# Elaboration Context Graph: A System to Support Understanding the Contexts in Elaboration Processes of Research Documents

Sho Onami[1], Ryo Onuma[2], Hiroki Nakayama[3], Hiroaki Kaminaga[4], Youzou Miyadera[5], Shoichi Nakamura[6]
Dept. Computer Science and Mathematics, Fukushima University, Fukushima, Japan[1, 2, 3, 4, 6]
Division of Natural Science, Tokyo Gakugei University, Tokyo, Japan[5]

***Abstract*—Elaborating research documents is carried out by repeatedly creating and editing documents while simultaneously performing tasks such as surveys, presentation of results, and discussion of research directions. Although indispensable for advancing research, such work is often challenging because it requires handling diverse documents. Effective execution therefore demands an accurate understanding of the elaboration contexts of research documents, including related artifacts, referenced documents, and the circumstances and history of past tasks, so that these can be applied in subsequent work. However, these contexts grow increasingly large and complex as research progresses, making them difficult to grasp and reducing task efficiency. This paper describes a method for generating an elaboration context graph by organizing documents involved in the elaboration process using work history data recorded on a PC. The graph visually represents the documents, screenshots capturing work scenes, and the relationships among them, thereby supporting the understanding of elaboration contexts. We further describe a system developed on the basis of this method. Finally, we report an experiment conducted with the prototype and discuss the system's effectiveness.**

***Keywords*—Elaboration contexts graph; elaboration work of research documents; elaboration contexts; understanding work circumstances and histories; screenshots**

## I. Introduction

Research activities generally involve various document creation and editing tasks, such as drafting and revising research plans, taking notes on ideas, creating presentation materials, and writing papers. These tasks are accomplished by creating and editing the documents while simultaneously accumulating various other tasks in parallel such as searching the Web, engaging in discussions with others, reviewing existing documents. To proceed successfully with these tasks, it is important to accurately organize and understand the elaboration contexts (i.e., the documents and artifacts used in the elaboration work of research documents along with the work circumstances and histories) and leverage them for subsequent work. However, since such research activities often extend over long periods, the involved documents become numerous and their elaboration contexts grow increasingly complicated.

To address these issues, there have been methods for document management [1–11] and support methods for recalling tasks on PCs [12–21]. However, these existing methods are limited in their scope of support and are insufficient for assisting with the understanding of the complicated elaboration contexts. On the other hand, there have been methods to support accumulation, understanding, and reuse of the artifacts and histories of intellectual work [22–28]. However, most of these methods require workers to register information into the system themselves, which imposes a significant burden during complicated document elaboration work.

In this research, we aim to develop a system that is responsible for collecting the history data of researchers' document elaboration work on their PCs, analyzing the accumulated history data, and generating and presenting the elaboration context graphs. This system enables to support the understanding of the elaboration contexts of research documents more proactively than conventional methods.

The rest of this paper is organized as follows. Section II clarifies direction of this research while reviewing the related work. Section III details the elaboration context graph and its generation algorithm. Section IV describes the design of a support system for understanding elaboration contexts and implementation of its prototype. Section V describes an experiment using the prototype. Section VI concludes this paper.

## II. Issues and Direction

### A. Understanding the Contexts in Elaboration Processes of Research Documents

When creating and editing documents such as research papers and presentation slides, a researcher often engages in various tasks in parallel, including reviewing existing documents and searching the Web. Moreover, documents created by a researcher him/herself, as well as those discovered and viewed in the past, often serve as important reference materials for subsequent work. In this research, we refer to the process of creating and editing research documents, which is performed through the long-term accumulation of such multiple tasks, as the "elaboration process of research documents."

In such elaboration work, the number of documents involved in the process—such as artifacts and reference materials—gradually becomes enormous as the project extends over a long period. To utilize these documents for future work,

it is essential to properly manage when and in which work they were used, as well as their current location. When managing these documents manually, it is common practice to assign each document an appropriate filename on the basis of its content, document type, and other relevant factors. Then, the documents are organized into folders in accordance with target work and project timelines. However, when attempting to utilize previously used documents, it is difficult to recall the target documents by their previous usage contexts, as well as locating the desired ones among the large number of documents, based solely on file and folder names. Additionally, valuable documents that a researcher has forgotten or are unaware of may become buried.

Furthermore, to effectively utilize past documents, it is crucial to accurately understand the work circumstances and histories surrounding the work at the time—such as what other documents were used alongside them, what references were consulted, and what was considered. However, understanding such elaboration contexts solely by folder management is quite difficult.

Additionally, these problems become even more serious when someone other than the original author attempts to understand the elaboration contexts and to make use of the involved documents. Finding documents useful for the latest work among those created and organized by others is even more difficult than finding documents one has created on one's own. Furthermore, it is nearly impossible to ascertain the context in which the document was edited or reviewed by others.

Therefore, there is strong need for an effective method to support utilizing previously used research documents while understanding their elaboration contexts.

### B. Related Work

There have been many reports on methods for document management [1–11]. Hermosa et al. pointed out that traditional document management processes involve multiple platforms, making it difficult to track documents when they are shared. They developed a document management system (DMS-TM [1]) to centrally manage documents originating from multiple departments within the university. This system generates a tracking code when a document is sent, and once the processing cycle is complete, stores it in the database using the tracking code for future use. Each department can utilize web-based document sharing functionality while continuing to use their existing environment. Although this system is expected to facilitate reuse of documents by preventing the loss, it only targets completed documents and does not provide support for accumulating and understanding the work histories.

In addition, Nishi et al. developed a method to support users in re-accessing documents used in their past work on a PC [2]. This method focuses on the folders used for past work rather than individual documents and groups these folders on the basis of specific perspectives. By presenting users with these folder groups as a list, this method supports the reuse of past documents. Although this method is expected to promote the utilization of existing documents, it does not take into account the relationship between documents and past work scenes that cannot be fully expressed by folders. Therefore, it is not sufficient as support for understanding the work histories and circumstances.

