

Structured Annotations to Support Collaborative Writing Workflow

by

Qixing Zheng

A.Sc., Gainesville College, 1999
B.Sc., The University of Georgia, 2001

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

The Faculty of Graduate Studies

(Computer Science)

The University of British Columbia

December 2005

© Qixing Zheng 2005

Abstract

Most co-authoring tools support basic annotations, such as edits and comments anchored at specific locations in the document. However, they do not support higher-level communication about a document such as commenting on the tone of a document, giving more explanation about a group of basic annotations, or having a document-related discussion. Such higher-level communication gets separated from the document, often in the body of email messages. This causes unnecessary overhead in the write-review-edit workflow inherent in co-authoring.

To address the problem, we first established user-centered requirements for annotation support. We conducted a small field investigation of email exchanges including document attachments, among three small groups of academics (3 to 5 people each). We categorized the higher-level communication from the email and developed a set of eleven requirements to support document annotations. We next developed document-embedded structured annotations called “bundles” that incorporate higher-level communication into a unified annotation model meeting the set of requirements. We also designed and implemented a high-fidelity prototype called the “Bundle Editor” that illustrates our structured annotation model.

Finally, we conducted a usability study with 20 participants to evaluate the annotation reviewing stage of co-authoring. The study showed that the annotation bundles in our high-fidelity prototype reduced reviewing time and increased accuracy, compared to a system that supports only edits and comments.

Contents

Abstract	ii
Contents	iii
List of Tables	vi
List of Figures	vii
Acknowledgements	viii
1 Introduction	1
1.1 Research Motivation	2
1.2 Research Contributions	4
1.3 Overview of the Thesis	4
2 Related Work	6
2.1 Collaborative Writing	6
2.2 Annotations	9
2.2.1 Annotation Definition	9
2.2.2 Annotation Model	10
2.3 Existing Co-authoring Systems	11
2.3.1 Research Systems	11
2.3.2 Three Commercial Systems	12
3 Requirements-Gathering through Field Investigation	14
3.1 A Small Field Investigation	14

3.2	Requirements for Structured Annotations	16
4	Structured Annotation Model	20
4.1	Model Elements	20
4.2	Identifying Types of Annotations	22
4.3	Linking Bundle Creation to the Co-authoring Process	23
5	A Prototype of Structured Annotations: The Bundle Editor	25
5.1	Major Interface Components	25
5.2	Functional Description of the Bundle Editor	27
5.2.1	Basic Functionality	27
5.2.2	The Four Primary Ways of Creating Bundles	30
5.2.3	Working with Bundles	31
5.3	Iterative Design and Low- to Medium- Fidelity Prototypes	32
5.4	Implementation	36
5.5	Pilot Testing	36
6	Evaluation of Structured Annotations	38
6.1	Methodology	39
6.1.1	Two Systems	39
6.1.2	Tasks	43
6.1.3	Measures	48
6.1.4	Experimental Design	50
6.1.5	Participants	50
6.1.6	Procedure	50
6.1.7	Hypotheses	51
6.2	Results	52
6.2.1	Testing Hypotheses	52
6.2.2	Other Effects	53
6.2.3	Self-reported Measures	55
6.2.4	Other Feedback	56
6.2.5	Summary of Results	57

6.3	Discussions	58
6.3.1	Bundle Concept Is Intuitive	58
6.3.2	Bundles Reduce Navigation Time	58
6.3.3	Bundles Improve Accuracy	59
6.3.4	Users Group Annotations	59
6.3.5	Scalability of Bundles	60
6.3.6	Cost/Benefit Tradeoff	60
6.3.7	Bundles Provide a More Pleasant User Experience	60
7	Conclusion and Future Work	62
7.1	Conclusions	62
7.2	Future Work	62
	Bibliography	65
A	Usability Study Questionnaire	71
B	Usability Study Documents	77
C	Usability Study Task Sets	86
C.1	Task Instructions in Black Hole Document	86
C.2	Task Instructions Music and Consumer Document	90
C.3	Task Instructions in the Practice Document	93

List of Tables

3.1	Evaluating requirements against current co-authoring systems. . .	18
6.1	Comparison of the Bundle System and the Simple System	43
6.2	Speed measures across five tasks. Df = (1,16). N=20.	53
6.3	Accuracy measures across five tasks. Df = (1,16). N=20.	54

