by using the relationships between slides. It is difficult for users to understand the relevant information between portions of slides in terms of the query. For example, a user may want to study slide 4 to further understand "vegetable" in the lecture content about Vegetable as food. Our method generates a snippet for slide 4 that captures portion P_4 of slide 4, along with portion P_2 of slide 2 that includes text on the indent levels, explaining "produce" with regard to "vegetable." Portions P_3 and P_5 include text on the indent levels, explaining "cabbage and spinach are green vegetables," with regard to "vegetable" for slides 3 and 5 (see Figure 15). In this case, slide 2 explains that "produce" has a *generalized* relationship with slide 4 with regard to the keyword "vegetable," and slides 3 and 5 both explain that "cabbage and spinach are green vegetables," implying a *detailed* relationship with slide 4 in terms of "vegetable." When the user browses the snippet for slide 4, consisting of portion P_4 from slide 4 and portions P_2, P_3, and P_5 from slides 2, 3, and 5, respectively, he or she is provided with more information on "vegetable" than just that in slide 4, and this enables the user to further his or her understanding easily. Therefore, our snippet-generation method is based on the context of slides to present snippets, which contain portions related to the user query in a detailed order to enable snippet comprehension at the conceptual level. This section describes how to calculate the degrees of these two measures for ranking slides and how to generate snippets, based on the relationships between slides related to the query.

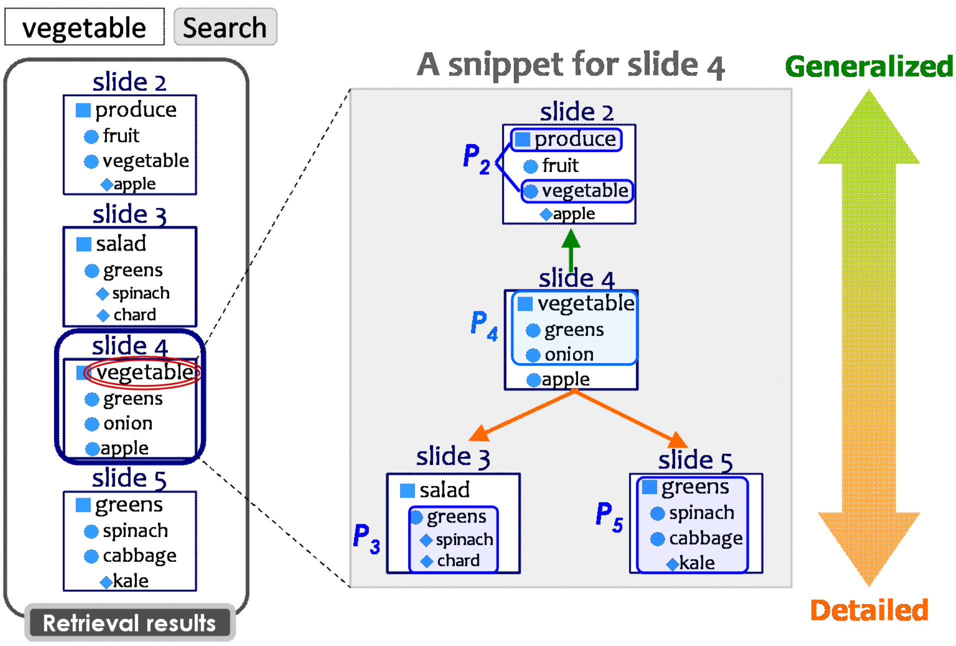


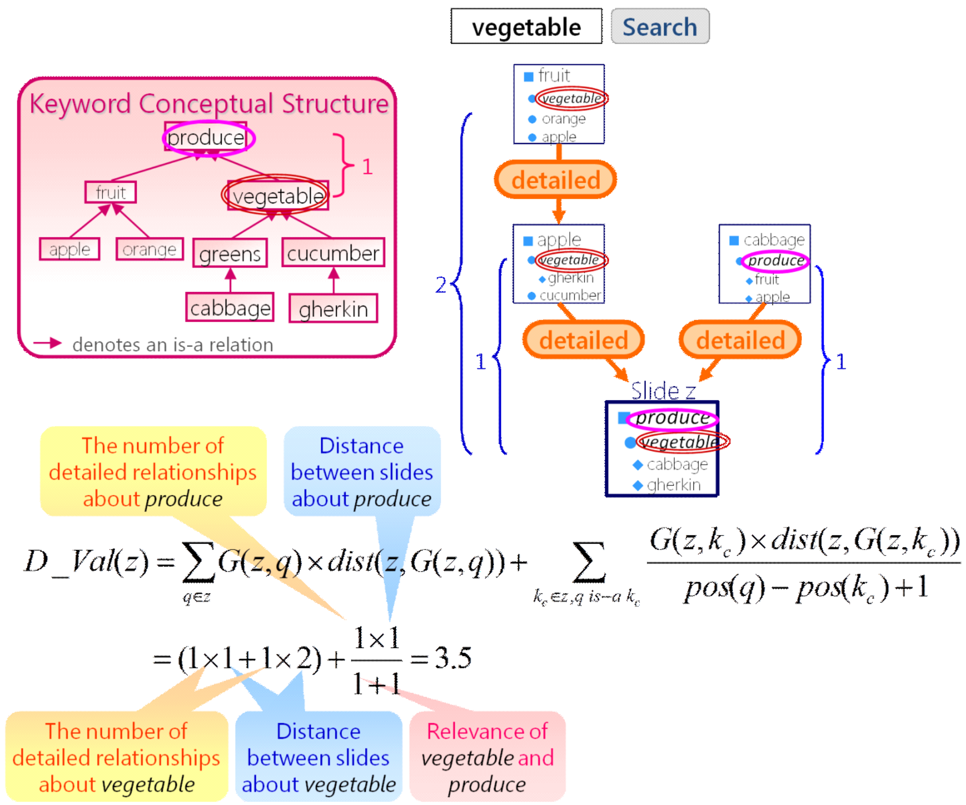
Figure 15: An Example of Snippet Generation

4.3.1 Slide Ranking based on the Measure of DETAIL

In an order of retrieved slides providing detailed information in terms of a user query, the user must have a deep understanding of the desired slides related to the query. Then, slide ranking by using the measure of $DETAIL$ can aid the user to understand the query with a detailed explanation well. If a slide provides detailed information regarding a query as compared to that provided by other slides, this slide is known as a specific slide, and it provides specific explanation about other slides with a high degree of $DETAIL$. As shown in Figure 16, slide z has a detailed relationship with other slides in terms of the content on "produce" and "vegetable" related to the query keyword "vegetable."

We consider the function of the degree of $DETAIL$ using the following indicators.

- The number of the target slide x has a detailed relationship with the generalized slide $G(x, q)$ with regard to the query keyword q.
- The generalized keyword k_c of q (means q is-a k_c) in x is extracted from the conceptual structure.
- The relevance of k_c and q is expressed in terms of the distance between the



$$D\_Val(z) = \sum_{q \in z} G(z,q) \times dist(z, G(z,q)) + \sum_{k_c \in z, q \text{ is-a } k_c} \frac{G(z,k_c) \times dist(z, G(z,k_c))}{pos(q) - pos(k_c) + 1}$$

$$= (1 \times 1 + 1 \times 2) + \frac{1 \times 1}{1+1} = 3.5$$

Figure 16: An Example of a Slide with a High Degree of DETAIL

position of k_c and q in the keyword conceptual structure.

- The distance between x and $G(x,q)$ for q indicates the number of detailed relationships existing between x and $G(x,q)$; the distance between x and $G(x,k_c)$ for k_c indicates the number of detailed relationships existing between x and $G(x,k_c)$.

We described these indicators as follows. If x has a detailed relation with $G(x,k_c)$ with regard to k_c, we can say that x includes detailed specific information regarding q. If the distance between the position of k_c and q is short in the keyword conceptual structure, then the intensity of k_c and q is high such that the value of relevance of x and $G(x,k_c)$ is high. Further, if the number of $G(x,q)$ and $G(x,k_c)$ is large and the number of detailed relationships between x and $G(x,q)$ with regard to q or that between x and $G(x,k_c)$ with regard to k_c is large, then the distance between them is long such that the value of $DETAIL$ of x is high.

The function of the degree of DETAIL is expressed as

$$D\_Val(x) = \sum_{q \in x} G(x,q) \times dist(x, G(x,q)) + \sum_{k_c \in x,\ q\ is-a\ k_c} \frac{G(x,k_c) \times dist(x, G(x,k_c))}{pos(q) - pos(k_c) + 1} \quad (21)$$

where $G(x,q)$ that extracts the number of x has a detailed relationship of $G(x,q)$ with regard to q. Further, $dist(x, G(x,q))$ is the distance between x and $G(x,q)$ with regard to q that extracts the number of detailed relationships between x and $G(x,q)$. It should be noted that $G(x,k_c)$ that extracts the number of x has a detailed relationship of $G(x,k_c)$ about k_c with regard to q. The function $pos(q) - pos(k_c) + 1$ is the relevance of k_c and q that extracts the distance between the position of k_c and q in the keyword conceptual structure. Further, the function $dist(x, G(x,k_c)$ extracts the distance between x and $G(x,k_c)$ in terms of the number of detailed relationships between x and $G(x,k_c)$. Then, the function calculates the relevance of x and $G(x,q)$ with regard to q, and the relevance of x and $G(x,k_c)$ for k_c with regard to q. If $G(x,q)$ and $dist(x, G(x,q))$ are large, the value of detail between x and $G(x,q)$ is high. If $pos(q)-pos(k_c)+1$ is small, the value of relevance of k_c and q is high. If $G(x,k_c)$ and $dist(x, G(x,k_c)$ are large, the degree of $DETAIL$ of x is high.

4.3.2 Slide Ranking based on the Measure of GENERALITY

In an order of retrieved slides providing general information in terms of a user query, the user must easily grasping the general-content of the desired slides related to the query. Then, slide ranking on the basis of the measure of $GENERALITY$ can aid the user to obtain a generalized explanation about the query, easily. If a slide provides general information regarding a query as compared to that provided by other slides, this slide is known as a general slide, and it provides explanation about other slides with a high degree of $GENERALITY$. As shown in Figure 17, slide w has a generalized relationship with other slides in terms of the content on "vegetable" and "greens" related to the query keyword "vegetable."

We consider the function of the degree of $GENERALITY$ using the following indicators.

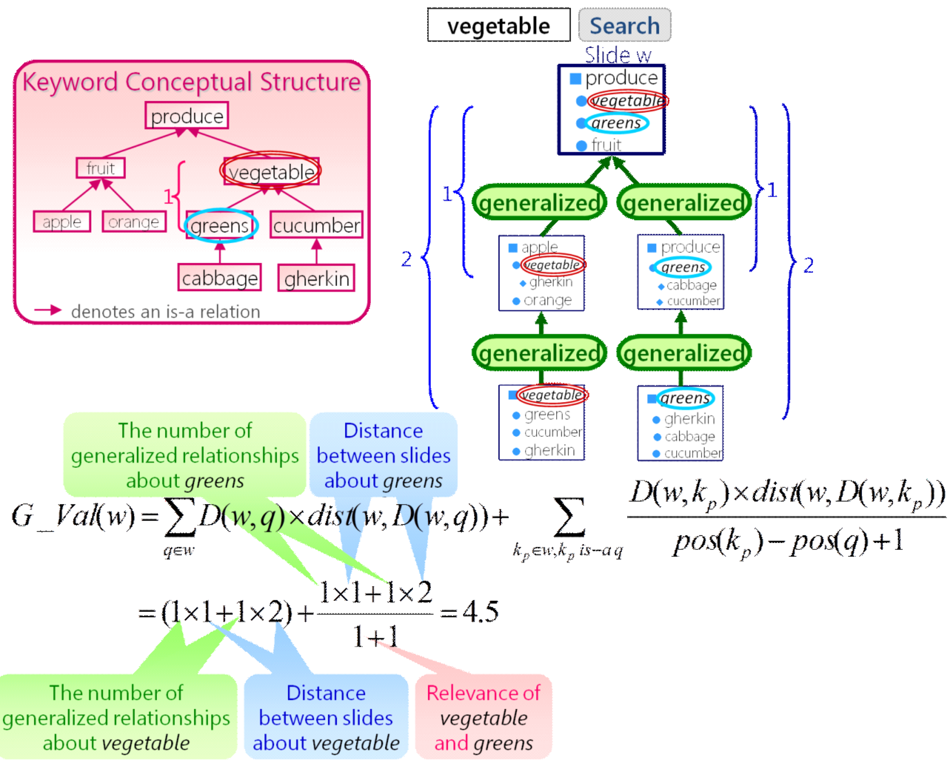


Figure 17: An Example of a Slide with a High Degree of GENERALITY

- The number of target slide x has a generalized relationship with the detailed slide $D(x, q)$ with regard to the query q.
- The specified keyword k_p of q (means k_p is-a q) in x is extracted from the keyword conceptual structure.
- The relevance of k_p and q is expressed in terms of the distance between the position of k_p and q in the keyword conceptual structure.
- The distance between x and $D(x, q)$ for q indicates the number of generalized relationships existing between x and $D(x, q)$; the distance between x and $D(x, k_p)$ for k_p indicates the number of generalized relationships existing between x and $D(x, k_p)$.

We described these indicators as follows. If x has a generalized relation of $D(x, k_p)$ with regard to k_p, we can say that x includes general information regarding q. If the distance between the position of k_p and q is short in the keyword conceptual structure, then the relevance of k_p and q is high such that the value of relevance of x and $D(x, k_p)$ is high. Further, if the number of $D(x, q)$

and $D(x, k_p)$ is large and the number of generalized relationships between x and $D(x, q)$ with regard to q or that between x and $D(x, k_p)$ with regard to k_p is large, then the distance between them is long such that the value of $GENERALITY$ of x is high.

The function of the degree of $GENERALITY$ is expressed as

$$\begin{aligned} G\_Val(x) &= \sum_{q \in x} D(x, q) \times dist(x, D(x, q)) + \\ &\sum_{k_p \in x,\ k_p\ is-a\ q} \frac{D(x, k_p) \times dist(x, D(x, k_p))}{pos(k_p) - pos(q) + 1} \end{aligned} \tag{22}$$

where $D(x, q)$ that extracts the number of x has a generalized relationship of $D(x, q)$ with regard to q. Further, $dist(x, D(x, q))$ is the distance between x and $D(x, q)$ with regard to q that extracts the number of generalized relationships between x and $D(x, q)$. It should be noted that $D(x, k_p)$ that extracts the number of x has a generalized relationship of $D(x, k_p)$ about k_p with regard to q. The function $pos(k_p) - pos(q) + 1$ is the relevance of k_p and q that extracts the distance between the position of k_p and q in the keyword conceptual structure. Further, the function $dist(x, D(x, k_p)$ extracts the distance between x and $D(x, k_p)$ in terms of the number of generalized relationships between x and $D(x, k_p)$. Then, the function calculates the relevance of x and $D(x, q)$ with regard to q, and the function calculates the relevance of x and $D(x, k_p)$ about k_p with regard to q. If $D(x, q)$ and $dist(x, D(x, q))$ are large, the value of generality between x and $D(x, q)$ is high. If $pos(k_p) - pos(q) + 1$ is small, the value of the relevance of k_p and q is high. If $D(x, k_p)$ and $dist(x, D(x, k_p)$ are large, the degree of $GENERALITY$ of x is high.

As can be seen, our method for retrieving users' desired slides and ranking slides into two types focuses on different measures can satisfy users' demands.

4.3.3 Identifying the Portions of Focused Slides

Although our method can retrieve slides related to a user query, the relevance of the information contained on the focused slides must be determined. Therefore, our method first identifies the portions of the focused slide related to a user query based on the keyword conceptual structure and document structure. Let

x be the slide number of the focused slide. When the query keyword q and other keywords in slide x satisfy Eqs. (16), (17), (18), (23), (24), (25), and (26), portion P of slide x is determined to be related to the query keyword q.

$$K_w(x,q) = \{k_h \mid k_h \in x, l(x,q) \geq l(x,k_h), q \text{ part-of } k_h\} \quad (23)$$

$$L_g(x,q) = \{s_n \mid l(x,k_u) \leq l(x,s_n) \leq l(x,q), k_u \in K_g(x,q) \cup K_w(x,q)\} \quad (24)$$

$$L_s(x,q) = \{s_t \mid l(x,q) \leq l(x,s_t) \leq l(x,k_v), k_v \in K_s(x,q) \cup K_p(x,q)\} \quad (25)$$

$$P = L_g(x,q) \cup L_s(x,q) \quad (26)$$

Here, $K_w(x,q)$ is a set of keywords that can be considered as a whole concept in terms of q in slide x. In Eq. (23), keyword k_h is included in the levels that have a hierarchical relationship with the level of q, and k_h belongs to the set of keywords $K_w(x,q)$ in slide x; $l(x,k_h)$ is less than or equal to $l(x,q)$ in the document structure, and q (e.g., "vegetable") has a part-of relation with k_h (e.g., "leaf") in the keyword conceptual structure (see Figure 13). When k_h does not exist in slide x, $K_w(x,q)$ will be empty. A set $L_g(x,q)$ consists of sentences from the levels that contain general information related to q in slide x. Sentence s_n belongs to the set of sentences $L_g(x,q)$ in slide x if the following condition is satisfied: s_n must be included in one of the indent levels ranging from the level of the sentence containing q to the level of the sentence containing keyword k_u, where k_u belongs to $K_g(x,q)$ or $K_w(x,q)$, and q is-a k_u or q is part-of k_u in slide x. The selection and extraction of the s_n is performed according to Eq. (16) or Eq. (23).

In Eq. (24), $l(x,s_n)$ is not greater than $l(x,q)$ in the document structure, so $L_g(x,q)$ will extract the sentences, s_n, containing q in levels ranging from $l(x,q)$ to $l(x,s_n)$. In addition, $l(x,s_n)$ is greater than or equal to $l(x,k_u)$ in the document structure, so $L_g(x,q)$ will also extract sentences containing k_u in levels ranging from $l(x,s_n)$ to $l(x,k_u)$. A set $L_s(x,q)$ consists of sentences from levels that contain specific information related to q in slide x. Sentence s_t belongs to the set of sentences $L_s(x,q)$ in slide x, where s_t is included in the indent levels of sentences from the level of the sentence containing q to the level of sentence containing k_v. The keyword k_v, which has an is-a or part-of relation with q, belongs to $K_s(x,q)$ or $K_p(x,q)$. This extraction is performed using Eqs. (17) or

(18). In Eq. (25), $l(x, s_t)$ is not greater than $l(x, k_v)$ in the document structure, so $L_s(x, q)$ will extract sentences containing q in levels ranging from $l(x, k_v)$ to $l(x, s_t)$. As $l(x, s_t)$ is greater than or equal to $l(x, q)$ in the document structure, $L_s(x, q)$ will also extract sentences containing q in levels from $l(x, s_t)$ to $l(x, q)$. Thus, Eq. (26) can be used to extract a portion P of slide x, and thus combine the sets of sentences from different levels, $L_g(x, q)$ and $L_s(x, q)$.

4.3.4 Determining the Relevant Portions of Related Slides

When slide x_g is a *generalized* slide that has a *generalized* relationship with the focused slide x, related to query keyword q, portion P_g of slide x_g provides the general content of portion P of the focused slide x related to q. Therefore, portion P_g of the generalized slide x_g is determined using the query keyword q from the focused slide x.

Portions of Generalized Slides

When slide x_g is a *generalized* slide that has a *generalized* relationship with the focused slide x, related to query keyword q, portion P_g of slide x_g provides the general content of portion P of the focused slide x related to q. Therefore, portion P_g of the generalized slide x_g is determined using the query keyword q from the focused slide x.

$$P_g = L_g(x_g, k_u) \cup L_g(x_g, q) \tag{27}$$

When the query keyword q in slide x_g satisfies Eqs. (16), (21), (24), and (27), then portion P_g of the generalized slide x_g is determined. This is because the amount of content in slide x_g that is generic to q is greater than that in slide x. A set $L_g(x_g, q)$ consists of sentences from levels that contain general information related to q in slide x_g, and satisfies the same conditions as the set $L_g(x, q)$ (these conditions apply to slide x and are given by Eq. (24)). In addition, when slide x_g contains the keyword k_u, which belongs to $K_g(x, q)$ or $K_w(x, q)$, then a set $L_g(x_g, k_u)$ is used to extract a further set of sentences. These come from levels that provide general information in terms of k_u, the more generalized concept related to q in slide x_g, and satisfy the same conditions as the set $L_g(x, q)$ (these conditions apply to slide x and are given by Eq. (24)). When slide x_g contains two or more k_u, as determined from the focused slide x, then we can extract two or more sets of sentences from $L_g(x_g, k_u)$. Thus, Eq. (27) can be used to determine the portion P_g of slide x_g that combines the sets of sentences from $L_g(x_g, k_u)$ and $L_g(x_g, q)$.

Portions of Detailed Slides

When slide x_d is a *detailed* slide that has a *detailed* relationship with the focused slide x in respect of the query keyword q, portion P_d of slide x_d provides specific, detailed information about portion P of the focused slide x related to q. Therefore, we determine portion P_d of the *detailed* slide x_d using the query keyword q from the focused slide x.

$$P_d = L_s(x_d, q) \cup L_s(x_d, k_v) \tag{28}$$

When the query keyword q in slide x_d satisfies Eqs. (17), (18), (25), and (28), then portion P_d is determined from the *detailed* slide, x_d. This is because the amount of content in slide x_d specific to q is greater than that in slide x. A set $L_g(x_g, q)$ consists of sentences from levels that contain specific information related to q in slide x_d, and satisfies the same conditions as the set $L_s(x, q)$ (these conditions apply to slide x and are given by Eq. (25)). Moreover, when slide x_d contains the keyword k_v, which belongs to $K_s(x, q)$ or $K_p(x, q)$, then a set $L_s(x_d, k_v)$ is used to extract an additional set of sentences. These are extracted from levels that provide specific information in terms of k_v, the more specified concept related to q in slide x_d, and satisfy the same conditions as the set $L_s(x, q)$ (these conditions apply to slide x and are given by Eq. (25)). When slide x_d contains two or more k_v, as determined from the focused slide x, we can extract two or more sets of sentences from $L_s(x_d, k_v)$. Eq. (28) can then be used to determine the portion P_d of slide x_d that combines the sets of sentences from $L_s(x_d, q)$ and $L_s(x_d, k_v)$.

As mentioned above, our method for generating snippets of the focused slides satisfies user demand by relating portions of the *generalized*, focused, and *detailed* slides to provide content varying from generalized to detailed based on a user query for specific content.

4.4 Evaluation

4.4.1 Prototype System

We built a presentation slide retrieval system (see Figure 18) implemented by a Slide-Ranking viewer and a Slide-KWIC browser in Microsoft Visual Studio 2008 C#, which aims to identify, for user queries, not only precise retrieve target slides but also their semantic relevance and their surrounding context for supporting comprehension.

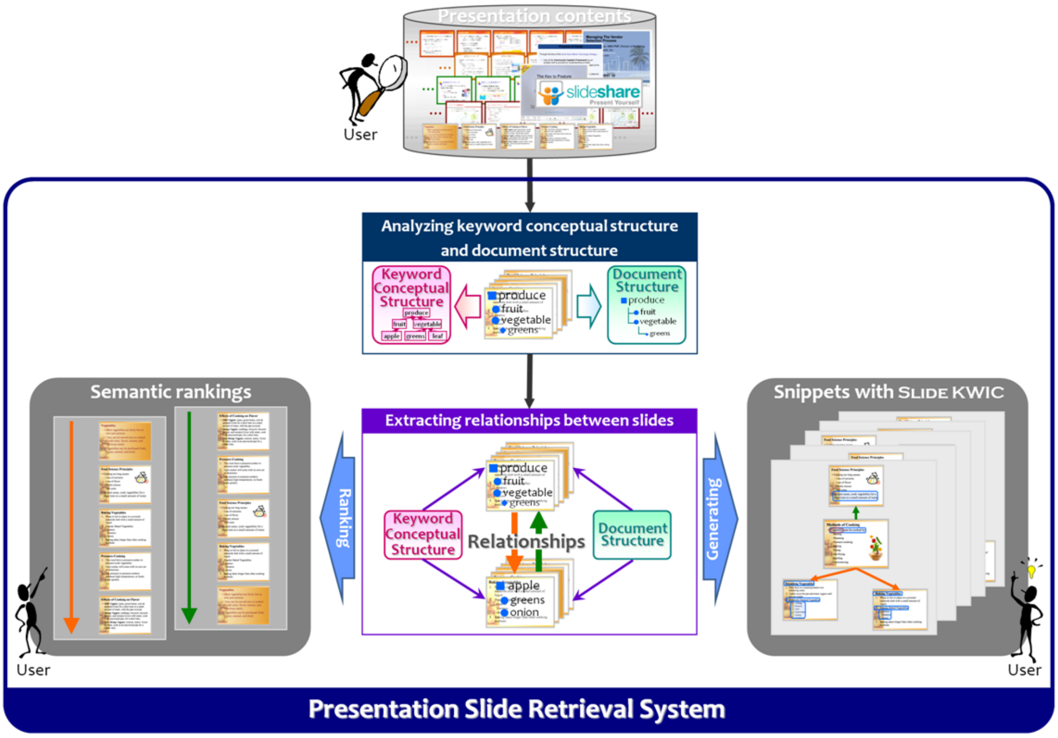


Figure 18: Presentation Slide Retrieval System Architecture

This system implements the analysis part, the determination part, and the application part. In the analysis part, we analyze the features of slide text that the keyword conceptual structure by using WordNet [29, 99, 26] extracts an is-a and a part-of relations between keywords; and the document structure by using the level position information about keywords in the slide by using the Office Open XML. The terms in the slides are extracted using a morphological analyzer *MeCab* [49], which is in SlothLib [64, 84] . In the determination part, all types of relationships between slides are extracted on the basis of the keyword conceptual structure and the document structure. There are two applications in the application part as follows:

- Slide-Ranking viewer

The application for retrieving slides in the semantic orders by a Slide-Ranking viewer as shown in Figure 19. Slide rankings are determined by calculating the two degrees of measures on the basis of the relationships between slides related to a user query. Using this retrieval system, a user can select the presentation content for studying. When the user enters a query of interest in the textbox and presses the "Search" button, the retrieved slides in two ranking types are presented in the Slide-Ranking viewer.

