Collaborative Graph Composition Is More Productive than Collaborative Text Composition

その他のタイトル	共同グラフ作成は共同テキスト作成より生産性が高い
学位授与年月日	2020-09-18
URL	http://hdl.handle.net/2261/00079943

THE UNIVERSITY OF TOKYO

MASTER THESIS

Collaborative Graph Composition Is More Productive than Collaborative Text Composition

（共同グラフ作成は共同テキスト作成より生産性が高い）

Author:

ZHANG Ziliang

Supervisor:

Prof. Koiti HASIDA

A thesis submitted in fulfillment of the requirements for the degree of Master of Information Science and Technology

Graduate School of Information Science and Technology

July 30, 2020

Declaration of Authorship

I, ZHANG Ziliang, declare that this thesis titled, “Collaborative Graph Composition Is More Productive than Collaborative Text Composition” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed.

Signed: ZHANG ZILIANG

Date: 2020/08/06

THE UNIVERSITY OF TOKYO

Abstract

Graduate School of Information Science and Technology

Master of Information Science and Technology

Collaborative Graph Composition Is More Productive than Collaborative Text Composition

by ZHANG Ziliang

Graphs can express semantic structures of documents more straightforwardly than texts. Preceding researches have proven that graph-document composition is more productive than text-document composition in single-author settings. We have developed a software application to support multiple-author composition of graph documents and thereby verified the superiority of graphs to texts in the productivity of collaborative composition.

Acknowledgements

I take this opportunity to express gratitude to all the faculty members and students of Hasida laboratory for their advice and help during the two years of my master's program. Especially, thanks to Professor Hasida, who has given me great help and guidance in my studies and research. He organizes laboratory meetings at least once a week to discuss my research in detail, gives me great help at every important stage, and guides me to complete this thesis. Thanks to Mr. Matsubara, Mrs. Sarada Balachandran Nair Sumadevi, and other contributors for their contributions and help in the development of the Semantic Editor. Thanks to Mr. Yao for discussing the experiment of synchronous collaborative authoring with me and contributing a part of experimental data. Thanks to Mr. Karilas for the help and many suggestions for the experiment. Finally, I would like to thank my beloved family for their support and help during my studies. Their encouragement made me take an important step in my life.

Contents

1. Introduction

According to a survey of international adult literacy by OECD (Organization for Economic Co-operation and Development), adult's ability of literacy is generally low. More than 70% of the surveyed countries have adult literacy ability scores at level 1 and level 2 of the five-level scoring standard, indicating very poor and weak, respectively (Darcovich, 2000; OECD Skills Studies, 2016; Thorn, 2009). Another reading skill test conducted by Arai et al. (Arai, et al., 2017) implies that junior high school students did no better than a dependency analysis machine in terms of reading skills.

On the other hand, graphs such as mind maps and concept maps are widely used to visualize semantic structures or relationships for education, business, and other purposes. Compared with text documents, the explicit graphical representation of graph documents can easily express nonlinear and complex content. Therefore, the composition of various complex documents such as contracts, manuals, and so forth could be improved by replacing traditional linear text documents with graph documents containing explicit semantic structures such as labelled links among concepts. Moreover, machines would perform better analysis given graph-structured documents (Devlin, Chang, Lee, & Toutanova, 2018). Since the construction of a semantic corpus is costly and time-consuming, composing graphs in daily life can help generate a semantic corpus for further NLP research, such as machine translation, information extraction, question answering, etc.

Yagishita et al. 's research (Yagishita, Munemori, & Sudo, 1998) has shown that graphs can improve the content quality in single-author document composition. However, multiple-author collaborative document composition has been much less systematically studied. The benefits of collaboration are apparent, and the important role of documentation in teamwork, information sharing, and consensus-building is irreplaceable. For instance, in business, creating a contract may involve multiple stakeholders and professionals. Similarly, in academia, co-authors may collaborate to compose academic papers (Beck, 1993). It is necessary to assess how well people behave in graph-based collaborative authoring and what requirements are necessary for this approach to be able to perform at a maximum efficiency level.

We have been developing a software application, Semantic Editor, to support the collaborative composition of the Resource Description Framework graph (RDF-graph) (Lassila, Swick, Wide, & Consortium, 1998) documents. Semantic Editor supports Diagrammatic Semantic Authoring (DSA) (Hasida, Decentralized, Collaborative, and Diagrammatic Authoring, 2017; ISO/CD 24627-3: Language resource management — Comprehensive Annotation Framework (ComAF) — Part 3: Diagrammatic semantic authoring (DSA), 2019) as a potential ISO standard to specify graphical/diagrammatic documents with explicit semantic structures addressed by RDF.

In this thesis, by comparing text and graph as the content carrier in collaborative authoring through experiments, we have established reliable evidence of the superiority of graphs in collaborative authoring. Though experiments, we confirmed that RDF-graph is more productive and more conducive to collaboration than text in collaborative authoring.

The thesis is structured as follows. In Chapter 2, we list the related work to our research. Chapter 3 explains the purpose and methodology of our research. Chapter 4 and Chapter 5 explains the synchronous collaborative authoring experiment and the asynchronous collaborative authoring experiment, respectively, including hypotheses, experiment process, and data analysis of the experiment. Chapter 6 discusses the phenomena revealed by the data based on the experiments in Chapters 4 and 5, summarizes the entire thesis, and presents conclusions.

2. Related Work

When we were studying RDF-graph based collaborative authoring, we also received much inspiration from other prior researches. In this chapter, we first introduce some other forms of graph documents, and point out their advantages and disadvantages, and then discusses the current difficulties and challenges of collaborative authoring. Finally, we discuss Yagishita's research (Yagishita, Munemori, & Sudo, 1998) on graph-based single-author composition.

2.1. Graph Documents

Graphs, due to their visual representation, have natural advantages when expressing nonlinear and complex content, and can intuitively illustrate the relationship between data. Graphs can present data that are too numerous or complicated to be described adequately in the texts and in less space. (Slutsky, 2014)

Mind map, a note-taking technique promoted by Buzan (Buzan & Buzan, 2006), uses links that can be marked with relationships to connect concepts. The mind map uses keywords and key concepts to express the relationships of all levels of themes with mutual affiliation and related hierarchical maps. Information is sorted and organized by priority. Links to colors and pictures are established to enhance people's ability to remember information. A previous study (Holland, Holland, & Davies, 2004) conducted on students showed that mind mapping is more efficient in understanding concepts in the art and design field. Moreover, positive subjective effects were observed on the participating students solving essay-writing problems. The benefits of mind maps are that they can enhance creativity and recall ability, better solve problems, and focus on topics, and improve organization and thought arrangements. However, the mind map cannot be used as a substitute for text because it does not support ontology, especially semantic relations, so that the language relationship between nodes cannot be fully expressed. Therefore, it can only be used as a note-type graph to help memorize. Moreover, because of its dependence on priority and hierarchy, it is usually limited to depicting the relationship of the hierarchical tree-like structure.

Another graph technique often used in the education field is the concept map introduced by McAleese (McAleese, 1998). Concept maps also express information in a structured way through the connection of nodes and labels. Moreover, the concept map introduces the ontology of the concept. The connection relationship between nodes represents the relationship between discourse/arguments, such as "causes" and "requires". A study by Willerman et al. (Willerman & Mac Harg, 1991) showed that the concept map can provide the classroom teachers with a meaningful and practical structured approach for using advance organizers in their classes.

The Resource Description Framework (RDF) data model (Lassila, Swick, Wide, & Consortium, 1998) is based on the idea of making statements about resources in expressions of the form subject–predicate–object, known as triples. In this study, we consider the RDF graph as a sort of concept graph, where each node is equivalent to a simple sentence or phrase, and each link represents a binary relationship between two nodes with semantic meaning. Each relationship is either directed (asymmetric) or undirected (symmetric). Thanks to the support of this semantic ontology, it is possible to create an RDF graph that expresses the same meaning for any text document. In this sense, graph documents can replace text documents.

In order to further utilize the advantages of graphs, we considered collaborative authoring of graphs. A collaborative system allows a group of users to work together in different locations and at different times (Gao, Gao, Xiong, & Lee, 2018). In all fields, the collaborative composition or co-authoring of documents has become increasingly important. Traditionally, document-based work collaboration has always been linear text documents. For example, the collaborative editing tool Google Docs[1] has been championed by many researchers and has become the main application for editing text documents. However, according to the statement of D'Angelo (d'Angelo, Di Iorio, & Zacchiroli, 2018), collaboration on common document parts happens often, but it happens asynchronously with authors taking turns in editing. The simultaneous editing of common document parts happens very rarely.

There are some other mind map and concept map tools, but we found that these tools lack support for ontology, making it impossible to generate semantic maps equivalent to text, or lack the function of providing operations for connection, or lack the support of multiple-author collaborative editing. To this end, we developed a software application, Semantic Editor, to support the collaborative composition of RDF-graph documents.

2.2. Collaborative Authoring

The benefits of cooperation are self-evident. Many work environments require collaborative authoring of documents. Academic paper is a good example. Co-authors may need to collaboratively compose and refine a document (Beck, 1993). Similarly, in business, creating a contract may involve multiple stakeholders and professionals. Therefore, multiple users need to compose a document collaboratively. Nowadays, many technologies are supporting collaborative authoring, making this kind of collaboration simple and easy to carry out. Regardless of whether participants are in a unified geographic location, documents can be quickly and effectively shared and co-edited. Traditionally, document-based work collaboration has been linear text documents. According to the research of Adler et al. (Adler, Nash, & Noël, 2004), the new technology supporting collaboration based on documents will also cause some problems. For example, multiple copies of the same document can lead to confusion, as group members make conflicting modifications to the document.

The research of Emigh et al. (Emigh & Herring, 2005), a genre analysis of two web-based collaborative authoring environments, Wikipedia and Everything2, reveals

[1] https://www.google.com/docs/about/

how users, acting through mechanisms provided by the system, can shape (or not) features of content in particular ways. Its research found that the greater the degree of post-production editorial control afforded by the system, the more formal and standardized the language of the collaboratively authored documents becomes, analogous to that found in traditional print encyclopedias.

Adler et al. have put forward several major challenges to collaborative authoring software. The RDF-graph editing tool, Semantic Editor, designed and used in this thesis, solved the management of time and space problem, allowing collaborators to collaborate in editing at different times and locations. Based on PLR (Personal Life Repository) (Hasida, Personal life repository as a distributed PDS and its dissemination strategy for healthcare services, 2014), it solved many other problems such as private and shared workspaces, simultaneity and locking, protection, and security, etc. Thus, it helps users express their ideas smoothly.

2.3. Graph in Single-Author Composition

In an experiment about B type KJ method (an idea-generating method), Yagishita et al. (Yagishita, Munemori, & Sudo, 1998) proposed an evaluation method for sentences of B type KJ method, which compared the Petri nets generated by the participants using KJ method and not using KJ method. They analyzed Petri nets with size, height, width, expansion degree, aggregation degree, aspect ratio, and other indicators as quality indicators, proved that graphs improve content quality in single-author document composition. However, in their research, the documents generated by KJ method cannot address the full content of document because they do not support any ontology, especially semantic relations. Therefore, their documents cannot be considered as a substitute for text documents.

Yagishita's work is inspiring for us. In our study, we use RDF graphs to address the full content of documents owing to ontologies. We also used similar methods to compare graph documents and text documents. We first invite participants to compose documents in the form of text or graph according to some given topics. Then we convert graph documents and text documents into corresponding Petri nets. Finally, we analyze the results according to some quality indicators to study the performance of graph documents and text documents in collaborative authoring.

3. Research Purpose and Methodology

Semantic Editor supports Diagrammatic Semantic Authoring (DSA) (Hasida, Decentralized, Collaborative, and Diagrammatic Authoring, 2017; ISO/CD 24627-3: Language resource management — Comprehensive Annotation Framework (ComAF) — Part 3: Diagrammatic semantic authoring (DSA), 2019) as a potential ISO standard to specify graphical/diagrammatic documents with explicit semantic structures addressed by RDF. Theoretical effectiveness of DSA has already been studied, but its merits in collaboration support have not been empirically evaluated. To investigate the merits of graphs for collaborative work, we have experimentally evaluated the quality of RDF-graph documents collaboratively composed by Semantic Editor in comparison with collaboratively composed text documents.

The purpose of this thesis is to verify whether RDF-graph, as a form of document and a carrier of content, is more productive and more conducive to collaboration than text in a collaborative authoring environment. We designed experiments to compare the content of the document composed by several groups of participants for a given topic when using RDF-graph and text, respectively. According to some quality indicators, the content components are quantitatively analyzed, and then the score differences between text and RDF-graph on these dimensions are compared to verify our hypotheses.

The experimental results are also expected to identify users' needs regarding collaborative work on graph composition and to clarify properties of graph documents, which will both be incorporated in future improvements of Semantic Editor.

We conducted two experiments, respectively, one for synchronous collaborative authoring and the other for asynchronous collaborative authoring. The detailed experimental process will be mentioned later. In order to complete such an experiment, in this chapter, we will clarify the tools we used for the collaborative diagrammatic composition of RDF graphs, the Semantic Editor, and the method of quantitative comparison.

3.1. Semantic Editor

Semantic Editor is an application software developed to support real-time collaborative diagrammatic composition of RDF graphs. It is programmed in Java and can be used on any computer running JVM (Java Virtual Machine). Semantic Editor uses PLR (Personal Life Repository) (Hasida, Personal life repository as a distributed PDS and its dissemination strategy for healthcare services, 2014) for protecting communication security. PLR is a decentralized, secure, low-cost, and scalable Personal Data Store (PDS) for socially sharing and utilizing personal and other data based on the data subjects' intention. PLR allows its users (individuals and organizations) to securely share their data directly (i.e., without any middleman) via end-to-end encryption. Its operation cost for both the application/service providers and the end-users is meager because the shared online storage may be Google Drive[2], OneDrive[3], and others.

The major functionalities of Semantic Editor are the following two:

- RDF-graph composition: The user can create, move, delete, and modify nodes and links between nodes. A node can contain a text, which is typically a simple sentence or phrase. A link represents a semantic relationship between the two connected nodes.
- Collaborative work: Real-time collaboration among multiple users is supported by data synchronization through public clouds.

To support large scale document, we introduced the idea of hypernode in this application. Any node in a graph can be a hypernode, which contains another graph as an embedded graph of the root one. Therefore, with the hierarchical organization of graphs and embedded graphs, users can compose a large scale document by the Semantic Editor more easily. The use of hypernode can be nested so that the Semantic Editor can support documents of any scale in hierarchical mode.