List of Figures

4.1	Annotation Model	22
5.1	The Bundle Editor	26
5.2	Bundle Creation	28
5.3	Navigating Bundles	29
5.4	Comment Display in the Document	33
5.5	Text Display Scheme	34
6.1	Bundle System	41
6.2	Simple System	42
6.3	Task Instruction Illustration	44
6.4	Navigation and Decision Time	49
6.5	Navigation Time Graph	54
6.6	Non-task Related Annotations Graph	55
6.7	Collaborative Reviewing Methods	57
6.8	Collaborative Reviewing Activities	58
B.1	Original docB	78
B.2	Annotated docB	79
B.3	Original docM	81
B.4	Annotated docM	82
B.5	Annotated Practice Document	84

Acknowledgements

First I would like to thank my supervisors, Dr. Joanna McGrenere and Dr. Kellogg Booth, for their guidance and support throughout the entire thesis project process. Joanna was always very helpful in helping me organize my thoughts and ideas about the project and providing detailed and thorough feedback. Kelly always had novel ideas to solve research problems, which inspired me greatly.

I would also like to thank Dr. Barry Po for the many hours he devoted to helping me with analyzing study results as well as providing valuable feedback at different stages of this research. I am also especially grateful to Dr. Steven Wolfman for his insightful comments on the project and for being the second reader of this thesis.

Finally, I would like to thank many friends that helped and supported me in the project. In particular, I thank Tristram Southey for his never-ending support and encouragement.

Financial support for this research was provided by NSERC, the Natural Sciences and Engineering Research Council of Canada, through its Discovery Grants program, and its Research Networks Grants program funding for NECTAR, the Network for Effective Collaboration Technology through Advanced Research. Facilities and research infrastructure were provided by the Canada Foundation for Innovation. The user study was conducted under Certificate B05-0494 that was issued by the University of British Columbia Behavioural Research Ethics Board.

QIXING ZHENG

UNIVERSITY OF BRITISH COLUMBIA

To my parents – Zheng, Dashui and Li, Weiwei.

To my love – Tristram Southey.

Chapter 1

Introduction

Co-authoring academic papers, books, business reports, and even web pages is common practice [39]. Word processors and other tools provide some support for collaborative authoring, but not as effectively as we might desire. Much of the effort in collaborative writing is spent reviewing and editing drafts [32]. Typical workflow involves co-authors annotating drafts and passing them back and forth. Basic annotations are edits (insertions and deletions) and comments on specific parts of the document. However, co-authors also communicate at a higher level about a document, for example, by suggesting changes to the document’s tone, clarifying previous annotations, or responding to other co-authors’ document-related questions. This higher-level communication is not currently well supported by collaborative authoring tools. We use the term “co-authoring” to refer to this entire writing-reviewing-editing cycle.

While the purpose of annotations ranges from strictly personal (fine-grain highlighting to aid memory) to more communal (comments or questions for co-authors) [23], Neuwirth et al. suggest that the most important purpose of shared annotations is to enable fine-grained exchanges among co-authors of a document [30]. We present a novel framework for co-authoring that fully integrates all annotations (basic edits and comments as well as higher-level communication) into a document and we introduce structured annotations that explicitly support workflow management within the co-authoring cycle.

1.1 Research Motivation

Let us first consider the following scenario:

Jen, John, and Mary are collaborating on a conference paper using Microsoft Word 2003 (MS Word). Jen reviews the first draft. She turns on the “Track Changes”¹ and “Comment” features in MS Word, then makes her changes and adds her comments to the document. She then saves the revised document and sends it to John and Mary as an email attachment. In the email, Jen summarizes the changes she made in the document. For instance, she advises that most of her changes are in the first four sections, and that whoever works on the document next needs to spend more time on the Results and Conclusion sections. Jen also points out an important global change she made in the document, replacing the word “Intrusive” with “Obtrusive” in some but not all instances. At the end of the email message, Jen includes some questions for John and Mary to address, such as what the title of the document should be.

After John receives Jen’s email, he notifies everyone that he will be the next person to edit the document. Like Jen, John reviews the annotated document in MS Word using the Track Changes and Comment functions. When he finishes, he saves the document as his revised version and sends it to Jen and Mary as an email attachment. He also describes in his email message a number of the changes he has made and recommends that Jen and Mary review particular comments first.

Finally, it is Mary’s turn to review the document. She too performs the same annotate-and-email steps.