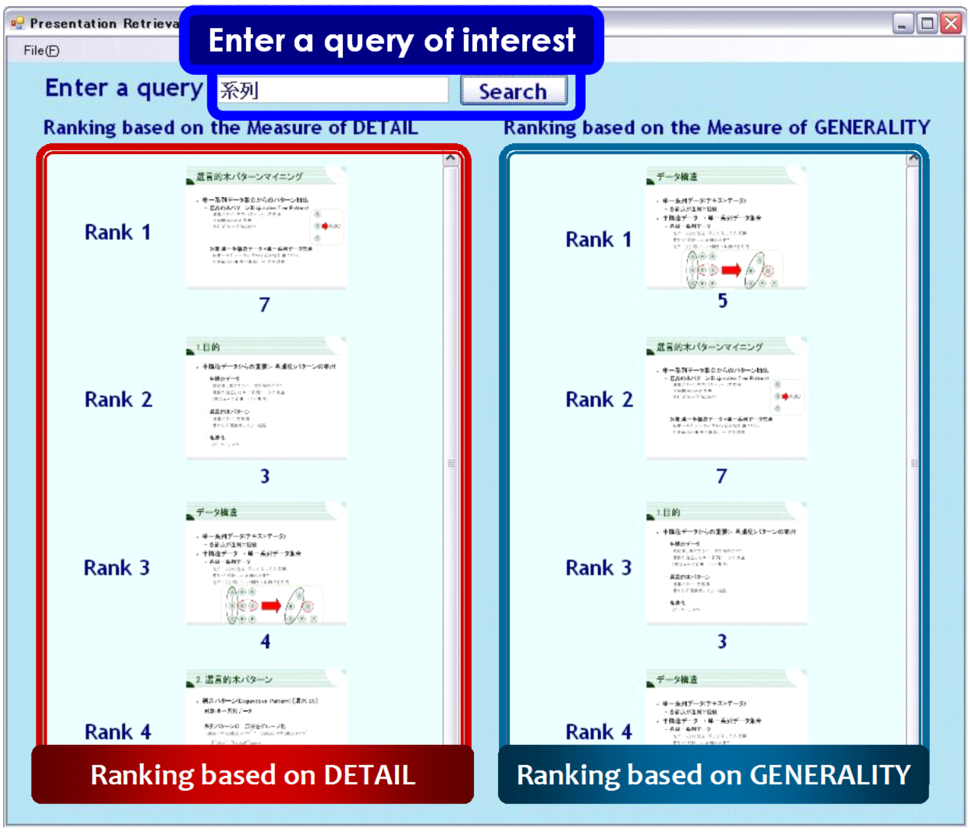


Figure 19: Screenshot of Slide-Ranking Viewer

- Slide-KWIC browser

The application for browsing the retrieval results by a Slide-KWIC browser as shown in the right part of Figure 20. Snippets are generated by identifying the portions of the retrieved slides with relevant portions of the related slides based on the relationships between slides. After a user selects the presentation content for studying, enters a query of interest in the textbox and presses the “Search” button, the retrieved slides are presented in the retrieval results part. When the user clicks any retrieved slide as a focused slide, the Slide-KWIC Browser for a snippet of the focused slide is presented in the other windows. Therefore, the focused slide with its portion that sentence are extracted in a listbox on the center position of the Slide-KWIC Browser window as a basic layer, and other related slides with their relevant portions in the listboxes are displayed in the high layer and the low layer of the focused slide.

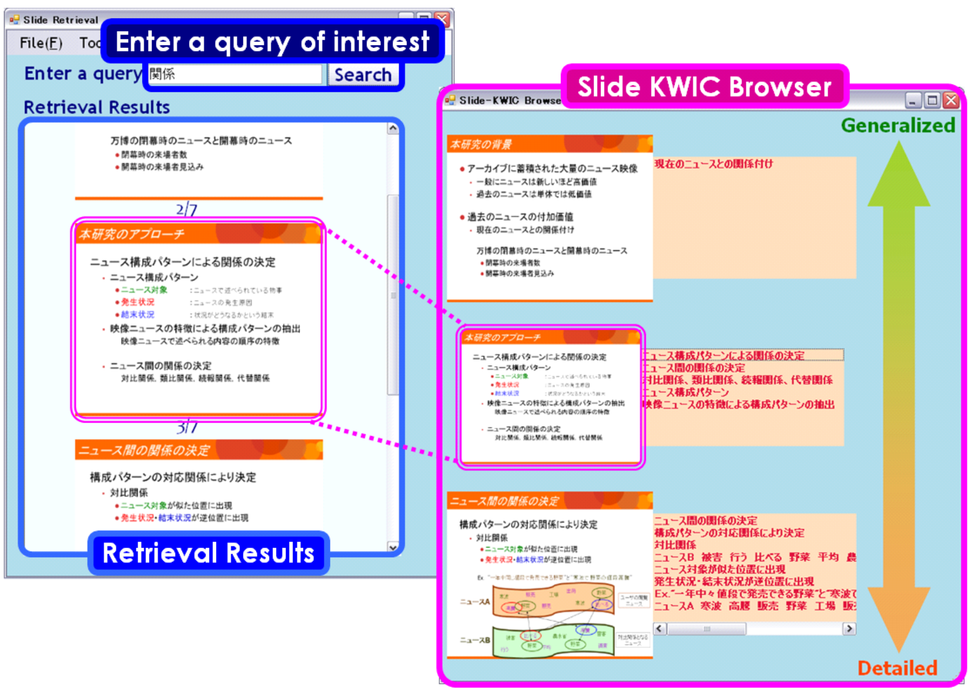


Figure 20: Screenshot of Slide-KWIC Browser

4.4.2 Experimental Dataset

In our experiments, we examined the proposed method of snippet generation for slide-browsing support based on the relationships between slides. We prepared a dataset using actual content, as shown in Table 4, consisting of (1) four actual academic presentations from a session of DEWS2006 in the DBSJ Archives [14], and (2) 36 actual lecture presentations [39] of four introductory courses from

Table 4: Experimental Dataset

(1) Academic Contents		
No.	**Title**	**Number of slides**
P-W	Mining Disjunctive Tree Patterns	22
P-X	A Web Archive Search Engine Based on the Temporal Relation of Query	15
P-Y	Video Archive Contents Browsing Method based on News Structure Patterns	7
P-Z	Improvement on Processing Rules Stored in Individual Metadata for Flexible Contents Management	30
(2) Lecture Contents		
No.	**Title**	**Average number of slides**
L-W	Methods of Education Research	16.4
L-X	Introduction to Psychology	29.6
L-Y	Social Statistics	24.1
L-Z	Introduction to Statistics	27.4

the lecture archives of the Social Informatics department at Aoyama Gakuin University. There were 5–15 students in the School of Human Science and Environment, University of Hyogo, taking the Social Informatics course and Information Media lab who participated in the following experiments. We assumed that the academic content in Informatics requires a certain level of expertise and is difficult to understand, and that introductory lectures provide a basic level of knowledge in Informatics and are thus easily understandable for the students who participated in the following experiments. We show and discuss the experimental results in the follow sections.

4.4.3 Experiment 1: Validity of Relationship Types

This experiment was designed to assess the generation of snippets based on relationships between slides related to a user query. Five participants freely described the relationships which existed between two slides, assessing 199 slide pairs containing keywords sampled at random from the four academic presentations in the dataset. Relationships between the slide pairs were determined if and when three or more participants described the same relationship. We calculated the coverage using the slide pairs, which were determined according to any relationship type identified by participants; we also defined the *others* relationship for those that could not be determined by our method, as shown in Figure 21.

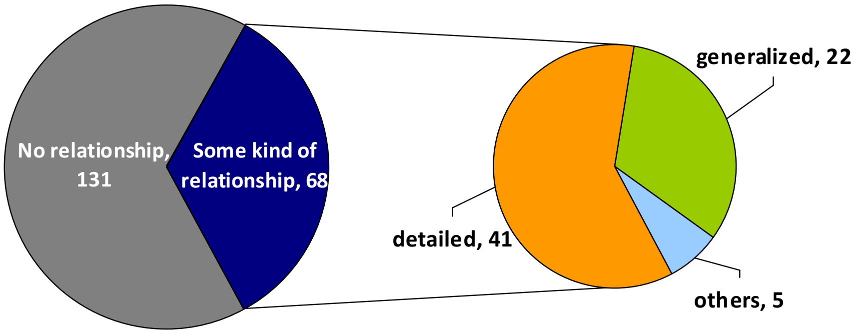


Figure 21: Coverage of Relationship Types

Table 5: Classification of Relationship Types

		Results determined using our system		
		detailed	*generalized*	*others*
Correct	*detailed*	**91**	17	58
	generalized	35	**65**	48
answers	*others*	6	6	91

The results and our findings were as follows:

- Coverage reached 92.6% (63/68). 68 slide pairs were determined to have some kind of relationship; 41 slide pairs were described as *detailed*, 22 slide pairs were described as *generalized*, and 5 slide pairs were classified as *others*. No relationship was determined for 131 slide pairs. We concluded that our defined relationships can account for slide relationships.
- Coverage reached a low of 60.3% (41/68) using our system. Of the 68 slide pairs that we had determined as having a relationship (see Figure 21), the experiment participants only agreed with our opinion of the relationship type on 41 occasions; we thus concluded that the slide relationships in presentations cannot be expressed comprehensively by using our method alone.

Table 5 lists the classification results. A correct answer was defined as a relationship between two slides where three or more participants described the same relationship. Participants did not have any particular bias [1)], and we

[1)] Five participants, i.e., A to E. The ratio of the same answers by A and B was 70.9%

consider that the correct answers can be defined using the answers obtained from participants. Only one type of relationship defined in our system was determined by the participants for any given slide pair, and then this answer was duplicated by more participants. For example, if three of five participants give the answer of "detailed" while each of the remaining two participants does the answers of "generalized" and "others," respectively, the correct answer becomes "detailed" and the numbers of relationship of "detailed," "generalized," and "others" are accumulated by 3, 1, and 1, respectively. We found that *detailed* includes "instance" relationships, where slides show the specific examples with their explanations, and *generalized* includes "parallel" relationships, where slide pairs describe information derived from a single topic on equal terms. However, these relationships did not occur frequently, and are thus difficult to define. We should therefore improve the definitions of *detailed* and *generalized* relationships. This experiment confirmed that the relationships between slides containing any keyword could be covered by using the concept of relationship types. In our method, we focused on *detailed* and *generalized* relationships at the conceptual level, but this should be expanded to determine other types of semantic relationships.

We used three representative keywords from each academic presentation to extract 678 slide pairs. We evaluated the validity of the rules for determining the two types of relationships by precision [1]), relative recall [12] [2]), and F-measure [3]) using the results obtained from four methods, and a correct answer was considered to be a slide pair where three or more participants found some relationships present in their free description. The four methods are: "Frequency", using the keyword frequency, "document", using the document structure only, "concept",

(141/199); by A and C was 55.3% (110/199); by A and D was 70.4% (140/199); by A and E was 69.8% (139/199); by B and C was 50.3% (100/199); by B and D was 61.8% (123/199); by B and E was 60.3% (120/199); by C and D was 50.8% (101/199); by C and E was 61.3% (122/199); by D and E was 72.9% (145/199)

1) Precision $= \frac{\text{Number of correct answers of relationships determined by a method}}{\text{Total number of relationships determined by a method}}$

2) Relative recall $= \frac{\text{Number of correct answers of relationships determined by a method}}{\text{Total number of correct answers of relationships determined by all four methods}}$

3) F-measure $= \frac{2 \times \text{Precision} \times \text{Relative recall}}{\text{Precision} + \text{Relative recall}}$

using the keyword conceptual structure only, and “proposed”, which used our proposed method.

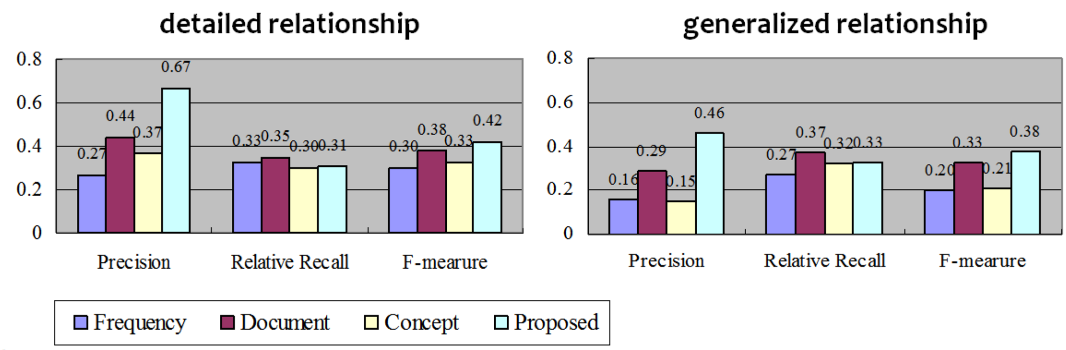


Figure 22: Performance Measure Graph

The results for the slide relationships found by the four methods are shown in Figure 22, and they can be explained as follows:

- The relative recall of *detailed* or *generalized* was low, and many correct answers were detected to have no relationships with our method. We consider the limitations of WordNet to be one factor for the low relative recall. Although WordNet is a large lexical database, it does not necessarily contain all concepts related to an experimental keyword, as there may be new concepts associated with a technical term or new words used in the academic presentation. For instance, while keywords such as “disjunction,” “mining,” and “preorder” frequently appear in the main content of the academic presentation P-W, our method based on WordNet cannot extract semantic relations for them, as they are not included in WordNet.
- For *detailed*, our method returned more than half of the correct answers. The precision of our method performed well. However, there was a little confusion between “detailed” and “instance.” “Detailed” means a more specific explanation of a term; “instance” means a specific explanation of a term through the use of cases or examples. Therefore, our method returned some results as “detailed” when participants labeled their correct answers as “instance.” We focused on whether the specific explanation of a term contained more information, but not how the specific explanation

was given, such as in examples.

- For *generalized*, even if the same slide set was considered, participants' answers differed in terms of "generalized" and "parallel." If participants answered "generalized," this means that they understood the content well. However, if the participants answered "parallel," it means that they understood only that the slides had a relationship. We consider *generalized* to be effective when a user can understand that slides have at least some relationship, but cannot determine the relationship type.

The graph in Figure 22 shows that the precision and F-measure of our system were higher than those of other methods for determining *detailed* and *generalized* relationships. This experiment confirmed that slides with academic content have some kind of relationship between each other. Our proposed relationships may provide an appropriate definition for using the semantic relations between keywords and the document structure of indents. Furthermore, we believe that a considerable number of slides in the academic presentations provide detailed explanations. However, we should enhance our method for extracting semantic relations between keywords to consider the semantic data model of keywords. In particular, for academic content containing a lot of technical terms, this method should not only involve the use of WordNet, but also include such aspects as the use of domain-specific dictionaries, such as the Handbook of Information Processing [1)] and the Medical Dictionary [2)]. As mentioned above, we can improve the accuracy of our method for determining relationships between slides.

4.4.4 Experiment 2: Validity of Ranking Types

We showed the participants the following ten rankings that are ten keywords sampled at random in slides from (1) 4 actual academic contents used in Experiment 1. Then, we let them rank the slides with regard to ten given keywords in the order of degree of $DETAIL$ and $GENERALITY$, respectively. For each given keyword, we then calculated the Spearman's rank correlation coefficient

[1)] Information Processing Society of Japan

[2)] http://www.medterms.com/script/main/hp.asp

[1] between the participant rankings and our system rankings. The Spearman's rank correlation coefficient ranges from -1 to 1, where -1 indicates that two rankings are completely reverse whereas 1 indicates that the rankings are exactly the same.

Six participants, i.e., A, B were M.S. students, C to F were B.S. students, participated in this experiment. The experimental results are listed in Table 7 and our findings were as follows:

- The degrees of our proposed measures, i.e., $DETAIL$ and $GENERALITY$, are greater than 0 and that on an average, the measure determined by the participants, i.e., $DETAIL$, shows the best performance. These results indicate that our proposed method that takes into account the relationships between slides about the given keyword based on the semantic relations between keywords and the document structure of indents can be successfully applied to the presentation content retrieval engine on the basis of semantic rankings.
- The degree of $GENERALITY$ on an average was low here, i.e., dataset (f) and dataset (i). We calculated the degree of $GENERALITY$ by using the *generalized* relationship between slides with regard to the given keywords and the specified keywords of the given keywords. In particular, the degree determined by participant D in dataset (f) and participant C in dataset (i) were too low. We consider that it is difficult for participant D and participant C to ascertain the specified keywords of the query keyword in the retrieved slides, which may reduce the performance.
- Although a slide has a *generalized* relationship with other slides with regard to the given keyword and it contains many specified keywords of the given keyword that has a *generalized* relationship with other slides that were not retrieved, the specified keywords of the given keyword were unknown by a participant. It can be seen that our method can extract many concepts of the given keywords by effectively using the keyword conceptual structure.

[1] Correlation coefficient $= 1 - \frac{6 \times \Sigma(\text{Rank number by our system} - \text{Rank number by each participant})^2}{\text{Quantity of slides} \times (\text{Quantity of slides}^2 - 1)}$

Table 6: Comparison between Spearman's Rank Correlation Coefficients Obtained by Subject Evaluation and Our System

Dataset	Given keyword	Ranking type	Participants						Average
			A	B	C	D	E	F	
(a)	"pattern"	*DETAIL*	0.77	0.63	0.80	0.80	0.63	0.77	0.73
(b)	"series"	*DETAIL*	1.00	0.70	0.70	0.70	0.70	0.60	0.73
(c)	"data"	*DETAIL*	1.00	0.40	0.40	0.20	0.80	0.80	0.60
(d)	"composition"	*DETAIL*	0.80	0.20	1.00	0.80	0.40	0.20	0.57
(e)	"contrast"	*DETAIL*	1.00	0.50	1.00	1.00	0.50	1.00	0.83
(f)	"relationship"	*GENERALITY*	0.80	0.80	0.20	-0.40	0.60	0.40	0.40
(g)	"aggregate"	*GENERALITY*	0.80	0.80	0.80	1.00	0.20	0.40	0.67
(h)	"group"	*GENERALITY*	0.20	0.80	1.00	1.00	-0.20	1.00	0.63
(i)	"geography"	*GENERALITY*	1.00	0.40	-0.40	0.40	0.80	0.40	0.43
(j)	"comparison"	*GENERALITY*	0.70	-0.10	0.70	0.70	0.90	0.60	0.58

From the results of this experiment, we find that we have to improve the determination of the ranking algorithm by using the relationships between slides containing keywords which the semantic relations do exist between them.

4.4.5 Experiment 3: Validity of Portion Identification

This experiment aimed to verify whether the proposed method is useful for identification of portions containing sentences relevant to a user query. Five participants freely captured portions containing sentences from different indent levels in the slides, and assessed three representative keywords from 40 actual presentations in the dataset to identify portions of 312 slides. A correct answer was defined as a portion where three or more participants found the sentences on the indent levels of the slides that they had captured. In this study, we evaluated the validity of the rules for identifying portions of slides in terms of the query keywords, using precision [1)], recall [2)], and F-measure [3)] to compare the results obtained by our method with those obtained from participants who gave correct answers for each academic presentation and in each lecture explaining different topics. In addition, we compared the portions obtained by our method and the portions of sentences containing the given keywords on indent levels with their anteroposterior (AP) levels.

[1)] Precision $= \frac{\text{Number of correct answers of portions extracted by our method}}{\text{Total number of portions extracted by our method}}$

[2)] Recall $= \frac{\text{Number of correct answers of portions extracted by our method}}{\text{Total number of correct answers by participants}}$

[3)] F-measure $= \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$

The results of the experimental identification of portions of academic and lecture presentations are listed in Tables 7 and 8, and can be explained as follows:

- The average F-measures for this experiment on academic and lecture presentations look similar. However, the average precision and recall of the lecture presentations were both higher than those for the academic presentations. We therefore concluded that it is difficult to understand the slides used in academic presentations that require some level of expertise, and we used WordNet, which does not contain all concepts related to some general words. For example, a slide with the query keyword "structure" was used to identify portions of it in presentation P-Y (see Figure 23). Sentence levels containing "news subject," "generation status," and "conclusion status" were correctly related to "news structure pattern" by participants. Our method, however, could not determine these keywords, as WordNet does not recognize "subject" or "status" as having a part-of relation with "structure."
- The average precision of all experimental portions from academic or lecture presentations was low; our method extracted a much greater number of portions than those for which participants concurred. We believe that when determining correct answers, the participants did not consider slide titles or figures in slides in terms of the given keywords when our method was used.
- Comparing the results of the two methods, the average precision and average F-measure of our method were both higher than those of the other method. Although the results of the two methods look similar, the other method did not extract some portions containing sentences in slides that explained the given keywords, and some sentences on the AP levels were extracted which were not related to the given keyword.

This experiment confirmed that our method can extract the appropriate portions of slides, using semantic relations between keywords and the document structure of indents. However, we want to use an enhanced method for extracting mathematical formulas related to the given keywords. Furthermore,

Table 7: Results of Identified Portions of Slides in Academic Contents

	Academic contents by our method				
	P-W	*P-X*	*P-Y*	*P-Z*	Average
Precision	69.6% (298/428)	60.4% (166/275)	57.7% (142/246)	66.1% (360/545)	63.5%
Recall	67.3% (298/443)	71.2% (166/233)	64.0% (142/222)	75.8% (360/475)	69.5%
F-measure	0.68	0.66	0.64	0.71	0.67
	Academic contents by the levels contain the given keywords with their AP levels				
	P-W	*P-X*	*P-Y*	*P-Z*	Average
Precision	52.8% (295/559)	47.4% (180/380)	53.8% (135/251)	56.1% (415/740)	52.5%
Recall	64.6% (295/443)	77.3% (180/233)	60.8% (135/222)	87.4% (415/475)	73.0%
F-measure	0.59	0.59	0.57	0.68	0.61

Table 8: Results of Identified Portions of Slides in Lecture Contents

	Lecture contents by our method				
	L-W	*L-X*	*L-Y*	*L-Z*	Average
Precision	71.3% (196/275)	60.3% (193/320)	63.4% (716/1130)	69.6% (400/575)	66.2%
Recall	53.7% (196/365)	70.7% (193/273)	81.9% (716/874)	82.0% (400/488)	72.1%
F-measure	0.61	0.65	0.71	0.75	0.69
	Lecture contents by the levels contain the given keywords with their AP levels				
	L-W	*L-X*	*L-Y*	*L-Z*	Average
Precision	65.3% (261/400)	50.0% (233/466)	52.3% (792/1513)	56.8% (420/740)	56.1%
Recall	71.5% (291/365)	85.3% (233/273)	90.6% (792/874)	86.1% (420/488)	83.4%
F-measure	0.68	0.63	0.66	0.68	0.67

we should consider how to identify the keywords at different levels in figures or tables to improve performance in this experiment. In general, we may also use the conceptual descriptions on the Wikipedia website [1], an encyclopedia providing a vast amount of structured world knowledge, to build a large ontology. Therefore, we can improve the accuracy of our method for identifying portions of slides by ceasing to use WordNet, and instead using domain-specific dictionaries for technical terms, or Wikipedia for general words.

[1] http://www.wikipedia.org/

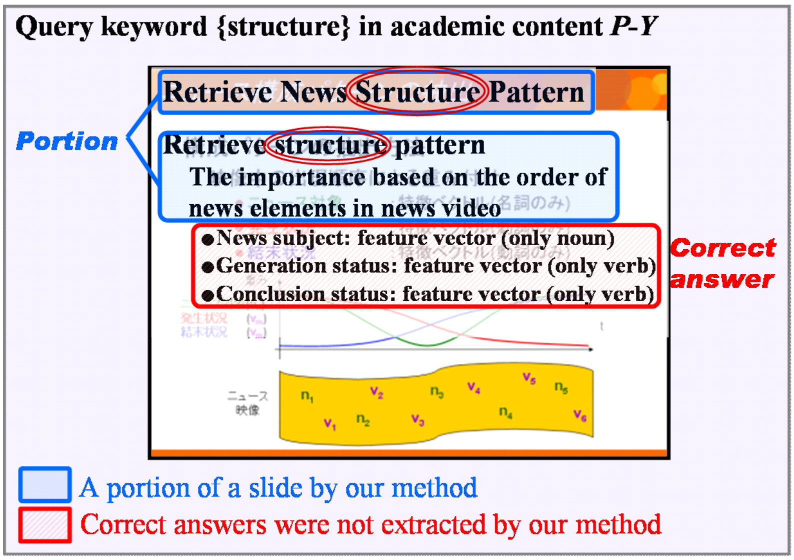


Figure 23: Inadequate Identification of a Portion of a Slide

4.4.6 Experiment 4: Validity of Snippet Generation

This experiment aimed to verify whether the proposed method is useful for generation of snippets for slides. We showed the participants 87 snippets, composed of portions of slides pertaining to the given keywords from the experimental dataset used in Experiments 1 and 3. Five participants took part in this experiment; the snippets presented a detailed explanation of the given keywords in order of the relevant portions in the slides. A correct answer was defined as three or more participants describing snippets of the focused slides with other slides as correct.