Semantic Editor uses an ontology of discourse and other relations to address the relationships among nodes. This ontology defines fundamental semantic and pragmatic relations, as shown in Figure 3-1. This ontology has been used in the experiment discussed later, but it is easy to replace the ontology employed in Semantic Editor. The connection relationship is divided into symmetrical relationship and asymmetrical relationship. Symmetrical relationship is undirected connection and asymmetrical relationship is directed connection. In the ontology we currently use in Semantic Editor, except for the symmetrical relationships listed below, all others are asymmetrical relationships:

- And
- Similar
- Equal
- Contrast
- Or
- Unlike
- Conflict
- Sametime

[2] https://www.google.com/drive/
[3] https://onedrive.live.com/

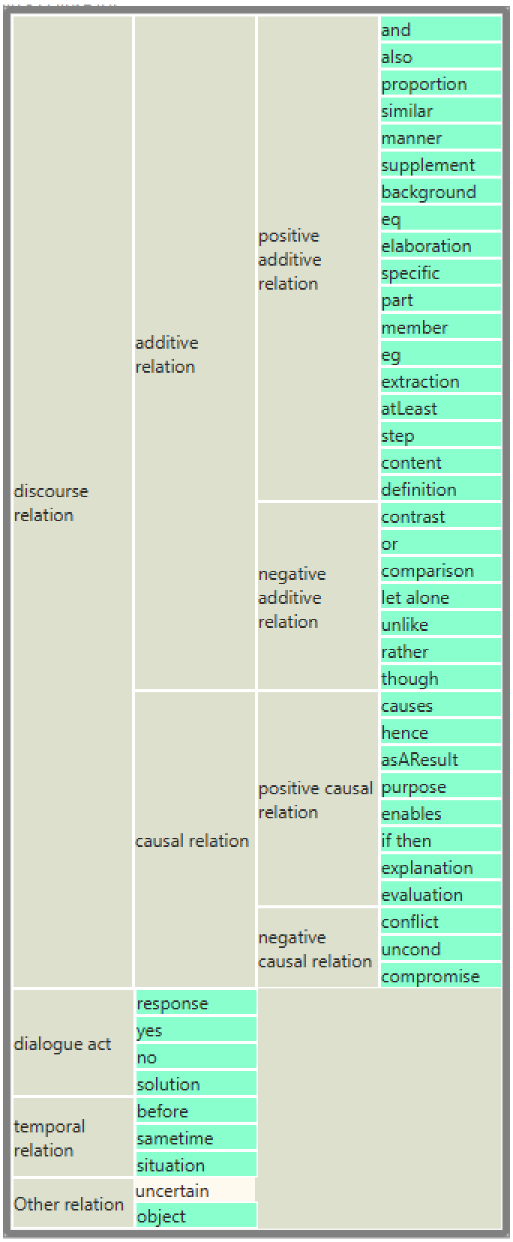

Figure 3-1: Ontology currently used in Semantic Editor

3.2. Document Evaluation

Although much work has been done regarding the quality of collaborative work, little work has been done on the quality of collaborative document composition, probably because document quality is hard to evaluate. There can be many dimensions, such as syntactic complexity, or textual cohesion, etc. In our study, to evaluate a document, the artistry is not a focus or a quality factor of the document. Instead, we are focusing on the richness of the explicit content. A method to evaluate such document quality has been proposed by Yagishita et al. (Yagishita, Munemori, & Sudo, 1998). They focus on some quality indicators (QIs) to evaluate the Petri net derived with or without KJ method. Similarly, to compare the content quality of text documents and graph documents, we need to first convert them into comparable standard format, i.e., the corresponding Petri nets. We convert text documents and graph documents into Petri nets with the following standard:

- A node in the Petri net contains a simple sentence or phrase, representing an information unit.
- There will not be two or more nodes in the Petri net that express the same information.
- The connection between nodes is the semantic relationship between information units.
- The semantic information embodied by net has a one-to-one relationship with text or graph; that is, no semantic information is lost or added during the conversion process.
- There will always be a node representing the main topic of text or graph.

In the conversion process, we will focus on the richness of the content, that is, we pay attention to whether there is new information to be discussed, and how the information is connected by semantics, instead of evaluating the content based on the artistic nature of the words and writing.

3.2.1. Graph-to-Net Conversion

The conversion of an RDF-graph document to net structure consists of three parts, and all of them require manual work. The first part is to replace the hypernodes with their inner content, thus, to include all content in one big graph. The second part is to convert the nodes in the graph to the nodes in the net. A node in the graph is usually converted to a node in the net. However, multiple nodes in the graph may be synonymous and hence converted to one node in the net. On the other hand, if a node contains multiple sentences that can be split into multiple information units, then it will be converted into multiple nodes in the net and connected by the designated semantic relationships. The third part of the graph-to-net conversion is to identify the root node and the descendant nodes in the net. The node containing the main topic will be regarded as the root, and the nodes on the discussion path that extended from the root node will be regarded as children nodes until this thread of discussion stops.

Figure 3-2 and Figure 3-3 show an example RDF-graph composed by Semantic Editor, and the corresponding converted net is shown in Figure 3-5. The red circled numbers in the RDF graph label an information unit and correspond to the nodes with the same number in the net. Figure 3-2 is the root graph of the document. In the root graph, node No.7, which is in blue font, is a hypernode containing an embedded graph shown in Figure 3-3. The first step of the conversion is to replace the hypernode with its inner content, and then we have the merged big graph shown in Figure 3-4. Based on the merged graph containing every information units, we have the converted Petri net shown in Figure 3-5, representing the same content as the RDF-graph.

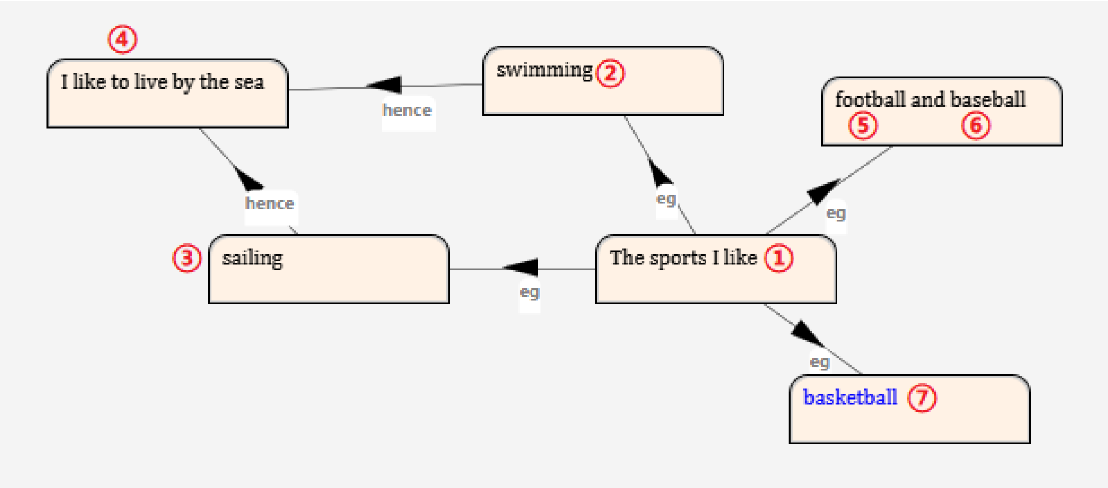

Figure 3-2 An RDF-graph document composed by Semantic Editor (root graph)

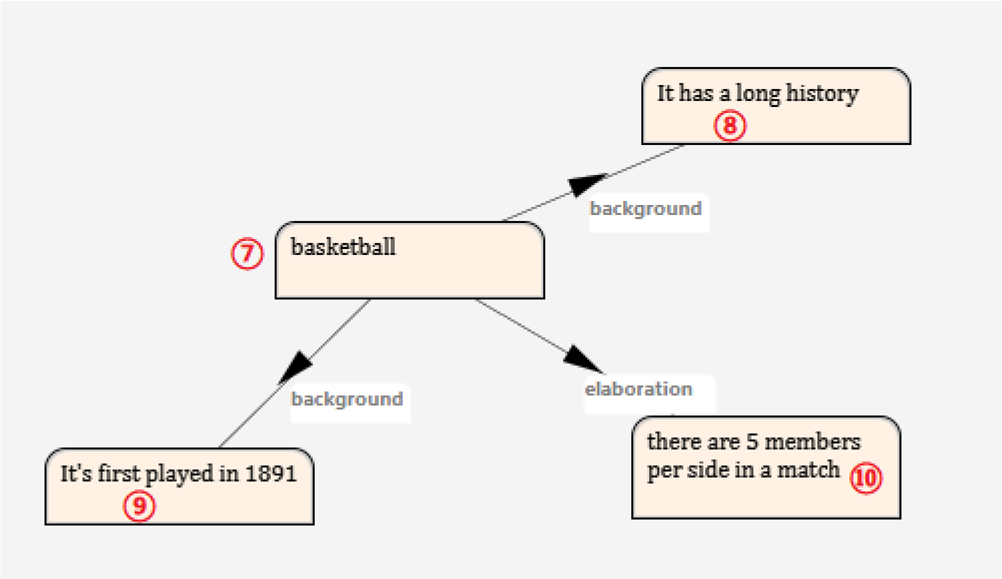

Figure 3-3 An RDF-graph document composed by Semantic Editor (embedded graph)

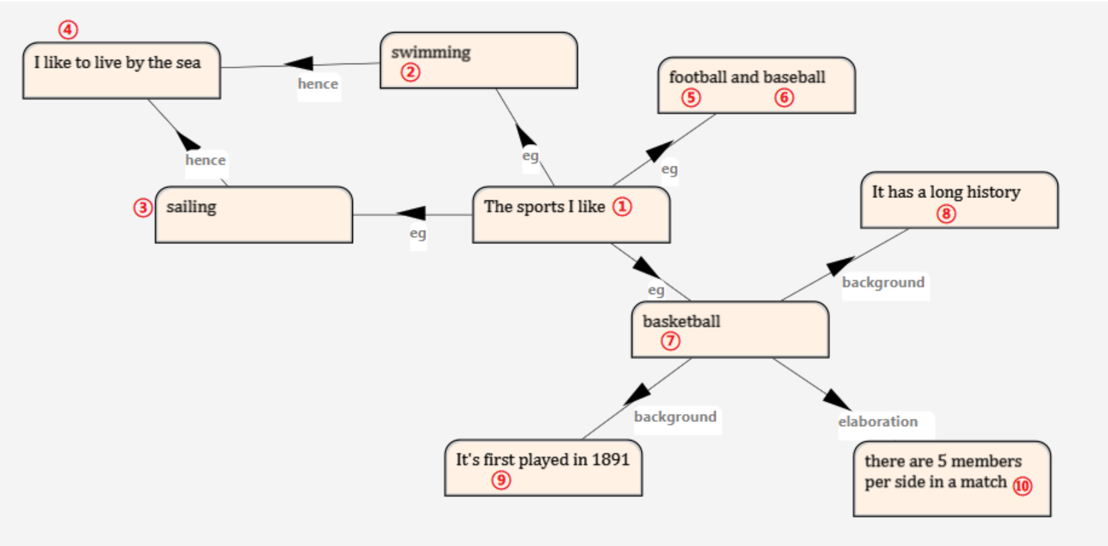

Figure 3-4 An RDF-graph document composed by Semantic Editor (merged)

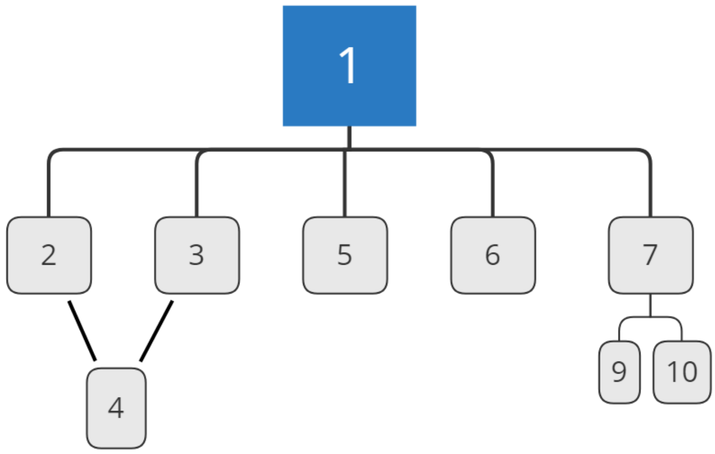

Figure 3-5 Corresponding net of the composed graph document

The title of this graph is the sports I like. Note that the No.5 and No.6 information units are written in the same node in the RDF graph, but apparently, they are two independent information units and should be split into two nodes in the net. This makes sense because there are five sports are mentioned in the document. Football and baseball should take similar positions as basketball does in this document. Moreover, the information unit No.8 expresses the duplicated meaning as unit No.9. Thus, No.8 is discarded in the converted net.

3.2.2. Text-to-Net Conversion

The text-to-net conversion consists of two parts involving manual work. The first part is to obtain nodes in the net. Here a sentence or a phrase in the text document is usually converted to a node in the net. As with the graph-to-net conversion, multiple sentences or parts of sentences may be synonymous and hence be converted to one node in the net. Moreover, a sentence may contain multiple information units to be converted to as many nodes in the net.

The second part of the text-to-net conversion is the same as that of the graph-to-net conversion, which is to identify the root node and the leaf nodes in the net.