¹With the Track Changes feature turned on, each insertion, deletion, or formatting change made in the document is tracked. When reviewing each tracked change, one can either accept or reject it.

As described in the scenario, co-authors often make basic annotations using their word processors and then send the revised document to co-authors via email as attachments, pathnames in a shared file system, or URLs on the web. This is usually done asynchronously, one author at a time. The annotate-and-email sequence is repeated until the document is completed. Higher-level communication, such as summarizing changes and more general document-related discussions, often takes place outside the document, usually in the bodies of emails used to circulate the drafts. This approach is problematic because it requires co-authors to maintain collaboration artifacts in different places (word processor files and emails) with no formal association between the two. This unnecessarily complicates workflow. Valuable information can be buried and easily forgotten or misplaced [40]. The number of emails can easily grow rapidly when co-authors rely on email for document-related discussions such as deciding on a title. Associating the correct emails with the correct version of the document quickly becomes overwhelming.

Even if the appropriate emails are located, depending on the nature of information communicated, it can be difficult to navigate between email and document content. For example, in the above scenario Jen explains in her email why the word “intrusive” is replaced with “obtrusive” in some but not all cases. Not only does Jen need to make the basic edits, she must provide a separate comment providing her rationale for the edits as well as navigation descriptions so other co-authors can find the edits. The navigation descriptions could be general (“whenever we describe haptic signals”) or very precise (“one is in the second sentence of paragraph 5 on page 3”) to help other co-authors to find changes. More precise descriptions require more effort from the originator, but make it easier for co-authors to locate the changes. However, no matter how precise the descriptions are in the email, co-authors still need to spend effort to find the relevant annotations. Moreover, this workload does not increase linearly as annotations are added, because co-authors often need to read annotations more than once to ensure they have found the right set described in the email. Moreover, there is no easy way to describe semantic groupings of

annotations using current tools, other than plain English text descriptions, so significant communication overhead exists [40] and the co-authoring workflow suffers.

1.2 Research Contributions

The goal of our research is to use structured annotations to uniformly support all annotation activities and facilitate workflow management within a co-authoring cycle. The focus of our research is small, distributed groups of co-authors collaborating asynchronously on editing and reviewing documents embedded with a large number of annotations. The research makes the following contributions:

1. Integrated all annotations fully within the document.
2. Developed a set of eleven requirements for supporting document annotations.
3. Created a comprehensive structured annotation model that satisfies the set of requirements.
4. Designed and implemented a hi-fidelity prototype incorporating the annotation model.
5. Conducted a user study and showed that structured annotations in our high-fidelity prototype reduced reviewing time and increased accuracy, compared to a system that only supports edits and comments.

1.3 Overview of the Thesis

Chapter 2 discusses related work and provides background for the research from three areas: general research in collaborative writing, research on shared annotation, and a survey of existing co-authoring tools. Chapter 3 describes the requirements gathering for document annotations that was performed through a

small field investigation that resulted in a set of eleven requirements for supporting annotations both inside and outside the document. Chapter 4 introduces document-embedded structured annotations called “bundles,” which are part of a unified annotation model that meets the requirements identified in Chapter 3. Chapter 5 describes a high-fidelity prototype for structured annotations called the “Bundle Editor.” In Chapter 6, we discuss the usability study with 20 subjects. The study evaluated the annotation reviewing stage of co-authoring and showed that the annotation bundles in our high-fidelity prototype reduced time and increased accuracy, compared to a system that supports only edits and comments. Chapter 7 summarizes the main results in the thesis and discusses several areas for future research.

Substantial portions of this thesis will be published in the 2006 Proceedings of the SIGCHI conference on Human factors in computing systems [36].

Chapter 2

Related Work

In this chapter, we first review the literature on collaborative writing and describe related research issues. We then provide a more focused discussion of research on annotation and survey existing co-authoring systems.

2.1 Collaborative Writing

Collaborative writing, or group writing, is any writing done in collaboration with one or more other persons [14]. Common stages in collaborative writing are planning, writing, reviewing and editing [27]. Baecker et al. [2] summarize the activities carried out in each stage of collaborative writing as follows.

Planning. Sketch out the main idea of the document, and gather component pieces such as references. Plan how the document will be written and define the roles of each author. Produce an outline of the document.

Writing. Translate the ideas generated and outlined in the previous stage into text, and edit portions of the text while writing.