The results are shown in Tables 9 and 10; the experimental results were as follows:

- The results depended on those from Experiments 1 and 2. However, in Experiment 1, we did not evaluate the determination of the relationship types in lecture content. For this experiment, the results for the academic and lecture content look similar. As in Experiment 2, the average precision and recall of the lecture presentations were both higher than those of the academic content; we concluded that there was no difference in the snippet generation between the slide relationships used in the academic and lecture content.

- The average recall of all experimental snippets from the academic and lecture presentations was low. When our method was used in Experiment 2, many of the correct answers were found to contain the sentences on indents in portions not extracted by our method. A snippet consists of portions of the focused slide and depends on identification of these portions, which is based on our method using WordNet, and so did not determine that some keywords have semantic relationships between them. This was one of the reasons why the recall was low. Therefore, these portions for generating snippets also need to be considered.
- The average precision of all experimental snippets from the academic and lecture presentations was high. The results indicate that our method can generate appropriate snippets of relevant portions of slides based on the relationships between these slides, and the method can then be successfully applied to support browsing slide retrieval by generating snippets at the conceptual level.
- A few experimental snippets identified portions that did not include detailed information related to the focused slides; that is, relationships did not exist between them. In addition, many of the relevant portions were not strongly related to the portion of the focused slide, which may have reduced the precision.

This experiment showed that our method can generate snippets of relevant portions of related slides via the query, by effectively using the relationships between the slides. The results of this experiment suggest that we need to improve the determination of the snippet-generation algorithm by using the relationships between slides, and extracting the portions of slides relevant to the query. Our method used WordNet, which will have had a bearing on the determination of the relationships between slides, and the identification of the portions of slides, due to the shortcomings already mentioned. Therefore, we plan to use domain-specific dictionaries or Wikipedia for extracting semantic relationships between keywords in the future work.

Table 9: Results of Generated Snippets from Academic Contents

	Academic contents by our method				
	P-W	*P-X*	*P-Y*	*P-Z*	Average
Precision	68.6% (175/255)	62.8% (76/125)	62.1% (59/95)	80.0% (108/135)	68.4%
Recall	66.0% (175/265)	67.0% (76/114)	57.0% (59/106)	66.7% (108/162)	64.2%
F-measure	0.67	0.64	0.60	0.73	0.67

Table 10: Results of Generated Snippets from Lecture Contents

	Lecture contents by our method				
	L-W	*L-X*	*L-Y*	*L-Z*	Average
Precision	67.3% (175/260)	69.2% (229/331)	72.8% (732/1005)	74.7% (396/530)	71.0%
Recall	63.4% (175/276)	69.2% (229/331)	73.2% (732/1000)	66.6% (396/595)	68.1%
F-measure	0.65	0.69	0.73	0.70	0.70

4.4.7 Experiment 5: Efficacy of Browsing Snippets

In this experiment, we verified how the proposed method can help users to browse by introducing snippets. When users browse slides containing information, the snippets presented by our system let users easily grasp the context of the focused slides in terms of the given keywords. We conducted this experiment with 15 participants, using four given keywords for 17 slide pages taken from two actual presentations: the academic content in P-Y contains seven slide pages, providing a level of expertise in Informatics that is important for the participants, and lecture material in L-Z containing 22 slide pages, providing basic knowledge in Informatics that is easy to understand for the participants. For evaluation purposes, we first prepared correct answers by asking three students which slide had the most detailed information related to each given keyword in each presentation from the experimental dataset. We defined a correct answer as when two or three students identified the same slide. Secondly, we provided two retrieval results for each given keyword using (a) the conventional method, where slides are retrieved by matching keywords, and (b) our method, where the corresponding snippets are generated by our system.

After providing these two retrieval results to the 12 students who did not take part in preparing the correct answers, we asked two questions in two steps

Table 11: Experimental Results of Efficacy of Browsing Snippets

Dataset	Browsing slides only	Browsing slides with their snippets
Expertise in *P-Y*	14/24	21/24
Prior knowledge in *L-Z*	20/24	18/24
Total	34/48	39/48

as follows:

Step 1. Presenting the slides retrieved by method (a).

$Q1$: Which slide do you think provides the most detailed information related to the given keyword in these retrieval results? Please write your answer as the slide number and the reason for your selection.

Step 2. Presenting the retrieval results for method (b), including snippets.

$Q2$: When you browsed the snippets for the slides presented in $Q1$, did you change your answer to $Q1$? If so, please write the changed slide number and your reason for changing. If not, please give the reason why you did not change it.

We analyzed these answers, and the results are shown in Table 11. The vertical columns show how many correct answers were given when browsing the slides only, and how many correct answers were given when the snippets were also given to the participants. The horizontal rows show the breakdown of correct answers by knowledge levels required for the presentations. The experimental results are as follows:

- The total number of correct answers from browsing slides with their snippets was more than that when browsing slides only. Therefore, we believe that users browsing slides with their associated snippets can grasp the context of the focused slides, in relation to the given keywords, more easily.
- Browsing slides with their snippets provided more correct answers than browsing the slides only in presentation P-Y, and browsing slides with their snippets provided more correct answers in P-Y than in L-Z. P-Y provides expertise that is difficult for participants to understand, while L-Z provides knowledge that should be easily understood by participants. We confirmed that snippets are more useful when users browse slides containing a higher level of knowledge, rather than when they browse slides containing

information they are already aware of.

- In the *L-Z* dataset, there were fewer correct answers when browsing slides with their snippets than when browsing only the slides. *L-Z* provides prior knowledge that is easily understood by participants, so they are able to select many correct answers by browsing slides only. In addition, our proposed snippet-generation model only has three layers and does not consider the relevance of the related slides; thus, two participants were a little confused about the snippets for slides. However, for a majority of participants, we confirmed that our snippet-generation model is helpful for users browsing slides with their snippets.
- Our snippet-generation method is based on the relationships between slides, and works by identifying relevant portions of the focused slides. We concluded that a few generated snippets have the effect of determining the relationships between slides.

This experiment showed that our method of browsing slides with snippets is more useful than browsing slides only. In particular, our snippet-generation method is helpful for browsing slides containing higher levels of expertise alongside snippets.

In this paper, we evaluated our method by conducting Experiments 1–5, using presentation content made in PowerPoint and containing a layer structure (the levels of indentation) in the slides. We confirmed that our method is useful by satisfying certain criteria related to levels of indentation that are used to structure content in slides. Our proposed method does not only focus on presentation content made using PowerPoint, but can also be applied to a variety of other important presentation formats, such as Apple Keynote [2], Google Docs [20], and Prezi [76]. This is because these presentation formats contain a layer structure, in common with PowerPoint. We consider our proposed method to be applicable to a variety of presentation software, and we plan to evaluate our method with other presentation formats in the future work. Moreover, we evaluated our method by using WordNet to extract the semantic relations between keywords, and experimental results suggested that we can improve the accuracy of our method by using domain-specific dictionaries for technical terms in academic content, or Wikipedia for general words in presentations.

As mentioned above, we are also aware of the limitation of our method in not focusing on visual effects in slides. At the present time, authors often focus on visual effects that are easily understandable, and more attractive than slides with simple text. We do not currently use font or visual information, but it would not be difficult to improve our method by considering such data. Future developments to this method could also consider visual elements of figures, and the color distribution and animation occurrence in slides, as we can acquire this information by analyzing XML files from the various presentation formats. Furthermore, our method can be extended to consider the document structure not only in the slides, but also in associated presentation data. Finally, it is possible to treat retrieval units for other applications to use our proposed system.

4.5 Summary

In this theme, we have built a retrieval system for e-Learning focusing on the relationships between slides on the basis of the features of the slide text. We analyzed the features of the slide text focused on the semantic information that we derived a keyword conceptual structure as an ontological structure, and structural information that we derived a document structure of indents in the slide text. We described the detail of this retrieval system that implements the application part providing a Slide-Ranking viewer supporting for retrieving slides in rankings; and a Slide-KWIC browser support for browsing the retrieved slides with snippets. We evaluated the classification of the relationship types and the validity of determining the relationship types by our proposed method with actual data. Trough the evaluate experiments, we confirmed that

- Slides in the academic content have some kinds of relationships between each other.
- A considerable number of slides in the academic content provide detailed explanations
- The performance of our method is better than other methods.

From this experiment results, we must to improve the coverage of the types of relationships and their definitions via our method, and we should expand it to determine other types as semantic relationships. Additionally, we should use an enhanced method for extracting the semantic relations between keywords; this method should not involve the use of WordNet only, such as involving the use of a large ontology construction.

Chapter 5 Outline Generation for Presentation Slides based on Expression Styles using Document Structure

5.1 Introduction

With the advent of usable presentation tools to create attractive slides, such as Microsoft PowerPoint [52], Apple Keynote [2], and OpenOffice Impress [1], presentations now play a socially important role in promoting understanding in many fields, including business and education. Many university instructors have used Web services such as myBrainshark [60] and SlideBoom [82] to store presentation slides that they use in lectures owing to the features for browsing, sharing, and reusing slides. Prezi [76] provides a service for editing, browsing, and sharing presentation data. Although useful and powerful support tools for creating slides and Web services for sharing slides are widely used, they are generally not effective in generating lecture slides with content that is understood by students; instructors are more effective with preparation of slides of this nature. In particular, lecture slides are often made from textbook chapters (hereafter known as 'chapters') with information to be conveyed by the instructors. It is important, therefore, to focus on how to express the information from textbooks that will appear in slide format. In order to address this problem, we can generate outlines that serve as slide layouts and express typical words from target chapters based on their roles in referred slides, focusing on difference in document structure (i.e. text structure within a chapter, slide structure within a slide). For example, 'vegetable' appears in the body of text in a slide entitled 'Agriculture Market', which relates to sections entitled 'Agriculture Market Analysis', 'Vegetable Production', and 'Vegetable Plants'; it is dispersed within a textbook chapter. 'Vegetable' is the title of another slide that is related solely to the section entitled 'Vegetable Production'.

Our approach creates semi-automatically an editable slide outline producing slides based on typical words. Main points or topics in a logic structure are presented by focusing on a hierarchical representation of the words to help presenters prepare slides easily and efficiently. We explored slide outline creation and found that a word might be expressed differently in various slides. For instance, a word may be the title of one slide or in the body of the text in another for presentation content. We also learned that various styles of presentation slides are typically created from the same document based on different expressions of words. As depicted in Figure 24, a teacher can take a target chapter from a textbook containing lectures to prepare slides by using generated outlines based on expression styles of referred slides.

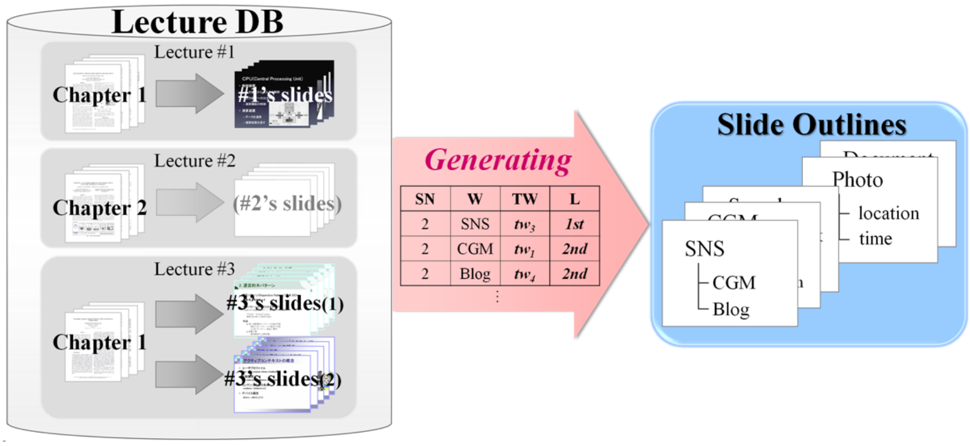


Figure 24: Conceptual Diagram of Outlines for Lecture Slides Generated from a Textbook

An example is shown in Figure 25. Chapter 2 is the target chapter, and Chapters 2 and 1 have the same text structure; we can extract the expression styles of words in referred slides as Presentation 1 (created based on Chapter 1 and specified by a presenter). We can, therefore, generate outlines for slides from Chapter 2 based on the expression styles of the words in Presentation 1 and an analysis of the differences between tendency of word appearance in Chapter 1 and Presentation 1.

As our initial motivation for this work, we envisioned an instructor who gives lectures in familiar or similar situations (e.g. students with the same knowledge level or same number of students in a class) and frequently makes presentation files to compose slides in the same expression. We recognized that

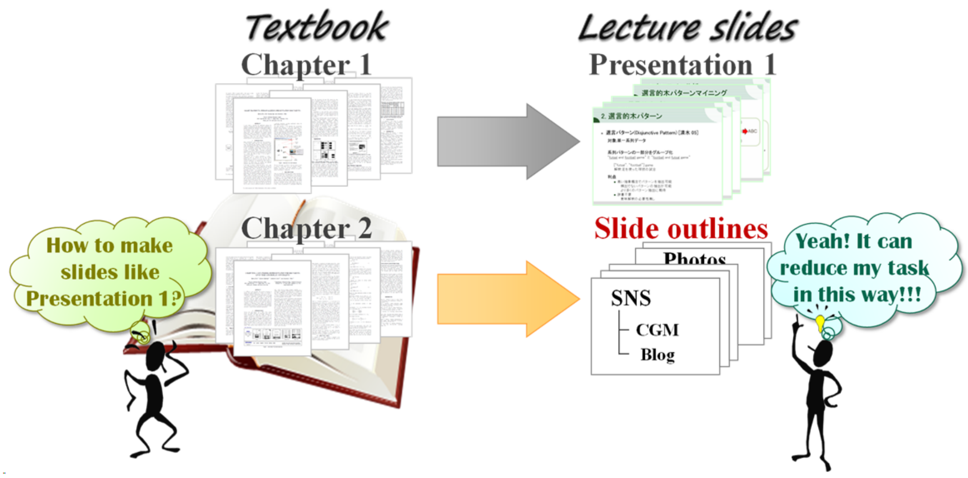


Figure 25: Outline Generation from a Target Chapter using a Chapter and Referred Slides

instructors must know how to create slides and how to present them. Thus, we focused on how to make presentation slides for routine lectures by referring to textbooks. We considered the conventional definition of expression style as the expression of content based on scenario construction in slides (i.e. textbook content). We newly defined and extended the meaning of expression styles based on word alignment to position levels of words in slides extracted from the slide structure, and it is used to express how typical words to be handled in the slides by considering how each word represented in each slide from chapters. Accordingly, we utilized the expression styles of referred slides from previous lectures specified by the instructor to generate slide outlines for the next lecture.

Level positions of words in slides and roles of words were analyzed. We defined slide structure based on indents within slides, focusing on titles and bulleted text in the slides. As mentioned above, expression styles were extracted from level positions of words in the referred slides. Two main features that we recognized as particularly helpful for generating slide outlines and arranging corresponding words from target chapters based on difference in document structure (i.e. text structure, slide structure) are as follows:

- When a word is dispersed throughout all sections of a chapter, it is used generally in the chapter; when a word occurs frequently in a certain section of a textbook chapter, it is associated with a specified description in the chapter.

- When a word appears in the slide title of a slide or in lines that are less indented, it is a topic of the slide; and when a word appears in the body of the text on a slide, it is used to explain the topic of that slide.

These two features are particularly helpful for concisely conveying information in slides from textbook chapters by characterizing the differences between the tendency of word appearance in chapters and their associated slides.

In this theme, we theorized that a presenter could prepare presentation slides for each lecture from chapters in a textbook by arranging words in slides according to the same expressions of words in referred slides created from chapters specified by the presenter. Hence, we made a model to generate outlines for arranging the corresponding words from a target chapter in slides based on the expression styles of the words in referred slides; corresponding words from the target chapter were analyzed according to differences between the tendency of word appearance in chapters and slides based on document structure (i.e. text structure, slide structure) in the same domain. Finally, we attempted to achieve our goal to generate outlines for slides in an experiment conducted with a real dataset.

5.2 Determination of Tendency of Word Appearance

For exploring difference in document structure (i.e. text structure, slide structure), we determined tendency of word appearance by calculating the dispersion and concentration of words within the text structure of a chapter and determining the generalized and detailed information of words within the slide structure of a slide. We defined the text structure of a chapter in terms of its logical units, or sections, which in turn consists of section heads and paragraphs (or subsections, which in turn consist of subsection heads and paragraphs). In our method, the text structure is extracted as a tree-shaped structure (hereafter 'tree'), consisting of the chapter head as a root node and paragraphs as leaf nodes. In addition, a parent-child relationship symbolizes the relationship between the chapter head and section head, section and subsection heads, and subsection heads and paragraphs (see Figure 26).

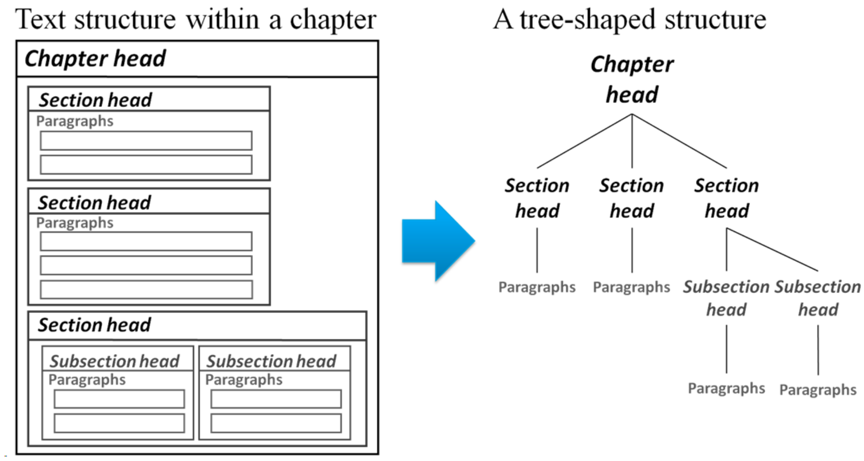


Figure 26: An Example of a Tree-shaped Structure of Text Structure within a Chapter

The content of a presentation includes a number of slides that have structured text information. We defined the slide structure from slides appearing in the outline pane [81], based on the indentions in the slide. The slide title is the first level. The first item of text within the body of a slide is considered to be the second level, and the depth of the sub-items increases with indention level (third level, fourth level, etc.). Objects that are outside of the text, such as figures or tables, are considered to be at the same indention level as the text in which they appear.

5.2.1 Determination of Tendency of Word Appearance in a Chapter

If a word is dispersed in a chapter, the tendency of word appearance of this word is deemed dispersion; it belongs to $W_{d_{ch}}$. In contrast, if a word appears as centered in a chapter, the tendency of word appearance of this word is deemed concentration; it belongs to $W_{c_{ch}}$. We explain the determination of $W_{d_{ch}}$ and $W_{c_{ch}}$ using a word q, and we calculate the degree of dispersion and concentration of q in the chapter ch. When q is dispersed at a high degree, q is determined to belong to $W_{d_{ch}}$; when q is centered at a high degree, q is determined to belong to $W_{c_{ch}}$.

$$
\begin{aligned}
W_{d_{ch}} &= \{q \mid min\Big(\frac{\sum_{i=1}^{n} dist(s_1, q_i)}{n}, \frac{\sum_{i=1}^{n} dist(s_2, q_i)}{n}, ..., \\
&\qquad \frac{\sum_{i=1}^{n} dist(s_j, q_i)}{n}, \frac{\sum_{i=1}^{n} dist(s_m, q_i)}{n}\Big) > \alpha\} \qquad (29) \\
W_{c_{ch}} &= \{q \mid min\Big(\frac{n}{\sum_{i=1}^{n} dist(s_1, q_i)}, \frac{n}{\sum_{i=1}^{n} dist(s_2, q_i)}, ..., \\
&\qquad \frac{n}{\sum_{i=1}^{n} dist(s_j, q_i)}, \frac{n}{\sum_{i=1}^{n} dist(s_m, q_i)}\Big) > \alpha\} \qquad (30)
\end{aligned}
$$

where q_i is the i^{th} word q, and s_j is the j^{th} section in a chapter ch. For deriving $W_{d_{ch}}$ and $W_{c_{ch}}$, we calculated the degree of dispersion and concentration of q based on the standard deviation concept. In Eqs. (29) and (30), the function $dist$ calculates the distance between q_i and s_j; that is, a number indicates how many sections there are between a section that contains q_i and s_j. In our method, the section s_j was the standard for measuring how many sections there were between the same word q and its appearance in the chapter ch. n is the number of times that q appears in ch, and m is the number of sections in ch.When q_i appears in s_j, the distance between them is one (1). The formula in Eq. (29) or Eq. (30) that returns the minimum value of q is extracted using the function min because there are unknown expectations. Thus, the highest degree of expectation is obtained for a position in the section with the lowest degrees of dispersion or concentration. When the minimum value is high in Eq. (29) or Eq. (30), it is highly possible that q will appear dispersed or centered in ch. $W_{d_{ch}}$ or $W_{c_{ch}}$ represents a bag of words in the chapter; if the formula is greater than a threshold α in Eq. (29), the tendency of word appearance of q is determined to be dispersed in ch, and q belongs to $W_{d_{ch}}$. If the formula is greater than a threshold α in Eq. (30), the tendency of word appearance of q is then determined to be centered in a range in ch, and q belongs to $W_{c_{ch}}$.

5.2.2 Determination of Tendency of Word Appearance in Slides

If a slide has more information in terms of a given word contained in a prior slide in a presentation file, the tendency appearance of this word becomes upper, and this word belongs $W_{u_{x,y}}$. In contrast, if a slide has generalized information in terms of a given word contained in a prior slide in the presentation file, the tendency appearance of this word becomes lower, and this word belongs to $W_{l_{x,y}}$. We have explained the determination of $W_{u_{x,y}}$ and $W_{l_{x,y}}$ using a word q, which is present in slides x and y. When q and the other words in slides x and y satisfy certain conditions, q is determined to belong to $W_{u_{x,y}}$ or $W_{l_{x,y}}$.

$$K(x,q) = \{k \mid k \in x, l(x,q) < l(x,k)\} \tag{31}$$

Here, $K(x,q)$ is a bag of words that can be considered to provide an explanation in terms of q in slide x. $l(x,q)$ is a function that returns the indent level of q in slide x. When q appears frequently in slide x, $l(x,q)$ will return the lowest possible value (i.e. the uppermost level at which q occurs in slide x). This because we consider that when q appears in an upper level, all the other levels in which q appears in the body of that slide are explanatory points related to a deeper occurrence of q. The word k is included in the levels that have a hierarchical relationship with the level of q, and k belongs to the bag of words $K(x,q)$ in slide x. $l(x,k)$ is greater than $l(x,q)$, in that k is a child of q in the slide structure. When k is not present in slide x, $K(x,q)$ will be empty.

Based on the above criteria, we computed the number of words related to q for slides x and y and compared their numbers using the following formulas:

$$W_{u_{x,y}} = \{q \mid |K(x,q)| < |K(y,q)|\} \tag{32}$$

$$W_{l_{x,y}} = \{q \mid |K(x,q)| > |K(y,q)|\} \tag{33}$$

where the function $|K(x,q)|$ extracts the total number of k, which belongs to $K(x,q)$ in slide x. $K(y,q)$ is also a bag of words in slide y, and the words satisfy the same conditions as $K(x,q)$ in Eqs. (32) and (33). Thus, Eqs. (32) and (33) can be used to count the number of words in $K(x,q)$ and $K(y,q)$, for slides x and y, respectively. Furthermore, $W_{u_{x,y}}$ and $W_{l_{x,y}}$ represent bags of words in slides x and y. If the number count for slide x is lower than that for slide y in Eq. (32), the word appearance tendency of q is determined to become upper in slides x and y, and q belongs to $W_{u_{x,y}}$. If the number count for slide x is greater than that for slide y in Eq. (33), the word appearance tendency of q is then determined to become lower in slides x and y, and q belongs to $W_{l_{x,y}}$.

Table 12: Patterns in Differences of Tendency of Word Appearance

	Tendency of word appearance	
Patterns	in a chapter	in slides
tw_1	*dispersion*	*upper*
tw_2	*dispersion*	*lower*
tw_3	*concentration*	*upper*
tw_4	*concentration*	*lower*

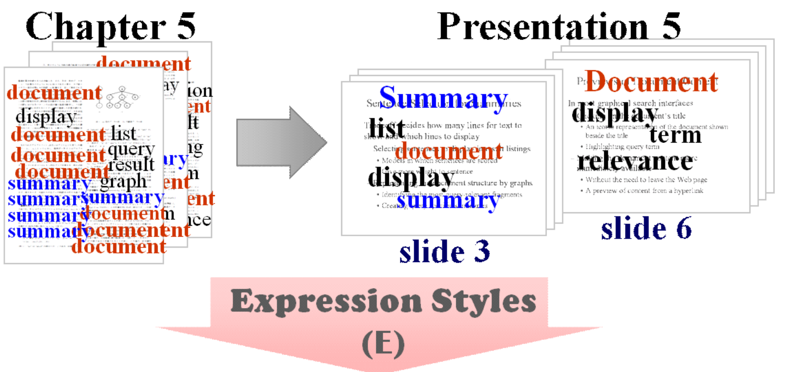
5.2.3 Patterns of Tendency of Word Appearance

Four patterns have been identified to explain the differences between the appearance tendency of a word q in the chapter and slides (Table 12). Explanations for each pattern follow.