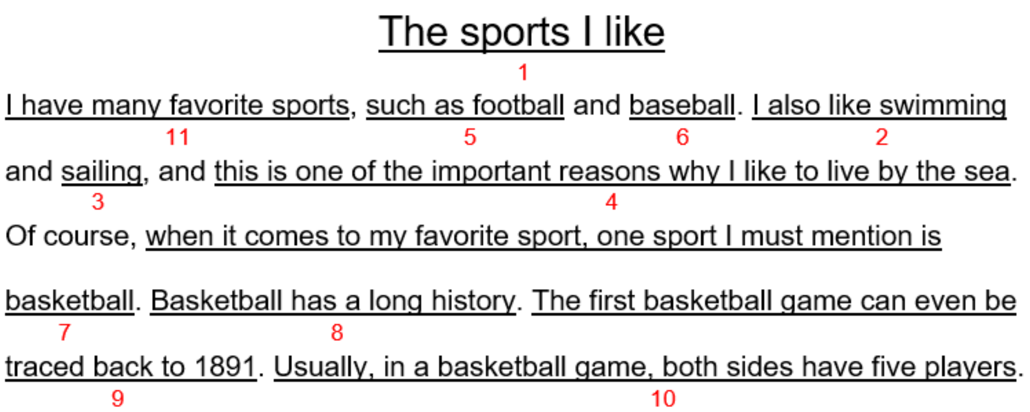
The sports I like
1
I have many favorite sports, such as football and baseball. I also like swimming
11 5 6 2
and sailing, and this is one of the important reasons why I like to live by the sea.
3 4
Of course, when it comes to my favorite sport, one sport I must mention is
basketball. Basketball has a long history. The first basketball game can even be
7 8
traced back to 1891. Usually, in a basketball game, both sides have five players.
9 10

Figure 3-6 An example text document

Figure 3-6 shows an example of a text document expressing the same content as Figure 3-2. To keep consistent with graph document, we label the information unit in the text document by the same numbers as in graph document, not in order instead. Of course, a similar Petri net derived from the text document should have the same structure as Figure 3-5, using a node to express an information unit in the document and connecting them by the semantic relationships. At the same time, unrelated or duplicated parts should be discarded. The corresponding net for this text document is shown in Figure 3-7, and it is the same as Figure 3-5.

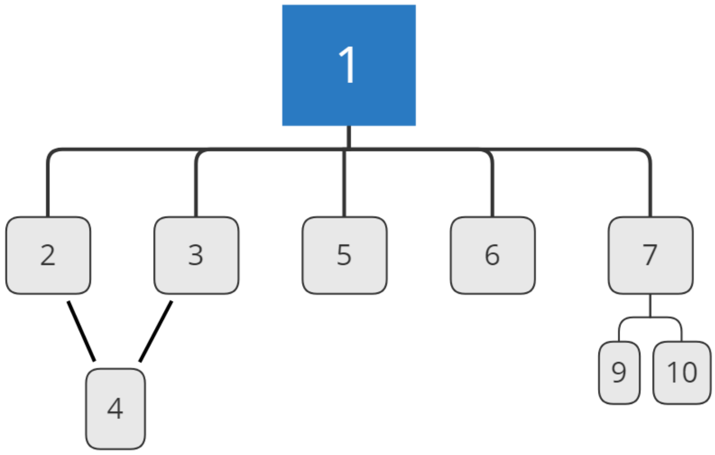

Figure 3-7 Corresponding net of the composed text document

3.2.3. Quality Indicators

Since we focus on the richness of the explicit content, we do quantitative analysis based on the converted net. The quality indicators are defined based on the richness parameters of the net, such as the number of nodes, the number of connections, and the number of descendant nodes of a node. In the experiment of asynchronous collaborative authoring, we will also analyze the connection between the content generated by the two asynchronous phases, such as the number of links between the two parts. The detailed definition of the quality indicators used in the two experiments will be discussed in section 4.2 and section 5.2, respectively.

4. Synchronous Collaborative Authoring

This chapter will show our experimental research on synchronous collaborative authoring. We designed experiments to allow participants to compose in text or RDF graph according to a given topic. Through the conversion process mentioned in section 3.2, we converted the obtained documents into the corresponding nets. We then quantitatively analyzed the obtained net based on the quality indicators to verify our hypotheses.

4.1. Experiment Design

18 research participants participated in our experiment, and they all met the following conditions:

- They understand what graph documents are.
- They can determine the relationship between nodes, such as causality, purpose, etc.
- They are willing to collaborate with others to compose documents.

The experiment consists of 18 sessions of collaborative document composition. In each session, two research participants will be invited to the designated experimental site, and at the same time, face to face to collaboratively compose a text document or a graph document to carry out a particular task. During the experiment, they can communicate in person.

We divided the 18 research participants into 9 groups, each with 2 people, and asked each group to participate in two document-composition sessions. In one session, they used Google Docs[4] to compose a text document addressing one task, and in the other session, they used Semantic Editor to compose an RDF-graph document addressing another task. For each task, a text document and an RDF-graph document were composed by different groups. For each group of participants, they never encountered the same task in two sessions.

Each session lasted a maximum of 30 minutes, within which the participants were asked to complete a document (either text or graph). Each participant used a PC. We provided the PCs, mouse devices, and keyboards to the participants. There was no bottleneck concerning the physical environment of the experiment.

[4] https://www.google.com/docs/about/

To encourage the research participants to create meaningful documents with many information units, we devised 9 tasks concerning topics familiar to the participants so that they could think of many relevant points. These tasks belong to three categories, each promoting a different writing style. Table 4-1 shows the details of the tasks and the schedule of the whole experiment.

Table 4-1 experiment tasks and schedule[5]

Categories	**Tasks**	**Texts**	**Graphs**
Agree or Disagree	1. Some people say that you should get the highest possible degree, not to work too early. Do you agree? Why?	Group 1	*Group 2
	2. Some people say that online shopping is a better way to shop. Do you agree? Why?	*Group 2	Group 1
	3. Some people say that everyone must learn a bit of programming now. Do you agree? Why?	*Group 3	Group 4
Advantages and Disadvantages	4. What are the advantages and disadvantages of living in cities?	Group 4	*Group 3
	5. If we could travel in time, what are the advantages and disadvantages?	*Group 5	*Group 6
Introduce Preferences	6. If you could get a superpower, what do you want to get, explain it?	*Group 6	*Group 5
	7. What sports do you like?	Group 7	Group 8
	8. Introduce Japanese food.	Group 8	Group 9

[5] The data of sessions marked with an asterisk (*) were collected by Zifan YAO, Hasida Lab, the University of Tokyo.

	9. Introduce tourist attractions in China.	Group 9	Group 7

4.2. Quality Indicators

Since we are concerned about the richness of explicit content, we use the following quality indicators (QIs) to evaluate the quality of document content quantitatively.

- **Size** is the number of nodes in the net. It is the number of information units or the number of ideas. Large documents are informative.
- **Height** is the maximum number of nodes in one thread from the start to the end. In-depth discussions tend to be high.
- **Expansion Degree** is defined based on the definition of downward degree. The downward degree of a node is the number of nodes linked downwardly to this node. In a net structure, the downward degree is equivalent to the number of children of a node. Expansion degree is defined as:

  if $T_{out} \neq \emptyset$, then

  Expansion Degree = $\sum_{t \in T_{out}} \{t_{out}(N) - 1\}$

  else **Expansion Degree** = 0

  T_{out} is the set of nodes with downward degree greater than or equal to 2.

  $t_{out}(N)$ is the downward degree of a node.

  The expansion degree indicates how many times in the entire document, the content of discussion has expanded out of additional branches to discuss the same issue from different perspectives. The larger the expansion degree, the more comprehensive the document.

- **Aggregation Degree** is defined based on the definition of upward degree. The upward degree of a node is the number of nodes linked upwardly to this node. In a net structure, the upward degree is the number of parents of a node. Aggregation degree is defined as:

  if $T_{in} \neq \emptyset$, then

  Aggregation Degree = $\sum_{t \in T_{in}} \{t_{in}(N) - 1\}$

  else **Aggregation Degree** = 0

  T_{in} is the set of nodes with upward degree greater than or equal to 2.

  $t_{in}(N)$ is the upward degree of a node.

  The aggregation degree indicates how many times in the entire document, the content discussed has been summarized. The larger the aggregation degree, the higher the summarizing ability of the document.

By converting both RDF-graph documents and ordinary text documents by the conversion process discussed in section 3.2, we get the corresponding Petri nets. Then we calculate the above quality indicators of the Petri nets to quantitively evaluate the quality of the graph document and the text document, respectively.

Figure 4-1 shows an example of the net structure. In this example, there are 10 nodes in the net, so that its size is 10. The height of the net is 3 because the furthest node from the root is 3 layers down, the path 1-2-6-10 is one of the possible paths from the root node to the furthest node.

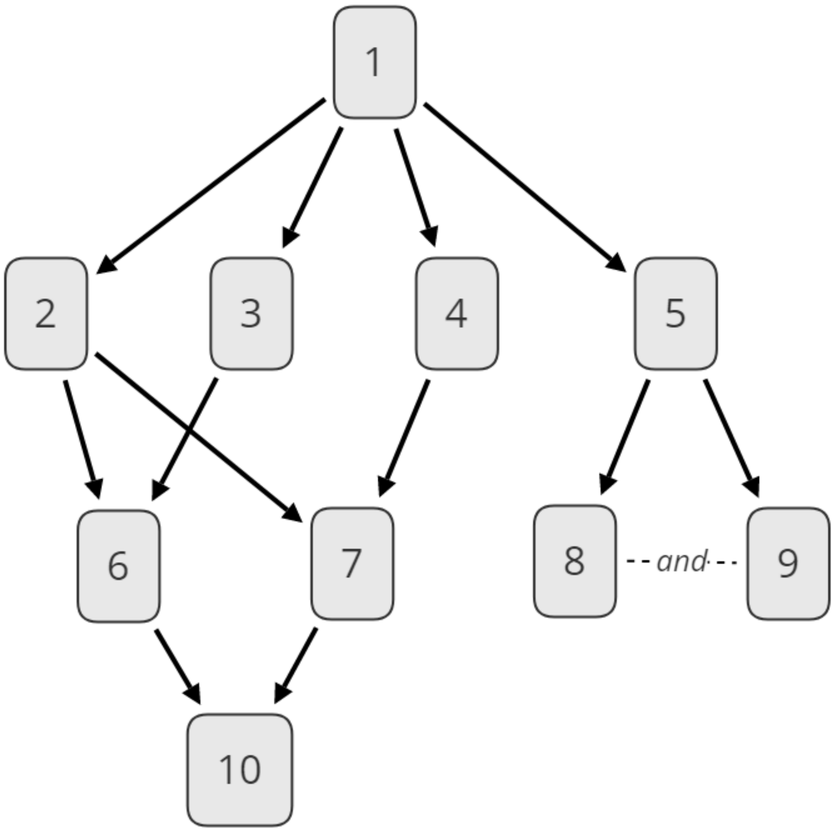

Figure 4-1 An example of converted net

- And
- Similar
- Equal
- Contrast
- Or
- Unlike
- Conflict
- Sametime

Table 4-2 shows the downward degree and the upward degree of the nodes in the example. By the definition of expansion degree, only the nodes with downward degree greater or equal to 2 should be counted; thus, the expansion degree of this net is $(4-1)+(2-1)+(2-1)=5$; similarly, by the definition of aggregation degree, only the nodes with upward degree greater or equal to 2 should be counted; thus, the aggregation degree of this net is $(2-1)+(2-1)+(2-1)=3$. This makes sense because, in the net, the content of discussion has expanded out of additional branches five times, there are in total five side-branch in this net. At the same time, the discussion is concluded three times, and there are three branches merged into another.

Note that the relationship between node No.8 and node No.9 is symmetric, and it does not contribute to any downward degree or upward degree since node No.8 and node No.9 are in an equal position, and there is no progressive relationship between them. In the ontology we currently use in Semantic Editor, symmetric relationships include:

- And
- Similar
- Equal
- Contrast
- Or
- Unlike
- Conflict
- Sametime

Table 4-2 Downward degree and upward degree of the nodes in the example

Node	Downward degree	Upward degree
1	4	0
2	2	1
3	1	1
4	1	1
5	2	1
6	1	2
7	1	2
8	0	1
9	0	1
10	0	2

4.3. Hypotheses

The purpose of the synchronous collaborative authoring experiment is to verify whether RDF-graph, as a form of document and a carrier of content, is more productive than text in a synchronous collaborative authoring environment. The general hypothesis of this experiment is:

RDF-graph is more productive than text in synchronous collaborative authoring.

According to the QIs we defined and discussed in section 4.2, we decompose this general hypothesis into the following four hypotheses, corresponding to the four structural QIs.

- *Hypothesis 1:* RDF-graphs are larger than texts in synchronous collaborative authoring.
- *Hypothesis 2:* RDF-graphs are taller than texts in synchronous collaborative authoring.
- *Hypothesis 3:* RDF-graphs are more expansive than texts in synchronous collaborative authoring.
- *Hypothesis 4:* RDF-graphs are more aggregative than texts in synchronous collaborative authoring.

4.4. Statistical Analysis

According to the conversion process we discussed in section 3.2 and the QIs we discussed in section 4.2, we calculate the scores of the QIs for text documents and graph documents and show the result in Table 4-3 and Table 4-4, respectively.

Table 4-3 QI scores of text documents

QIs	Size	Height	Expansion	Aggregation
Task 1	11	6	2	0
Task 2	15	3	10	1
Task 3	13	3	6	0
Task 4	17	5	7	0
Task 5	13	5	2	1
Task 6	9	3	5	1
Task 7	20	3	9	1
Task 8	20	4	10	0
Task 9	23	3	13	0

Average	15.67	3.89	7.11	0.44

Table 4-4 QI scores of graph documents

QIs	Size	Height	Expansion	Aggregation
Task 1	17	9	2	0
Task 2	18	5	6	0
Task 3	25	6	15	1
Task 4	24	3	16	2
Task 5	20	4	8	0
Task 6	31	4	16	0
Task 7	24	6	10	0
Task 8	20	6	13	4
Task 9	27	7	14	0
Average	22.89	5.56	11.11	0.78

Before we compare the QI scores, to eliminate the differences in the difficulty among the nine tasks, we need to normalize the data so that the average value of the graph and the text be 1.0 for each the structural quality indicator (size, height, expansion degree, and aggregation degree) for each task. In a task, if the text size and the graph size are t and g, respectively, then the normalized size of text and graph are defined as follows.

if t + g = 0, then

normalized size of text = 1

normalized size of graph = 1

else

normalized size of text = 2t / (t + g)

normalized size of graph = 2g / (t + g)

The normalized sizes, heights, widths, expansion degrees, and aggregation degrees of texts and graphs for the 9 tasks are shown in Figure 4-2 to Figure 4-5, respectively. The average values of normalized QIs are shown in Figure 4-6.