Reviewing and editing. Make changes and generate comments about the written text to make the document more coherent. Ensure that there are no grammar errors in the document and that all formatting requirements have been satisfied.

There are two modes of collaborative writing. One is called “synchronous writing,” which is a tightly-coupled collaboration among group members. In synchronous writing, collaborators write at the same time. The other mode

of collaborative writing is called “asynchronous writing”, in which co-authors usually by access and modify shared documents at different times, with only one person working on the document at a time. Asynchronous writing is often done in the document editing and reviewing stages [39]. Typically, the asynchronous writing process is sequential, so the document is passed from one co-author to another in turn. In our research, we focus on the editing and reviewing stages of asynchronous writing, which is when most collaboration occurs [32].

Collaborative writing is a broad research area with many interesting and challenging research issues, some of which are outlined below. Most of the these issues fall in the general area of Computer Supported Collaborative Work (CSCW).

Synchronous vs. asynchronous communications: how to effectively support synchronous and asynchronous communication and integrate them in collaboration. For example, Chandler [9] discussed some of the characteristics of asynchronous collaboration and applied them to a case study involving a team composing a mission statement. The focus of Anchored Conversations research [10] is on tightly coupled synchronous collaborative work between distributed group members. Its main contribution is to anchor text chats into documents so that collaborators can have conversations within the existing work context. Jackson and Grossman [18] described an integrated synchronous and asynchronous collaboration system that solved the traditional workgroup barriers of time and space. Rhyne and Wolf [37] argued that the binary distinction of synchronous and asynchronous communication was unnecessary and harmful, and presented a model that included both synchronous and asynchronous collaboration software as submodels.

Version control and consistency maintenance: various document control methods and consistency management algorithms, including problems of merging two versions of a document. For example, Whitehead [41] talked about two application-layer network protocols, WebDAV and DeltaV, which provide capabilities for remote collaborative authoring, metadata management, and version control for the Web. Dourish and Bellotti [12] explored application

semantics for consistency management in a collaboration toolkit. Munson and Dewan [26] described a flexible object merging framework that allows definition of a merge policy based on the particular application being used and the context of the collaborative activity. Neuwirth et al. [29] introduced a software system, *flexible diff*, that finds and reports differences (i.e., “diffs”) between versions of a text.

Group awareness, notification, and roles: awareness and notification of changes to document content, of group progress, and of group members’ current activities. Huang and Mynatt [17] identified many potential benefits of an awareness system that displays information within a small, co-located group in which the members already possess some awareness of one another’s activities. Mendoza-Chapa et al. [24] presented a comparative analysis of workspace and conversational awareness support in collaborative writing systems. Neuwirth et al. [31] explored the interactions among co-authors in collaborative writing. These interactions are influenced by the presence, knowledge, and actions of other co-authors. Brush and Borning [4] introduced a lightweight group awareness technique called “Today” messages, which are short daily status emails to keep group members aware of work progress and reducing the need for face-to-face meetings. Group awareness is directly related to the roles members of a collaborative writing group play. Dourish [13] talked about “different mechanisms, informational, role restrictive, and shared feedback, that current CSCW systems use to support group awareness.” Jaeger and Prakash [19] described the requirements of role-based access control for collaborative systems.

Shared annotations: exploring annotation models, interface design for annotations, and robust annotation positioning in evolving documents. Weng and Gennari [40] presented an activity-oriented annotation model that resembles the rich functionality of physical annotations for an enhanced collaborative writing process. Wojahn et al. [42] compared three types of interface design for annotations: split-screen interface, interlinear interface, and aligned interface. Brush [5] took a first step toward examining, from a user’s perspective, what an annotation system should do when a document changes. A good overall refer-

ence in this area is Brush's Ph.D. dissertation, "Annotating Digital Documents for Asynchronous Collaboration" [3].

Workflow management: as noted in Chapter 1, there are workflow problems in the reviewing and editing stages. Workflow problems also exist in other stages of collaborative writing and in systems that support collaborative work in general. This research area focuses on finding better ways to support collaborative workflow and reduce the workload. Allen [1] first introduced workflow in collaborative work. Woods et al. [43] described different characteristics of information overload. Phelps [35] introduced the use of wizards to better facilitate workflow in collaboration.

These issues are not isolated. They usually overlap. Our research investigates these and related problems in shared annotation and workflow management.