- tw_1: $q \in W_{d_{ch}} \cap W_{u_{x,y}}$; the tendency of word appearance of q in the chapter is dispersion and the tendency of word appearance of q in the slides becomes upper.
- tw_2: $q \in W_{d_{ch}} \cap W_{l_{x,y}}$; the tendency of word appearance of q in the chapter is dispersion and the tendency of word appearance of q in the slides becomes lower.
- tw_3: $q \in W_{c_{ch}} \cap W_{u_{x,y}}$; the tendency of word appearance of q in the chapter is concentration and the tendency of word appearance of q in the slides becomes upper.
- tw_4: $q \in W_{c_{ch}} \cap W_{l_{x,y}}$; the tendency of word appearance of q in the chapter is concentration and the tendency of word appearance of q in the slides becomes lower.

Based on the above, we can determine which words should be described in the chapter and the corresponding slides. Further, we can regulate how they are described. In addition, the patterns of the tendency of word appearance provide an indication of the level of explanation needed for words appearing in slides, and whether these words appear dispersed in multiple sections or concentrated in a specified portion of the chapter. In the example shown in Figure 27, 'document' is dispersed in all sections in Chapter 5, but it is also a title for slide

6 in Presentation 5. When this word is dispersed in the chapter, it assumes an upper position in the slides as tw_1. Slide 6 delves into the topic of 'document' in detail and summarizes the information regarding its appearance in all sections of Chapter 5. On the other hand, 'summary' repeatedly appears in a certain section and slide 3 includes this word as its title (Presentation 5). When the tendency of word appearance of 'summary' is concentration in the chapter, it assumes an upper position in slides as tw_3. Slide 3 offers specialized information regarding 'summary' and refers to a concentrated section from Chapter 5.



Slide No. (SN)	Word (W)	Tendency of Word Appearance (TW)	Level (L)
3	summary	tw_3	*1st*
3	document	tw_1	*3rd*
3	display	tw_4	*2nd*

⋮

Figure 27: An Example of Expression Styles

5.3 Outline Generation

5.3.1 Extraction of Expression Styles

To generate outlines for slides, a slide layout is used; it consists of words based on expression styles by using presentation slides made from chapters in textbooks and specified by presenters. Therefore, we have defined the expression style E to denote tendencies of word appearance TW of each word that belongs to W, and the expression of presentation is represented by the level positions L of each word in W that belongs to a corresponding slide in SN. They refer to a table is shown in Figure 27 as follows:

$$E = (SN, W, TW, L) \tag{34}$$

$$W = \{q \mid q \in TW\} \tag{35}$$

$$TW = \{tw_1, tw_2, tw_3, tw_4\} \tag{36}$$

Here, E can be considered as a database with four indexes: SN, W, TW, and L in Eq. (29). SN denotes the slide number in a presentation. In Eqs. (35) and (36), W is a set of words that belong to TW, and TW is a set of appearance tendencies of words in chapters and their corresponding slides illustrate all four patterns (Section 5.2.3). L denotes the level positions of the words in slides via slide structure. For this study, we extracted expression styles to understand what words used in slides were from chapters, and how about their expressions in slides.

5.3.2 Extraction of Corresponding Words

For this study, we considered texts from textbook chapters with text structures similar to trees (Figure 26, preface of Section 5.2). Our definition of a subtree is a tree consisting of a node (section head or subsection head) but not a root node (chapter head) and all of its descendants (subsection head or paragraphs) in a tree T_A (T_B) of text structure within a referred chapter A (a target chapter B). Therefore, we extracted a set of subtrees of T_A and a set of subtrees of T_B. When a word z belonged to root nodes of respective subtrees belonging to T_A and a word z' belonged to root nodes of respective subtrees belonging to T_B, consistency among them caused us to determine that z' in B corresponds to z in A. We have defined a set of subtrees as belonging to $PT(T_A(z))$, and their respective root nodes contained z and $PT(T_A(z))$, which belonged to subtrees of T_A. A set of subtrees belonging to $PT(T_B(z'))$ had root nodes that contained z' and $PT(T_B(z'))$, which belonged to subtrees of T_B. Then, we extracted z' in B by matching $PT(T_A(z))$ and $PT(T_B(z'))$ as part of a structure-matching method [44]. Words such as z in T_A and z' in T_B were not the same, and the structure-matching method helped identify non-linguistic matches and differentiate between seemingly identical structures in different contexts.

For each word, there were many words in the subtrees $PT(T_A(z))$ of T_A and

$PT(T_B(z'))$ of T_B to be compared, and the number of the same structures in the subtrees belonging to $PT(T_A(z))$ and $PT(T_B(z'))$ will be larger. Based on the structure-matching method, we extracted a pair C of z in A and z' in B for the following formula:

$$C = \{(z, z') \mid \frac{1}{2}\Big(\frac{sum(PT(T_A(z)), PT(T_B(z')))}{N_A} + \frac{sum(PT(T_A(z)), PT(T_B(z')))}{N_B}\Big) > \beta, z \in W\} \quad (37)$$

where the function *sum* extracts the total number of the same structures in the subtrees belonging to $PT(T_A(z))$ and $PT(T_B(z'))$. N_A is the total number of the subtrees belonging to $PT(T_A(z))$, and their respective root nodes contain z in T_A. N_B is the total number of subtrees belonging to $PT(T_B(z'))$, and their respective root nodes contain z' in T_B. We calculated the similarity of $PT(T_A(z))$ and $PT(T_B(z'))$ with the above formula. If the formula was greater than a threshold β so that $PT(T_A(z))$ and $PT(T_B(z'))$ were similar, z' was determined to be the corresponding word of z. Thus, z' was the candidate word for using the expression styles of z in A, and we extracted it as appropriate based on word appearance tendencies of z' in B, as compared to the same tendencies in z' and z. Finally, we are able to generate outlines for slides by using the expression style of z' in the same expression style as z, according to Eqs. (34), (35), and (36). In this way, the number of outlines was determined to be the same as the number of referred slides, and outlines were arranged in the same order as the referred slides.

5.3.3 Generation of Outlines for Slides

Presentations consist of slides that rely on a combination of words and images to drive home a point. The way these elements are combined creates the design layout, which is crucial to making slides understandable and memorable. In this paper, outlines were used to design different slide layouts to communicate key points from chapters in lectures. We considered key points as the roles of words from the chapters expressed in slides, and we focused on hierarchical representation of the words, considering the difference in document structure (i.e. text structure, slide structure) that we focused on the word appearance tendencies in chapters so that their slides could be extracted by our proposed method. Therefore, we created slide outlines constructed of different layouts to express words in slides as specified by a presenter (referred slides). Based on the expression styles drawn from referred slides, we can generate outlines for slides from a target textbook chapter by extracting the words from the chapter that correspond to the words in the referred chapter. Therefore, we can semi-automatically generate editable outlines for slides from the target chapter; we simply generate outlines according to the referred slides in the same order and with the same number of slides as the referred slides. In this paper, our generated outlines for making slides focused on hierarchical representation of words to organize the content. Thus, presenters can arrange selected outlines to make a presentation file.

For example, a presenter wants to prepare presentation slides for a lecture regarding Chapter 6 in a textbook. Our method generates outlines for slides for Chapter 6, referring to slides in Presentation 5 from Chapter 5 (Figure 28). In Chapter 5, the word 'document' appears in all sections. The word also appears in the title of slide 6 in Presentation 5, so the expression of 'document' in slide 6 is title (1st level). In Chapter 6, the word 'query' appears in all sections that correspond to 'document' in Chapter 5. The outline for slide y generated from Chapter 6 shows that 'query' appears in the title of slide y, which explains 'query expansion'. 'Query' in slide y has the same expression style as 'document' in slide 6. When the presenter creates slides referring to the outlines, such as slide y, the information for 'query' in slide y is constructed in the same way as for the level position of 'document' in slide 6. It is based on the same expression style by arranging the words to express 'query' in the title of slide y. The generated outlines can be used to create slide layouts that construct words according to the same roles of the words in the referred slides, and the outlines then enable the presenter to make slides easily.

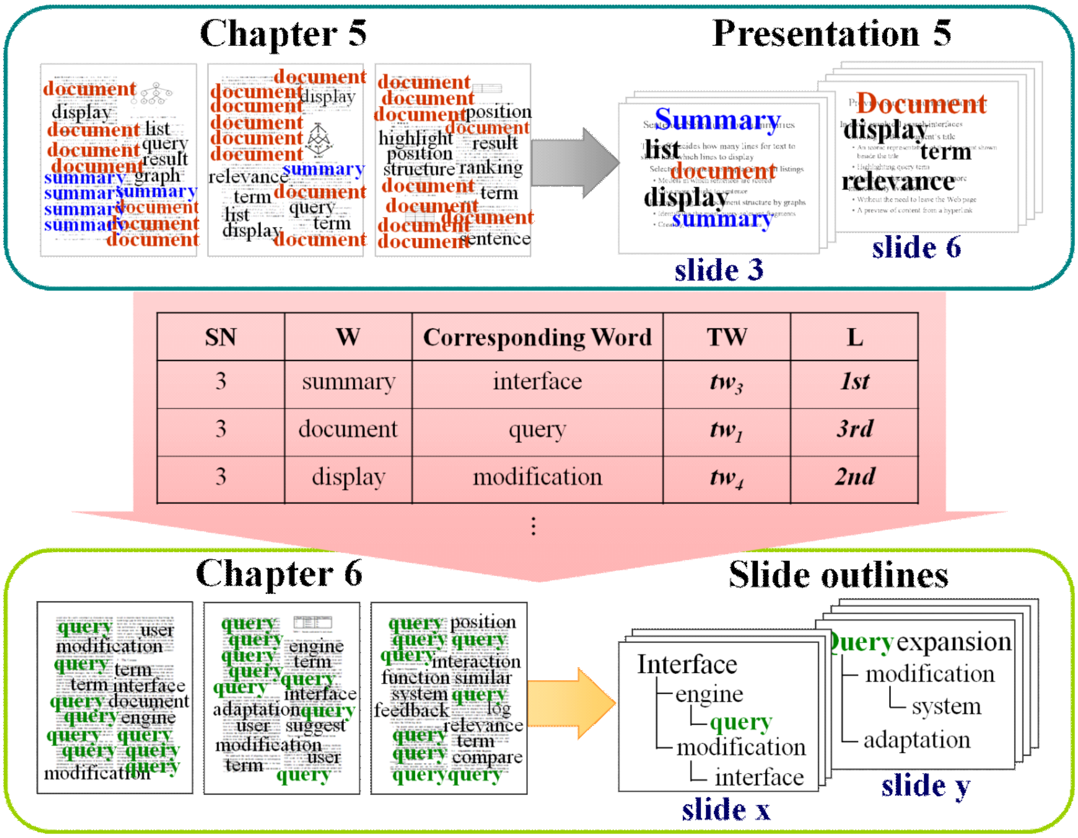


Figure 28: An Example of Outline Generation

5.4 Evaluation

5.4.1 Implementation

Based on the method described above, we built a tool to support outline generation, using Microsoft Visual Studio 2010 C#. The tool has three stages: analysis, determination, and generation. In the analysis stage, we analyze the features of a slide and a chapter. The slide structure and information on the indent level of words are constructed by using Office Open XML files from PowerPoint in Microsoft Office 2007. (In our implementation, we developed a PowerPoint parser, but parsers for Keynote, Open Office Impress, and so on can also be developed. Therefore, we can also use content made by other presentation formats). The text structure of a chapter and information on logic units is constructed using its original LaTeX file. When the chapter is a PDF file, we should convert PDF files into XML files using [68]. The words in the slides can be extracted using the morphological analyzers [49] and [84, 64].

In the determination stage, all expression styles of words in referred slides are extracted based on slide structure, and the patterns of word appearance tendency in chapters and their slides are extracted based on the text structure and the slide structure. Corresponding words from a target chapter can be extracted based on tendency of word appearance in that chapter and the referred chapter by matching subtrees with respective root nodes containing the words in the text structures of the target chapter and the referred chapter. Thus, in the generation stage, slide outlines are generated by arranging the corresponding words from the target chapter based on the expression styles of the words in the referred slides specified by a user.

After a user selects the chapter from a textbook for preparing presentation slides, he/she specifies a presentation file from a chapter in the same textbook for reference. The prototype tool has a function to generate slide outlines as layout structures based on Office Open XML Formats in PowerPoint 2007.

5.4.2 Dataset

The aim of this evaluation was to verify whether our proposed method is useful for slide outline generation. We first prepared two presentation datasets from a textbook called *Search User Interfaces* [25]:

- Dataset 1: S_{A_1} from Chapter A and S_{B_1} from Chapter B were made by the same person (P_1).
- Dataset 2: S_{A_2} from Chapter A and S_{B_2} from Chapter B were made by the same person (P_2).

Because P_1 and P_2 are characterized by single authorship, we assumed that words in S_{A_1} and S_{B_1} had the same expression styles in Dataset 1; the words in S_{A_2} and S_{B_2} had the same expression styles in Dataset 2, and A and B had the same text structure in the same textbook. S_{A_1} and S_{B_1} each contain 10 separate slides, not counting the cover slide in Dataset 1; S_{A_2} contains 11 slides, and S_{B_2} contains 10 slides (Dataset 2). We used A and S_{A_1}, S_{A_2} to generate respective outlines for presentation files O_1, O_2 from B based on our method using different expression styles of slides. S_{A_1} and S_{A_2} are called referred slides; the respective slides in S_{B_1} and S_{B_2} serve as correct answers regardless of whether the hierarchical representation of the words in the respective outlines generated from B are correct or not.

5.4.3 Validity of Outline Generation

We generated 10 outlines in O_1 from B in the same order and with the same number of slides as S_{A_1} and 11 outlines in O_2 from B in the same order with the same number of slides as S_{A_2}. We extracted the corresponding words from B and arranged them in slide outlines based on the expression styles of the words in S_{A_1} and S_{A_2}, respectively. Finally, we compared the generated slide outlines in O_1 and O_2 with the respective correct answers in S_{B_1}'s slides and S_{B_2}'s slides (see Figures 29 and 30). For evaluating the generated outlines, we conducted

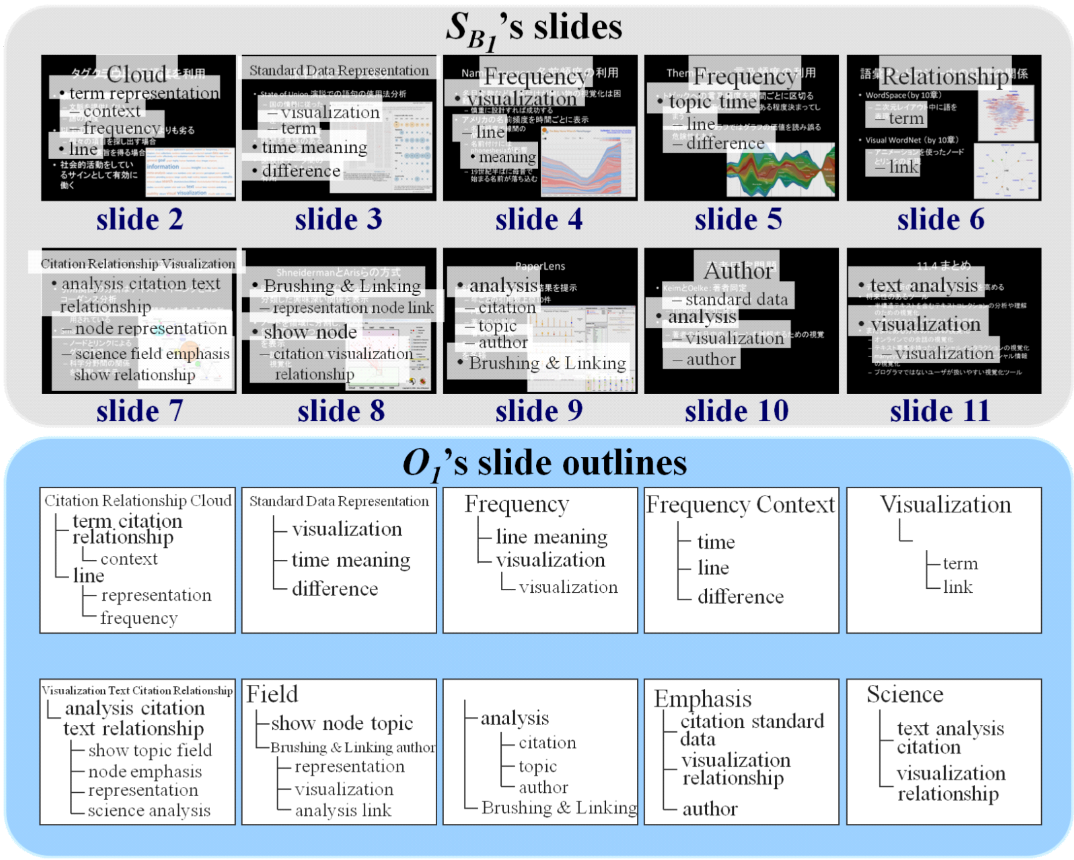


Figure 29: Generated Slide Outlines in O_1 Compared with Slides in S_{B_1}

an evaluation based on two aspects:

(a) measuring the coverage of the words in the generated outlines in O_1 and O_2 that also appear in S_{B_1} and S_{B_2}, respectively. In this way, we calculated the

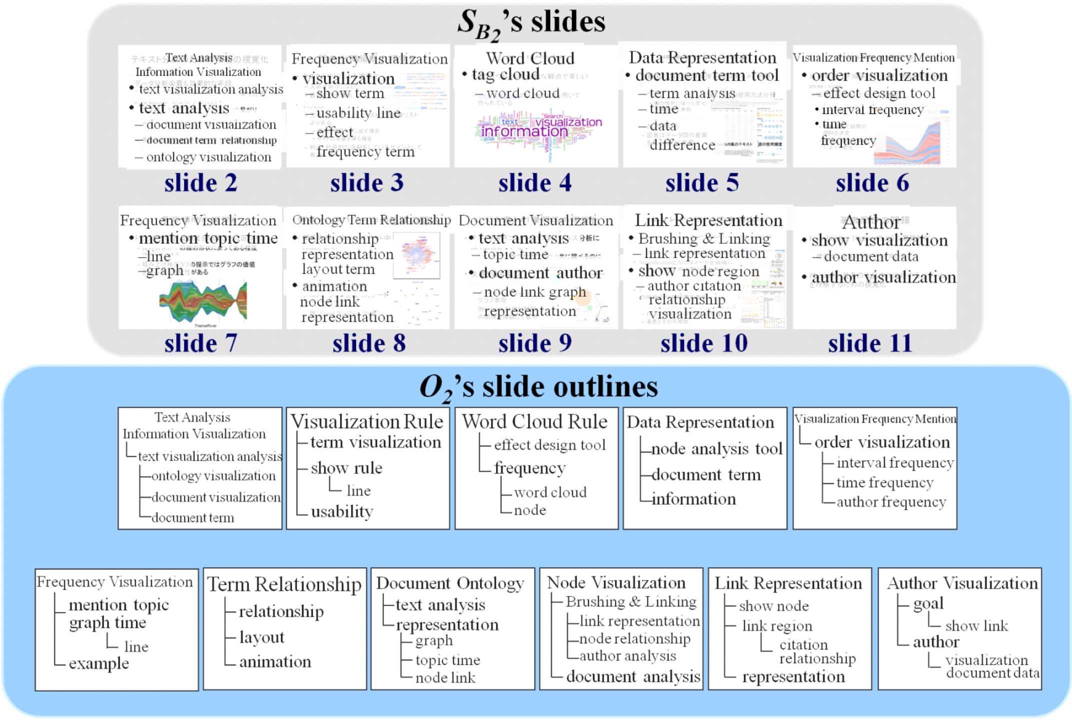


Figure 30: Generated Slide Outlines in O_2 Compared with Slides in S_{B_2}

coverage as a 'recall' of the words in generated outlines without assessing slide structure.

(b) measuring the accuracy of structures in the generated outlines in O_1 and O_2 by comparing them with S_{B_1} and S_{B_2} based on the hierarchical relationships between two words in the generated outlines. In this way, we calculated the accuracy of structures to evaluate whether the generated outlines in O_1 and O_2 maintained them as they were in S_{B_1} and S_{B_2}, respectively.

In the experimental results of our evaluation, (a) the coverage of the extracted words in O_1 reached 33.8% (25/74); in O_2, coverage reached 37.8% (31/82). The average for these was 35.8%. (b) The accuracy of the structures in O_1 was 42.3% (254/25∗24); for O_2, accuracy was 44.8% (417/31∗30). The average accuracy for both structures was 43.6%. The results of (a) indicate that presenters using the method proposed in this study can extract corresponding words by a conventional method (structure-matching method). However, sometimes we extracted words that corresponded to multiple words in textbook chapters, and only a small number of the extracted words were correct. In addition, we considered the figure captions for determining the words in the chapters. S_{A_1} and S_{B_1}, as well as S_{A_2} and S_{B_2}, were written by the same person; several words from these files appear in related slides and in captions of figures included in the chapters. Such variables contributed to a lower rate of coverage for responses.

Experimental results of (b) showed that our proposed method is effective in arranging words in generated outlines based on their expression styles. The rate of accuracy for the structures in generated outlines was low due to the fact that it was dependent on the small number of the extracted corresponding words in (a). Using our method, we determined that the hierarchical relationships between some words in O_1 and O_2 were not in consistency with those in S_{B_1} and S_{B_2}. Figure 31 presents an example of adequate results; an outline in O_1 was generated from B. At the sentence level, the words ‘citation’, ‘topic’, and ‘author’ appear at the third level, and ‘analysis’ and ‘Brushing & Linking’ at the second level in the body of a slide as correct answers in S_{B_1} made by P_1. Our method arranged the corresponding words ‘citation’, ‘topic’, and ‘author’ from B at the third level. Further, ‘analysis’ and ‘Brushing & Linking’ from B appeared at the second level in the body of the generated outline and in S_{B_1}. The hierarchical relationships of these words were the same in S_{B_1}’s slide and in the generated outline in O_1. Figure 32 presents an example of inadequate

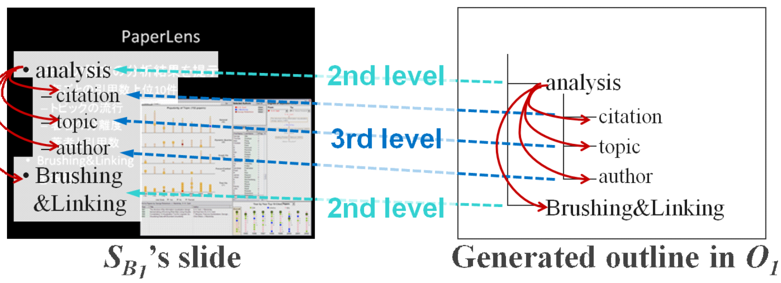


Figure 31: Adequate Generation of an Outline in O_1

results. An outline in O_1 was generated from B. At the sentence level, the words ‘visualization’ and ‘term’ appear at the third level, and ‘time meaning’ and ‘difference’ appear at the second level in the body of a slide as correct answers in S_{B_1} made by P_1. Our method, however, arranged the corresponding words from B, ‘visualization’, ‘time meaning’, and ‘difference’ on the second level in the body of the generated outline in O_1; the hierarchical relationships of ‘visualization’ and ‘time meaning’, as well as ‘visualization’ and ‘difference’ were not the same as in S_{B_1}’s slide.

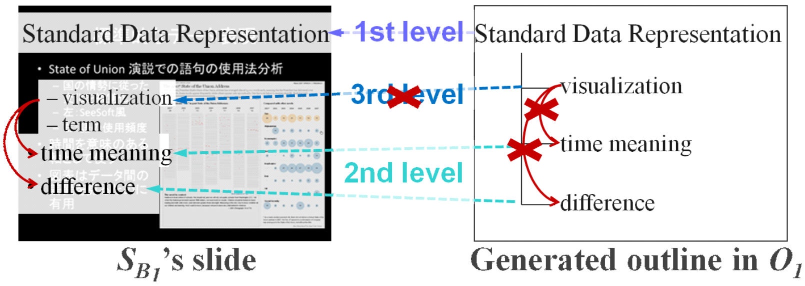


Figure 32: Inadequate Generation of an Outline in O_1

Based on the explanation above, we found that there were only 26.2% (16/61) of the same words in O_1 and O_2 (see Figures 29 and 30). Although O_1 and O_2 were generated from the same chapter (B), they contained their respective typical words to express the content from B. In addition, the same words were characterized by different expression styles in O_1 and O_2. For instance, ‘visualization’ was the main topic of B, both in O_1 and O_2; however, ‘visualization’ was a topic in the title of two outlines in O_1 and almost half of all outlines in O_2. Furthermore, ‘cloud’ was described at the first half of generated outlines in O_1, but it was not described at the first half of generated outlines in O_2. Therefore, we confirmed that our method can provide different outlines from the same chapter according to different expression styles and as determined by strategies of presenters/lecturers.