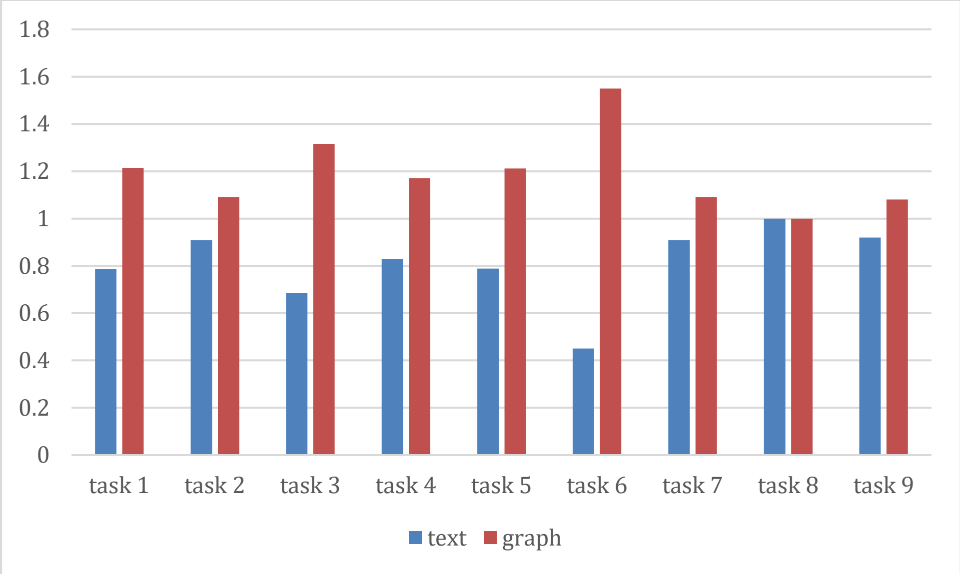

Figure 4-2 Normalized sizes

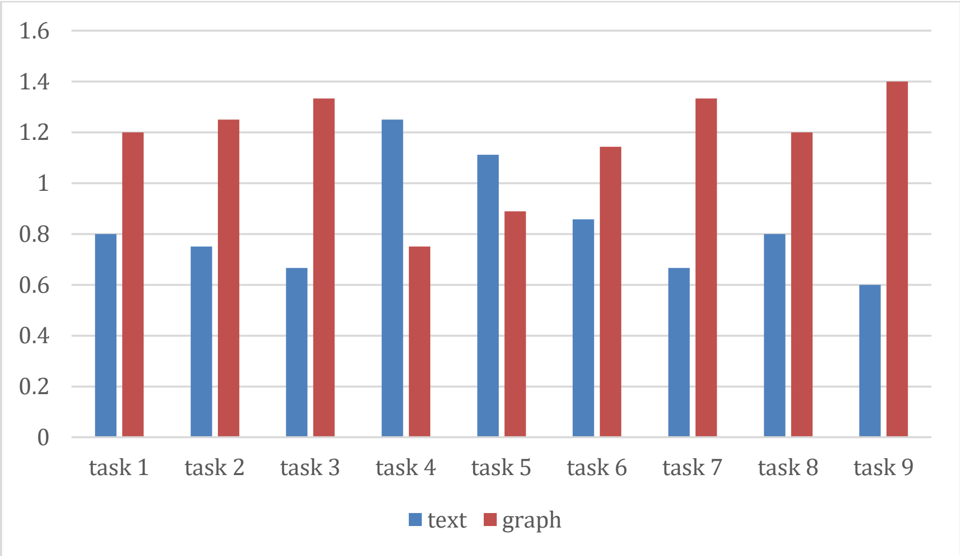

Figure 4-3 Normalized heights

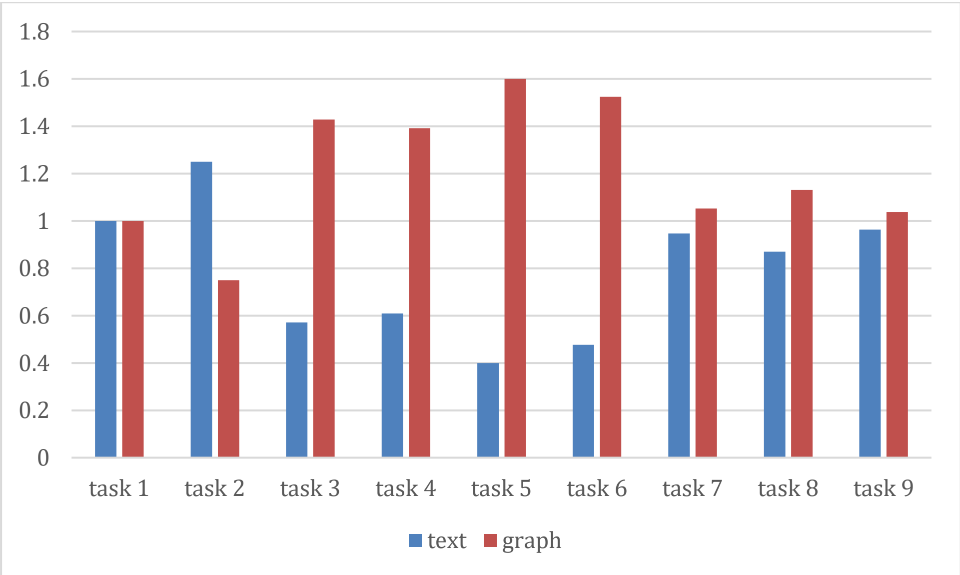

Figure 4-4 Normalized expansion degrees

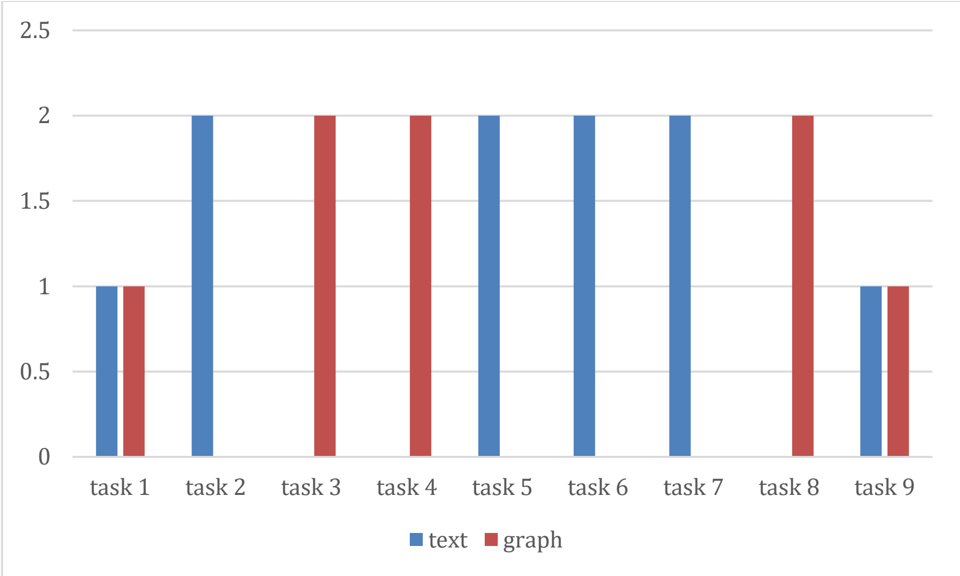

Figure 4-5 Normalized aggregation degrees

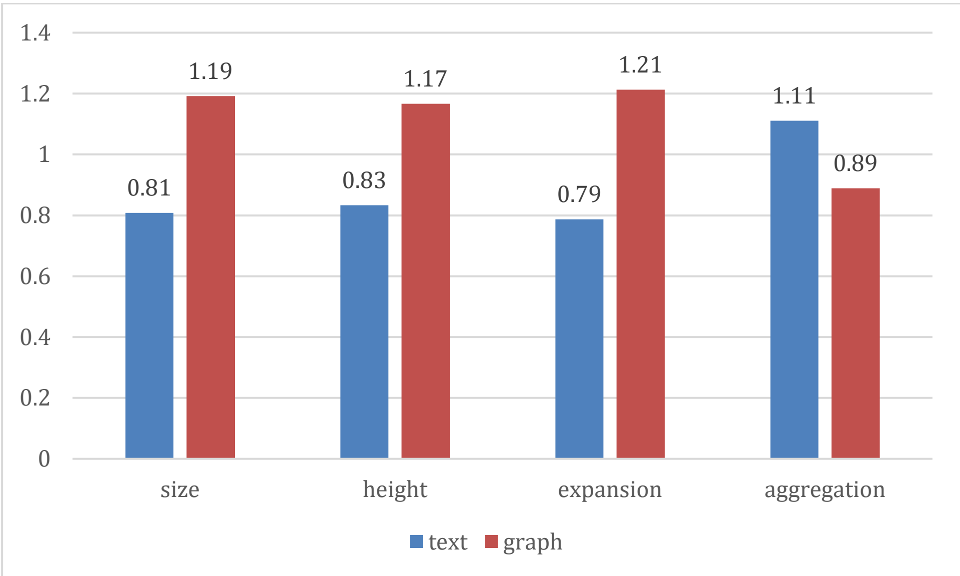

Figure 4-6 Average of normalized QIs

From the data, we can see that the aggregation degree data is very extreme, whether it is text or graph, the normalized aggregation degree is usually 0 or 2. Comparing Table 4-3 and Table 4-4, we can see that the aggregation degree value tends to zero, which makes the aggregation degree not statistically significant, we will see the detailed results from the hypothesis tests.

4.5. Hypothesis Tests

Based on the data discussed above, we test the four hypotheses we mentioned in section 4.3. We apply one-tailed paired t-test to these hypotheses. Table 4-5 summarizes the text-graph comparisons with respect to the four structural QIs. The average aggregation degree of texts is higher than that of RDF-graphs, but the p-value is apparently too large, so that we cannot conclude that text has an advantage in aggregation degree than RDF-graph. Thus hypotheses 1, 2, and 3 survive under significance level α=5%, strongly supporting the original hypothesis, which is to the effect that RDF-graph is more productive than text in synchronous collaborative authoring. In the meanwhile, hypotheses 4 is not established from our data. In this case, there is not enough evidence to show that RDF-graphs are more aggregative than texts in synchronous collaborative authoring. However, texts do not show any advantage over graphs in aggregation degree, either.

Table 4-5 Comparison of text and graph

	Size	Height	Expansion	Aggregation
Text Mean	0.81	0.83	0.79	1.11
Graph Mean	1.19	1.17	1.21	0.89
P-Value	0.0040	0.0244	0.0276	0.3644

4.6. Summary

In the experiment of synchronous collaborative authoring, we confirmed that RDF-graph is more productive than text in synchronous collaborative authoring. Among them, we use size, height, expansion degree, and aggregation degree as QIs to quantitatively evaluate the quality of the document and find that the size, height, and expansion degree all strongly support the original hypothesis, that is, RDF-graphs are larger, taller, and more expansive than texts in synchronous collaborative authoring. The data of aggregation degree seems to be extreme, and there are many zero cases. The normalized average aggregation degree of texts is higher than that of RDF-graphs, but the p-value is too large, so that we cannot conclude that text has an advantage in aggregation degree than RDF-graph. As a result, we do not have enough evidence to prove that RDF-graphs are more aggregative than texts in synchronous collaborative authoring. Nevertheless, at the same time, texts do not show any advantage over graphs in aggregation degree, either.

4.6.1. In-Depth Document Analysis

The documents composed in our experiment show different structural characteristics for different tasks. In some cases, the participants conducted an in-depth discussion on just one aspect of the task. As their thought focused on this specific aspect, the document they composed present a tall and narrow net structure. For example, in tasks 1 and 2, the graphs are not more expansive but taller or larger than the texts. On the other hand, participants sometimes seemed to think about all aspects of the task, so that the discussion on each aspect was shallow due to the time constraint. Hence the resulting document has a wide and short net structure. For example, in tasks 4 and 5, the graph documents are not taller but more expansive and larger than the text documents. This suggests that graph documents are probably larger than the text documents for the same task even if the graphs are high-narrow or wide-short.

The documents in our experiment are obviously different from the ones in Yagishita's experiment (Yagishita, Munemori, & Sudo, 1998). The documents from their experiment mainly focused on finding solutions for given problems, so the process of discussion was very aggregated, and participants would discuss from multiple divergent aspects for some specific points. The discussion of our text is open and does not require participants to analyze specific issues in-depth or draw clear conclusions; thus, it does not show a significant difference in the aggregation degree between texts and graphs. Specifically, in our experiment, the aggregation degrees of texts or graphs are very low. In many sessions, the aggregation degree is just zero.

The document qualities might be biased due to the composition abilities of groups of participants and the complexities of tasks. However, the effect of variances has been calculated during the hypothesis test, and the results strongly support our hypotheses. Since each group composed both a text document and a graph document, working on different tasks in different sessions, this is an unbiased sample. This sort of experimental design favors neither text documents nor graph documents and can weaken or eliminate the impact of differences in the capabilities of different groups. This backs up the statistical significance of the superiority of graphs to texts in our experiment.

4.6.2. Analysis of Other Indicators

In Yagishita's experiment, the net always had a simple hierarchical structure, and no complex cycle was involved in the net. Thus, they used width as a QI, which is defined as the maximum number of nodes at the same level in the hierarchical graph. However, in our experiment, there are cycles in the net; hence the idea of width is not suitable here. However, we calculated expansion degree and aggregation degree to show the comprehensiveness of the document. These two indicators can be a substitute for the width in such a cyclic structure. Expansion degree and aggregation degree are also defined in Yagishita's experiment; however, unlike our definition, they calculated the average downward degree or upward degree of nodes. We name these two indicators as average expansion degree and average aggregation degree, defined as follows.

- **Average Expansion Degree** is defined based on the definition of downward degree. The downward degree of a node is the number of nodes linked downwardly to this node. In a net structure, the downward degree is equivalent to the number of children of a node. Average expansion degree is defined as:

  if $T_{out} \neq \emptyset$, then

  $$\textbf{Average Expansion Degree} = \frac{\sum_{t \in T_{out}} \{t_{out}(N) - 1\}}{|T_{out}|}$$

  else **Average expansion Degree** = 0

  T_{out} is the set of nodes with downward degree greater than or equal to 2.

  $t_{out}(N)$ is the downward degree of a node.

- **Average Aggregation Degree** is defined based on the definition of upward degree. The upward degree of a node is the number of nodes linked upwardly to this node. In a net structure, the upward degree is the number of parents of a node. Average Aggregation degree is defined as:

if $T_{in} \neq \emptyset$, then

$$\textbf{Average Aggregation Degree} = \frac{\sum_{t \in T_{in}} \{t_{in}(N) - 1\}}{|T_{in}|}$$

else **Average Aggregation Degree** = 0

T_{in} is the set of nodes with upward degree greater than or equal to 2.

$t_{in}(N)$ is the upward degree of a node.

We do not think the average expansion degree and average aggregation degree can explain the comprehensiveness of the document or be QIs for the experiment, but we still calculate the value and show the result in Table 4-6. Moreover, if we normalize these two scores and apply one-tailed paired t-test to these data, we found out that in our experiment, there is not enough evidence to show a difference in average expansion degree or average aggregation degree between texts and graphs. The result in shown in Table 4-7.