2.2 Annotations

In the editing and reviewing stages of collaborative writing, annotations are added to a document so that co-authors can exchange ideas about the document [30].

2.2.1 Annotation Definition

Annotations have evolved from paper-based to digital. The term "annotation" itself carries many different meanings. Marshall [23] classified paper-based annotations into four categories, depending on their content types (whether they are explicit or implicit to another reader) and locations (whether the annotation's anchor is a point or a range). For example, a scribbled note at the end of a paragraph is an explicit annotation that might apply to the entire paragraph (a range) or to the point between paragraphs, whereas highlighted text and circled words are each examples of implicit range annotations. Similar to Marshall's definition, Brush et al. [6] defined digital annotations as markings made on a

document at a particular place, with each annotation having two components: an anchor and content. Fish et al. defined annotations to be hypertext nodes [15] that are linked to the base document. Ovsianokov et al. [34] proposed the idea of “clumps,” which are comments that can anchor at multiple places in the document. None of these definitions extend beyond simple editing (insert, delete, or replace) and comment annotations.

2.2.2 Annotation Model

Just as there is no standard definition for annotation, there is no standardized annotation model. In particular, there is no agreed-upon convention for structuring annotations. Previous research [8, 20, 22, 28] has identified various attributes of annotations, some of which may not apply to certain types of annotations:

- **Class:** insert, delete, comment, question, reply, etc.
- **Type:** text, graphics, voice, etc.
- **Title:** highlights what the annotation is about
- **Context:** the surrounding text where an annotation is located
- **ID:** unique identification number for an annotation
- **Timestamp:** when an annotation is created
- **Annotator/creator:** identifies the creator of an annotation
- **Anchor:** the concrete location of the portion a document to which an annotation refers
- **Status:** reviewing status of an annotation, e.g., new, read, accepted, or rejected
- **Priority:** an indication of an annotation’s importance

Weng and Gennari developed an eleven-attribute annotation model [40] that uses annotation status to support awareness of in-progress reviewing and revision activity. Their model has three major differences compared to previous annotation data models such as the Resource Description Framework (RDF) [25], which is used to describe a website’s metadata. First, the status information of an annotation supports progress tracking and provides cross-role feedback among reviewers and authors. Second, annotations have extended activity-oriented properties such as rating, category of problems, response deadline, etc. Finally, Weng and Gennari’s model allows for rich contextual information for an annotation, including both a versioned text anchor in the document and contextual threaded discussions for the annotation. Their model is the only one we are aware of that allows annotations to be anchored to the entire document; most models assume that annotations will be anchored at a particular location within the document. Ovsianokov’s [34] model is the only one that allows anchors to have multiple locations.

2.3 Existing Co-authoring Systems

Various tools support collaborative authoring. Brush [3] reviewed some of these annotation systems, focusing on issues such as online discussion in educational settings, notification strategies, and annotation re-positioning techniques in an evolving document. We review systems from the point of view of how well they support collaborative authoring workflow.

2.3.1 Research Systems

Co-authoring systems, or more specifically annotation systems, fit within the broader research area of collaborative writing. The classic collaborative writing systems such as PREP [30], Quilt [15], and SASSE [2] all support basic annotations, but do not support annotation grouping. In contrast, the recent Anchored Conversations system [10] allows text chats to be anchored into documents so

that co-authors can have conversations within their work context. Although this is a real-time conversation tool rather than a shared annotation tool, it is an attempt to integrate higher-level communication within the document.

2.3.2 Three Commercial Systems

Noel and Robert studied 42 users in May 2001 [32] and found that most individuals used word processors and e-mail as their main co-authoring tools. Eighty-three percent of the subjects used Microsoft Word 2003 (MS Word). MS Word **integrates** edit and comment annotations into the document and assigns annotation attributes automatically. Annotations can be filtered by author or by type (formatting changes, comments, insertions, or deletions). All annotations are listed in a reviewing pane below the document pane in the main window. Annotations are ordered by their positions in the document. MS Word **incorporates** edits into the document once they are accepted by one co-author, meaning other co-authors might not know of these edits once the document is saved. Word has a Web Discussion function for collaboration, but Cadiz et al. note that it is limited in terms of where annotations can be anchored [7].