5.4.4 Discussion

In the evaluation described in the previous subsection, we confirmed that our prototype tool generates outlines with a presentation slide structure that is as expressive as existing presentation slides made by the creator of the first set of slides, as shown in Figures 29 and 30. While the conventional methods generate slides by summarizing content in textbook chapters with limited formats, such as the text structures of textbook chapters, our method can generate slides with more variety in layout of referred slides because such freedom of expression of existing slides are specified by users. Specifically, our prototype tool generated different slide outlines from the same resource (e.g. textbook chapters) according to different expression styles based on difference in document structure (i.e. text structure, slide structure) for effectively organizing content in the outlines.

Although we confirmed that our method generates outlines with expression styles of the referred slides based on the slide structure, we encountered three main problems. The first problem is that the expression styles, which are based on slide structure, are not regarded as visual effects in slides. In the current age of savvy technology, presenters often focus on visual effects that are easily understandable and more attractive than slides with simple text. We did not initially build the use of font or visual information into our methodology, but it would not be difficult to improve our method by considering such data. Future modifications to this method could include enhancements with visual elements, such as figures, as well as color distribution and animation occurrence in slides. We can acquire the relevant information to produce such enhancements by analyzing XML files from the various presentation formats.

The second problem that we encountered is that the hierarchical relationships between words in the body of text in slides do not yield sufficient semantic representation. Therefore, we need to consider semantic relationships (e.g. *compared-with*, *oppose*, etc.) between the words that can be referred by the Rhetorical Structure Theory [45]; additionally, we must determine how to utilize these relationships to generate outlines effectively. The third noteworthy problem is that our outline-generation algorithm does not organize content in slides based on the expression styles of phrases (instead of words). We recognize that instructors often extract phrases from textbook chapters to produce a presentation slideshow for lectures. Determining the expression styles of phrases may offer better support for generating outlines.

5.5 Summary

In this theme, we proposed a method to generate outlines that provides support for slide-making based on expression styles of words in referred slides specified by users. We described our methodology in detail, including how expression styles are extracted with the use of slide structure. Additionally, we explained how we can analyze differences between the words within text structures of textbook chapters and slide structures of their slides by extracting tendency of word appearance from each. To generate outlines for slides from a target chapter, we extracted the words in the target chapter that corresponded to the words in a referred chapter; then, we used the same expression styles of the words in the referred slides to arrange the corresponding words in slide outlines. Through our evaluation, we confirmed that some of the outlines were successfully generated by our semi-automated prototype tool that makes slides by referring to existing slides. For future research, we need to extend the definition of slide outlines; in other words, we plan to improve our algorithm of outline generation for presentation slides from textbook chapters to consider the changes in context between word use in chapters and word use in slides.

Chapter 6 Dynamic Word Clouds: Context-based Word Clouds of Presentation Slides

6.1 Introduction

Presentation slides (e.g., PowerPoint [52], Keynote [2]) are now one of the most frequently used tools for educational purposes. A huge amount of slide-based lecture material, often prepared from teaching material used in actual classes at universities or other educational institutions, is freely shared on Web sites such as iTunes U [27] and MIT OpenCourseWare [55]. Thus, not only students who missed a lecture or presentation, but also anyone interested in the topic can study the presentation on their own. Therefore, techniques are in demand that will efficiently find one or more appropriate slides with content worth learning from the vast numbers of presentations available. Although many techniques for searching and recommending presentation slides have been proposed, some problems remain from the viewpoint of understandability for users browsing search engine results. One problem is a search engine does not consider context when matching user query words within presentation content, leading to a large number of candidate results. Another problem is the difficulty of general quick browsing, that is, when browsing slide titles only, users cannot grasp specifics of the content (see Figure 33(a)). In addition, word clouds of slides based on word frequency can destroy the implicit relevant information between slides and decrease the relevance of words in slides to the overall context (see Figure 33(b)). This makes it difficult to understand the context of words in candidate presentations when choosing relevant files.

As depicted in Figure 34, we present a quick browsing method that 1) generates context-based word clouds for each slide by weighting the words within the context of the presentation (i.e., the intra-slide structure and inter-slide structure) and 2) determines transitions between the word clouds based on relationships between the words in serial slides. In order to achieve our goal, we derive the intra-slide structure that slide structure by focusing on the level

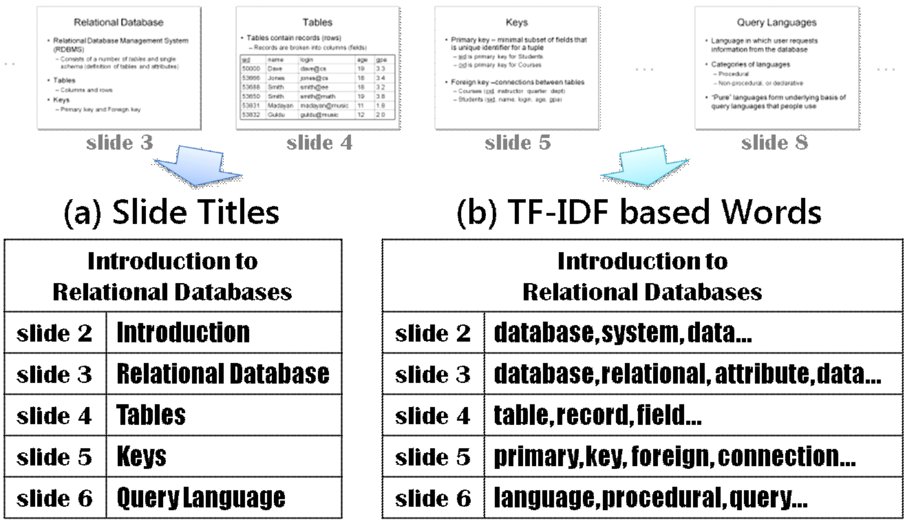


Figure 33: General Quick Browsing Methods

of indentation in the slide text, and determine the inter-slide structure that relationships between slides by considering words that appear at different indentation levels in the structure of other slides. For example, ‘Keys’ appears in the body of text in a slide entitled ‘Relational model’, which related to the slide entitled ‘Relational database’ and ‘Relational model’ appears in the body of text in it. ‘Keys’ is the title of other slide that is related to the slide entitled ‘Relational model’.

In this theme, we define presentation context to mean the context for the slide in a presentation, represented by the relevant information on the slide and allowing for relevant information from the rest of the presentation that is not included on the slide. We define two types of presentation context for a slide: link context and structural context, based on the relationships between slides and slide structure, respectively. Using presentation context, we can generate context-based word clouds of slides by weighting words in the presentation. There are two concepts that are particularly helpful when quick browsing presentation content:

Presentation flow: link or break [75] Often, presentations are formed of a chain of slides such that one slide links to the next. Sometimes, however, a slide will move from the point in a previous slide to a completely different point. In this case, there is a break between them.

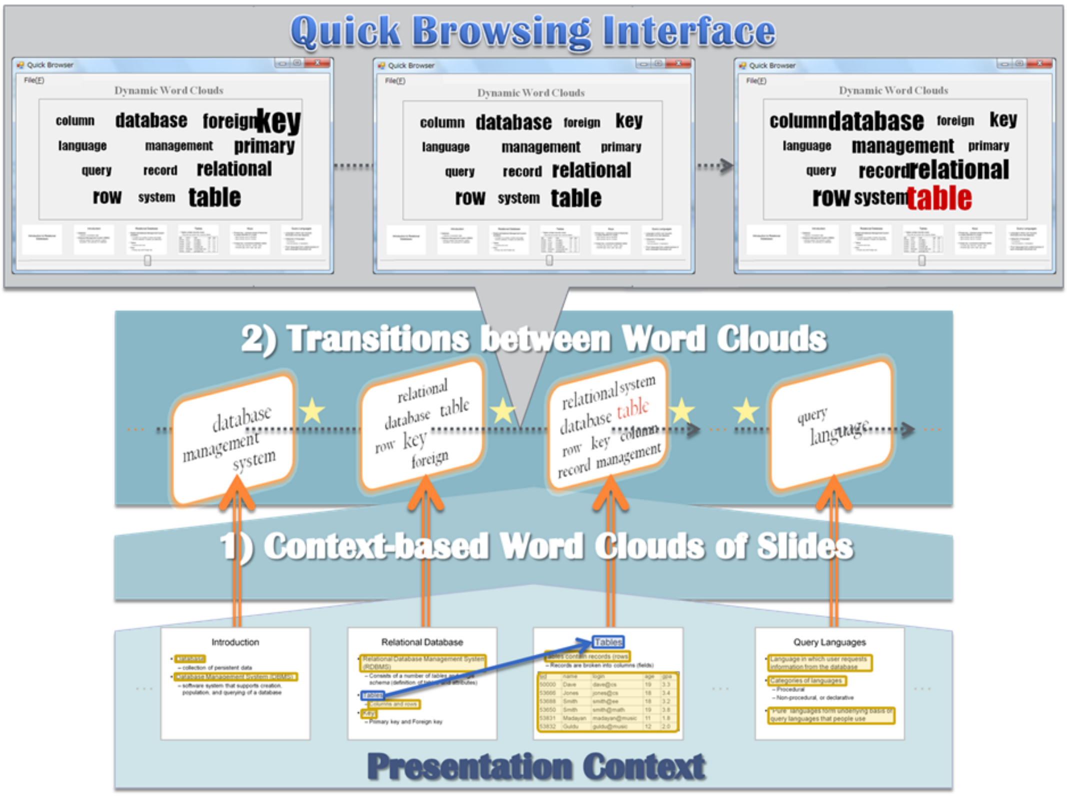


Figure 34: Conceptual Diagram of Our Proposed Quick Browsing Method

Highlight points: semantics This occurs when one slide describes a point from a previous slide in detail.

In this case, our approach presents words interactively from one slide to another as a streaming word cloud reflecting the flow of points in the slides, helping users to select relevant presentations from search engine results easily and effectively.

6.2 Determination of Presentation Contexts for Slides

We determine two types of presentation context for a slide: link context and structural context, based on the relationships between slides and slide structure, respectively. We define the slide whose presentation context is discussed to be the target slide. We construct the slide structure based on the indentations in the slide text. The slide title is the first indentation level; the first item of text within the slide body is the second indentation level, and the depth of the sub-items increases with indentation level (third level, fourth level, etc.). Non-text objects, such as figures or tables, are considered to be at the same indention level as the surrounding text.

6.2.1 Determination of Link Context for Slides

The link context for a target slide consists of links and anchors (similar to hyperlinks in Web pages) related to the text body and titles of other slides. They refer to words in the title of the target slide and titles of other slides that contain words in the body of the target slide. They also indicate from what type of content the target slide is referred. We extract the link context of the target slide by finding the same words at different levels in the target and other slides.

For a given bag of words M in the title and a given bag of words N in a level in the body of the target slide, words in the titles and levels in the body of other slides are extracted: $\cdots$, T_2, B_2, T_1, B_1. Here, T_i is the title of slide i and B_i are the words in a level of the body of slide i. If B_i corresponds to M, B_i can be considered as a link anchor. Then, B_i links to the target slide such that the words in B_i and its slide title T_i belong to the link context for the target slide, while the words in B_i are similar to that in M. This is calculated using the Simpson similarity coefficient [80], as $Sim(B_i, M) = |B_i \cap M|/min(|B_i|, |M|)$. When $Sim(B_i, M)$ exceeds a predefined threshold, the words in B_i and M are similar. Meanwhile, if N corresponds to T_i, N can be considered as a link anchor. Then, N links to the slide titled T_i in that the words in T_i belong to the link context, while the words in N are similar to that in T_i.

In Figure 35, the link context for slide y (in blue portions) shows that slide y explains "Relational Database," which is referred to on slide x as a subheading of 'Introduction,' and the subheading 'Tables' in slide y is described in slide z.

6.2.2 Determination of Structural Context for Slides

The structural context for the target slide consists of lower, current, and upper levels of the target levels corresponding to the link context in the target slide, and lower, current, and upper levels of the link context in other slides based on slide structure. When the target slide does not have a link context, we take the title of the target slide as the target level, and then we extract the structural context for the target slide that consists of the levels below the title.

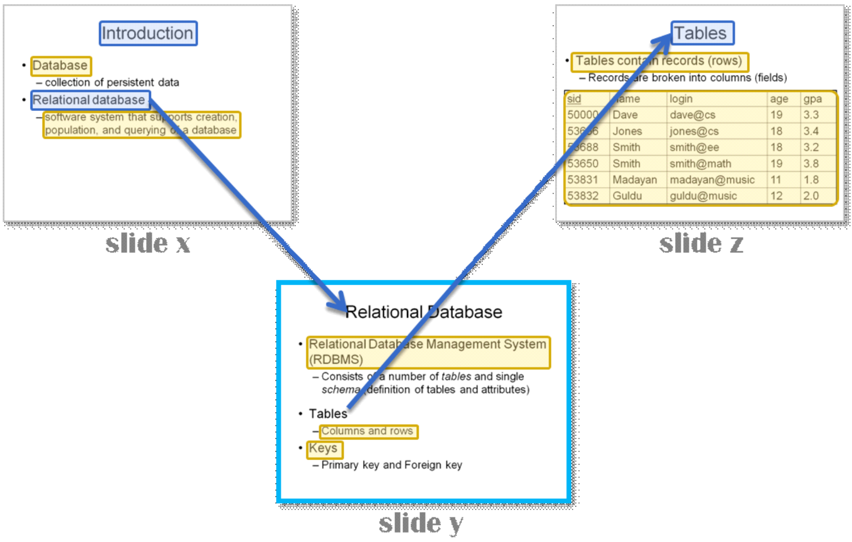


Figure 35: Presentation Contexts for Slide y

For a given bag of words N at a level in the body of the target slide, words in the lower, current, and upper level of N are extracted: $\cdots$, l_3, l_2, l_1. Here, l_j represents a bag of words at a particular indention level j. When l_j and N are the current level in the target slide, the words at level l_{j+1} are at the lower level of N and l_{j-1} are at the upper level of N. The lower, current, and upper levels of the link context in other slides are extracted in the same way.

In our method, when the target level in the target slide corresponds to the link context in more than one slide, we just extract the link context of the slide nearest to the target slide. The link context and structural context are extracted within a minimal range of surrounding information, containing just enough words to characterize the presentation context. Therefore, the presentation context expresses presentation flow and highlight points well.

Figure 35 illustrates the structural context for slide y (in yellow portions), where 'Tables' is related to 'RDBMS,' 'Keys,' and 'Columns and rows' at the upper, current, and lower levels in slide y and 'Tables' includes a link at a lower level, 'Tables contain records (rows),' in slide z; 'Relational Database' includes a link to a lower level at 'RDBMS' in slide y, and 'Relational Database' is related to current and lower levels at 'Database' and 'Software system' in slide x.

6.3 Presentation of Context-based Word Clouds

6.3.1 Generation of Word Clouds of Slides

To present a streaming context-based word cloud that reflects the semantics of the words, slide word clouds are generated from words extracted from the presentation context by weighting the words to determine font size. For each type of presentation context, we calculate the degree of the words that 1) appear close to the target slide and 2) appear frequently near the target slide but less frequently around other slides.

Let us consider each word of target slide S as a relevant object, denoted by o. The degree of o for the presentation context $P(S)$ is defined as follow:

$$W(o, P(S)) = \frac{density(o, P(S))}{dist(o, S)} \tag{38}$$

Here, $density(o, P(S))$ is the density of o for the presentation context $P(S)$, and $dist(o, S)$ is the distance between o and S. Intuitively, $density(o, P(S))$ means how densely the same word as o appears in $P(S)$. If the same word as o appears frequently in $P(S)$ but less frequently in other presentation contexts, $density(o, P(S))$ becomes large. Suppose that S is the kth slide among all slides (the target slide). The density of o in $P(S_k)$ is calculated as follows:

$$density(o, P(S_k)) = \frac{N_{A(o,P(S_k)}}{N_{A(o,U)}} \tag{39}$$

where $A(o, P(S_k))$ is a set of relevant objects representing the same word as the object o in the presentation context $P(S_k)$, and $A(o, U)$ is the set of relevant objects in the presentation context of all slides: $U{=}P(S_1) \cup P(S_2) \cup \ldots$. $N_{A(o,P(S_k))}$ and $N_{A(o,U)}$ represent the number of objects in $A(o, P(S_k))$ and $A(o, U)$, respectively. Because it is difficult to identify the set U due to mutual dependencies between the presentation contexts, we approximate the set U as the set of relevant objects of all slides.

The distance $dist(o, S)$ indicates the strength of the associations between the relevant object o and the target slide S, and is defined for each type of presentation context as follows:

Distance in link context The number of link relationships from the target slide S to the relevant object o.

Distance in structural context The number of parent, brother, and child nodes to be followed from the target levels in the target slide S to the relevant object o.

We generate context-based word clouds of slides by extracting the words based on the ratio of the degree of each word and the highest degree of the word in each slide. We also sort the font sizes of the words into three groups as follows:

$$STag(c, P(S)) = \{c \mid \tfrac{W(c,P(S))}{W_{max}(P(S))} \geq \theta_1, \tfrac{W(o,P(S))}{W_{max}(P(S))} < \theta_2\} \quad (40)$$

$$MTag(c, P(S)) = \{c \mid \tfrac{W(c,P(S))}{W_{max}(P(S))} \geq \theta_2, \tfrac{W(o,P(S))}{W_{max}(P(S))} < \theta_3\} \quad (41)$$

$$LTag(c, P(S)) = \{c \mid \tfrac{W(c,P(S))}{W_{max}(P(S))} \geq \theta_3\} \quad (42)$$

In Eqs. (40), (41), and (42), $W(c, P(S))$ is the degree of c and $W_{max}(P(S))$ is the highest degree of the word in S using Eqs. (38) and (39). Including too many words in each word cloud does not help users to browse them effectively, so we extract c in S such that the ratio of $W(c, P(S))$ and $W_{max}(P(S))$ is greater than a threshold (i.e., 0.25). $STag(c, P(S))$, $MTag(c, P(S))$, and $LTag(c, P(S))$ are the groups of weighted words to be displayed in small, medium, and large font size such that the ratios satisfy Eq. (40), Eq. (41), and Eq. (42), respectively. In this theme, we empirically set the values of the thresholds to be $\theta_1 = 0.25$, $\theta_2 = 0.50$, and $\theta_3 = 0.75$. Although, in general, the word position is important for word clouds, in this work, our dynamic word cloud instead places the word randomly so that the user is not biased to any specific terms based on their placement position. Furthermore, we provide an intuitive interface by changing font sizes only.

6.3.2 Determination of Transitions between Word Clouds

Based upon the presentation contexts for slides, we present dynamic word clouds with visual effects that reflect the relationships between words interactively. For this purpose, we use relationships between words in the word clouds similar to the relationships between slides defined in our previous work [33], which fall into four types based on the presentation contexts for slides:

Detailed **relationship** Titles of other slides belonging to the link context for the target slide in the word clouds. The word clouds of the other slides have more information about the link context than the word cloud of the target slide.

Generalized **relationship** Bodies of other slides belonging to the link context for the target slide in the word clouds. The word cloud of the target slide contains the words about the link context in the outline given in the word cloud of the other slides.

Parallel **relationship** Titles of other slides in the word clouds belonging to the link context that link to the current levels in the target slide, these word clouds of other slides are parallel with each other.

Independent **relationship** Slides do not have a link context for each other in their word clouds.

To present dynamic word clouds, the transitions discussed here explain the kinds of visual effects added to the relationship types, reflecting presentation flow or highlights. Presentation flow consists of many chains of serial slides such that each chain and each transfer switch between chains must be presented. For one chain of serial slides, *detailed*, *generalized* or *parallel* relationships exist between them. For a transfer switch between different chains of slides, the *independent* relationship exists between them. Highlights are the words belonging to the link context of one chain in detail. The effects for three types of transitions between the words in the generated word clouds are as follows:

Font size changes A *shrinking* or *expanding* effect is set between serial slides when presenting one chain in the presentation flow. When the font sizes of the words in the current word cloud are smaller than those in the previous one, a *shrinking* effect is implemented. When the font sizes of the words in the current word cloud are larger than those in the previous one, an *expanding* effect is applied after a *shrinking* effect. Users can easily understand that they are following a chain of slides where the words are mentioned.

Color changes A *coloring* effect is set between serial slides to highlight detailed points in one chain of the presentation flow. For a current word cloud, the words belonging to the next word cloud, which are described in detail on the next slide, are drawn in red. Users easily see that these words are highlighted in one chain. When the highlighted words in the current word cloud are not detailed in the next word cloud, these words are drawn the default color (black).

Switching A *dissolve* effect is applied to a transfer switch between different chains in the presentation flow. The current word cloud disappears and the next word cloud appears gradually in its place. Users easily grasp that a transfer switch has occurred.

6.4 Application

6.4.1 Prototype System

In this paper, based on the method described above, we built a novel quick browsing interface to support users to quickly gain a broad understanding of presentation contents (see Figure 36). The font size of each word is set to be the degree of the word in presentation contexts. This interface also uses color to visualize the detailed points.

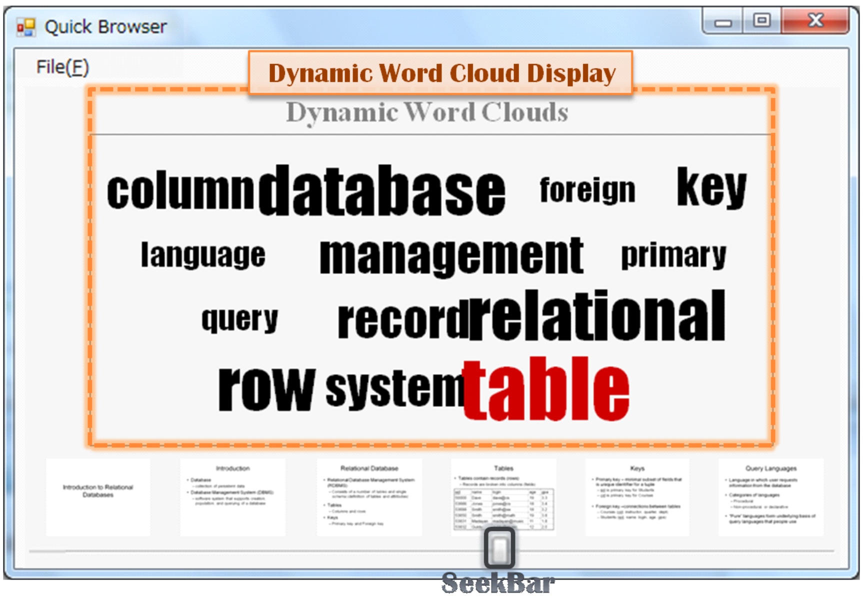


Figure 36: Screenshot of Quick Browsing Interface

Users can specify any presentation content for quick browsing from the results, and the browser presents all words from all word clouds of slides with an initial font size (i.e., 20pt) in a dynamic word cloud display. When a user moves a seekbar to turn over slides, and the weighted words belonging the word clouds of slides with their font sizes (i.e., small: 30pt, medium: 40pt, large: 50pt) are dynamically presented with visual transitions in the word cloud display. We also considered that figures or tables are important visual objects in presentations, we then attempt to build an interface is shown in Figure 37, it has both a dynamic word cloud display and a figure (table) display. In the future, we plan to attempt to build other kinds of interfaces to express the presentation content well.

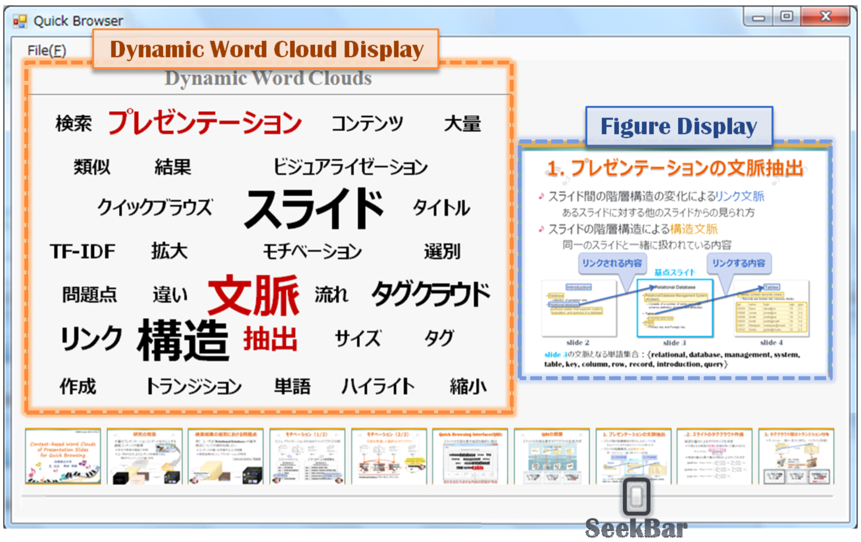


Figure 37: Screenshot of Our Presentation for This Work in Japanese

6.4.2 Validity of Context-based Word Cloud Generation

We confirmed our context-based word cloud generation method by using four presentations from our dataset, P_A, P_B, P_C, and P_D. In here, P_A [70] and P_B [3] are online lecture contents related to database; P_C and P_D are academic presentation contents from DEWS workshops for members of the society [14]. We show an example of extracted weighted words with their values and determined sizes for generating context-based word clouds of P_A named “Introduction to Relational Databases” (see Figure 38).