Table 4-6 Average expansion degree and average aggregation degree

	Text		Graph	
QIs	**Average expansion**	**Average aggregation**	**Average expansion**	**Average aggregation**
task1	1.00	0.00	1.00	0.00
task2	1.67	1.00	2.00	0.00
task3	1.50	0.00	1.88	1.00
task4	1.75	0.00	2.67	1.00
task5	1.00	1.00	1.33	0.00
task6	2.50	1.00	2.00	0.00
task7	1.80	1.00	1.43	0.00
task8	1.67	0.00	2.17	1.00
task9	4.33	0.00	2.33	0.00
Average	1.91	0.44	1.87	0.33

Table 4-7 Comparison of average expansion and average aggregation

	Average expansion	Average aggregation
Text Mean	0.98	1.11
Graph Mean	1.02	0.89
P-Value	0.3788	0.3644

4.6.3. Improvement Plan

In this experiment, we have been able to prove that graphs have advantages in real-time face-to-face simultaneous collaborative authoring. However, through observation, we still found some new problems.

First of all, in the experiment, whether it is text or graph, a typical collaboration process is that two participants spend a period of time for oral discussion, then divide the whole task into two, and then complete their corresponding part of the task, respectively. We found that neither text nor graph contributed much to the collaboration and communication per se. The process of collaboration and communication was completed by both participants through oral discussion. Besides, the so-called collaboration in such a way of dividing task makes the content of the document split. In the actual composition process, the interaction between the two participants on their content is low. However, this process is not quantitively evaluated. This is one of the driving factors of the second experiment, the experiment of asynchronous collaborative authoring.

Second, 30-minute experiments are often slightly shorter for co-authoring. In this experiment, sometimes participants will spend more time discussing, but only a small part of the time for writing. We should increase the experiment time in the following experiment.

Third, the participants in this experiment are all school students, which may introduce some kind of bias. In the following experiment, we need to invite more participants with different identities to conduct the experiment to balance the participants' background difference, eliminate bias.

Summarizing the experience of this experiment, we completed the experiment of asynchronous collaborative authoring, and report the details of it in Chapter 5.

5. Asynchronous Collaborative Authoring

In Chapter 4, we conducted the experiment of synchronous collaborative authoring and proved that RDF-graph is more productive than text in synchronous collaborative authoring. However, people involved in co-authoring do not always have the opportunity to co-edit in the same place face to face at the same time. At the time of writing this thesis, the current COVID-19 pandemic on a global scale is enough to make people deeply realize that non-simultaneous asynchronous collaborative authoring should be paid more attention.

In section 4.6, we summarized the problems observed in the experiment of synchronous collaborative authoring. In order to solve these problems and study the help of RDF-graph and text for people's cooperation, we designed an experiment of asynchronous collaborative authoring and proposed some new QIs suitable for asynchronous collaborative authoring to test our hypotheses quantitatively.

5.1. Experiment Design

In this experiment, a total of 10 participants participated in our experiment, and they all met the following conditions:

- They understand what graph documents are.
- They can determine the relationship between nodes, such as causality, purpose, etc.
- They are willing to collaborate with others to compose documents.

In this experiment, we prepared 10 tasks, and each task will be composed by different participants in the form of text or graph. For each form of the document, it was composed by two participants in two phases. Among them, the first participant will write a draft based on the task topic, and the second participant will complete the entire document based on the task topic and combined with the draft left by the first participant. For any task, 4 sessions are corresponding to it, namely text phase 1, text phase 2, graph phase 1, and graph phase 2. For each experimental participant, they will participate in these four sessions on different four days, and the task he/she faces is different every time. Moreover, for each task, its corresponding 4 sessions are completed by four different participants. In total, there are 40 sessions in this experiment, and each session lasted a maximum of 30 minutes.

For any participant, before participating in the experiment, he/she will be told that this is a two-stage experiment, and he/she will need to collaborate with an anonymous participant to complete a document. For the first author of each form, i.e., the one who conducts text phase 1 or graph phase 1, he/she needs to leave a draft for the second participant based on the task topic. There are no requirements for the genre and format of the draft. For the second participant, he/she will complete the entire document according to the task topic and the draft left by the first participant. The second participant will be informed that he/she has the right to make any additions, deletions, modifications, and other operations to the draft, but the final document must be a complete and unobstructed text document or graph document.

Obviously, this experiment is not synchronized, there is no need for face-to-face collaboration at the same time, and there is no additional communication between the two collaborators. The primary device of this experiment is an Amazon EC2[6] virtual machine (VM) located in Tokyo. Table 5-1 shows the detailed configurations of the virtual machine instance.

Table 5-1 Configuration of experiment machine

Configuration	**Details**
VM service	Amazon EC2
Availability Zone	ap-northeast-1a
Physical location	Tokyo
Instance type	t2.large
Number of vCPUs	2
Processor	Intel® Xeon® CPU E5-2676 v3 @ 2.40G Hz
Architecture	x86_64
Memory Size	8192 MiB
Storage Size	30 GiB
Storage Type	gp2
IOPS	100
Platform	Windows Server 2019 64-bit Operating System

Each participant in the experiment uses their own PC to connect to the VM instance for operation. Participants will prepare their own PC, mouse, keyboard, and other necessary facilities and have good network access to the VM instance. In order to unify the device environment of experiment participants to the greatest extent. Whether it is a text document or a graph document, participants will connect to the VM instance to perform operations. For the text document, the participants will compose by Google Docs through the VM. For the graph document, we will prepare the operating environment of the Semantic Editor in the VM instance in advance, and the participants will compose by the Semantic Editor through the VM.

[6] https://aws.amazon.com/ec2/

Before the experiment, we explained the experiment process and conducted simple operation training to the participants through the remote desktop software TeamViewer[7]. We conducted a simple training on the basic operations of Google Docs and Semantic Editor, including operations such as addition, deletion, modification, etc. Moreover, during the experiment, we observed the operation of participants through the TeamViewer. If the participant used a Windows system, the experiment participant would use Remote Desktop Services[8] to connect to the VM instance, and the observation would be done through the TeamViewer. For macOS users, the observer logged in to the VM instance through Remote Desktop Services to observe, and the participant performed remote desktop operations through the TeamViewer. Throughout the experiment, the participants' network conditions and operation fluency were good.

In order to encourage participants to compose rich documents as much as possible and avoid bias caused by different professional backgrounds of participants, we selected 10 fictional task topics, each of which is not based on any specific professional knowledge or only based on some well-known common sense. The 10 tasks are described as follows, and Table 5-2 shows the schedule of the whole experiment.

1. After the quarantine of the COVID-19, you and your friends plan to have a party. Please describe what you will do at the party.
2. The entrance examination of the University of Tokyo is affected by the COVID-19, so the school decided to hold an online examination. But to prevent cheating, we need more measures to help. The school is super rich and can use any existing technology or device for this exam. Please describe how this examination could be held.
3. You and your friend plan to develop a VR game, the player will play an alien on another planet, describe what a player can experience in the game.
4. In the year of 3030, humans still live on earth, and the techniques are much well developed. Describe a daily life for someone living in 3030.
5. You and your friend plan to climb the mount Himalayas, please describe what you need to prepare for this.
6. You and your friend found a new island in the Pacific Ocean, and it is big enough to establish a new country. You and your friend decide to establish a new country and will rule this island, describe what you will do.
7. You and your friend found a way to redesign the human physiological structure. Describe what you will design for the next generation of human beings.
8. You and your friend devised a machine that can make anything invisible for a day. Describe how you will use this machine.
9. You and your friend won a billion dollars in the lottery, describe how you will spend this money.

[7] https://www.teamviewer.com/en/
[8] https://docs.microsoft.com/en-us/windows-server/remote/remote-desktop-services/welcome-to-rds

10. You and your friend got a superpower, describe what it is and how you will make use of it.

Table 5-2 Experiment schedule

Tasks	Text phase 1	Text phase 2	Graph phase 1	Graph phase 2
1	Participant 1	Participant 2	Participant 4	Participant 3
2	Participant 4	Participant 3	Participant 1	Participant 2
3	Participant 3	Participant 4	Participant 2	Participant 1
4	Participant 2	Participant 1	Participant 3	Participant 4
5	Participant 5	Participant 6	Participant 8	Participant 7
6	Participant 8	Participant 7	Participant 5	Participant 6
7	Participant 7	Participant 8	Participant 10	Participant 9
8	Participant 10	Participant 9	Participant 7	Participant 8
9	Participant 9	Participant 10	Participant 6	Participant 5
10	Participant 6	Participant 5	Participant 9	Participant 10

5.2. Quality Indicators

Since we are concerned about the richness of explicit content and the collaboration level between phase 1 and phase 2 of a task, we use the following quality indicators (QIs) to quantitatively evaluate the quality of document content and the collaboration level. The definitions of size, height, expansion degree, and aggregation degree are the same as section 4.2, but for convenience, we also list them along with new QIs defined here.

- **Size** is the number of nodes in the net. It is the number of information units or the number of ideas. Large documents are informative.
- **Height** is the maximum number of nodes in one thread from the start to the end. In-depth discussions tend to be high.
- **Expansion Degree** is defined based on the definition of downward degree. The downward degree of a node is the number of nodes linked downwardly to this node. In a net structure, the downward degree is equivalent to the number of children of a node. Expansion degree is defined as:

if $T_{out} \neq \emptyset$, then

Expansion Degree = $\sum_{t \in T_{out}} \{t_{out}(N) - 1\}$

else **Expansion Degree** = 0

T_{out} is the set of nodes with downward degree greater than or equal to 2.

$t_{out}(N)$ is the downward degree of a node.

The expansion degree indicates how many times in the entire document, the content of discussion has expanded out of additional branches to discuss the same issue from different perspectives. The larger the expansion degree, the more comprehensive the document.

- **Aggregation Degree** is defined based on the definition of upward degree. The upward degree of a node is the number of nodes linked upwardly to this node. In a net structure, the upward degree is the number of parents of a node. Aggregation degree is defined as:

  if $T_{in} \neq \emptyset$, then

  Aggregation Degree = $\sum_{t \in T_{in}} \{t_{in}(N) - 1\}$

  else **Aggregation Degree** = 0

  T_{in} is the set of nodes with upward degree greater than or equal to 2.

  $t_{in}(N)$ is the upward degree of a node.

  The aggregation degree indicates how many times in the entire document, the content discussed has been summarized. The larger the aggregation degree, the higher the summarizing ability of the document.

- **Connectivity** is the number of links composed in phase 2 that directly connect any nodes composed in phase 2 with any nodes composed in phase 1. The connectivity shows the interaction level of the content composed in two phases. The tighter collaborated composition will result in higher connectivity.

By converting both RDF-graph documents and ordinary text documents by the conversion process discussed in section 3.2, we get the corresponding Petri nets. Then we calculate the above quality indicators of the Petri nets to quantitively evaluate the quality of the graph document and the text document, respectively.

Figure 5-1 shows an example of the net structure. In this example, the blue nodes and links are composed in phase 1, and the red nodes and links are composed in phase 2. There are 10 nodes in the net so that its size is 10. The height of the net is 3 because the furthest node from the root is 3 layers down, the path 1-2-6-10 is one of the possible paths from the root node to the furthest node. The connectivity of the net is 4, because there are 4 links directly connect nodes composed in phase 1 with nodes composed in phase 2, and these four links are 1-4, 1-5, 2-7, and 7-10.

Note that if we only consider the content composed in phase 1, the size and height are 5 and 3, respectively. However, after the composition of phase 2, the content is enriched, and the QIs are enlarged.

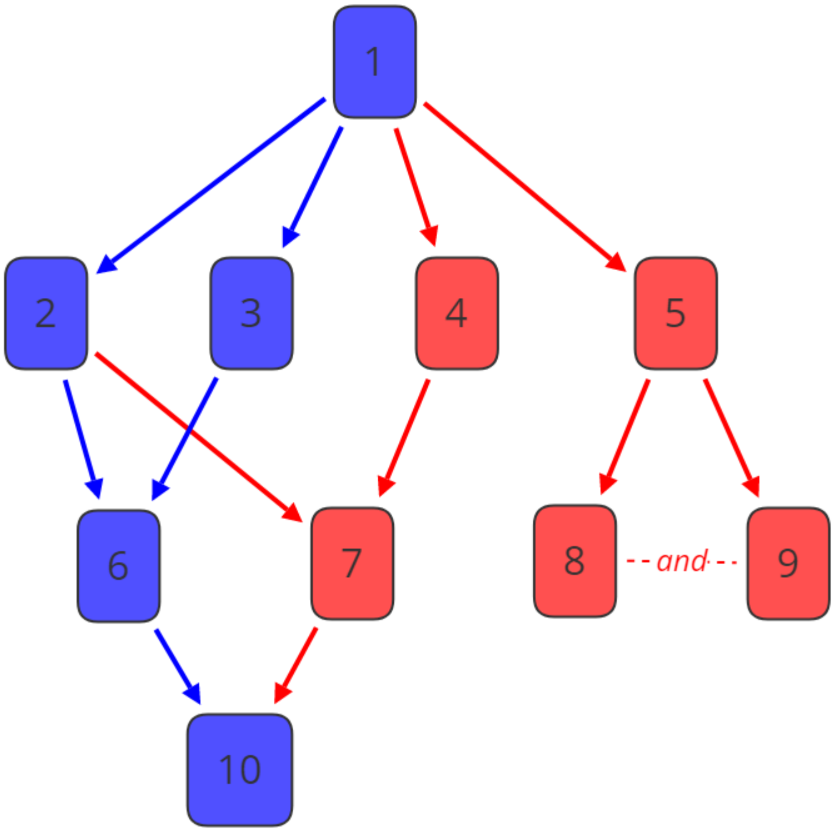

Figure 5-1 An example of converted net

- And
- Similar
- Equal
- Contrast
- Or
- Unlike
- Conflict
- Sametime

Table 5-3 shows the downward degree and the upward degree of the nodes in the example. By the definition of expansion degree, only the nodes with downward degree greater or equal to 2 should be counted; thus, the expansion degree of this net is $(4-1)+(2-1)+(2-1)=5$; similarly, by the definition of aggregation degree, only the nodes with upward degree greater or equal to 2 should be counted; thus, the aggregation degree of this net is $(2-1)+(2-1)+(2-1)=3$. This makes sense because, in the net, the content of discussion has expanded out of additional branches five times, there are in total five side-branch in this net. At the same time, the discussion is concluded three times, and there are three branches merged into another.