In contrast to Word, annotations in Adobe Acrobat Professional 7.0 (Acrobat) do not alter the original document because they are not incorporated into the document. This must be done manually after reading the annotations. Status indicators and more sophisticated filtering by annotation type, reviewer, or status are provided. The reviewing pane in Acrobat uses a threaded display, not simple document order, so replies to an annotation are listed indented and below the original annotation.

Recently, numerous web-based collaborative authoring tools have been developed [3]. XMetal Reviewer, a new product by Blast Radius Inc. [44], is such a system. Designed for reviewing XML documents, it combines many of the advantages of MS Word and Acrobat. Basic annotations are integrated within the document, and global comments appear at the top of the document. Insertions and deletions can be incorporated into the document rather than

kept as annotations, but this can always be reversed. This makes annotations persistent, unlike in MS Word where accepted changes lose their identities as annotations once they are accepted. XMetal Reviewer facilitates discussion by letting co-authors reply to each other's annotations in real-time and in context to reduce miscommunication. Annotations can be filtered by type, author, or status. XMetal is server-based to support collaboration among a large group of people, which could be a drawback for small groups that want a lightweight solution.

In all three systems, annotations can only be grouped using system-defined filters such as filter-by-author or filter-by-status. Because comments about a specific aspect of a document may be scattered throughout the document, it would be useful to be able to gather them together. In a similar vein, there is only a partial record of the co-authors' annotating processes. Some systems keep track of editing sessions but do not otherwise capture ordering or relationships between individual annotations. This limitation was identified by Weng and Gennari [40], who noted that “[a]nnotations should be activity oriented.”

Chapter 3

Requirements-Gathering through Field Investigation

In this chapter, we describe the requirements gathering phase of our research on document annotation. A small field investigation was conducted to better understand the nature of annotation activities during co-authoring and to validate our own experiences with co-authoring. Eleven design requirements for supporting document annotations were developed based on our literature review and field investigation. We evaluated three current co-authoring systems against these requirements.

3.1 A Small Field Investigation

We analyzed the email exchanges and document attachments of three small groups of academics (3 to 5 people in each group). We examined both the email content and the annotations in the attachments to find the relationships between the two. Each group had co-authored a conference paper, approximately 8 pages in length. They had all finished the conference paper by the time we sent out an email request for collecting collaborative writing data. After consulting with their group members, one author from each writing group voluntarily forwarded all relevant email exchanges including document attachments to the author of this thesis.¹ We analyzed a total of 158 email exchanges across the three groups

¹The request was made after the groups had completed their co-authoring activities, so some messages may not have been captured because not all messages were sent to every co-author, and even those that were may not have been saved by the co-author who forwarded

(52 emails per group on average). While many of the emails included document attachments (Microsoft Word or LaTeX files), our analysis focused on the text content of the email and its relationship to the document. Below we categorize the most frequently occurring content, and provide the percentage of the 158 emails to which each category and sub-category apply. Note that these are not exclusive categories. Most emails fall into more than one category, so the percentages do not sum to 100.

To-do item(s) describe what remains to be done, or should be done next (89%). The ordering of the items implicitly prioritizes the work, and sometimes co-authors give explicit direction on priorities. These often include collaborators' available times to work on the paper.

Summaries of edits that a co-author has made to the document (92%) often appear together with to-do lists. Co-authors often summarize edits about issues that arise at multiple places in the document (78%), such as global word replacements or spelling changes throughout a document.

Discussions about the document often include parts of the text copied into an email to provide context (64%). These include two subcategories: **questions** are sometimes directed at a particular co-author (53%); **general comments** (41%) pertain to the entire document (comments on the tone of the document or suggestions about document structure).

Comments-on-comments are comments about one or more previous comments. These most often concern comments that have not yet been addressed (31%) or advice to co-authors on how to process the referred-to comments (34%).

Information expressed as text embedded in email constitutes what we referred to at the outset of this thesis as "higher-level communication." Co-authors devote a lot of effort to describing how annotations relate to each other because text is inefficient for expressing annotation location, type, or context, especially when an issue arises at multiple places in the document. Currently, the messages. We believe we obtained most of the relevant messages. The difficulty inherent in this collection process underscores one of our assumptions, which is the unreliability of email as an archival record of collaborative annotation activity.

co-authors must describe associated annotations by writing comments (either internal to the document or externally in email). There is no way to directly annotate multiple annotations using electronic tools. Recognizing this limitation, we developed a list of requirements to build an annotation model that would unify all document-related communication by adding structure to annotations.