In this example, slide 3 entitled “Relational Database” that ‘relational’ and

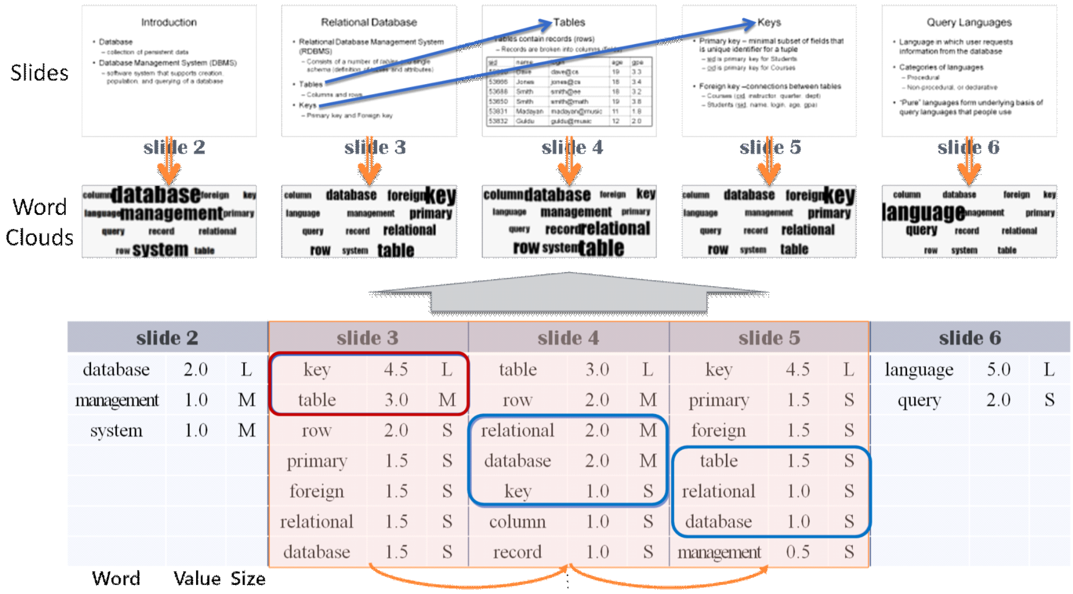


Figure 38: An Example of Extracted Words for Generating Context-based Word Clouds of Slides

'database' are important in general. However, in our method, we considered the context of slide 3 that 'key' and 'table' have high value in slide 3. In addition, for slides 4 and 5, we can extract weighted words such as 'relational' and 'database' that are not included in slides 4 and 5, but these words are related to them. Therefore, we considered that users can grasp the flow of slides 3 to 5 about 'tables and keys in relational databases' well.

6.4.3 Application Examples

When a user wants a presentation about 'relational database overview: tables and keys,' he/she can specify any presentation content for quick browsing from the results. An example of browsing the dynamic word clouds of two presentations from our dataset, P_A and P_B, is shown in Figure 39. In this case, the user browses the flow of slide 3 to slide 5 in P_A named "Introduction to Relational Databases," and the flow of slide 5 to slide 7 in P_B named "An Introduction to Relational Databases." When the user moves the seekbar to go from slide 3 to 5 in P_A, the font sizes of 'table,' 'database,' 'relational,' and 'row' are increased in slide 4, and the font sizes of 'key,' 'foreign,' and 'primary' are increased in slide 5. In particular, 'table' and 'key' are drawn in red in slides 4 and 5, respectively.

On the other hand, when the user moves the seekbar to go from slide 5 to 7 in P_B, the font sizes of 'table,' 'relationship,' and 'entity' are increased in slide 6, and the font sizes of the words such 'SQL,' 'language,' and 'data' are increased in slide 7. In particular, 'metadata,' 'relationship,' and 'SQL' are also drawn in red in slides 5, 6, and 7, respectively.

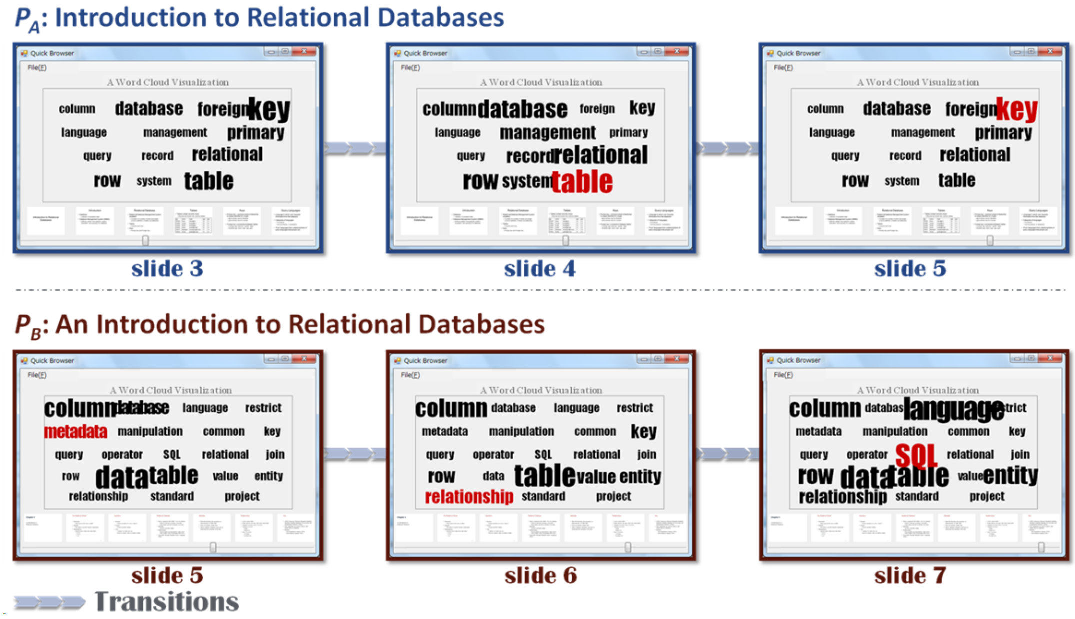


Figure 39: Examples of Transitions between Word Clouds of P_A and P_B

In the case of P_A, we found that 'table' and 'key' are core points in the flow of slides 3 to 5, the presentation explains 'tables and keys in relational databases,' detailed relationships exist between 'table' in slides 3 and 4, and 'key' in slides 3 and 5. There is also a parallel relationship between 'table' and 'key' in slides 4 and 5. For P_B, 'metadata,' 'relationship,' and 'SQL' are core points in the flow of slides 5 to 7, the presentation explains 'characteristics of data in relational database,' and a parallel relationship exists among 'metadata,' 'relationship,' and 'SQL' in slides 5, 6, and 7. Therefore, P_A is worth learning in that it better meets the user's needs. Although we confirmed that our proposed method enables a user effectively and easily select presentations with contents that meet his/her needs, we encountered difficulties when presentations (i.e., P_A and P_B) had a similar title.

We need to consider how best to present the differences in similar presentations (e.g., different topics or same topics with different context information, etc.). Additionally, we must also provide a word cloud visualization that compares presentations simultaneously with the differences clearly marked.

6.5 Summary

In this theme, we proposed a quick browsing method for presentation content that uses dynamic word clouds to present words interactively with visual effects to help a user visually understand the context of content within a presentation. We described how presentation context can be determined from slide structure and the relationships between slides. In order to generate context-based word clouds of slides, we extracted and weighted words from presentation context, and then presented dynamic word clouds with visually transitions that highlighted the semantic relationships between slides. Finally, we confirmed our context-based word cloud generation method with four presentation contents and shown several application examples.

Chapter 7 iPoster: A Collaborative Browsing Platform for Presentation Slides based on Semantic Structure

7.1 Introduction

Slide-based visual presentation support, such as Microsoft PowerPoint or Apple Keynote, is now one of the most frequently used tools for educational purposes currently. However, this format has been criticized repeatedly because of the limitations it imposes on presenters or authors [91]. Enormous amounts of slide-based educational materials that are often based on the collaborative learning teaching materials (i.e., textbooks), are freely shared on Web sites such as Coursera [13] and SlideShare [83]. Thus, students can browse the presentations anywhere, such as on a public display connected to a computer in a classroom or on their own tablets. However, the current slideshow mode of presentations merely permits fluid navigation of linear structures, even while it is being presented to a diverse audience. Moreover, in CSCW tools, group awareness plays an important role in enhancing the effectiveness of the application [8]. In a case of a linear text document, it is challenging to map collaborators to one another, based on mutual interests. Canvas presentations are attempts to mitigate the problems posed by slideware. For instance, Prezi [76] provides an infinite canvas with a zoomable user interface (ZUI) [6] as an alternative to the traditional slides. This interface permits the canvas format to support the creation of expressive layouts. These layouts can be zoomed out, allowing the slide arrangement to be presented in its entirety to the audience [41]. The canvas model was also adopted by pptPlex [72]. In order to effectively support collaborative learning, presenters or authors will be required to create and deliver presentations in a nonlinear fashion. However, this will be time-consuming and pose challenges in designing.

As depicted in Figure 40, we present a collaborative browsing platform that generates a meaningfully structured presentation by transporting the presentation slides. It promotes user interaction and communication and is called the

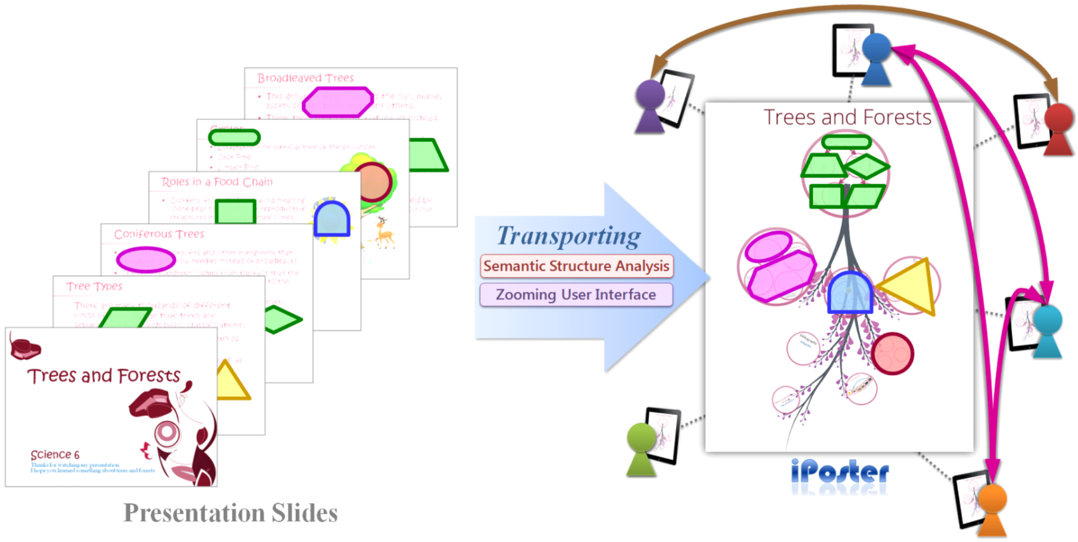


Figure 40: Conceptual Diagram of a Collaborative Browsing Platform based on iPoster

"iPoster," or an interactive poster. Users can access an iPoster on their tablets using a zooming metaphor in cyberspace. They can interactively browse the iPoster through user operations, and connect with each other's tablets. A collaborative browsing platform based on iPoster, which can share and navigate information, matches each user's specific requirements by analyzing the operations of the users. Further, it detects other users who have similar requirements by mapping the similarity in their operations and conveys their interests to each other. iPoster can be implemented by 1) analyzing the semantic structure of textual and graphic elements in slides and the semantic relationships between them; and 2) employing the zooming user interface for organizing elements in structural layouts, using zooming and panning transitions based on a basic idea of Prezi. In semantic structure analysis, we first extract elements by examining the presentation context of the particular element in the slides. The semantic relationships between these elements are determined using implicit hyperlinks in slides, based on a slide structure. Specifically, we derive the slide structure by focusing on the itemized sentences of bullet points present in the slide text. There are various types of structural layouts for constructing an iPoster, such as tree structure, stacked Venn, and pyramid structure. In this paper, in order to provide an overview of the content, we utilize a tree structure, combined with a stacked Venn for an iPoster. Finally, our iPoster is generated based on semantic relationships, using a ZUI for collaborative browsing, which can raise the collaborative awareness, and interaction, besides enabling users to understand the educational presentations easily and efficiently.

7.2 Semantic Structure Analysis of Presentation Slide

In this section, we describe a semantic structure analysis model for extracting elements and determining the semantic relationships between them. Preliminary ideas regarding this model are given in an algebraic query model [73] as well.

7.2.1 Element Extraction

The two most salient and dominant elements in a presentation slide are the set of textual elements and the set of graphic elements. These are based on the itemized sentences of bullet points in the slide text. We define the slide title as the 1st level, the first item of text within the slide body as the 2nd level, and the depth of the sub-items increases with the indentation levels (3rd level, 4th level, and so on). Non-text objects such as figures or tables are considered to be at the same indention level as the surrounding text.

We define textual elements as topics that focus on the nouns in slides. Based on the presentation context, a topic can be described as a learning point with multiple nouns that frequently appears at the higher levels (such as in the slide title) in neighboring slides. Initially, we extract noun phrases using a language analysis toolkit MSR Splat [58, 77] based on the XML files of slides.

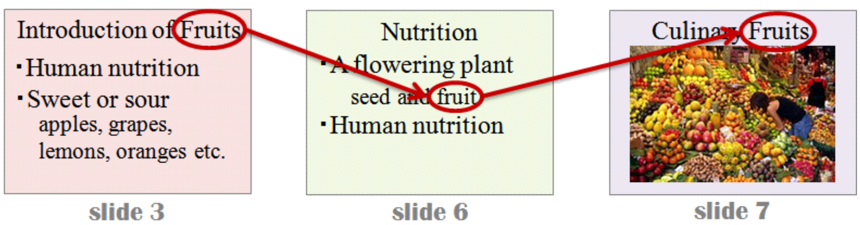


Figure 41: An Example of “fruit” Appears at Different Levels in Slides

The topics that appear in the title of a slide and the body of other slides can be considered to indicate its context in a presentation (see Figure 41). Then, we extract topics by locating the same noun phrases in different slides, at varied levels. If a noun phrase k appears at different levels in slides s_i and s_j, then k is a candidate for being one of the topics T in the presentation. The steps to determine T using k is explained here, which is presented both, in s_i and s_j.

$$T = \{(k, s_i, s_j) \mid l_{max}(k, s_i) \neq l_{max}(k, s_j)\} \tag{43}$$

where, T is a bag of noun phrases that can be considered as candidates for topics. $l_{max}(k, s_i)$ is a function that returns the highest level of k in the slide s_i. For instance, when the highest level is the title, i.e., the first indentation level, of s_i, then $l_{max}(k, s_i)$ returns 1; and when the highest level is the third indentation level of s_j, then $l_{max}(k, s_j)$ returns 3. When k appears at different levels, k is determined as a candidate for topics provided $l_{max}(k, s_i)$ is not equal to $l_{max}(k, s_j)$. Then, the weight of k in T is defined using the levels of k, and the distance between slides s_i and s_j, as follows:

$$I(k) = \frac{1}{l_{max}(k, s_i)} + \sum_{k, s_i, s_j \in T} \left(\frac{1}{l_{max}(k, s_j)} \cdot \frac{1}{dist(s_i, s_j)} \right) \tag{44}$$

where $l_{max}(k, s_i)$ indicates the weight of k in s_i, i.e., it returns the highest level of k in slide s_i by Eq. (43). $dist(s_i, s_j)$ corresponds to the strength of the association between s_i and s_j, and it denotes the distance between s_i and s_j. Thus, if k appears at a high level in s_i and s_j, and the distance between s_i and s_j is short, the weight $I(k)$ of k is high.

When compared to pure textual elements, images are more attractive, appealing and informative from a psychological standpoint. Based on the study of search results presentation [40], it can be noted that summaries with images assist in quicker understanding of the results, thereby helping in arriving at relevant judgments faster. Therefore, we define graphic elements as images corresponding to the topic candidates in slides, given that the noun phrases in the surrounding text of the images are similar to the topic candidates. We considered that the images used to describe the content in slides, and a slide title can be a subject of the content. This is calculated using the Simpson similarity coefficient [80]. The surrounding text can be selected from any portion of the slide, from its title to its body (i.e., from the high level to the low level). When the similarity exceeds a predefined threshold, the noun phrases in the surrounding text and the topic candidates are considered similar. Then, the images are recognized as the corresponding images of the topic candidates.

7.2.2 Determination of Semantic Relationships between Elements

Semantic relationships between elements are determined from a document tree of a presentation to enable users obtain relevant information between the key elements at a glance, for a quick understanding of the content.

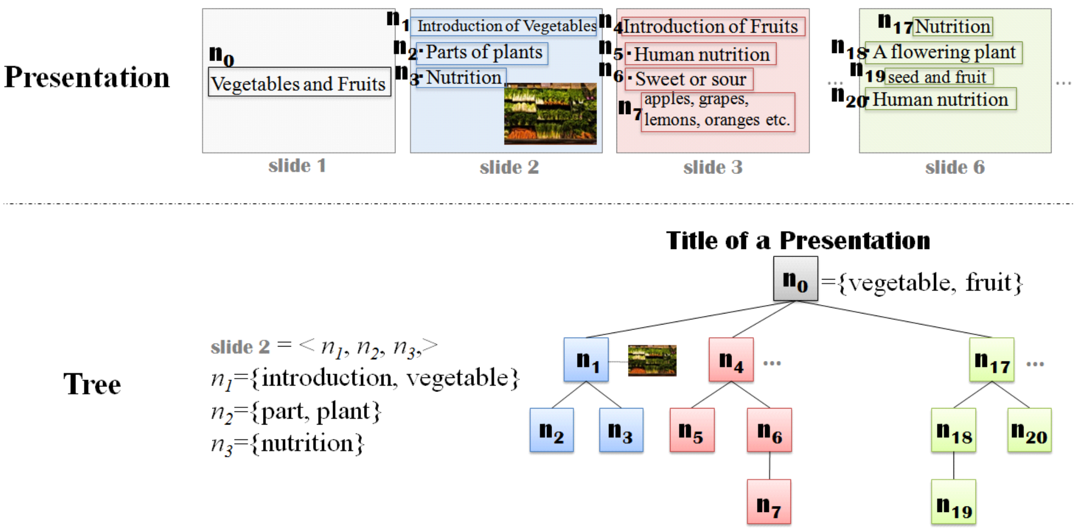


Figure 42: An Example of a Presentation and Its Tree Representation

Basic Definitions and Algebra

The presentation shown in Figure 42 is represented as a rooted ordered tree $D = (N, E)$ with a set of nodes N and a set of edges $E \subseteq N \times N$. There exists a distinguished root node from which the rest of the nodes can be reached by traversing the edges in E. Each node, except the root, has a unique parent node. Each node n of the document tree is associated with a logical component, such as $< title >$ or $< sections >$, based on the bullet points of slides using an XML file in the given presentation. There is a function $words(n)$ that returns the representative noun phrases of the corresponding component in n. A partial tree of the document tree D with a given noun phrase as its root is defined as a fragment f. It can be denoted as $f \subseteq D$. A slide is a fragment by the slide title. In Figure 42, $< n_1, n_2, n_3 >$ is the set of nodes in slide 2 and a fragment of the sample document tree.

To formally define the semantic relationships between the noun phrases from the extracted elements, we first define operations on fragments, and sets of fragments using a pairwise fragment join [73]. Let F_x and F_y be two sets of fragments in a document tree D of a given presentation, then, the pairwise fragment join of F_x and F_y, denoted as $F_x \bowtie F_y$, is defined to extract a set of fragments. This set is yielded by computing the fragment join of every combination of an element in F_x and an element in F_y, in pairs, as follows:

$$F_x \bowtie F_y = \{f_x \bowtie f_y \mid f_x \in F_x, f_y \in F_y\} \tag{45}$$

Figure 43 illustrates an example of operation for pairwise fragment join. It refers the sample document tree in Figure 42. For the given two noun phrase x $= nutrition$ and $y = fruit$, where $F_x = \{< n_3 >, < n_5 >, < n_{17}, n_{18}, n_{19}, n_{20} >, < n_{20} >\}$, $F_y = \{< n_4, n_5, n_6, n_7 >, < n_{19} >\}$, $F_x \bowtie F_y$ produces a set of fragments $\{< n_3 > \bowtie < n_4, n_5, n_6, n_7 >, < n_5 > \bowtie < n_4, n_5, n_6, n_7 >, < n_{17}, n_{18}, n_{19}, n_{20} > \bowtie < n_4, n_5, n_6, n_7 >, < n_{20} > \bowtie < n_4, n_5, n_6, n_7 >, < n_3 > \bowtie < n_{19} >, < n_5 > \bowtie < n_{19} >, < n_{17}, n_{18}, n_{19}, n_{20} > \bowtie < n_{19} >, < n_{20} > \bowtie < n_{19} >\}$ on applying Eq. (45).

Semantic Filters

We determine semantic relationships between the given noun phrases, x and y, from the extracted elements using the set of fragments produced by taking pairwise fragment join as semantic filters. For this, we define four types of semantic filters by considering the horizontal and vertical relevance, as well as the structural semantics from the document tree of the given presentation.

Horizontal distance Logically interrelated slides of a presentation are typically close to each other. Therefore, is such presentations, the horizontal distance between nodes in different slides of a document tree is a reasonable measure of the inter-relationship between nodes. Specifically, when the

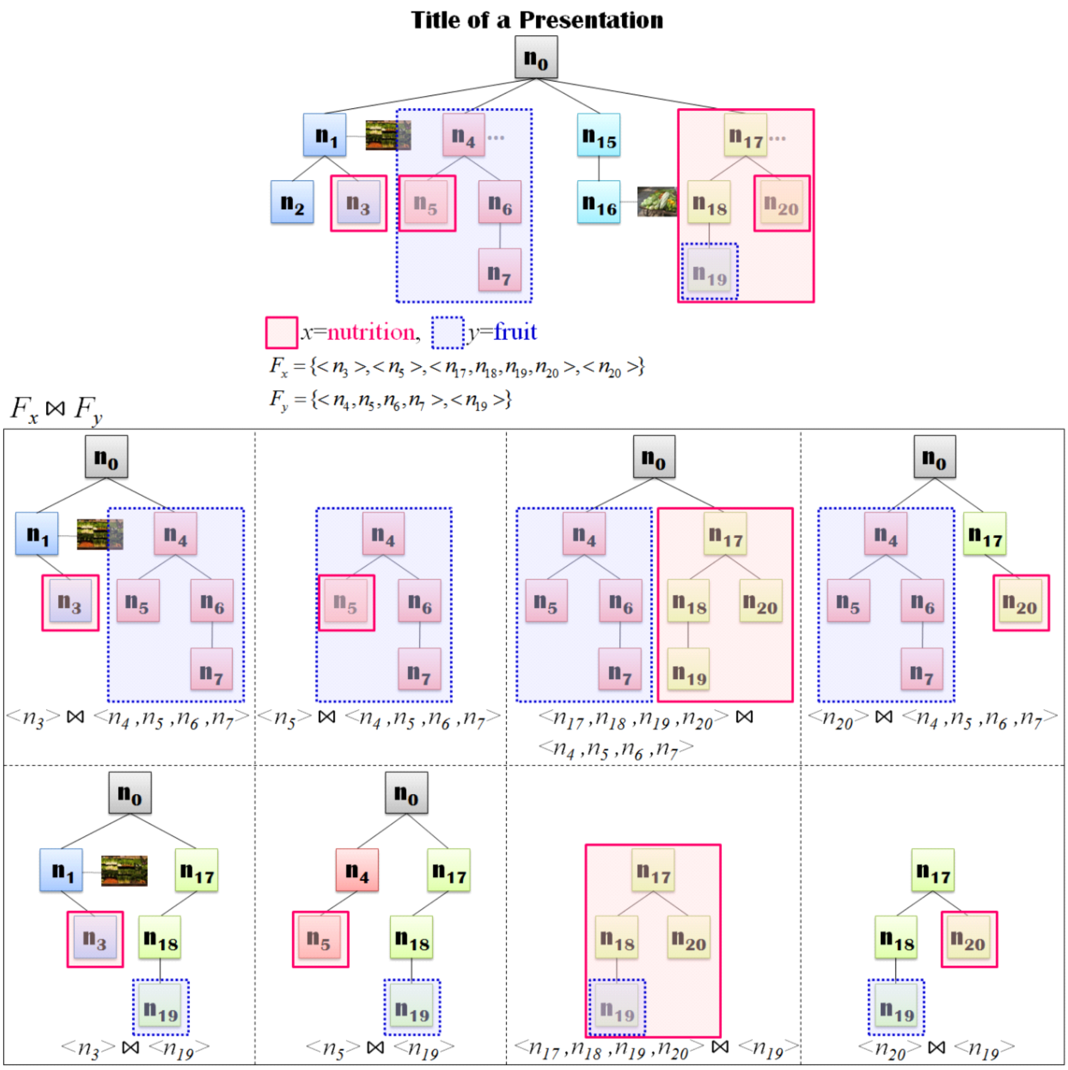


Figure 43: An Example of Pairwise Fragment Join

horizontal distance between the nodes in slides containing x and y exceeds a certain threshold, x is irrelevant to y. Supposing, $hdist(t_i, t_j)$ denotes the distance between the nodes of the slide titles t_i and t_j in slides containing x and y, we set the threshold value α at $|N|/2$, i.e., half the total number of nodes N in the document tree, for normalizing various presentations. If $hdist(t_i, t_j)$ does not exceed α, then the distance between two slides containing x and y is short (i.e., relevant); contrarily, if $hdist(t_i, t_j)$ exceeds α, the distance between two slides containing x and y is long (i.e., irrelevant).