Note that the relationship between node No.8 and node No.9 is symmetric, and it does not contribute to any downward degree or upward degree since node No.8 and node No.9 are in an equal position, and there is no progressive relationship between them. In the ontology we currently use in Semantic Editor, symmetric relationships include:

- And
- Similar
- Equal
- Contrast
- Or
- Unlike
- Conflict
- Sametime

Table 5-3 Downward degree and upward degree of the nodes in the example

Node	Downward degree	Upward degree
1	4	0
2	2	1
3	1	1
4	1	1
5	2	1
6	1	2
7	1	2
8	0	1
9	0	1
10	0	2

Note that if we only consider the content composed in phase 1, both the expansion degree and the aggregation are 1. However, after the composition of phase 2, the content is enriched, and the QIs are enlarged.

5.3. Hypotheses

The purpose of the asynchronous collaborative authoring experiment is to verify whether RDF-graph, as a form of document and a carrier of content, is more productive and more conducive to collaboration than text in an asynchronous collaborative authoring environment. The general hypothesis of this experiment is:

RDF-graph is more productive and more conducive to collaboration than text in asynchronous collaborative authoring.

According to the QIs we defined and discussed in section 5.2, we decompose this general hypothesis into the following five hypotheses, corresponding to the five structural QIs.

- *Hypothesis 1:* RDF-graphs are larger than texts in asynchronous collaborative authoring.
- *Hypothesis 2:* RDF-graphs are taller than texts in asynchronous collaborative authoring.
- *Hypothesis 3:* RDF-graphs are more expansive than texts in asynchronous collaborative authoring.
- *Hypothesis 4:* RDF-graphs are more aggregative than texts in asynchronous collaborative authoring.
- *Hypothesis 5:* RDF-graphs are more conducive to collaboration than texts in asynchronous collaborative authoring.

5.4. Statistical Analysis

According to the conversion process we discussed in section 3.2 and the QIs we discussed in section 5.2, we calculate the scores of the QIs for text documents and graph documents and show the result in Table 5-4 and Table 5-5, respectively.

Table 5-4 QI scores of text documents

QIs	Size	Height	Expansion	Aggregation	Connectivity
Task 1	18	5	11	1	4
Task 2	17	4	9	4	9
Task 3	23	4	12	3	8
Task 4	15	7	4	0	1
Task 5	21	7	10	5	5
Task 6	21	3	10	0	4
Task 7	18	4	10	2	4
Task 8	24	9	9	1	1
Task 9	19	4	10	0	4
Task 10	29	13	8	1	4
Average	20.50	6.00	9.30	1.70	4.40

Table 5-5 QI scores of graph documents

QIs	Size	Height	Expansion	Aggregation	Connectivity
Task 1	32	6	17	7	6
Task 2	59	6	37	3	15
Task 3	62	9	23	6	9
Task 4	35	7	22	11	8
Task 5	44	4	29	7	17
Task 6	26	18	13	13	14
Task 7	53	6	25	9	6
Task 8	37	10	11	3	1
Task 9	57	9	27	5	7
Task 10	33	7	17	8	9
Average	43.80	8.20	22.10	7.20	9.20

Before we compare the QI scores, to eliminate the differences in the difficulty among the nine tasks, we need to normalize the data so that the average value of the graph and the text be 1.0 for each the structural quality indicator (size, height, expansion degree, aggregation degree, and connectivity) for each task. In a task, if the text size and the graph size are t and g, respectively, then the normalized size of text and graph are defined as follows.

if t + g = 0, then

normalized size of text = 1

normalized size of graph = 1

else

normalized size of text = 2t / (t + g)

normalized size of graph = 2g / (t + g)

The normalized sizes, heights, widths, expansion degrees, aggregation degrees, and connectivity of texts and graphs for the 10 tasks are shown in Figure 5-2 to Figure 5-6, respectively. The average values of normalized QIs are shown in Figure 5-7.

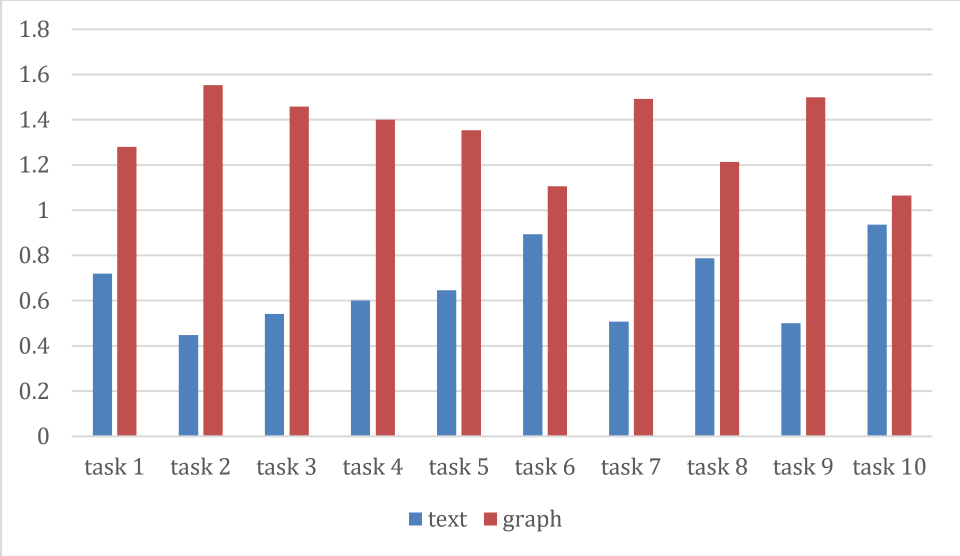

Figure 5-2 Normalized sizes

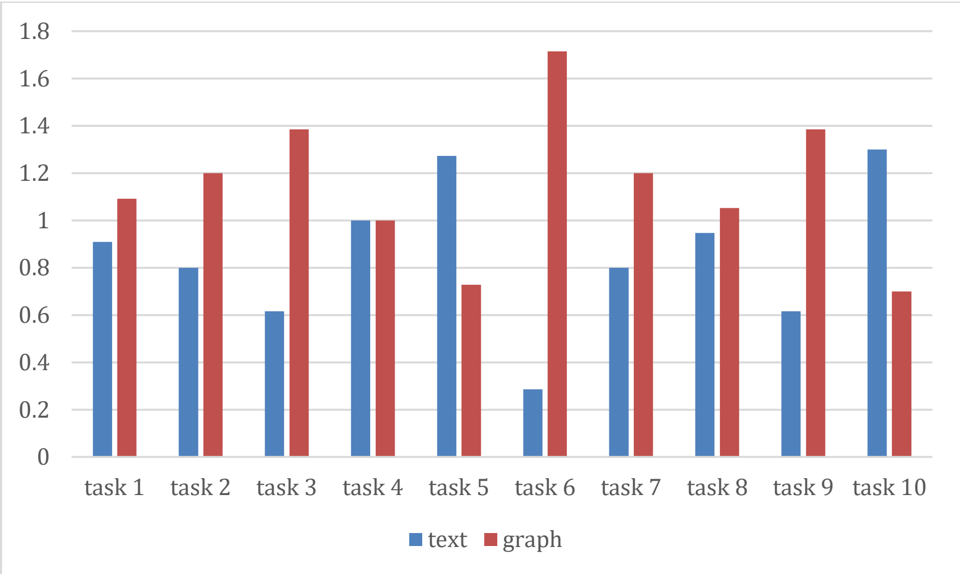

Figure 5-3 Normalized heights

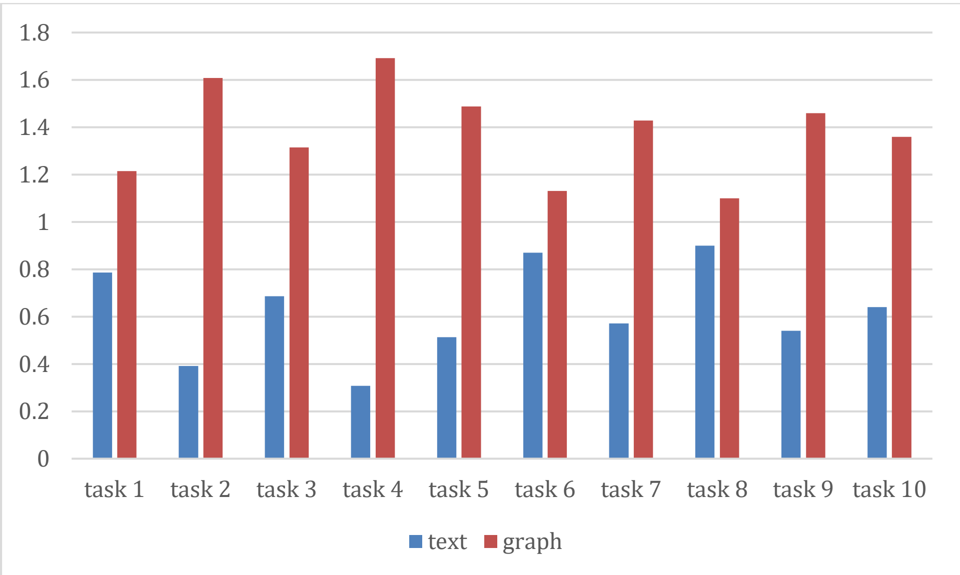

Figure 5-4 Normalized expansion degrees

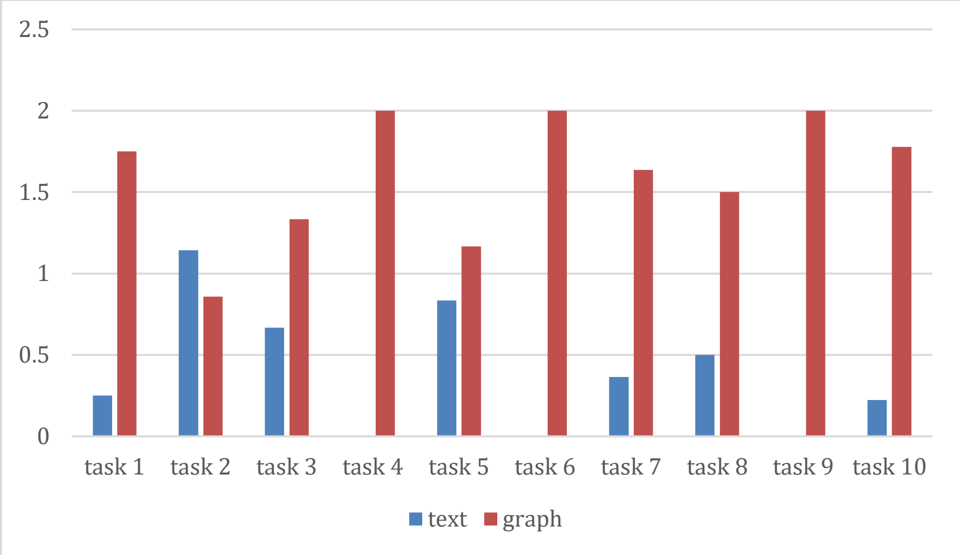

Figure 5-5 Normalized aggregation degrees

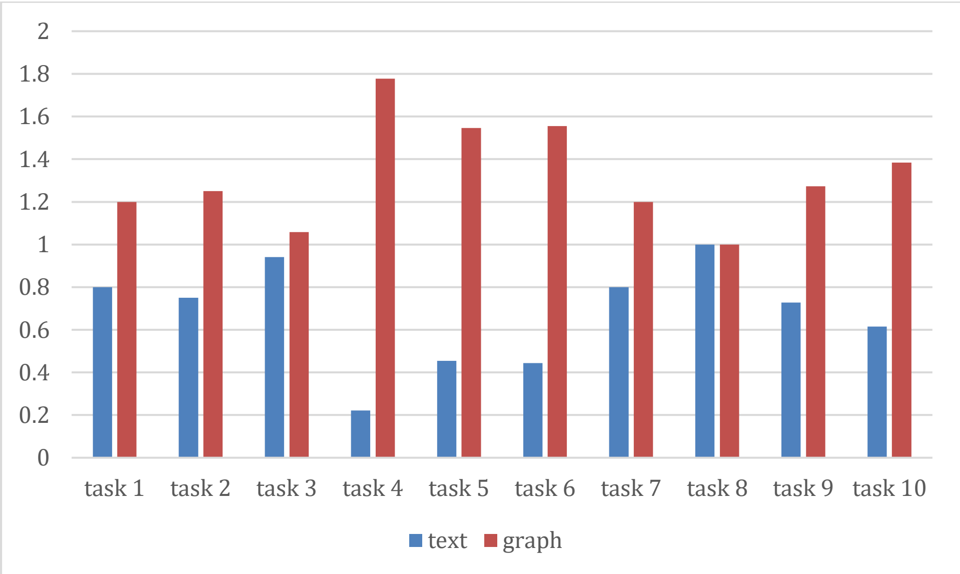

Figure 5-6 Normalized connectivity

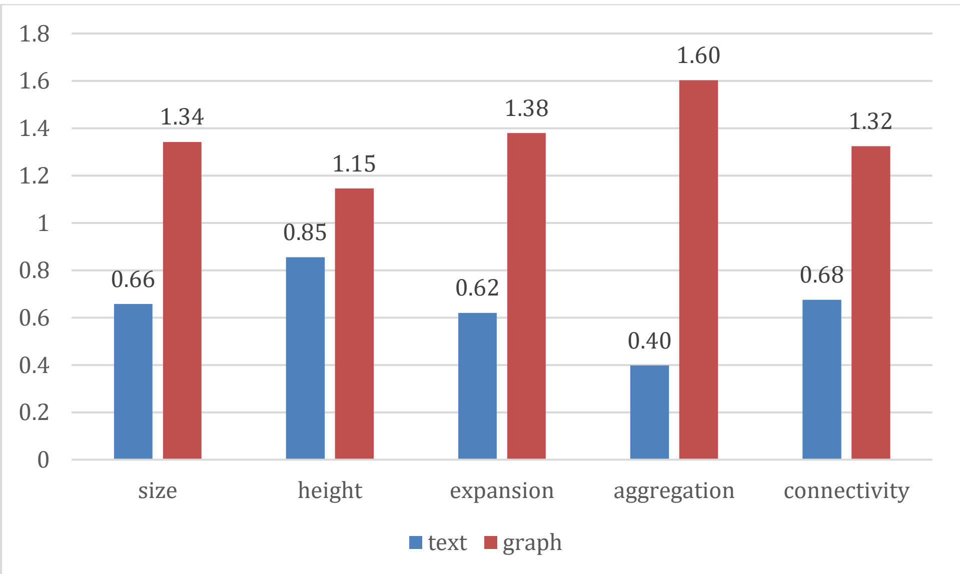

Figure 5-7 Average of normalized QIs

5.5. Hypothesis Tests

Based on the data discussed above, we test the five hypotheses we mentioned in section 5.3. We apply one-tailed paired t-test to these hypotheses. Table 5-6 summarizes the text-graph comparisons with respect to the five structural QIs. Thus hypotheses 1, 3, 4, and 5 survive under significance level α=5%, strongly supporting the original hypothesis, which is to the effect that RDF-graph is more productive and more conducive to collaboration than text in asynchronous collaborative authoring. However, in the meanwhile, hypotheses 2 is not established from our data under significance level α=5%. In this case, there is not enough evidence to show that RDF-graphs are taller than texts in asynchronous collaborative authoring.