Moreover, there have been support methods for recalling past work on a PC [12–21]. Hu et al. analyzed the effects of viewing time-lapse videos on the self-reflection of knowledge workers such as engineers, programmers, editors, and researchers, with the aim of improving their productivity [12]. Although the focus on time-lapse video as a means to recall past work is interesting, only a specific period of work can be observed because it is centered on video viewing, and it does not account for discovering used documents. Therefore, it is currently difficult to utilize this method to support understanding of contexts in the elaboration process of research documents.

Additionally, Meyer et al. developed a task management tool (called AIRbar [13]) aimed at encouraging knowledge workers, who handle numerous tasks daily, to reflect on their work and improve their awareness of the achievement status of priority tasks as well as influencing factors. This tool provides prompts that encourage users to specify priority tasks while self-confirming, and helps them reflect on their own work on a PC by visualizing the progress and the time spent on respective tasks. Although a practical effect of helping users understand their own work status can be expected, it targets daily work reviews and does not consider understanding the histories of long-term work.

These methods focused on supporting specific parts of the work processes. On the other hand, there have been methods to support accumulation, understanding, and reuse of the artifacts and histories of intellectual work [22–28].

Zloch et al. [22] presented a framework of the research knowledge graph (RKG) that expresses research artifacts such as datasets, publications, and methods, along with their relationships. They examined methods to promote the organization and reuse of research artifacts using RKG and discussed its possible applications. The results suggested the potential effectiveness of RKG in enabling researchers themselves to manage and reuse knowledge resources in their research work. However, regarding the registration of elements for building RKG, it was limited to discussing the advantages and challenges of manual input, semi-automation, and full automation. Therefore, methods for addressing the long-term elaboration work of research documents have not been established yet. Moreover, they dealt with only the research artifacts, and understanding the elaboration contexts has not been sufficiently considered.

Furthermore, Devezas et al. [23] proposed a wiki-based environment for researchers to manage key elements of their research, such as literature, datasets, experimental methods, and their results. This environment is promising as a method for documenting detailed circumstances and histories surrounding the work. It can facilitate researchers' reflection on their own activities and promote information sharing among research project members. However, researchers themselves must describe the overview and histories of the work, which imposes a significant burden on them and makes it difficult to apply the environment to long-term research projects.

Thus, most existing information systems rely on workers themselves to register information, placing a heavy burden on them and raising concerns about the potential loss of important work histories due to omissions of registration. Consequently, there is a strong need for an effective method to accumulate documents and contexts involved in complex elaboration work with minimal burden and to support understanding of them.

### C. Issues in Elaborating Research Documents

On the basis of the above background, this research focuses on the following three issues to be overcome in understanding the contexts during the elaboration work of research documents.

(Issue 1) It is difficult to manage and understand the tasks and documents involved in the elaboration work of research documents due to their diversity.

(Issue 2) Understanding research documents used in the past and their relationships by associating them with the work scenes is important, but difficult.

(Issue 3) Observing and understanding the elaboration contexts of research documents becomes increasingly time-consuming since the contexts grows more complex as the work extends over a longer period.

### D. Approach

To resolve these issues, this research first constructs a system framework to monitor the elaboration work of research documents on a PC by a researcher and automatically acquire the work history data (Solution to problem 1). Next, we develop a method to generate a “elaboration context graph” based on the analysis of the work history data. This graph visually represents the elements involved in the elaboration process—such as artifacts, documents used, and screenshots embodying the work scenes—and the relationships between them. (Solution to problem 2). Finally, based on the method, we develop a system to generate the elaboration context graph and visually present it to users so that they can intuitively observe the circumstances and histories in past elaboration work along with the artifacts and used documents (Solution to problem 3).

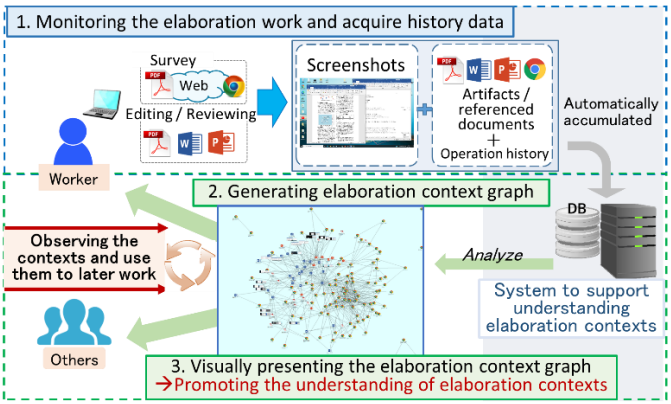


Fig. 1. Overview of support for the understanding of elaboration contexts.

This facilitates the understanding of elaboration contexts, which have previously been difficult, and enables their utilization in subsequent work. An overview of elaboration context understanding support is shown in Fig. 1.

## III. Method for Generating Elaboration Context Graph

### A. Elaboration Context Graph

Most existing information systems intending to accumulate the artifacts and processes of intellectual work have forced workers themselves to register the documents used in the work and the relationships between them into the system. However, there are limits to forcing users to register information every time they work on complex and long-term work of research document elaboration.

Therefore, this research proposes a novel method to support the recall and understanding of the elaboration contexts by having the system monitor researchers' work on their PCs and automatically generating an “elaboration context graph” that visually represents the documents involved in the work and the relationships between them on the basis of analyzing the accumulated history data.