3.2 Requirements for Structured Annotations

We derived eleven design requirements for annotation systems that reflect co-authoring workflow. The first seven requirements are based on our literature review of annotation models and on current annotation systems. The last four requirements address communication that currently occurs outside the document, which were identified in our field investigation.

- R1.** Support basic annotations such as edits and comments with specific anchors. This allows co-authors to exchange fine-grained information within the document context. All the current annotations systems we examined support this requirement.
- R2.** Provide an easy way to incorporate changes specified in annotations into the document. This saves co-authors the effort required to manually incorporate changes after reading the annotations.
- R3.** Preserve the original annotations in case co-authors want to refer back to them later. This avoids the loss of annotation history that happens in many systems when the document is saved after changes are accepted.
- R4.** Support both a separate annotation list view as well as the ability to view annotations integrated within a document. Most of current annotation systems support dual views.
- R5.** Monitor reviewing annotation status to help co-authors keep track of the reviewing process. This allows co-authors to quickly identify an annota-

tion's reviewing progress and apply proper reviewing effort to it accordingly.

- R6.** Support document-related discussion with annotation reply functions and threaded display of annotations. This encourages co-authors to reply to each other's comments in the document and also helps them see the relationships between the reply annotation and earlier annotations.
- R7.** Support flexible and uniform filtering. This allows co-authors to review more focused and smaller sets of annotations by treating annotations as objects and annotation attributes as fields so uniform filtering according to one or more attributes can be applied to any set of annotations to retrieve a smaller set of annotations.
- R8.** Allow annotations to be directed to specific co-authors. We found from the field investigation that co-authors often direct document-related comments, especially questions, to specific co-authors in email messages.
- R9.** Support general comments that anchor to the entire document, or to single or disconnected sets of points or ranges within a document. Co-authors often include general comments in emails because there is no designated locations in the document for general comments.
- R10.** Allow users to prioritize annotations. Quite often, in an email message, co-authors advise other co-authors which annotations should be reviewed first or given priority.
- R11.** Support annotation of groups of annotations. The most frequently occurring in the email content we collected were summaries of edits, to-do lists and comments-on-comments, all of which are examples of annotations of groups of annotations.

We evaluated the three systems discussed previously in the Related Work chapter (Microsoft Word 2003, Adobe Acrobat Professional 7.0, and XMetal

	MS Word	Acrobat	XMetal
R1: basic anchors	Yes	Yes	Yes
R2: incorporated	Limited	Limited	Yes
R3: reversible edits	Limited	No	Yes
R4: dual views	Yes	Yes	Yes
R5: status	Limited	Yes	Yes
R6: discussions	Yes	Yes	Yes
R7: filtering	OR only	OR only	OR only
R8: specify receiver(s)	No	No	Yes
R9: general comments	No	No	Yes
R10: prioritization	No	Limited	No
R11: grouping	No	No	No

Table 3.1: Evaluating requirements against current co-authoring systems.

Reviewer) against these requirements. The results are summarized in Table 3.1, which suggests that current tools fail to support some of the requirements.

Of the three systems, XMetal best meets the requirements. However, it and the other systems fail to support two important requirements, which relate to the workflow overhead described in Chapter 1.

None of the systems supports annotating groups of annotations (**R11**). Email categories such as summaries of edits and comments on comments identified in the field investigation are instances of grouped annotations. For example, summaries of edits are basically comments on a group of edits in the document. Using current co-authoring systems, there is no way to annotate the edits directly, so co-authors have to rely on additional media to exchange this information.

The second requirement that all three systems fail to fully support is the need for flexible filtering (**R7**). All three systems only support OR filtering, which means co-authors can only filter document-embedded annotations on one attribute (e.g., author or status) at a time. Flexible filtering should at least allow

co-authors to do AND filtering, such as finding Jen's (creator attribute) unread (status attribute) comments (class attribute), in just one filter operation.

Chapter 4

Structured Annotation Model

In this chapter, we present a comprehensive annotation model that introduces both mandatory and optional attributes of an annotation. We then discuss existing and new types of annotations applying the model. Finally, we link structured annotations with workflow. This forms the theoretical framework underlying the high-fidelity prototype described in Chapter 5.

4.1 Model Elements

Using the requirements listed in section 3.1 as a guide, we constructed a comprehensive model of annotations that encompasses the behaviors we observed in the field investigation. In the model, every annotation has a set of attributes. Depending on the purpose of the annotation, some of the attributes can be empty.