Vertical distance Logically, indentations of slides are typically close to each other. Therefore, when the distance between the slides containing x and y is long, and x and y are at the low levels in slides, they can be less relevant in the document tree. When vertical distance between the nodes in slides containing x and y exceeds a certain threshold, and x and y are at the low level in the slides, x is irrelevant to y. Supposing, $vdist(r, q)$ denotes the distance between the root node r and the node containing each given noun phrase q (e.g., x or y), we set the threshold value β at $ave(depth)$, which is an average of the depth of levels in the document tree, for normalizing various presentations. If $vdist(r, q)$ does not exceed β, then the level of the node containing x or y is high (i.e., relevant); contrarily, if $vdist(r, q)$ exceeds β, the level of the node containing x or y is low (i.e., irrelevant).

Hierarchy For judging the semantics of x and y, we compare the levels of x and y in the fragments based on the theory of hierarchical semantics. When $l(x) < l(y)$, it denotes that the level of x is higher than the level of y; x is a superordinate concept of y (y is a subordinate concept of x). Contrarily, $l(x) > l(y)$ denotes that the level of x is lower than the level of y; x is a subordinate concept of y (y is a superordinate concept of x). When $l(x) = l(y)$, this denotes that the level of x is same as the level of y; they have coordinate concept with each other.

Inclusion We can consider the inclusion relationships between the fragments of x and y. When $f_x \subseteq f_y$, it denotes that the fragment of x is included in the fragment of y, i.e., f_x is a partial tree of f_y. Contrarily, when $f_x \supseteq f_y$, it denotes that the fragment of x includes the fragment of y, i.e., f_y is a partial tree of f_x.

Semantic Relationship Type

We determine five types of semantic relationships between the given noun phrases, x and y, by combining the semantic filters of Table 13. For measuring the relevance between x and y, we focus on the *horizontal distance* and the *vertical distance*. Here, when the *horizontal distance* between them is long, the *vertical distance* should be short. We determine hierarchical relationships, x *shows* y, x *describes* y, and x *likewise* y, by focusing on *hierarchy*. In x *shows* y, $l(x) < l(y)$ means x is a superordinate concept of y (y is a subordinate

Table 13: Semantic Relationships with Semantic Filters

Types	Horizontal distance	Vertical distance	Hierarchy	Inclusion
x *shows* y	$< \alpha$	*either*	$l(x) < l(y)$	*either*
x *shows* y	$\geq \alpha$	$< \beta$	$l(x) < l(y)$	*either*
x *describes* y	$< \alpha$	*either*	$l(x) > l(y)$	*either*
x *describes* y	$\geq \alpha$	$< \beta$	$l(x) > l(y)$	*either*
x *likewise* y	$< \alpha$	*either*	$l(x) = l(y)$	*either*
x *likewise* y	$\geq \alpha$	$< \beta$	$l(x) = l(y)$	*either*
x *part-of* y	$< \alpha$	*either*	*either*	$f_x \subseteq f_y$
x *part-of* y	$\geq \alpha$	$< \beta$	*either*	$f_x \subseteq f_y$
x *has-a* y	$< \alpha$	*either*	*either*	$f_x \supseteq f_y$
x *has-a* y	$\geq \alpha$	$< \beta$	*either*	$f_x \supseteq f_y$

concept of x). In x *describes* y, $l(x) > l(y)$ means x is a subordinate concept of y (y is a superordinate concept of x). Then, *show* and *describe* are functionally interchangeable, when x *describes* y is from the viewpoint of y *shows* x. In x *likewise* y, $l(x) = l(y)$ means x and y have coordinate concept with each other. We determine inclusion relationships, which are x *part-of* y and x *has-a* y, by focusing on *inclusion*. In x *part-of* y, $f_x \subseteq f_y$ means that the concept of x is included in the concept of y. In x *has-a* y, $f_x \supseteq f_y$ means that the concept of x includes the concept of y. Then, *part-of* and *has-a* are functionally interchangeable, when x *has-a* y is from the viewpoint of y *part-of* x. When x and y fail to match these determinations of semantic relationships, x and y are independent. Therefore, a numbers of semantic relationships between x and y are formed from a set of fragments produced by taking the pairwise fragment join; a semantic relationship is determined by majority.

In this work, the semantic relationships follow a transitivity law, e.g., iff x *shows* y, y *shows* z, then it is assumed that x *shows* z.

7.3 iPoster: Interactive Poster Generation

We generate an iPoster possessing the following two features: (1) Providing an overview of elements from the slides, retaining this feature of traditional posters; and (2) Utilizing a zooming user interface, reflecting the semantics of the elements and promoting user interaction.

7.3.1 Determination of Element Layouts

For providing an overview of elements from slides, we attempt to determine the element layouts by utilizing a tree structure combined with a stacked Venn, based on the semantic relationships between the elements. When hierarchical relationships exist between two elements, i.e., either *show*, *describe*, or *likewise* exists between the elements, they reveal a hierarchy between those elements, as applied to a tree structure. *Show* or *describe* maps a parent-child relationship in the tree structure; for instance, if x *shows* y (y *describes* x), then we mark x in a parent area and y in a child area, suggesting that the layer of x is higher than the layer of y. Additionally, *likewise* maps a sibling relationship in the tree structure; for instance, if x *likewise* y, then we locate x and y in the same layer. Inclusion relationships between two elements, i.e., *part-of* and *has-a*, reveals a logical relationship of inclusion and exclusion applied, as to a stacked Venn. For instance, x *part-of* y (y *has-a* x), we conceive an area of x that is included in an area of y, and that the area of y is larger than the area of x.

7.3.2 Determination of Transitions between Elements

To utilize a zooming user interface for navigating through presentations, the transitions discussed here explain the kinds of visual effects that are applied to the semantic relationship types, to reflect the meaning of the elements from the slides. We animate the zooming and panning transitions for navigating through elements in the structural layout based on the semantic relationship types; this can help users to visually understand the overview and details of the contents within a presentation.

Transitions for *show* (*describe*) When the hierarchical relationship, *show* (*describe*), between two elements is not included in an inclusion relationship, i.e., *part-of* (*has-a*), then, firstly the view must be zoomed-out from the focused element to an overview of the tree structure, following which, it must be zoomed-in to the target element. In addition, when *show* (*descibe*) between two elements is included in the inclusion relationship, the transitions between two elements includes zooming-out from the focused element to the whole element area in the stacked Venn, and zooming-in to the target element. Therefore, the transitions include passing through the overview or the whole element area, which helps users to easily grasp the super-sub relation existing between them.

Transitions for *likewise* When the hierarchical relationship, *likewise*, exists between two elements, the transitions between the two elements include zooming-out from the focused element to an area enclosing both the elements and their parent element, and then zooming-in to the target element. Therefore, the transitions indicate the presence of the parent element; thereby elucidating to the user the existence of a subservient relationship.

Transitions for *part-of* **(***has-a***)** When the inclusion relationship, *part-of* (*has-a*), exists between two elements, the transition between the two elements pans from the focused element to the target element. Therefore, this simple and direct transition between the two elements helps users to easily understand that they are dependent on each other, and that there exists an inclusion relationship between them.

In addition to the above, the transitions between two independent elements include zooming-out from the focused element to all elements, and then zooming-in to the target element. Therefore, these transitions help the users to easily know that they are irrelevant with respect to each other in an iPoster.

7.4 Collaborative Browsing Platform based on iPoster

Based on the method described above, we build a novel collaborative browsing platform that aids users to interactively gain a broad understanding of the presentation slides, based on the users ' operations and our semantic structure analysis. We generated an interactive poster using an online lecture material called "Trees and Forests [90]." As depicted in Figure 44, our iPoster provides an overview of "Trees and Forests," containing key points such as "Forests," "Forest Ecosystem," "Forests and Humans," "Ecosystem Members," "Products," and "Food Chain." Users A, B, C, D, and E are interactively browsing our iPoster that is shared in the cyberspace, operating on their tablets from anywhere.

In this case of collaborative browsing with iPoster, 1) iPoster can share the most important topics with each other and represent information that meets

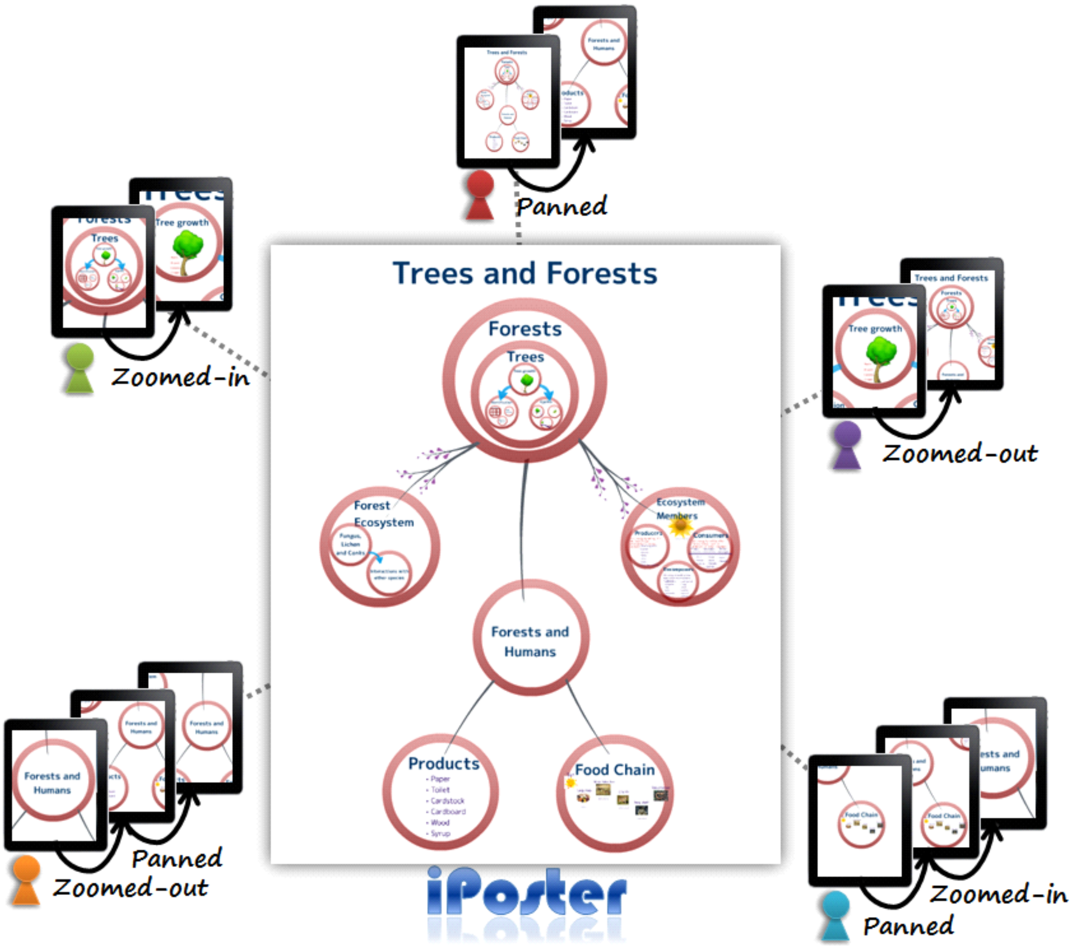


Figure 44: An Example of a Collaborative Browsing Platform based on iPoster

each user's specific requirements on certain topics, 2) Users can detect other users who have the similar requirements on certain topics and share their interests with each other through their tablets. An example is shown in Figure 45, the area of "Tree growth" is highlighted on iPoster sharing on all tablets, because of the area of "Tree growth" is zoomed-in on the tablets of B, D, and E by their zoom-in operation. Because "Identification" and "Cycles" *describes* "Tree growth," and are included in "Trees," we assumed that the users want to get details of "Tree growth" with their zoom-in operation. Therefore, the iPoster represents the transitions between the area of "Tree growth" and the areas of its details (including "Identification" and "Cycles") on the tablets of B, D, and E. As shown in Figure 45, on the tablets of B, D, and E, the iPoster firstly zooms-out from the area of "Tree growth" as shown in **i**, to the

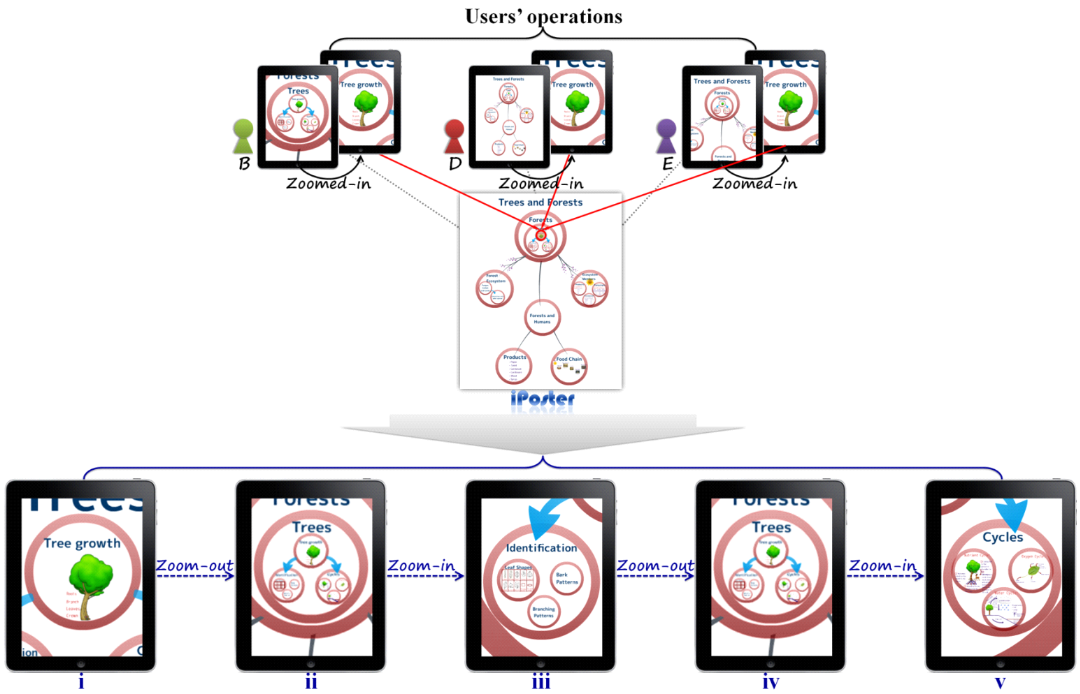


Figure 45: An Example of Showing Details of "Tree growth" based on Zooming Transitions

whole area of "Trees" as shown in **ii**; this conveys to B, D, and E about the whole concept "Trees," which contains "Tree growth," after that, the iPoster zooms-in to the area of "Identification" as shown in **iii**. This enables B, D, and E to understand that "Identification" is a detail of "Tree growth." Next, the iPoster zooms-out from the area of "Identification" as shown in **iii** to the whole area of "Trees" again as shown in **iv**, following which it zooms-in to the area of "Cycles" as shown in **v**. This enables B, D, and E to comprehend that "Cycles" is a detail of "Tree growth" as well. In general, "Tree growth" is a rather uncommon subject for content; however, in this work, we supposed that it was a topic which is worthy to know, considering that many users focused on it. In addition, we can represent the relevant information based on semantic structure analysis, by deriving the users' requirements from their operations.

On the other hand, when A pans from the area of "Products" to the area of "Forests and Humans," and C pans from the area of "Food Chain" to the area of "Forests and Humans," areas including "Forests and Humans," "Products," and "Food Chain" are represented on their tablets and they can share their

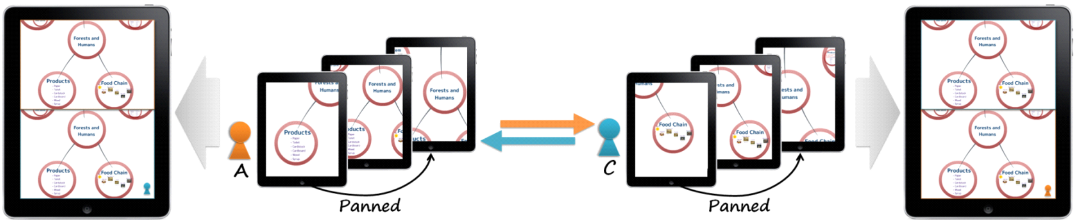


Figure 46: An Example of Sharing Screens of Users A and C with Each Other

screens with each other using their tablets (see Figure 46). In this case, since "Products" *describes* "Forests and Humans," "Food Chain" *describes* "Forests and Humans," and "Products" *likewise* "Food Chain," based on our semantic structure analysis, we considered that A and C have similar needs concerning the topic "Forests and Humans" by panning from its coordinate subtopics, "Products" and "Food Chain." Then, we display the whole area of "Forests and Humans" and share their screens with each other in order to support them to compare their interests, and to promote their communication.

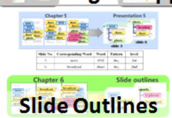
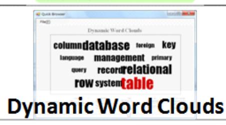
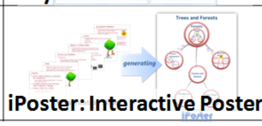
7.5 Summary

In this paper, we built a collaborative browsing platform for presentation slides based on interactive poster generation, called the "iPoster," for presenting elements (i.e., textual and graphic elements) in a meaningfully structured layout with automatic transitions, such as zooms and pans, to promote user interaction. Especially, we introduced a semantic structure analysis model for extracting elements and determining the semantic relationships between the elements of the slides. In order to generate an iPoster in a zoomable canvas, we initially placed the elements in a tree structure combined with a stacked Venn. We then attached the zooming and panning transitions between the elements, based on the semantic relationship types. iPoster enables users to interactively and collaboratively browse, and understand educational presentations easily and efficiently using their tablets.

Chapter 8 Conclusions

In this doctoral dissertation, in order to management of presentation contents, we studied on a structural and semantic analysis of relevant information, structural information, and contextual information from presentation contents. For this, we proposed three approaches: a) we explored semantic relationships inside between slides or scenes; b) we analyzed expression styles of existing slides; c) we presented presentation context intuitively. As a whole, we could confirm our approaches enable us to advance next-generation presentation contents and furthermore to conduct structural and semantic analysis for presentation content management that support for retrieval to readers or searchers, generation to presenters or authors, and grasping overviews to readers. We proposed our five methods based our approaches which are summarized in Table 14 and as follows.

Table 14: Summary of Our Methods

Theme	Approach	Method	Application	Usage
(1)	Semantic Relationships	Scene Combination for Slides with Recorded Videos	 Scene Combination	Retrieve
(2)	Semantic Relationships	Semantic Slide Ranking and Snippet Generation	 Slide Ranking & Snippets	
(3)	Expression Styles	Outline Generation for Presentation Slides	 Slide Outlines	Generate
(4)	• Semantic Relationships • Presentation Context	Dynamic Word Clouds of Presentations	 Dynamic Word Clouds	Overview
(5)	• Semantic Relationships • Presentation Context	iPoster: A Collaborative Browsing Platform for Slides	 iPoster: Interactive Poster	

(1) Scene Combination for Slides with Recorded Videos We developed a system of automatically generating learning channels for readers to extract scenes and combined scenes from slides with their recorded video based on semantic relations. The system analyzed the type of semantic relation on the basis of the metadata of structural information, such as indents and texts in slides, and the set of keywords in the text of the speech in the video. In this way, our newly generated learning channels let users easily focus on either highly detailed slides or introductory slides without needing to examine all of the data. We described our method and the semantic relations between the scenes and discussed a prototype system with evaluation experiments focusing on precious of our proposed method. In this case, we could show that there is a possibility to utilize semantic relations analyzed by exploiting heterogeneous media features of presentation contents.

(2) Semantic Slide Ranking and Snippet Generation We built a slide retrieval system for searchers involving i) semantic ranking and ii) snippet generation, and we discussed how to present the retrieval results to users by considering what rank orders of the slides related to a query and what portions of the slides are relevant to the query, on the basis of the relationships between slides. These methods are based on the keyword conceptual structure of the semantic relations that implicitly exist between keywords, and the document structure of the indent levels in the slides. With our novel i) slide ranking method and ii) snippet-generation method, not only precise retrieve target slides but also the semantic ranking of them, thus ranking either highly detailed slides or generalized slides in an order to help users easily learn through slides; and the relevant portions of them in the presentation by focusing on portions from either detailed or generalized slides, thus giving their surrounding context to help users easily determine which slides to learn are useful or not. Finally, we cloud show the effectiveness of our methods with evaluation experiments focusing on precious of our proposed method that enable the users to browse slide rankings and snippets of the retrieved slides efficiently and effectively.

(3) Outline Generation for Presentation Slides Although most slides generated by conventional methods follow structured document summaries (e.g. academic papers), our method has been designed to generate outlines for lecture slides from textbook chapters. We aimed to organize slide layouts from target chapters based on the expression styles of referred slides by presenters or authors. Therefore, we analyzed level positions of words in the referred slides and arranged words from target chapters to generate slide outlines based on difference in document structure (i.e. text structure within a chapter, slide structure within a slide). To achieve this, we extracted differences between tendency of word appearance in chapters and their associated slides (referred slides). This method generated slide outlines by using the expression styles of the corresponding words from the target chapters in the same layout as that of the referred slides. Finally, we could show the possibility to generate slides outlines by reusing the expression styles of words in the referred slides.

(4) Dynamic Word Clouds of Presentations We developed a quick browsing tool to help readers easily and effectively grasp overviews of presentations. For the purpose, we provided a word cloud visualization that summarizes information to help the users visually understand the context of each presentation. Words important to the "presentation context," is first extracted based on components of the presentation (i.e., slide structure and links between slides). In order to generate word clouds of slides, we weighted extracted words from presentation context, and then presented transitions that highlighted the semantic relationships between slides. Finally, our word cloud visualization could show the words are interactively presented with visual effects in presentations with some application examples.

(5) iPoster: A Collaborative Browsing Platform for Presentation Slides Recently, zoomable presentations as a substitute to the traditional presentations that allows users to zoom in and out of the presentation media. Then, we built a collaborative browsing platform for presentation slides based on interactive poster generation, called the "iPoster," for presenting elements (i.e., textual and graphic elements) in a meaningfully structured layout with automatic transitions, such as zooms and pans, to promote user interaction. Especially, we introduced a semantic structure analysis model for extracting elements and determining the semantic relationships between the elements of the slides. In order to generate an iPoster to provide an overview of presentation slides, we initially placed the elements in a tree structure combined with a stacked Venn. We then attached the zooming and panning transitions between the elements, based on the semantic relationship types. Finally, iPoster could enable readers to interactively and collaboratively browse, and understand educational presentations easily and efficiently using their tablets.

In future work, we will further study on **Next-generation Presentation Data Analytics** and **Advanced Management**. In order to deal with various presentation contents, we will consider how to theoretically analyze salient features based on cognitive science from next-generation presentation contents, and what are differences in presentation contents with multiple languages. In addition, by analyzing the archived presentation data in many fields, we will comprehend implications of concepts in various levels of knowledge. Furthermore, by stepping into analyses for human-computer interaction (HCI) and collaborations through presentation contents, we will develop advanced applications which are efficient and useful in education area, such as self-learning navigation, social learning, and gamification.

Acknowledgments

First and foremost, I would like to give thanks for is God having planned to have me born into this world. I would like to extend my utmost thanks to my doctoral thesis supervisor, Professor Kazutoshi Sumiya at University of Hyogo, for his guidance and support over the past three years. This dissertation would not have been possible without his deep involvement and encouragement. I also would like to thank my thesis committee: Professor Mamiko Okada at University of Hyogo, Professor Hidetoshi Ikeno at University of Hyogo, and the external examiner, Professor Haruo Yokota at Tokyo Institute of Technology. They took time from their busy schedule to supervise my doctoral thesis by giving valuable comments and advices with their expert knowledge in different fields; cognitive science, systems biology, and computer science.

This dissertation has much support from other professors, researchers. Particularly, I would like to mention Dr. Daisuke Kitayama at Kogakuin University and Professor Setsuko Miyamoto at University of Hyogo for their fruitful comments and suggestions, and kind encouragement. Besides the dissertation work, I have been very fortune to work with many bright professors, researchers, and students. I learned many from all of them, and they are always my source of inspiration. A special thanks to my mentors Dr. Koji Yatani and Dr. Darren Edge during my internship at HCI Group in Microsoft Research Asia.

I want to acknowledge faculties, visitors and students I had the pleasure to meet in school, including Dr. Ryong Lee at Korea Institute of Science and Technology Information (KISTI), the secretary of information media laboratory, Ms. Noriko Fukui, and all other members of the laboratory for their valuable discussions and advices, cooperation, and warm encouragement.

Finally, I am deeply indebted to my family has forever been a constant source of encouragement and spiritual support, in particular my parents, for the countless practical favors and emotional boosts over the years.

References

[1] Apache OpenOffice Impress: http://www.openoffice.org/product/impress.html.

[2] Apple Inc.: Keynote: http://www.apple.com/apps/iwork/keynote/.

[3] Atilim University: http://www.atilim.edu.tr/m̃rehan/Chapter%203.ppt.

[4] Beamer, B. and Girju, R.: Investigating Automatic Alignment Methods for Slide Generation from Academic Papers, *Proc. of the 13th Conference on Computational Natural Language Learning (CoNLL 2009)*, pp. 111–119 (2009).

[5] Bederson, B. B. and Hollan, J. D.: Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics, *Proc. of the 7th Annual ACM Symposium on User Interface Software and Technology (UIST 1994)*, pp. 17–26 (1994).

[6] Bederson, B. B. and Hollan, J. D.: Pad++: a Zooming Graphical Interface for Exploring Alternate Interface Physics, *Proc. of the 7th ACM Symposium on User Interface Software and Technology (UIST 1994)*, pp. 17–26 (1994).