The elaboration context graph is an attribute graph that represents various documents used in the elaboration work as well as screenshots of the PC desktop representing work scenes as nodes, and the relationships between them as edges. We currently target the documents created during the elaboration process such as papers and presentation slides, the documents viewed in parallel such as published papers and Web pages, and screenshots of the PC desktop (collectively referred to as "research documents").

By using the elaboration context graph, users can easily grasp the process of work by tracing the related documents with a document of interest as starting point. Additionally, screenshots serve as intuitive clues for looking back the work scenes, enabling users to understand the work circumstances and the documents used at that time.

### B. Formalization of Elaboration Context Graph

Since the elaboration context graph treats various documents as nodes, the relationships between them are varied - documents referenced in parallel, derivations or developments from other documents, and citations. Moreover, extracting these relationships requires the analysis of various types of history data. In contrast, most existing methods for automatic graph generation are limited to handling only a single type or a very small number of data types. Therefore, we develop an algorithm for generating the graph to enable the system implementation of the elaboration context graph. As a preparation for this, we formalize the elaboration context graph to firmly handle to diverse documents and their relationships.

The elaboration context graph is defined as follows:

- **[Definition 1]** Elaboration context graph $G$ is represented as $G = (V, E)$. $V$ is a finite set of graph nodes, and $v$ $(\in V)$is a node in the graph that represents a research document. $E$ is a finite set of undirected connectors, and $e$ $(\in E)$ denotes a relationship defined between two research documents.

Each research document and relationship has the following attributes, respectively.

- **[Definition 2]** The research document attributes in the elaboration context graph $G$ are defined as $att_V = (kind_V, name_V, lastReadDate_V)$. These are attributes attached to node $v$ $(\in V)$ and their meaning are as follows:
  - $kind_V$ expresses the type of research document. $kind_V$ = (draft paper | presentation slides | referenced paper | Web page | note | screenshot). The type of research document $v$ $(\in V)$ is denoted as $kind_V(v)$.
  - $name_V$ expresses the file name of a research document or the title in case of a Web page and is represented by any string of characters. The name of a research document $v$ $(\in V)$ is represented as $name_V(v)$.
  - $lastReadDate_V$ expresses the last viewed date and time of a research document. The last viewed date and time of a research document $v$ $(\in V)$ is denoted as $lastReadDate_V(v)$.
- **[Definition 3]** The undirected connector attributes of the elaboration context graph $G$ are $att_E = (kind_E)$. These are attributes attached to the undirected connector $e$ $(\in E)$ and its meaning is as follows:
  - $kind_E$ means the kind of undirected connector. $kind_E$ = (Concurrent use | copy & paste | download | document derivation and development | work scene). The kind of undirected connector $e$ $(\in E)$ is denoted as $kind_E(e)$.
- **[Definition 4]** Each research document has individual sub-attributes by the type of document as follows:
  - $att_{draft\ paper,\ presentation\ slides,\ note}$ = (*fileFormat*(MS-Word | MS-PPT | PDF | text file), *filename, filePath*(Local path on the PC), *createTime*(Creation or first access date), *lastModifiedDate, screenshots*(List of screenshots related to the document)).
  - $att_{referenced\ paper}$ = (*fileFormat* (MS-Word | PDF), *filename, filePath* (Local path on the PC), *AcquisitionDate,* screenshots(List of screenshots related to the document)).
  - $att_{Web\ page}$ = (URL, pageTitle, firstTime(first access date)).
  - $att_{screenshot}$= (*filename*, *filePath* (Local path on the PC), *AcquisitionDate, documents*(List of documents shown in the screenshot)).

Research document attributes represent metadata on the node, while sub-attributes represent metadata on the research document itself. These definitions enable to design an algorithm for generating the elaboration context graph that represents the documents involved in the work, screenshots embodying the work scenes, and the relationships between them.

### C. Algorithm for Generating Elaboration Context Graph

Based on the above definitions, we generate an elaboration context graph in the following three steps (Algorithm 1).

*1) Accumulation of the work history data:* While the worker is elaborating research documents on a PC, the system monitors their work and accumulates the histories of the operations and occurrences of events. Specifically, we target the operations such as switching active windows, opening and saving documents, keystrokes in the editing documents, copying and pasting, and web browsing. Such diverse history data is required for generating nodes and edges with the attributes defined in Section III (B).

We developed a system framework to automatically accumulate such history data as the function of the system mentioned in Section IV. The framework consists of the resident programs and the database. The resident programs acquire the history data by monitoring the worker's operations on a PC as well as occurrences of events. In addition, desktop screenshots showing the work scenes are captured at regular intervals. The acquired history data is successively registered into the database. The resident programs impose no extra burden on the worker since they work in the background.

In this research, we target MS Word, MS PowerPoint, Adobe Acrobat Reader, and Google Chrome as applications that handle documents, and treats the document formats such as .docx, .pptx, .pdf, and web pages processed by these applications.

*2) Node generation:* On the basis of analyzing the accumulated work history data, we generate nodes $V$ which represent the research documents involved in the elaboration work.

First, we retrieve all documents opened or viewed during the work period by analyzing access logs to documents and web page browsing history. In order to exclude documents that have not substantially contributed to the elaboration work, we calculate the total active time of each document and adopt only those documents exceeding a certain threshold as nodes $v$.

Next, the attribute $att_V$ is assigned to each node in accordance with the definition of the research documents mentioned above. We initially assign the sub-attributes such as $name_V$ and $lastReadDate_V$ on the basis of the work history data and the metadata of each document.

We also extract the format of each document from the work history data and set the type of research document $kind_V$ for each node accordingly. However, since both the draft paper and the referenced paper share the same format (MS-Word and PDF), we distinguish them based on whether editing was performed. Specifically, in the case of a document with a certain number of keystrokes for editing and its copies saved under different names, the type of research document $kind_V$ of that node is set to "Draft Paper"; otherwise, $kind_V$ is set to "Referenced Paper".