Table 5-6 Comparison of text and graph

	Size	**Height**	**Expansion**	**Aggregation**	**Connectivity**
Text Mean	0.66	0.85	0.62	0.40	0.68
Graph Mean	1.34	1.15	1.38	1.60	1.32
P-Value	0.0001	0.0844	0.0001	0.0004	0.0011

5.6. Summary

In the experiment of asynchronous collaborative authoring, we confirmed that RDF-graph is more productive and more conducive to collaboration than text in asynchronous collaborative authoring. Among them, we use size, height, expansion degree, aggregation degree, and connectivity as QIs to quantitatively evaluate document quality, found that size, expansion degree, aggregation degree, and connectivity strongly support the original hypothesis, that is, RDF-graphs are larger, more expansive, more aggregative, and more conducive to collaboration than texts in asynchronous collaborative authoring. For height, the final normalized average score of graphs is still higher than texts, but the p-value in the hypothesis test is not small enough, which makes us unable to accept the hypothesis. Nevertheless, at the same time, text did not show any advantages over graph in height.

5.6.1. In-Depth Document Analysis

Compared with the synchronous collaborative authoring mentioned in Chapter 4, asynchronous collaborative authoring has increased the experiment time. For each document, the composition time has been increased from the original 30 minutes to one hour. This allows participants to have sufficient time to express their ideas. In this experiment, there is not enough evidence to show that RDF-graphs are taller than texts in asynchronous collaborative authoring. This is mainly due to the time-based storytelling in the composition of text. In task 5, the topic is to ask participants to pretend to climb the Himalayas and describe the preparations that need to be made for it. In the composition of text document, participants told a chronological story of pre-preparation and interaction, intensively described the historical development of the fictional story. This makes the height of the document increase significantly. In task 10, the task asked the participant to assume a superpower and describe how to use the power. In the composition of text document, participants described in detail the background story on how to obtain this ability and created a fictional chronological short story about superheroes. In this way, the fictional chronological story makes the text score on the height indicator higher. However, even so, for normalized height, the average score of graph is still higher than that of text, but the p-value is not small enough in the t-test hypothesis test to accept the original hypothesis. Combining size and expansion degree, we can see that under a specific topic, text has advantages for fictional chronological storytelling, but even so, the richness and expansion of the entire document produced are not as good as graph. Unlike Yagishita's experiment (Yagishita, Munemori, & Sudo, 1998), our experiment is still open and does not require participants to analyze specific issues in depth or draw clear conclusions. However, the aggregation shown by the graph document is still higher than that of the text document. The aggregation degree of some tasks in the text document is zero, such as task 4, task 6, and task 9, but in the graph document, the aggregation degree never is zero. Generally speaking, in terms of the richness of document and the breadth of discussion content, graph has a clear advantage over text. Moreover, on the height indicator, graph has no disadvantage to text.

In this experiment, five participants were school students, and the other five participants were staff members from various industries. The average sample composition of experimenters and the rotating experiment schedule minimized the deviation caused by the participants' writing ability and task complexity. For each experiment participant, they will participate in these four sessions on different four days, and the task he/she faces is different every time. For each task, its corresponding 4 sessions are completed by four different participants. Therefore, this is a fair experiment. This sort of experimental design favors neither text documents nor graph documents and can weaken or eliminate the impact of differences in the capabilities of different groups. This backs up the statistical significance of the superiority of graphs to texts in our experiment.

5.6.2. Analysis of Other Indicators

In addition to the five QIs related to the hypotheses, we also studied some other QIs. These QIs cannot be said to have any apparent connection with the quality of the document, but some of the phenomena they reflect are also worthy of our attention. Here, we show the definition and statistical results of these QIs as follows.

- **Average Expansion Degree** is defined based on the definition of downward degree. The downward degree of a node is the number of nodes linked downwardly to this node. In a net structure, the downward degree is equivalent to the number of children of a node. Average expansion degree is defined as:

  if $T_{out} \neq \emptyset$, then

$$\textbf{Average Expansion Degree} = \frac{\sum_{t \in T_{out}} \{t_{out}(N) - 1\}}{|T_{out}|}$$

else **Average expansion Degree** = 0

T_{out} is the set of nodes with downward degree greater than or equal to 2.

$t_{out}(N)$ is the downward degree of a node.

- **Average Aggregation Degree** is defined based on the definition of upward degree. The upward degree of a node is the number of nodes linked upwardly to this node. In a net structure, the upward degree is the number of parents of a node. Average Aggregation degree is defined as:

if $T_{in} \neq \emptyset$, then

$$\textbf{Average Aggregation Degree} = \frac{\sum_{t \in T_{in}} \{t_{in}(N) - 1\}}{|T_{in}|}$$

else **Average Aggregation Degree** = 0

T_{in} is the set of nodes with upward degree greater than or equal to 2.

$t_{in}(N)$ is the upward degree of a node.

- **Node Deletion Rate** is the percentage of nodes that were generated in phase 1 but were deleted in phase 2.

$$\textbf{Node Deletion Rate} = \frac{\left|\{n \mid n \in N_{p1}, n \notin N_{final}\}\right|}{\left|N_{p1}\right|}$$

N_{p1} is the set of nodes generated in phase 1.

N_{final} is the set of nodes in the final composed document.

- **Link Deletion Rate** is the percentage of links that were generated in phase 1 but were deleted in phase 2.

$$\textbf{Link Deletion Rate} = \frac{\left|\{l \mid l \in L_{p1}, l \notin L_{final}\}\right|}{\left|L_{p1}\right|}$$

L_{p1} is the set of links generated in phase 1.

L_{final} is the set of links in the final composed document.

- **Node Modification Rate** is the percentage of nodes that were generated in phase 1 but were modified in phase 2.

$$\textbf{Node Modification Rate} = \frac{|N_{mod}|}{\left|N_{p1}\right|}$$

N_{p1} is the set of nodes generated in phase 1.

N_{mod} is the set of nodes that were generated in phase 1 but were modified in phase 2. The content of a modified node should be roughly the same as the original node, with only minor modifications. Otherwise, it shall be deemed that a new node is created after the original node is deleted.

- **Link Modification Rate** is the percentage of links that were generated in phase 1 but were modified in phase 2.

$$\textbf{Link Modification Rate} = \frac{|L_{mod}|}{|L_{p1}|}$$

L_{p1} is the set of links generated in phase 1.

L_{mod} is the set of links that were generated in phase 1 but were modified in phase 2. The node connected by the modified link should be the same as the node connected by the original link, only the link label or link direction was modified. Otherwise, it should be regarded as a new link created after the original link is deleted.

Table 5-7 and Table 5-8 show the statistical result of the QIs defined above of text document and graph document, respectively.

Table 5-7 QI scores of text documents

QIs	Average Expansion	Average Aggregation	Node Deletion Rate	Link Deletion Rate	Node Modification Rate	Link Modification Rate
Task 1	1.83	1.00	40.0%	75.0%	35.0%	0.0%
Task 2	3.00	2.00	8.3%	8.3%	58.3%	25.0%
Task 3	2.00	3.00	38.5%	38.5%	15.4%	0.0%
Task 4	1.00	0.00	96.0%	100.0%	0.0%	0.0%
Task 5	2.50	2.50	0.0%	0.0%	6.7%	6.7%
Task 6	2.00	0.00	0.0%	0.0%	0.0%	0.0%
Task 7	3.33	2.00	0.0%	0.0%	0.0%	0.0%
Task 8	2.25	1.00	0.0%	0.0%	0.0%	0.0%
Task 9	2.00	0.00	0.0%	0.0%	54.5%	9.1%
Task 10	1.33	1.00	0.0%	0.0%	0.0%	0.0%
Average	2.13	1.25	18.3%	22.2%	17.0%	4.1%

Table 5-8 QI scores of graph documents

QIs	Average Expansion	Average Aggregation	Node Deletion Rate	Link Deletion Rate	Node Modification Rate	Link Modification Rate
Task 1	2.43	2.33	0.0%	0.0%	0.0%	0.0%
Task 2	2.18	1.00	0.0%	12.9%	12.9%	0.0%
Task 3	1.44	1.50	6.8%	9.1%	0.0%	0.0%
Task 4	2.20	1.57	0.0%	5.3%	10.5%	0.0%
Task 5	2.42	1.40	0.0%	15.4%	7.7%	0.0%

Task 6	1.63	1.63	0.0%	10.5%	0.0%	0.0%
Task 7	1.92	1.29	0.0%	0.0%	0.0%	0.0%
Task 8	1.38	1.00	0.0%	0.0%	0.0%	0.0%
Task 9	1.59	1.25	0.0%	0.0%	0.0%	0.0%
Task 10	1.89	1.14	0.0%	0.0%	0.0%	7.1%
Average	1.91	1.41	0.7%	5.3%	3.1%	0.7%

Since there are often zero cases in these data, we are not here to do the statistical test of the average value. However, the deletion rate and modification rate of the text document can explain some problems. The text document showed an extremely high deletion rate and modification rate in some tasks. Taking task 4 as an example, its deletion rate is very high. The node deletion rate and link deletion rate reached 96% and 100%, respectively. The second participant almost completely deleted the draft left by the first participant and wrote a new document. This is indeed the case. The topic of this task requires participants to imagine the life of human beings on some day in 3030. The first participant planned a grand story background in the draft and planned to start with the development of human society to estimate the state of society in 3030. In the draft, the first participant gave a detailed introduction to the development of human society as the background of the entire topic. However, this excessive attention to the background of the story led to a certain degree of disconnection from the topic, so that the second participant could not complete such a grand content of the historical background in a short time, so the draft was almost completely deleted in the final document. We call this phenomenon of paying too much attention to the background of the story as a background trap. Due to the existence of background traps, participants cannot produce richer and broader thinking content when composing text documents.

Since we did not make genre and format restrictions on the draft composed in phase 1, some participants only listed some ideas in the text draft in phase 1 but did not form smooth sentences or paragraphs. This makes the second participant have to reorganize these contents into smooth language and then express them as a whole. We can see this more clearly by referring to the node modification rate and link modification rate in Table 5-7. Take task 2 as an example. The higher modification rate indicates that the second participant has made extensive changes to the draft. This is actually the process of reorganizing the listed ideas into a smooth language. This organizational process not only cannot make the content richer, but on the contrary, due to the requirements of language fluency and the limitation of the author's writing ability, some content initially in the draft will be discarded or lost. A typical example is the text part of task 1, as shown in Table 5-9. This task requires participants to organize a party after the quarantine of the COVID-19 and describe the specific process of the activity. In the original draft, the first participant suggested organizing everyone to play board games, mentioning the work of preparing board game cards, drinks, and snacks. However, the preparation of drinks and snacks was ignored entirely in the final document. On the contrary, in terms of word count, the second participant has greatly expanded the content, and also mentioned the reasons for the party location. However, obviously, this paragraph also includes many sentences that are not highly relevant to the party process. The existence of these sentences only makes the article more fluent but does not help the article's information richness.

Table 5-9 Part of the text document of task 1[9]

Phase 1	**Final work**
Process on the day of the event: 1. Play board games at 2 pm-5 pm Items needed: board game cards; drinks; snacks	Activity process: 1. 2 pm-5 pm board games Item needed: board game cards Today’s 10-player board game will have a wealth of characters and game modes. Everyone has been playing through the cloud using APPs for a long time. Such a real face-to-face game can be described as a war after three months. Everyone will be very emotional by then. In particular, my voice might be overshadowed by the communication of people in the store, so I chose an outdoor cafe. Everyone is playing very well, 3 hours of game time will pass by in a flash, and everyone will feel that they are still not enough.

Figure 5-8 and Figure 5-9 respectively show the changes in the QIs of text document and graph document in two stages. Obviously, for both text document and graph document, QIs have increased in phase 2. The situation shown in Figure 5-10 and Figure 5-11 can better show the strength of text document and graph document in their respective QI growth. We can see that for the text document, although the various QIs have increased in phase 2, the normalized size and normalized expansion degree have decreased. This is because, in terms of size and expansion degree, the growth intensity of the text document in phase 2 is not as good as the graph document, which makes the relative value of the text document drop.

[9] The original document was written in Chinese, which was translated into English by the author of this thesis.