[7] Bergman, L., Lu, J., Konuru, R., MacNaught, J. and Yeh, D.: Outline Wizard: Presentation Composition and Search, *Proc. of the 15th ACM International Conference on Intelligent User Interfaces (IUI 2010)*, pp. 209–218 (2010).

[8] Bouamrane, M. and Luz, S.: Meeting Browsing: State-of-the-art-review, *Multimedia Systems*, Vol. 12, No. 4-5, pp. 439–457 (2007).

[9] Carenini, G., Ng, R. and Zhou, X.: Summarizing Email Conversations with Clue Words, *Proc. of the 16th International World Wide Web Conference (WWW 2007)*, pp. 91–100 (2007).

[10] Chen, Y., Wang, L., Dong, M. and Hua, J.: Exemplar-Based Visualization of Large Document Corpus, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 6, pp. 1161–1168 (2009).

[11] Choudary, C. and Liu, T.: Summarization of Visual Content in Instructional Videos, *IEEE Transactions on Multimedia*, Vol. 9, pp. 1443–1455 (2007).

[12] Clarke, S. C. and Willet, P.: Estimating the recall performance of Web search engines, *Aslib Proceedings*, Vol. 49, No. 7, pp. 184–189 (1997).

[13] Coursera: https://www.coursera.org/.

[14] DBSJ Archives: http://www.dbsj.org/Japanese/Archives/dbjapan/.

[15] Don, A., Zheleva, E., Gregory, M., Tarkan, S., Auvil, L., Clement, T., Shneiderman, B. and Plaisant, C.: Discovering Interesting Usage Patterns in Text Collections: Integrating Text Mining with Visualization, *Proc. of the 16th ACM Conference on Information and Knowledge Management (CIKM 2007)*, pp. 213–222 (2007).

[16] Dredze, M., Wallach, H., Puller, D. and Pereira, F.: Generating Summary Keywords for Emails using Topics, *Proc. of International Conference on Intelligent User Interfaces (IUI 2008)*, pp. 199–206 (2008).

[17] edubase Stream: http://stream.edubase.jp/.

[18] Gaskins, R.: PowerPoint at 20: Back to Basics, *Communication of the ACM*, Vol. 50, No. 12, pp. 15–17 (2007).

[19] Good, L. and Bederson, B. B.: Zoomable User Interfaces as A Medium for Slide Show Presentations, *Journal of Information Visualization*, Vol. 1, No. 1, pp. 35–49 (2002).

[20] Google Docs: https://docs.google.com/.

[21] Ham, F. V., Wattenberg, M. and Viegas, F. B.: Mapping Text with Phrase Nets., *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 6, pp. 1169–1176 (2009).

[22] Hanaue, K. and Watanabe, T.: Supporting Design and Composition of Presentation Document Based on Presentation Scenario, *Proc. of the 2nd International Symposium on Intelligent Decision Technologies (KES IDT 2010), SIST 4*, pp. 465–473 (2010).

[23] HandsOut: http://handsout.jp/.

[24] Hayama, T., Nanba, H. and Kunifuji, S.: Alignment between a Technical Paper and Presentation Sheets Using a Hidden Markov Model, *Proc. of the 2005 International Conference on Active Media Technology (AMT 2005)*, pp. 102–106 (2005).

[25] Hearst, M. A.: Search User Interfaces, *Cambridge University Press*, pp.

281–296 (2009).

[26] Isahara, H., Bond, F., Uchimoto, K., Utiyama, M. and Kanzaki, K.: Development of the Japanese WordNet, *Proc. of the 6th International Language Resources and Evaluation (LREC 2008)*, pp. 2420–2423 (2008).

[27] iTunes U: http://www.apple.com/jp/education/itunes-u/.

[28] Iwata, T., Yamada, T. and Ueda, N.: Probabilistic Latent Semantic Visualization: Topic Model for Visualizing Documents, *Proc. of the 14th ACM SIGKDD International Conference and Knowledge Discovery and Data Mining (KDD 2008)*, pp. 263–271 (2008).

[29] Japanese WordNet: http://nlpwww.nict.go.jp/wn-ja/.

[30] Kan, M.: SlideSeer: A Digital Library of Aligned Document and Presentation Pairs, *Proc. of the 7th ACM/IEEE-CS Joint Conference on Digital libraries (JCDL 2007)*, pp. 81–90 (2007).

[31] Kaye, J., Lillie, A., Jagdish, D., Walkup, J., Parada, R. and Mori, K.: Nokia Internet Pulse: A Long Term Deployment and Iteration of a Twitter Visualization, *Proc. of International Conference on Human-Computer Interaction (CHI 2012)*, pp. 829–844 (2012).

[32] Kitayama, D., Otani, A. and Sumiya, K.: A Scene Extracting Method based on Structural and Semantic Analysis of Presentation Content Archives, *Proc. of the 7th International Conference on Creating, Connecting and Collaborating through Computing (C5 2009)*, pp. 275–280 (2008).

[33] Kitayama, D., Otani, A. and Sumiya, K.: An Extracting Method of Semantic Relations between Scenes for Presentation Contents (in Japanese), *IPSJ Transaction on Database (TOD)*, Vol. 2, No. 2, pp. 71–85 (2009).

[34] Kobayashi, T., Nakano, W., Yokota, H., Shinoda, K. and Furui, S.: Presentation Scene Retrieval Exploiting Features in Videos Including Pointing and Speech Information, *Proc. of International Symposium on Large-scale Knowledge Resources (LKR 2007)*, pp. 95–100 (2007).

[35] Kurohashi, S., Shiraki, N. and Nagao, M.: A Method for Detecting Important Descriptions of a Word Based on Its Density Distribution in Text (in Japanese), *Information Processing Society of Japan (IPSJ)*, Vol. 38,

No. 4, pp. 845–854 (1997).

[36] Kushki, A., Ajmal, M. and Plataniotis, K. N.: Hierarchical fuzzy feature similarity combination for presentation slide retrieval, *EURASIP Journal on Advances in Signal Processing archive*, Vol. 2008, No. 188 (2008).

[37] Laufer, L., Halacsy, P. and Fischer, A.: Prezi Meeting: Collaboration in a Zoomable Canvas Based Environment, *Proc. of CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*, pp. 749–752 (2011).

[38] Le, H. H., Lertrusdachakul, T., Watanabe, T. and Yokoda, H.: Automatic Digest Generation by Extracting Important Scenes from the Content of Presentations, *Proc. of the 19th International Conference on Database and Expert Systems Application (DEXA 2008)*, pp. 590–594 (2008).

[39] Lecture Archives from Aoyama Gakuin University:
http://homepage3.nifty.com/˜terao/lecture/aoyama/lec_aoyama_top.html.

[40] Li, Z., Shi, S. and Zhang, L.: Improving Relevance Judgment of Web Search Results with Image Excerpts, *Proc. of the 17th International World Wide Web Conference (WWW 2008)*, pp. 21–30 (2008).

[41] Lichtschlag, L., Hess, T., Karrer, T. and Borchers, J.: Canvas Presentation in the Wild, *Proc. of CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*, pp. 537–540 (2012).

[42] Lichtschlag, L., Karrer, T. and Borchers, J.: Fly: A Tool to Author Planar Presentations, *Proc. of the 27th International Conference on Human Factors in Computing Systems (CHI 2009)*, pp. 547–556 (2009).

[43] Liu, S., Zhou, M. X., Pan, S., Song, Y., Qian, W., Cai, W. and Lian, X.: TIARA: Interactive, Topic-Based Visual Text Summarization and Analysis, *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 3, No. 25, pp. 1–28 (2012).

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

[45] Mann, W. and Thompson, S.: Rhetorical Structure Theory: Towards a Functional Theory of Text Organization, *Text*, Vol. 8, No. 3, pp. 243–281

(1988).

[46] Mathivanan, H., Jayaprakasam, M., Prasad, K. G. and Geetha, T. V.: Document Summarization and Information Extraction for Generation of Presentation Slides, *Proc. of International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom 2009)*, pp. 126–128 (2009).

[47] McCallum, A., Wang, X. and Corrada-Emmanuel, A.: Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email, *Journal of Artificial Intelligence Research*, Vol. 30, pp. 249–272 (2007).

[48] McCloud, S.: Reinventing Comics: How Imagination and Technology Are Revolutionizing an Art Form, *Harper* (2000).

[49] MeCab: http://mecab.sourceforge.net/.

[50] Mehler, A., Mehler, E., Dehmer, M. and Gleim, R.: Towards Logical Hypertext Structure - A Graph-Theoretic Perspective, *Proc. of I2CS '04, Guadalajara/Mexico, Lecture Notes in Computer Science, Berlin-New*, Springer, pp. 136–150 (2004).

[51] Mejova, Y., Schepper, K. D., Bergman, L. and Lu, J.: Reuse in the Wild: An Empirical And Ethnographic Study of Organizational Content Reuse, *Proc. of the 2011 annual conference on Human factors in computing systems (CHI 2011)*, pp. 2877–2886 (2011).

[52] Microsoft Corp.: PowerPoint: http://office.microsoft.com/en-us/powerpoint/.

[53] Miller, G. A.: WordNet: A Lexical Database for English, *Communications of the ACM (CACM)*, Vol. 38, No. 11, pp. 39–41 (1995).

[54] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. J.: Introduction to WordNet: An On-line Lexical Database, *International Journal of Lexicography*, Vol. 3, No. 4, pp. 235–244 (1990).

[55] MIT OpenCourseWare: http://ocw.mit.edu/index.htm.

[56] Miyamoto, M., Sakai, H. and Masuyama, S.: Research on Automatic Generation of Presentation Slides from a LaTeX Manuscript of a Paper (in Japanese), *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol. 18, No. 5, pp. 752–760 (2006).

[57] MPMeister: http://www.ricoh.co.jp/mpmeister/.

[58] MSR Splat: http://research.microsoft.com/en-us/projects/msrsplat/.

[59] Murai, S. and Ushiama, T.: Review-Based Recommendation of Attractive Sentences in a Novel for Effective Browsing, *International Journal of Knowledge and Web Intelligence*, Vol. 3, No. 1, pp. 58–69 (2012).

[60] myBrainshark: http://www.brainshark.com/mybrainshark.

[61] NAIST Digital Library: http://library.naist.jp/library/archive_top/index-e.html.

[62] News of Kyoto University: http://www.kyoto-u.ac.jp/ja/news_data/h/h1/news7/2010/100604_1.htm.

[63] Novak, J. and Canas, A.: The Theory Underlying Concept Maps and How To Construct and Use Them, *Institute for Human and Machine Cognition (IHMC)* (2008).

[64] Ohshima, H., Nakamura, S. and Tanaka, K.: SlothLib: A Programming Library for Research on Web Search (in Japanese), *The Database Society of Japan (DBSJ Letters)*, Vol. 6, No. 1, pp. 113–116 (2007).

[65] Okamoto, H., Kobayashi, T. and Yokota, H.: Presentation Retrieval Method Considering the Scope of Targets and Outputs, *Proc. of International Workshop on Challenges in Web Information Retrieval and Integration*, pp. 46–53 (2005).

[66] Parker, I.: Absolute PowerPoint: Can a Software Package Edit our Thoughts?, *The New Yorker* (2001).

[67] Pattanasri, N. and Tanaka, K.: Construction of an Entailment Ontology for Enhancing Comprehension of Search Results Inside e-Learning Materials, *Proc. of the 6th International Conference on Creating, Connecting and Collaborating through Computing (c5 2008)*, pp. 124–130 (2008).

[68] pdftohtml: http://pdftohtml.sourceforge.net/.

[69] Perer, A. and Smith, M.: Contrasting Portraits of Email Practices: Visual Approaches to Reflection and Analysis, *Proc. of Advanced Visual Interfaces (AVI 2006)*, pp. 389–395 (2006).

[70] Portland State University: http://web.cecs.pdx.edu/˜howe/cs410/lectures/Relational_Intro_1.ppt.

[71] PowerPoint Is Evil: http://www.wired.com/wired/archive/11.09/ppt2.html.

[72] pptPlex: http://www.microsoft.com/en-us/download/details.aspx?id=28558.

[73] Pradhan, S.: An Algebraic Query Model for Effective and Efficient Retrieval of XML Fragments, *Proc. of the 32nd international conference on Very large data bases (VLDB 2006)*, pp. 295–306 (2006).

[74] Pradhan, S., Tajima, K. and Tanaka, K.: A Query Model for Retrieving Relevant Intervals within a Video Stream, *Proc. of the 6th IEEE Int'l Conference on Multimedia Computing and Systems (ICMCS 99)*, Vol. 2, pp. 788–792 (1999).

[75] Presentation Flow: Link Or Break: http://philpresents.wordpress.com/2011/11/23/presentation-flow-link-or-break/.

[76] Prezi: http://prezi.com/.

[77] Quirk, C., Choudhury, P., Gao, J., Suzuki, H., Toutanova, K., Gamon, M., Yih, W., Vanderwende, L. and Cherry, C.: MSR SPLAT, a language analysis toolkit, *Proc. of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstration Session (NAACL HLT 2012)*, pp. 21–24 (2012).

[78] Sharmin, M., Bergman, L., Lu, J. and Konuru, R.: On Slide-Based Contextual Cues for Presentation Reuse, *Proc. of the 17th ACM International Conference on Intelligent User Interfaces (IUI 2012)*, pp. 129–138 (2012).

[79] Shibata, T. and Kurohashi, S.: Automatic Slide Generation Based on Discourse Structure Analysis, *Proc. of the 2nd International Joint Conference on Natural Language Processing (IJCNLP 2005)*, pp. 754–766 (2005).

[80] Simpson, E. H.: Measurement of Diversity, *Nature*, Vol. 163, p. 688 (1949).

[81] Slide Layouts in PowerPoint 2007: http://presentationsoft.about.com/od/powerpoint2007/ss/2007slidelayout_8.htm.

[82] SlideBoom: http://www.slideboom.com/.

[83] SlideShare: http://www.slideshare.net/.

[84] SlothLib: http://www.dl.kuis.kyoto-u.ac.jp/SlothLibWiki/.

[85] Smith, M. A. and Kanade, T.: Video Skimming and Characterization

through the Combination of Image and Language Understanding Techniques, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR07)*, pp. 775–781 (1997).

[86] Spicer, R., Lin, Y., Lelliher, A. and Sundaram, H.: NextSlidePlease: Authoring and Delivering Agile Multimedia Presentations, *ACM Transactions on Multimedia Computing, Communication and Applications*, Vol. 8, No. 4, pp. 53:1–53:20 (2012).

[87] Spicer, R. P., Lin, Y., Kelliher, A. and Sundaram, H.: NextSlidePlease: Authoring and Delivering Agile Multimedia Presentations, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, Vol. 8, No. 4, pp. 53:1–53:19 (2012).

[88] Strobelt, H., Oelke, D., Rohrdantz, C., Stoffel, A., Keim, A. D. and Deussen, O.: Document Cards: A Top Trumps Visualization for Documents, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 6, pp. 1145–1152 (2009).

[89] Tanaka, K. and Yoshikawa, M.: Towards abstracting complex database objects: Generalization, reduction and unification of set-type objects (extended abstract), *Proc. of the 2nd International Conference on Database Theory (Lecture Notes in Computer Science)*, Vol. 326, pp. 252–266 (1988).

[90] Trees and Forests: http://teacherweb.com/AB/GilbertPatersonMiddleSchool/MsDavid/Tree-Types-2b-Posting-version.ppt.

[91] Tufte, E.: The Cognitive Style of PowerPoint, *Graphics Press, Cheshire* (2003).

[92] VideoLectures.NET: http://videolectures.net/.

[93] Viegas, F., Golder, S. and Donath, J.: Visualizing Email Content: Portraying Relationships from Conversational Histories, *Proc. of the 2006 Conference on Human Factors in Computing Systems (CHI 2006)*, pp. 979–988 (2006).

[94] Wang, D., Li, T., Zhu, S. and Ding, C.: Multi-document Summarization via Sentence-level Semantic Analysis and Symmetric Matrix Factor-

ization, *Proc. of the 31st Annual International ACM SIGIR Conference (SIGIR 2008)*, pp. 307–314 (2008).

[95] Wang, Y., Kitayama, D., Lee, R. and Sumiya, K.: Automatic Generation of Learning Channels by Using Semantic Relations among Lecture Slides and Recorded Videos for Self-Learning Systems, *Proc. of the 11th IEEE International Symposium on Multimedia (ISM 2009)*, pp. 275–280 (2009).

[96] Wang, Y. and Sumiya, K.: Semantic Ranking of Lecture Slides based on Conceptual Relationship and Presentational Structure, *Proc. of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010)*, pp. 2801–2810 (2010).

[97] Watanabe, T., Ishiguro, Y. and Hanaue, K.: Automatic Composition of Presentation Slides, Based on Semantic Relationships among Slide Components, *Proc. of the 4th International Conference on Intelligent Interactive Multimedia Systems and Services (KES IIMSS 2011), SIST 11*, pp. 261–270 (2011).

[98] Wattenberg, M. and Viegas, F.: The Word Tree, an Interactive Visual Concordance, *Proc. of Information Visualization Conference (InfoVis 2008)*, pp. 1221–1228 (2008).

[99] WordNet: http://wordnet.princeton.edu/.

[100] Yokota, H., Kobayashi, T., Okamoto, H. and Nakano, W.: Unified Contents Retrieval from an Academic Repository, *Proc. of International Symposium on Large-scale Knowledge Resources (LKR 2006)*, pp. 41–46 (2006).

[101] Yoshiaki, Y., Masashi, T. and Katsumi, N.: A Support System for Making Presentation Slides (in Japanese), *Transactions of the Japanese Society for Arti cial Intelligence*, Vol. 18, No. 4, pp. 212–220.

Appendix

論文リスト

学術論文（査読あり）

1. Automatic Generation of Learning Channels by Using Semantic Relations among Lecture Slides and Recorded Videos for Self-Learning Systems
   Yuanyuan Wang, Daisuke Kitayama, Ryong Lee, and Kazutoshi Sumiya
   In Proc. of the 11th IEEE International Symposium on Multimedia (ISM 2009), pp. 275–280, San Diego, California, USA, December 2009
2. Semantic Ranking of Lecture Slides based on Conceptual Relationship and Presentational Structure
   Yuanyuan Wang and Kazutoshi Sumiya
   In Proc. of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010), pp. 2801–2810, Barcelona, Spain, September 2010
3. Slide KWIC: Snippet Generation for Browsing Slides based on Conceptual Relationship and Presentational Structure
   Yuanyuan Wang and Kazutoshi Sumiya
   In Proc. of the 9th International Conference on Creating, Connecting and Collaborating through Computing (C5 2011), pp. 40–47, Kyoto, Japan, January 2011
4. A Browsing Method for Presentation Slides Based on Semantic Relations and Document Structure for e-Learning
   Yuanyuan Wang and Kazutoshi Sumiya
   Journal of Information Processing, Vol. 20, No. 1, pp. 11–25, December 2011
5. Skeleton Generation for Presentation Slides based on Expression Styles
   Yuanyuan Wang and Kazutoshi Sumiya
   In Proc. of the 5th International Conference on Intelligent Interactive Multimedia Systems and Services (KES IIMSS 2012), SIST 14, pp. 551–560, Gifu, Japan, May 2012
6. A Method for Generating Presentation Slides based on Expression Styles

using Document Structure
Yuanyuan Wang and Kazutoshi Sumiya
International Journal of Knowledge and Web Intelligence, Vol. 4, No.1, pp. 93–112, March 2013

7. Dynamic Word Clouds: Context-based Word Clouds of Presentation Slides for Quick Browsing
Yuanyuan Wang and Kazutoshi Sumiya
In Proc. of the 6th International Conference on Intelligent Interactive Multimedia Systems and Services (KES IIMSS 2013), pp. 108–117, Sesimbra, Portugal, June 2013
8. iPoster: A Collaborative Browsing Platform for Presentation Slides based on Semantic Structure
Yuanyuan Wang, Kota Tomoyasu, and Kazutoshi Sumiya
In Proc. of the 1st Workshop of Quality, Motivation and Coordination of Open Collaboration (QMC 2013), Kyoto, Japan, November 2013 (to appear)

学術報告 (査読あり)

1. Semantic Ranking of Lecture Slides based on Conceptual Relationship and Presentational Structure
Yuanyuan Wang and Kazutoshi Sumiya
日本データベース学会，情報処理学会データベースシステム研究会，電子情報通信学会データ工学研究専門委員会 第2回データベース，Web，情報マネジメントに関する若手研究者国際ワークショップ (iDB Workshop 2010), pp. 65–71, 2010年8月.
2. A Browsing Method for Presentation Slides Based on Semantic Relations and Document Structure
Yuanyuan Wang and Kazutoshi Sumiya
日本データベース学会，情報処理学会データベースシステム研究会，電子情報通信学会データ工学研究専門委員会 第3回データベース，Web，情報マネジメントに関する若手研究者国際ワークショップ (iDB Workshop 2011), pp. 86–97, 2011年8月
3. A Generation Method of Presentation Slides based on Expression Styles

using Slide Structure

Yuanyuan Wang and Kazutoshi Sumiya

日本データベース学会，情報処理学会データベースシステム研究会，電子情報通信学会データ工学研究専門委員会 第4回データベース，Web，情報マネジメントに関する若手研究者国際ワークショップ (iDB Workshop 2012), 2012年8月

学術報告 (査読なし)

1. スライドと映像のメタデータを用いたシーンの意味的関係に基づくプレゼンテーション管理システム

   王 元元, 北山 大輔, 角谷 和俊

   電子情報通信学会データ工学研究専門委員会，日本データベース学会，情報処理学会データベースシステム研究会 第1回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2009) 論文集，E9-4, 2009年3月

2. プレゼンテーションコンテンツのための概念的構造と表層的構造に基づくスライドの関係判定方式

   王 元元, 北山 大輔, 角谷 和俊

   電子情報通信学会データ工学研究専門委員会，日本データベース学会，情報処理学会データベースシステム研究会 第2回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2010) 論文集，C1-1, 2010年3月

3. プレゼンテーションコンテンツのための概念構造に基づくスライドの関係判定方式

   王 元元, 北山 大輔, 角谷 和俊

   情報処理学会 創立50周年記念 (第72回) 全国大会講演論文集，pp. 869–870, 2010年3月

4. Slide-KWIC Browser: Snippet Generation Based on Semantic Relations and Document Structure for Presentation Slides

   王 元元, 角谷 和俊

   電子情報通信学会データ工学研究専門委員会，日本データベース学会，情報処理学会データベースシステム研究会 第3回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2011) 論文集, B7-3, 2011年2月

5. A Snippet Generation Method for Retrieving Slides in Presentation Content Sharing Sites

Yuanyuan Wang and Kazutoshi Sumiya
日本データベース学会 第2回ソーシャルコンピューティングシンポジウム講演論文集 (SoC 2011), p. 26, 2011年6月

6. 電子ポスターのためのスライド間の意味的関係を用いた閲覧支援
友安 航太, 王 元元, 角谷 和俊
第5回ユビキタスウェアラブルワークショップ論文集, p. 12, 2011年12月
7. Generating Slide Skeletons based on Expression Styles for Presentation Contents
王 元元, 角谷 和俊
電子情報通信学会データ工学研究専門委員会, 日本データベース学会, 情報処理学会データベースシステム研究会 第4回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2012) 論文集, D5-3, 2012年3月
8. スライド間の意味的関係を用いたプレゼンテーションコンテンツ閲覧方式
友安 航太, 王 元元, 角谷 和俊
電子情報通信学会 総合大会 情報・システムソサイエティ (ISS) 特別企画「学生ポスターセッション」予稿集, p. 225, 2012年3月
9. ポスターとスライドの構造に基づくズーミングを用いたポスター閲覧方式
友安 航太, 王 元元, 角谷 和俊
情報処理学会研究報告, Vol. 2012-DBS-155, No. 1, 2012年11月
10. Context-based Word Clouds of Presentation Slides for Quick Browsing
王 元元, 角谷 和俊
電子情報通信学会データ工学研究専門委員会, 日本データベース学会, 情報処理学会データベースシステム研究会 第5回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2013) 論文集, A10-4, 2013年3月
[学生プレゼンテーション賞, 優秀インタラクティブ賞]
11. スライドの構成要素と意味的関係に基づくインタラクティブポスター生成システム
友安 航太, 王 元元, 角谷 和俊
電子情報通信学会データ工学研究専門委員会, 日本データベース学会, 情報処理学会データベースシステム研究会 第5回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2013) 論文集, C8-4, 2013年3月.
12. A Word Cloud Visualization for Comparing Presentation Contents

Yuanyuan Wang and Kazutoshi Sumiya
日本データベース学会 第4回ソーシャルコンピューティングシンポジウム
講演論文集 (SoC 2013), pp. 47–52, 2013年6月

技術雑誌

1. 画像と異種マルチメディアとの融合＜第1回＞
北山 大輔，若宮 翔子, 王 元元, 角谷 和俊
日本工業出版株式会社　画像ラボ 2014年01号, pp. 41–45, 2014年1月
2. 画像と異種マルチメディアとの融合＜第2回＞
北山 大輔，若宮 翔子, 王 元元, 角谷 和俊
日本工業出版株式会社　画像ラボ 2014年02号, pp. 59–67, 2014年2月

紀要

1. 書道Webコンテンツの開発とデータベースシステムの設計
王 元元, 岳 五一, 劉 俊
甲南大学情報教育センター紀要, No. 6, CD-ROM, 28 pages, 2008年3月