Algorithm 1: Graph Generation

```
[Input]
- DAH: Document access and save history
- AWH: Active window switching history
- CPH: Copy & paste history
- WDH: Web browser's download history
- SSH: Screenshot acquisition history
[Output]
- G = (V, E): Elaboration context graph
[Global variable]
- Threshold_node: Threshold for document active time to adopt nodes
- Threshold_AWC: Threshold for count of switching active window
- Threshold_CUT: Threshold for concurrent use time
- Threshold_SS: Threshold for the number of related documents to
      adopt screenshot nodes
[Local variable]
- DN: Document node set
- SSN: Screenshot node set
[Method]
1   // Node generation
2   // Generate a set of used documents
3   UD = extraction UsedDocuments (DAH);
4   for( ∀d ∈ UD)
5       if(d.totalActiveTime >= Threshold_node)
            DN.add(d, att_V(d)= getAttribute(d));
6   // Edge generation
7   // Generate a set of all pairs of document nodes
8   pairsDN=generate_ pairsDN (DN)
9   for( ∀d1,d2 ∈ pairs_DN)
10      //Count activation switches between two documents
11      Count_AWC=count_ActiveChanges(d1,d2,AWH);
12      // Calculate the concurrent use time for two documents
13      Seconds_CUT=Calculation_ConcurrentUseTime(d1, d2);
14      //Analyze document pair whose activation switching and
        concurrent use time exceeds a threshold
15      If (Count_awc > Threshold_Count_AWC &&
            Seconds_ CUT > Threshold_Seconds_ CUT)
            E.add((d1, d2), kind_E(d1, d2)="ConcurrentUse" );
16      //Copy & paste
17      If(judge_copyandpaste_documents(d1,d2,CPH))
            E.add((d1, d2), kind_E(d1, d2)="Copy&Paste" );
18      // Document derivation and development
19      If (confirmation_DocumentDevelopment(d1,d2, DAH))
            E.add((d1, d2), kind_E (d1, d2)="Derivation" );
20      // Downloaded document and its source page
21      If (confirmation_DocumentDownload (d1,d2, WDH))
            E.add((d1, d2), kind_E (d1, d2)="Download" );
22  // Nodes and Edges of screenshots
23  for( ∀ss ∈ SSH){
24      ss_docs = extraction_ DocumentsRelatedtoScreenshots(ss);
25      //Decide which screenshots to adopt for the node
26      if (length(ss_docs)>= Threshold_screenshots
                and checkState ActiveWindow(ss, AWH))
27          SSN.add(ss);
28          for( ∀d∈ss_docs)
                E.add((ss, d), kind_E (ss, d)="WorkScene");
29      // Other screenshots
30      else for( ∀d∈ss_docs) att_V(d).screenshots.add(ss);
31  V = DN ∪ SSN;
32  return G = (V, E);
```

Additionally, we extract promising screenshots from those captured during the work. Specifically, we select screenshots containing windows that have remained active for a certain period of time, or those captured at the moment an active window switched. Then, we identify the documents visible in each screenshot on the basis of analyzing the work history data, and adopt screenshots linked to multiple documents as node $v$. All other screenshots are treated as sub-attributes of the nodes for the relevant documents.

*3) Edge generation:* We generate edges $E$ by analyzing the relationships between nodes in five perspectives defined as the kind of edge $kind_E$.

First, we extract a pair of documents that are repeatedly activated alternately on the basis of analyzing the history of window activation. We then assign an edge $e$ connecting these documents and set its type $kind_E$ to "Concurrent use".

Next, we extract the relationships where documents are derived or evolved from other documents (e.g., different versions of a document) on the basis of analyzing the history of document duplication and saving under different names. We then assign an edge $e$ connecting such documents and set its type $kind_E$ to "document derivation and development".

Moreover, we extract a pair of documents that are in a citation or referenced relationship on the basis of analyzing the history of copying and pasting the elements such as strings between different documents. We then assign an edge $e$ connecting these documents and set its type $kind_E$ to "copy & paste".

We also extract documents obtained from the web and the URLs of their source web pages on the basis of analyzing the download history of web browsers. We then assign edges $e$ connecting them, and set their type $kind_E$ to "download".

Furthermore, we assign an edge $e$ connecting the screenshot to the documents being used at the time it was captured, and set its type $kind_E$ to "work scene" using the results of the analysis performed when generating the nodes for the screenshots of the PC desktop representing the work scenes.

The formalization and development of the generation algorithm based on it enable the automatic generation of the elaboration context graph that visually represents the documents involved in the long-term elaboration process, work scenes, and their relationships, which was previously impossible. Additionally, it provides a path for future inclusion of new documents as targets.

## IV. Support System for Understanding Elaboration Contexts of Research Documents

### A. System Requirements

We develop a system to support understanding the contexts in elaboration processes of the research documents on the basis of the method for generating the elaboration context graph proposed in Section III. We initially organize the system requirements as follows on the basis of findings gained through observing actual elaboration work of research documents.

- Req.1) The system should be able to record the work history data and screenshots related to the research document elaboration work without imposing a burden on the user.
- Req.2) The system should be able to analyze the research documents involved in the elaboration work and their relationships.
- Req.3) The system should be able to present the research documents involved in the elaboration work and their relationships, as well as the work circumstances and histories, in a way capable for intuitive observation.
- Req.4) The system should be able to present information that allows users to easily review the work scenes related to a document and examine the details of the work scenes they focus on.
- Req.5) The system should be able to present information that allows users to easily review the contents and metadata of the research document of interest.
- Req.6) The system should be able to present information that allows users to easily identify the location of the research documents they need, as well as other documents managed together with them.
- Req.7) The system should be able to provide assistance in discovering the intended documents from among countless research documents.