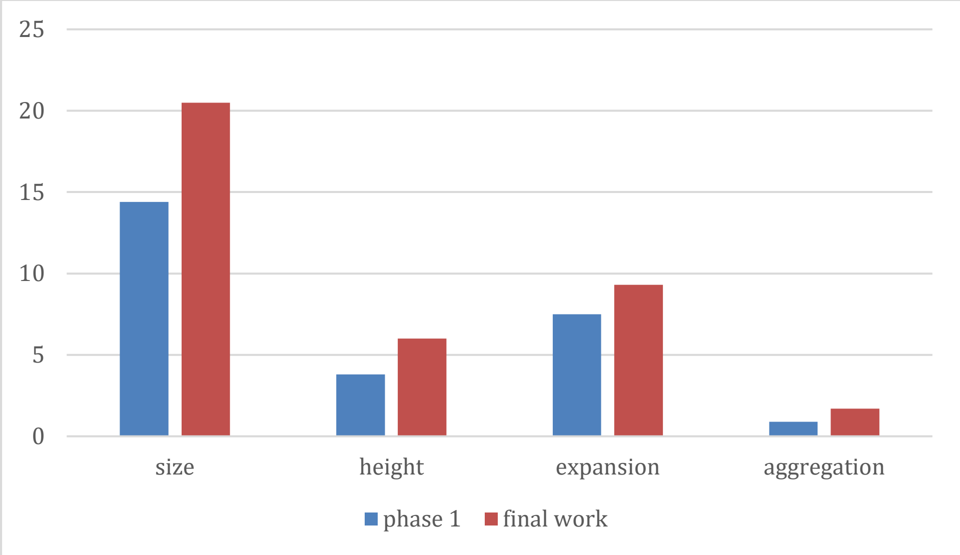

Figure 5-8 Comparison of raw QIs in the two stages of text document composition

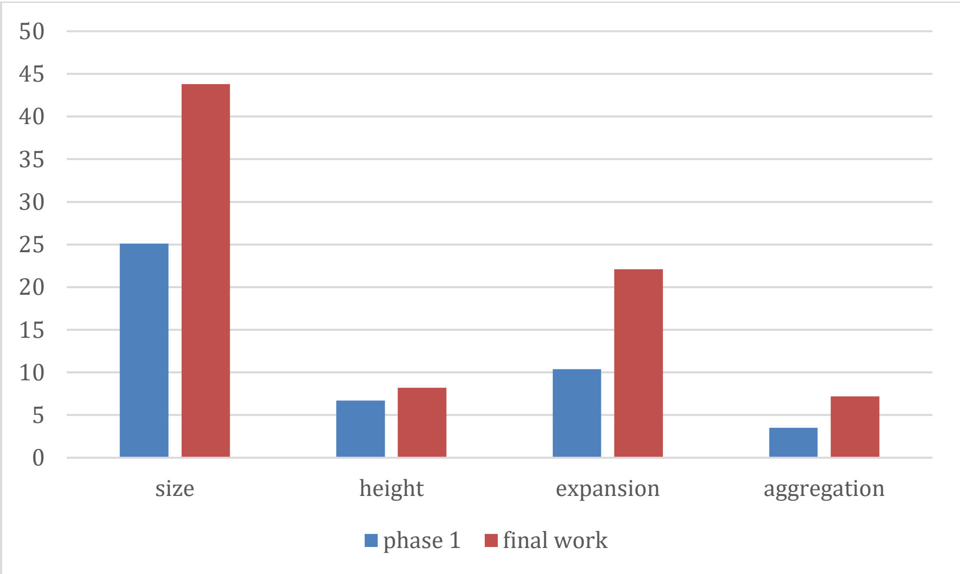

Figure 5-9 Comparison of raw QIs in the two stages of graph document composition

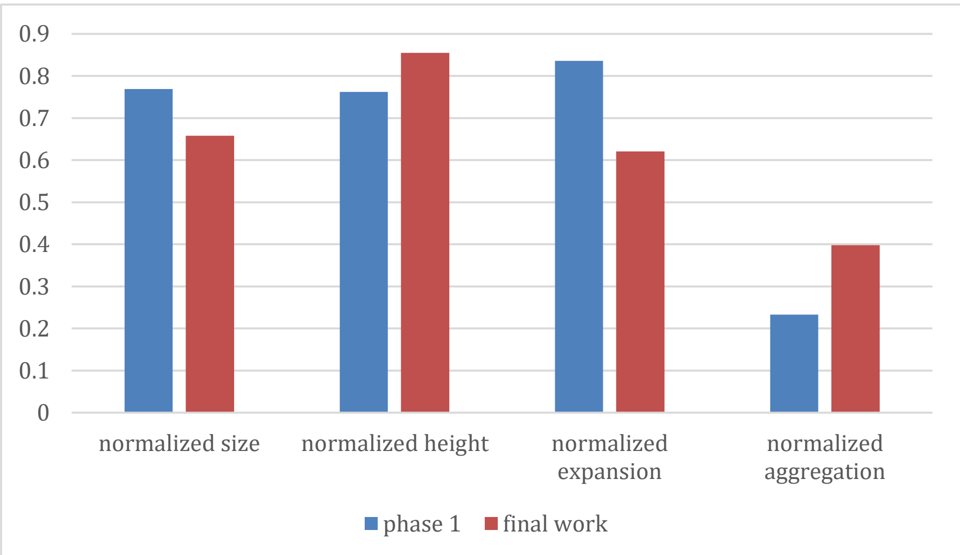

Figure 5-10 Comparison of normalized QIs in the two stages of text document composition

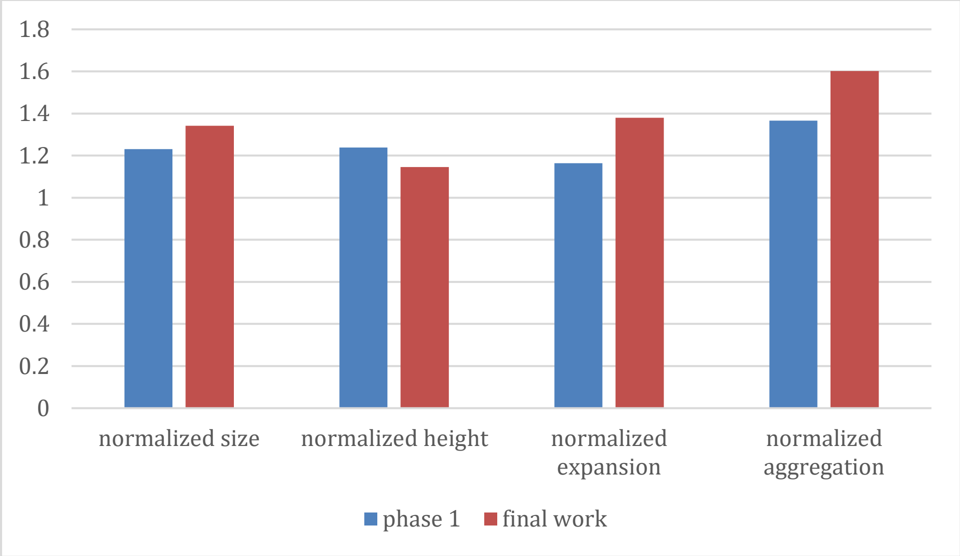

Figure 5-11 Comparison of normalized QIs in the two stages of graph document composition

Due to the existence of zero data, the normalized aggregation degree of text and graph have risen at the same time, but overall, in the final work, graph document also has an advantage in aggregation.

The only normalized QI that increases in the second stage of the text document is height. This is due to the text document's feature of fictional chronological storytelling discussed before. When using text for composition, it is easier for participants to complete a chronological fictional story, this makes the text score on the height indicator higher. However, at the same time, it is precisely because of this feature of text, the content created by two participants is more likely to be detached from each other. When the second participant gets the draft of a chronological fictional story left by the first participant, a typical behavior is to continue writing the story in chronological order, but the two parts of the story are not very connected. It shows that the two people do not have a strong collaboration in the content they create. This can be clearly seen from the comparison of connectivity in Figure 5-7 and Table 5-6. On the contrary, graph is beneficial to the collaboration of participants in content. Higher connectivity indicates that the content composed by the two participants is closely related to each other, rather than two separate parts that are disconnected from each other.

6. Discussion and Conclusion

In this paper, to verify whether RDF-graph, as a form of document and a carrier of content, is more productive and more conducive to collaboration than text in a collaborative authoring environment, we designed two experiments to quantitatively compare the performance of graph and text in collaborative authoring, performed hypothesis tests based on experimental data. After a quantitative comparative analysis, it is concluded that RDF-graph is more productive and more conducive to collaboration than text in collaborative authoring.

In the experiment of synchronous collaborative authoring, we confirmed that RDF-graph is more productive than text in synchronous collaborative authoring. Among them, we use size, height, expansion degree, and aggregation degree as QIs to quantitatively evaluate the quality of the document and find that the size, height, and expansion degree all strongly support the original hypothesis, that is, RDF-graphs are larger, taller, and more expansive than texts in synchronous collaborative authoring. The data of aggregation degree seems to be extreme, and there are many zero cases. The normalized average aggregation degree of texts is higher than that of RDF-graphs, but the p-value is too large, so that we cannot conclude that text has an advantage in aggregation degree than RDF-graph. As a result, we do not have enough evidence to prove that RDF-graphs are more aggregative than texts in synchronous collaborative authoring. Nevertheless, at the same time, texts do not show any advantage over graphs in aggregation degree, either.

In the experiment of asynchronous collaborative authoring, we confirmed that RDF-graph is more productive and more conducive to collaboration than text in asynchronous collaborative authoring. Among them, we use size, height, expansion degree, aggregation degree, and connectivity as QIs to quantitatively evaluate document quality, found that size, expansion degree, aggregation degree, and connectivity strongly support the original hypothesis, that is, RDF-graphs are larger, more expansive, more aggregative, and more conducive to collaboration than texts in asynchronous collaborative authoring. For height, the final normalized average score of graphs is still higher than texts, but the p-value in the hypothesis test is not small enough, which makes us unable to accept the hypothesis. Nevertheless, at the same time, text does not show any advantages over graph in height.

In the experiment of asynchronous collaborative authoring, we learned through analysis that due to the constraints of language fluency and cohesion in text, people pay too much attention to the environmental description and background description, and how to smoothly and naturally lead to core topics. This may be one of the reasons why the text document is not as rich as the graph document.

In the experiment, we also found that the drafts left by using the text document are difficult to be reused by the sequels. The collaborators usually need to reorganize the language in the drafts and modify or discard part of the content to satisfy the fluency of the article. Alternatively, for chronological- storytelling-style drafts, sequels often follow the chronological order and continue to write stories, but this leads to the separation of the two parts. As a result, the two authors did not collaborate fully in content. In graph composition, a participant can modify or augment nodes composed by another participant more easily. A participant can understand the logical location of the content from another participant more quickly and easily. Thereby, the users understand how the content from other participants relates to their own. In text composition, it is difficult to understand how a partner’s texts relate to their own unless the texts are read verbatim.

Graphs are more self-explanatory than texts. For the same task, participants composing texts tend to segment tasks and then spend time thinking about how to organize language and rhetoric so that the whole arguments are consistent. Collaborative graph composition seems to enable the participants to consider and discuss the central theme from different angles, therefore having better control over the entire discourse flow in the graph document. Besides, collaborative graph composition allows participants to focus more on the content itself than on rhetoric and eloquence. In summary, from the perspective of collaboration and content richness, graph as a document format has significantly better effects than text.

With respect to the completeness of conveying ideas and information, the graph as a carrier is as reliable as text. Besides, due to graphs' simplicity and more intelligible internal structure, graphs help people focus on the main topic itself and be better aware of the logical relationships among the nodes. This is particularly important when people collaboratively compose academic articles, business contracts, regulations, and so forth.

Even though Semantic Editor supports essential functions for collaborative graph composition, however, its technical flaws are a significant hindrance to the accuracy of our study. Ideally, we should use the most maturated graph editor and the most maturated text editor for the sake of fair comparison between graphs and texts in terms of collaboration support, but Semantic Editor is far less matured as a graph editor than Google Docs is matured as a text editor. Graphs may prove far better, therefore, in an ideal setting.

The sample size in our two experiments was relatively small, and the participants were mostly school students or well-educated employees. Therefore, future experiments should collect more data from a broader and more diverse population. Although our results support that graphics are superior to text in collaborative document writing, it does not mean that everyone can better utilize the performance of the system on graphics, especially considering the relatively small scale of this experiment. More insight into the superiority of graphs to texts for collaborative works is required. Future research is necessary for a more extensive and more diverse population. In order for participants to better grasp the composition of graphics, a longer adaptation period is required. We also want to compare the intelligibility of graph and text through experiments to clarify whether graph document is more readable than text document.

List of Figures

List of Tables

Bibliography

Adler, A., Nash, J., & Noël, S. (2004). Challenges in collaborative authoring software. *submitted to a special Issue of the IEEE Transactions on Professional Communication "Expanding the boundaries of E-collaboration*.

Arai, N., Todo, N., Arai, T., Bunji, K., Sugawara, S., Inuzuka, M., . . . Ozaki, K. (2017). Reading Skill Test to Diagnose Basic Language Skills in Comparison to Machines. *CogSci.*

Beck, E. (1993). A survey of Experiences of Collaborative Writing In Computer supported Collaborative Writing, ed. M Sharpies. London: Springer-Verlag.

Buzan, T., & Buzan, B. (2006). *The mind map book.* Pearson Education.

d'Angelo, G., Di Iorio, A., & Zacchiroli, S. (2018). Spacetime Characterization of Real-Time Collaborative Editing. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1-19.

Darcovich, N. (2000). *Literacy in the information age: final report of the International Adult Literacy Survey.* Publications de l'OCDE.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Emigh, W., & Herring, S. (2005). *Collaborative authoring on the web: A genre analysis of online encyclopedias.* IEEE.

Gao, L., Gao, D., Xiong, N., & Lee, C. (2018). CoWebDraw: a real-time collaborative graphical editing system supporting multi-clients based on HTML5. *Multimedia Tools and Applications, 77*(4), 5067-5082.

Hasida, K. (2014). Personal life repository as a distributed PDS and its dissemination strategy for healthcare services. *2014 AAAI Spring Symposium Series.*

Hasida, K. (2017). Decentralized, Collaborative, and Diagrammatic Authoring. *The 3rd International Workshop on Argument for Agreement and Assurance (AAA).*

Holland, B., Holland, L., & Davies, J. (2004). An investigation into the concept of mind mapping and the use of mind mapping software to support and improve student academic performance.

ISO/CD 24627-3: Language resource management — Comprehensive Annotation Framework (ComAF) — Part 3: Diagrammatic semantic *authoring (DSA)*. (2019). Retrieved from ISO: https://www.iso.org/standard/74582.html

Lassila, O., Swick, R., Wide, W., & Consortium, W. (1998). Resource Description Framework (RDF) Model and Syntax Specification.

McAleese, R. (1998). The knowledge arena as an extension to the concept map: Reflection in action. *Interactive Learning Environments, 6*(3), 251-272.

OECD Skills Studies. (2016). *The Survey of Adult Skills: Reader's Companion, Second Edition*. Retrieved from OECD Publishing, Paris: http://dx.doi.org/10.1787/9789264258075-en

Slutsky, D. (2014). The effective use of graphs. *Journal of wrist surgery, 3*(02), 67-68.

Thorn, W. (2009). International adult literacy and basic skills surveys in the OECD region.

Willerman, M., & Mac Harg, R. (1991). The concept map as an advance organizer. *Journal of research in science teaching, 28*, 705--711.

Yagishita, K., Munemori, J., & Sudo, M. (1998). A proposal and an application of an evaluation method for sentences of B type KJ method based on contents and structures. *Trans. IPSJ, 39*(7), 2029-2042.