### B. Function Design

To satisfy the above system requirements, this system is equipped with the following functions.

- Function for acquisition and accumulation of work history data: The function acquires and accumulates the history data and screenshots during the user's elaboration work on a PC. To avoid placing an unnecessary burden on the user, it starts automatically when the PC is turned on and automatically collects history data in the background (Solution to Req.1).
- Function for generating an elaboration context graph: The function extracts the documents involved in the elaboration work and their relationships on the basis of analyzing the accumulated work history data. It then generates an elaboration context graph, in which documents and screenshots representing work scenes serve as nodes, and their relationships as edges (Solution to Req.2).
- Function for displaying an elaboration context graph: The function visually presents the generated elaboration context graph in a way that allows users to interactively and intuitively observe the elaboration contexts by simple mouse operations (Solution to Req.3).
- Function for reviewing work scenes: When the user selects a document node, the function displays a list of desktop screenshots showing the work scenes at the time the selected document was being edited or viewed. Furthermore, when the user selects a screenshot node, the function enlarges it (Solution to Req.4).
- Function for displaying details of selected document: When a user selects a document node in the graph, the function displays the contents of that document along with attributes such as the last modified date and time, and the folder location (Solution to Req.5 and 6).
- Function for searching documents: The function searches for document nodes within the elaboration context graph that contain keywords entered by the user, and highlights the document nodes containing those words (Solution to Req.7).

### C. Implementation of Prototype

Based on the above design, we implemented a prototype of the support system as a desktop application for Windows. This system consists of modules, including a work history acquisition unit and a node/edge generation unit, as well as a base system that manages fundamental functions such as coordination among these modules and user authentication. Moreover, a database was built to manage the work history data and screenshots as well as generated context graphs.

This system was implemented on Windows 11 using Python as the development language. Moreover, MySQL was used as the DBMS. The system configuration is shown in Fig. 2.

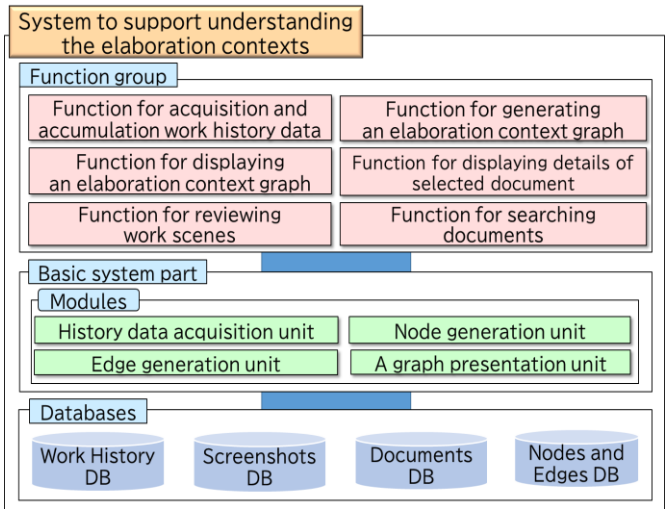


Fig. 2. System configuration.

### D. Assistance in Understanding Elaboration Contexts Using the System

An interface of the support system for understanding the elaboration contexts is shown in Fig. 3. When the user attempts to review their own or others' elaboration work of research documents, the system presents the elaboration context graph along with related information in response to the user's request.

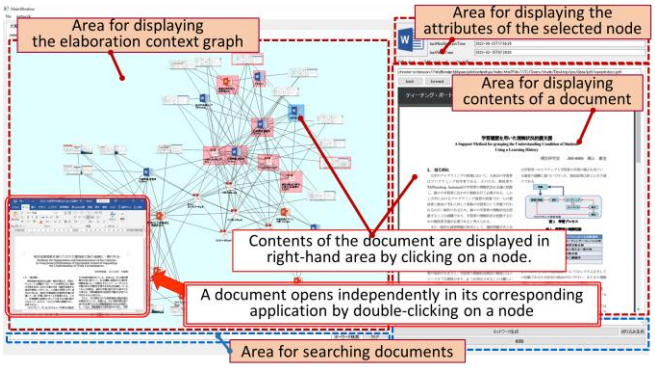


Fig. 3. An interface of the support system.

In the area for displaying the elaboration context graph (left side of Fig. 3), users can zoom in and out on the graph. Moreover, the user can interactively manipulate the document nodes. For instance, when a node on the graph is clicked, the metadata such as document name, and last access/modified date and time are displayed (upper right in Fig. 3). The content of the selected research document is also displayed depending on node selection (middle right of Fig. 3). In addition, double-clicking a node, the corresponding research document opens independently in its associated application, such as MS Word. Furthermore, the user can display a list of screenshots related to the node he/she has selected by switching tabs (Fig. 4). Clicking a screenshot of interest from the list, the screenshot is enlarged, allowing the user to observe the work scene at that time. The user can also check the folder containing the selected document, allowing to easily access other documents in the same folder.

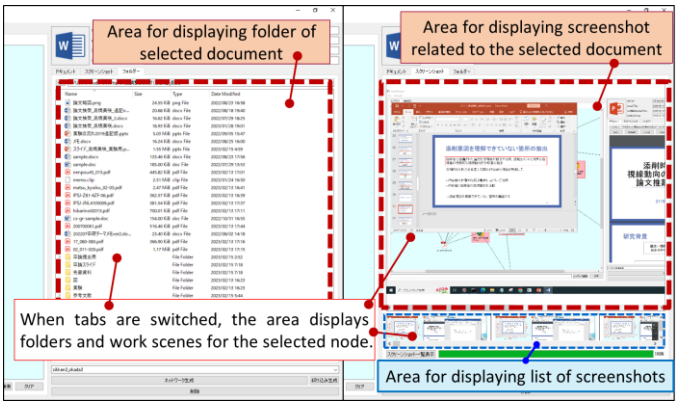


Fig. 4. Function for displaying details of selected document/function for reviewing work scenes.

By utilizing these functions, the user can observe the circumstances and histories of the elaboration work such as: What tasks were performed concurrently while creating a certain document, - What tasks were completed and what artifacts were obtained before starting to create a certain document, - How the reference to a certain document influenced the direction of subsequent document creation.

Additionally, when the user enters keywords into the search prompt, the nodes of documents containing those keywords in their file names or main text are highlighted on the graph. This makes it easy to locate the intended documents (lower left of Fig. 3).

Thus, assistance provided by the system enables the user to realistically observe the contexts in elaboration work of research documents.

## V. Experiment and Discussion

### A. Outline of Experiment

We conducted an experiment to verify the effectiveness of the prototype system in supporting the understanding of the elaboration contexts of research documents.

*1) Preparatory session:* First, two fourth-year undergraduate students majoring in computer science (participant A and B) searched the web for papers related to their own research themes and summarized the following points in a way of reinforcing the slides that they had previously created. The slides on which participants performed this work are called “reinforced slides”.

- Overview of main related works and their connections to the participant’s own research.
- Future issues to be addressed in the participant’s research.
- Clues for progressing the participant’s own research such as ideas for system design and promising methods of analysis.

The work time was set at approximately 90 minutes, and the support system collected the work history data during the participant's work.

*2) Main session:* After the preparatory session was completed, four third-year undergraduate students majoring in computer science (referred to as Subjects I, J, K, and L) observed the research documents used by participants A and B, who were senior students, in order to understand the research overview and the contexts in the work conducted by senior students. Specifically, we prepared a folder containing research documents used by senior students, including the collected literature and documents created by other students, and provided it to the subjects. Using the reinforced slides of senior students as a starting point, the subjects reviewed these documents and performed the following tasks.

- Task I) Understand the senior student's research overview and key points, as well as the new content they added to the reinforced slides.
- Task II) Select five or more documents that are helpful to the subjects themselves in understanding the research of senior students from the provided folder.
- Task III) Select up to three documents that the subject inferred “particularly referenced by the senior students in advancing their own research,” from the senior student's folder as in Task II).

Subjects performed these tasks in two sets (each set targeting the work history of a different senior student) while changing whether support was provided by the system. Table I shows the combinations of senior students targeted and the presence or absence of the support when these tasks were conducted. We recorded the work performed by the subjects and conducted interviews with them after the tasks were completed.

TABLE I. COMBINATIONS OF SENIOR STUDENTS TARGETED AND SUPPORT CONDITIONS

| | First | Second |
|---|---|---|
| Subject I | Without support (Participant A) | With support (Participant B) |
| Subject J | Without support (Participant B) | With support (Participant A) |
| Subject K | With support (Participant A) | Without support (Participant B) |
| Subject L | With support (Participant B) | Without support (Participant A) |

After completing the tasks, we examined the effectiveness of the support provided by the system in the following perspectives.

- Whether the support provided by the system can reduce the burden on workers, such as the time required for tasks.
- Whether differences exist in characteristics of subjects' behavior, such as the number of documents viewed during tasks and work tendencies, depending on the presence or absence of support.

Whether the support provided by the system enables the subjects to easily grasp the key points of the work performed by senior students.

### B. Results and Discussion

First, we confirmed that the series of processes—from acquiring the history data, its analysis, to the generation of the elaboration context graph—can be performed as intended. Fig. 5 shows the overviews of the elaboration context graphs for participants A and B generated in the experiment.

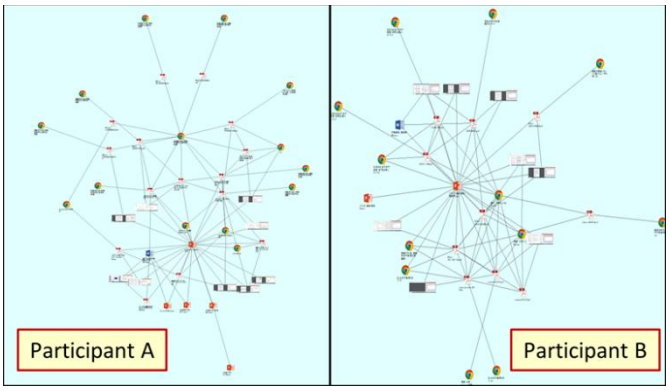


Fig. 5. Overview of the elaboration context graphs generated in the experiment.

Table II shows the duration of Tasks I, II, and III for subjects I through L. These results indicate that the time required for the tasks is reduced in many cases when support is provided. This is thought to be because the support reduced the burden of unnecessary tasks such as simple checking. On the other hand, subject K took slightly longer to complete Task II when support was provided. Upon closely observing the work circumstances using recorded videos, it was found that time was spent on carefully reading several papers. Therefore, it is not considered that there was any wasted time, nor that the results negate the effectiveness of the support.

Next, Table III shows the number of documents viewed by subjects I through L in Tasks I, II, and III. It was found that the number of documents reviewed decreased in most cases where support was provided. It is thought that this is because the subjects had no choice but to browse documents in the working folder blindly in case without support, whereas with support, they could smoothly progress their work by preferentially browsing documents of interest on the context graph presented by the system. On the other hand, Subject I viewed more documents in Tasks I and III when support was provided. Upon closely observing the work circumstances using recorded videos, it was confirmed that task itself proceeded smoothly similar to other subjects in case with support. In addition, it appeared that subject I was using their spare capacity to review additional relevant documents, which can be considered a positive tendency.

Furthermore, we even more carefully reviewed the videos and observed the work circumstances in detail. As a result, it was found that the workload and task difficulty appeared to be somewhat reduced when using the support system, not only in the case mentioned above. On the other hand, without support, subjects appeared to struggle to identify and review the documents that senior students had referenced among the vast number of documents available. From these results, it can be inferred that the elaboration context graphs and the details of documents presented by the system offered substantial support for performing the tasks. This trend of reduced workload is consistent with the above-mentioned decreases in task duration and the number of documents viewed.

Moreover, Table IV shows the accuracy in Task I, grasping the new contents added by senior students to the reinforced slides, for subjects I to L. For all subjects, it can be confirmed that the accuracy with support is equivalent to or higher than that without support. Taken together with the reductions in task duration and document views, these results indicate that subjects with support appeared to better understand the work performed by senior students, even when their accuracy was similar. Additionally, careful observation of the work using video revealed that subjects appeared to struggle while consulting numerous documents without support. In contrast, with support, subjects were observed to take notes on key points smoothly and without hesitation, using the provided documents and screenshots of actual work scenes as references. These findings are consistent with the accuracy of grasping key points of the task.

Additionally, in post-experiment interviews, we obtained comments such as: "When I used the system, documents related to the task are easier to find and review, making the work significantly easier compared to when no support was available." and "The relationships between documents were visually clear, and viewing screenshots alongside them made it easy to confirm the key points of the task."

These results indicate the system's effectiveness in both aspects of reducing the burden of understanding past work circumstances and reviewing documents, and of grasping key points of past work. Although this experiment was conducted on a limited scale and requires further verification, it suggests the fundamental effectiveness of our approach for understanding the elaboration contexts of research documents.

We also received feedback such as "Since the graph can only be scrolled using the scroll bar, it would be better if it could also be moved by dragging with the mouse." This provides valuable hints for future improvements.

TABLE II. RESULTS OF THE TASK DURATION

| | Task I | | Task II | | Task III | |
|---|---|---|---|---|---|---|
| | First | Second | First | Second | First | Second |
| Subject I | 96 min | 61 min | 72 min | 35 min | 18 min | 15 min |
| Subject J | 52 min | 45 min | 34 min | 23 min | 17 min | 17 min |
| Subject K | 48 min | 64 min | 115 min | 96 min | 28 min | 40 min |
| Subject L | 36 min | 42 min | 70 min | 72 min | 29 min | 46 min |

Underline: With support by system.

TABLE III. RESULTS OF THE NUMBER OF DOCUMENTS VIEWED

| | Task I | | Task II | | Task III | |
|---|---|---|---|---|---|---|
| | First | Second | First | First | Second | First |
| Subject I | 6 | 13 | 31 | 14 | 6 | 13 |
| Subject J | 17 | 12 | 32 | 15 | 19 | 13 |
| Subject K | 14 | 21 | 34 | 37 | 14 | 31 |
| Subject L | 4 | 5 | 12 | 20 | 8 | 9 |

Underline: With support by system.

TABLE IV. ACCURACY IN GRASPING SENIOR STUDENTS' REINFORCED SLIDES IN TASK I

| | First | Second |
|---|---|---|
| Subject I | 0.46 | 1.00 |
| Subject J | 1.00 | 1.00 |
| Subject K | 1.00 | 1.00 |
| Subject L | 0.92 | 0.51 |

Underline: With support by system

## VI. CONCLUSION

In this research, we developed a system to support the understanding of the contexts in the work of elaborating research documents. Specifically, we developed a method for generating the elaboration context graphs on the basis of analyzing documents involved in the elaboration work and their relationships, using the work history data. We also developed a prototype system to support the understanding of the elaboration contexts based on the method. Finally, we conducted an experiment using the prototype and discussed its effectiveness based on the results.

Future work includes further verification of the effectiveness of the support system for understanding the elaboration contexts through long-term experiments.

## ACKNOWLEDGMENT

This work was supported in part by Japan Society for the Promotion of Science under Grant-in-Aid for Scientific Research (B) (No. 24K00446).

## REFERENCES

[1] Karen R. Hermosa , Vince Cymel B. Loteriña and Joshua A. Sebuc , "Document Management System With Tracking and Monitoring in Manuel S. Enverga University Foundation-Candelaria Inc.," *International Journal of Advanced Research in Computer Science*, Vol.14, No.6, pp.31-43, 2023.

[2] R. Nishi, and Y. Nomura, "Virtual Folder System by Discovering and Classifying Task Folders," *IPSJ Journal*, Vol.62, No.2, pp.527-537, 2021.

[3] N. Saito, A. Kanai, S. Tanimoto, "An Organizational Knowledge Using Method by Extracting Activity Plan from Documents in a Shared Folder," *IPSJ Journal*, Vol.58, No.1, pp.153-165, 2017.

[4] N. L. Adam, M. Akmal Hakim Che Mansor, M. F. Pauzi and S. Cik Soh, "Document and Event Record Management System: A Prototype," *Proc. IEEE 12th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pp.223-227, 2022.

[5] I. Hristoski, J. Stojanov and Ž. Stojanov, "Design Considerations for a Generic Graph Database in Archival Document Management," *Proc. of the 13th International Conference on Applied Internet and Information Technologies (AIIT 2023)*, pp.360–367, 2023.

[6] P. Riba, A. Dutta, S. Dey, J. Lladós and A. Fornés, "Improving Information Retrieval in Multiwriter Scenario by Exploiting the Similarity Graph of Document Terms," *Proc. of the 14th IAPR International Conference on Document Analysis and Recognition*, pp.475-480, 2017.

[7] S. Ferilli, D. Di Pierro and D. Redavid, “Holistic Graph-Based Document Representation and Management for Open Science,” *Proc. of the 27th International Conference on Theory and Practice of Digital Libraries (TPDL 2023)*, pp.3-7, 2023.

[8] O. E. Famous, S. E. Adewunmi, V. Yemi-Peters, “A Comparative Review of Five Leading Document Management Systems Based on Information Security,” *Iconic Research and Engineering Journals,* Vol.7, No.7, pp.199-208, 2024.

[9] Jordan, S., Sternad Zabukovšek, S., Šišovska Klančnik, I. “Document Management System – A Way to Digital Transformation,” *NGOE Journal* , Vol.68, Issue 2, pp.43-54, 2022.

[10] Sambetbayeva, M., I. Kuspanova, A. Yerimbetova, S. Serikbayeva, and S. Bauyrzhanova, “Development of Intelligent Electronic Document Management System Model Based on Machine Learning Methods,” *Eastern-European Journal of Enterprise Technologies*, Vol.1, No.2, pp. 68-76, 2022.

[11] A. Kulkarni, C. Ramanathan and V. E. Venugopal, "Ontology Mediated Document Retrieval for Exploratory Big Data Analytics," *Proc. 2023 IEEE 17th International Conference on Semantic Computing (ICSC)*, pp.100-103, 2023.

[12] Donghan Hu and Sang Won Lee, “Exploring the Effectiveness of Time-lapse Screen Recording for Self-Reflection in Work Context,” *Proc. of the 2024 CHI Conference on Human Factors in Computing Systems*, No.1029, pp.1-14, 2024.

[13] Andre N Meyer, Nimra Ahmed, Isabelle Cuber, Sebastian Richner, Elaine M. Huang, Gail Murphy and Thomas Fritz, “'Stick to' Three: Fostering Awareness, Intentions, and Reflections on the Top Daily Tasks,” *Proc. of the 2025 ACM Designing Interactive Systems Conference,* pp.3647-3662, 2025.

[14] Donghan Hu and Sang Won Lee “ScreenTrack: Using a Visual History of a Computer Screen to Retrieve Documents and Web Pages,” *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems*, pp.1-13, 2020.

[15] L. Jin, L. Feng, G. Liu and C. Wang, "Personal Web Revisitation by Context and Content Keywords with Relevance Feedback," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No.7, pp.1508-1521, 2017.

[16] Rule, A., Tabard, A., & Hollan, J., “Using Visual Histories to Reconstruct the Mental Context of Suspended Activities.” *Human–Computer Interaction*, Vol.32, Issue.5–6, pp.511-558, 2017.

[17] Bon Adriel Aseniero, Charles Perin, Wesley Willett, Anthony Tang and Sheelagh Carpendale, “Activity River: Visualizing Planned and Logged Personal Activities for Reflection,” *Proc. of the 2020 International Conference on Advanced Visual Interfaces*, No.4, pp.1-9, 2020.

[18] H. Ikuta, K. Takashima, and K. Nishimoto “Efficient Collecting Method and Availability of Text Fragments Deleted during Document Creation Process,” *IPSJ Journal*, Vol.59, No.12, pp.2299-2314, 2018.

[19] Yoana Ahmetoglu, Akeisha Iskandar, Shermin Taoh, Andy Ying, Duncan P. Brumby and Anna L. Cox, “"I Work Much Better by Doing Less": How Task Duration Feedback Affects Optimistic Planning Bias in Academic Work,” *Proc. of the 4th Annual Symposium on Human-Computer Interaction for Work*, No.10, pp.1-17, 2025.

[20] Zhicheng Huang and Jeff Huang, “IRCHIVER: An Information-Centric Personal Web Archive for Revisiting Past Online Sensemaking Tasks,” *Proc. of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval*, pp.231-246, 2025.

[21] C. Sato, and M. Takaku, “Support for reflection of information gathering behavior using Web browsing history,” *Journal of Japan Society of Information and Knowledge*, Vol.30, No.2, pp.220-229, 2020.

[22] M. Zloch, D. Dessì, J. D’Souza, L. J. Castro, B. Zapilko, S. Karmakar, B. Mathiak, M. Stocker, W. Otto, S. Auer, and S. Dietze, ”Research Knowledge Graphs: The Shifting Paradigm of Scholarly Information Representation,” *Proc. of the European Semantic Web Conference (ESWC) 2025, Lecture Notes in Computer Science*, Vol.15719, pp.140-154, 2025.

[23] J. Devezas and S. Nunes, “Managing Research the Wiki Way: A Systematic Approach to Documenting Research,” *arXiv preprint*, arXiv:2104.01849, pp.1-11, 2021.

[24] P. Chakraborty and D. K. Sanyal, “A Personal Knowledge Graph for Researchers,” *Proc. of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation*, pp.129-132, 2024.

[25] L. John, A. M. Ghanmi, T. Wittenborg, S. Auer and O. Karras, “ExtracTable: Human-in-the-Loop Transformation of Scientific Corpora into Structured Knowledge,” *Proc. of the 29th International Conference on Theory and Practice of Digital Libraries*, pp.470-487, 2025.

[26] M. C. Nagan, “Using a Lab Wiki for Longevity and Knowledge Transfer in Undergraduate Research,” *International Journal of Quantum Chemistry*, Vol.120, No.20, 2020.

[27] F. Limani, A. Latif, and K. Tochtermann, “Scholarly Artifacts Knowledge Graph: Use Cases for Digital Libraries,” *Proc of the 15th International Conference on Metadata and Semantics Research (MTSR) 2021, Communications in Computer and Information Science*, Vol. 1537, pp. 245-256, 2022.

[28] L. John, A. M. Ghanmi, T. Wittenborg, S. Auer and O. Karras, “ExtracTable: Human-in-the-Loop Transformation of Scientific Corpora into Structured Knowledge,” *Proc. of the 29th International Conference on Theory and Practice of Digital Libraries*, pp.470-487, 2025.