Mandatory attributes include the **creator** of the annotation, a **timestamp**, **reviewing status** (unread/read and accepted/rejected), and an **anchor/reference** (the annotation's location and range relative to the document content or related to other annotations). Multiple non-contiguous ranges are permitted as the anchor for a single annotation. As a special case, the anchor can be the entire document. An annotation can also refer to one or more previous annotations, which is indicated the option substructure attribute discussed later in this section. All the mandatory attributes have default values. For example, the default

value for creator and timestamp are the current machine name and current time. A newly created annotation has a default status of “unread,” and will become “read” when the annotation is actively selected by a co-author. The default value for the anchor is calculated depending on where the annotation is placed. It can be a single point in the document, a range within the document, or a set of points and ranges.

Optional attributes include the **name** of the annotation (a short text string), a list of **recipients** (those co-authors who can view the annotation¹), a free-form text **note**, **modification** (insertion, deletion, or replacement of text), a **priority**, and **substructure** (a list of other annotations to which the annotation refers, in effect providing additional “anchors” to these annotations²). Each annotation must have at least one of the name, comment, modification, or substructure attributes in addition to having the four mandatory attributes. The name, note, modification, and substructure attributes are null (not present) by default. The default value for recipients is the list of authors.³

Three key annotation attributes in our model are anchor, note, and substructure. These make integrating high-level communications into documents possible, hopefully eliminating the need to use auxiliary channels such as email.

We classify annotations into two categories: **single annotations** that have no substructure, and **bundled annotations** that have substructure. The latter are called “bundles.” The most distinctive feature of our annotation model is the addition of structure to annotation. The anchor for an annotation captures how it relates to the document and the substructure captures how it relates to other annotations in the document.

¹We leave for future work an exploration of the mechanisms for specifying which co-authors have access to annotations and how this might affect workflow.

²These could include links to “permanent” external objects using URLs, or to attachments associated with the document itself, but the current model employs only simple links to previous annotations in the document itself.

³Again, future work could extend this attribute from simple recipient lists to more generalized lists of recipients that include primary recipients, secondary recipients, and anonymous recipients, mimicking the **To:**, **Cc:** and **Bcc:** fields in email, and permission or capabilities to reply, modify, or delete the annotation.

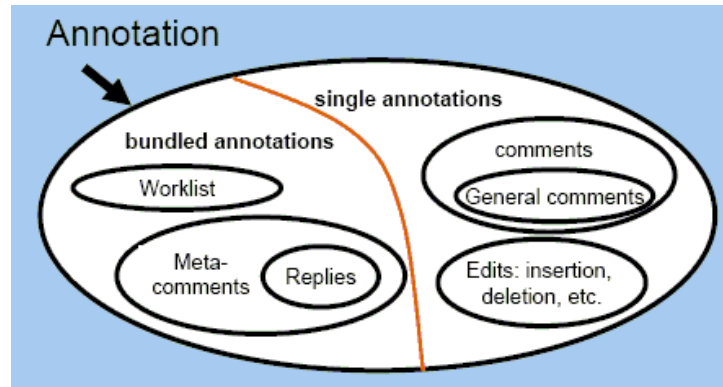


Figure 4.1: A Venn diagram illustrating how different types of annotations fit into the annotation model

We can annotate a group of annotations by creating a bundle that has a set of previous annotations as its substructure. A bundle refers to each of the annotation in its substructure and may have a note attached with it that further defines the the relationship between the annotation and its substructure.

Our definition distinguishes bundles as those annotations that have non-null substructure, but we will sometimes use the term “bundle” to refer generically to any type of annotation in our in our model. This will be true when we introduce the Bundle Editor, which handles both single annotations and bundled annotations.

4.2 Identifying Types of Annotations

Within the annotation model there are some special cases that correspond to common annotation types in traditional systems. A simple **edit** has an anchor to a range of text and the modification attribute. A **comment** has only the note attribute with an anchor into specific document content, and a **general comment** is a comment whose anchor is the entire document. These are all single annotations (Figure 4.1).

A number of interesting new types of annotation also arise from our model. A **meta-comment** is a comment that has substructure, indicating the list of annotations to which it refers, and these in turn may have anchors into the document. Meta-comments can have their own document anchors (in which case they are not “pure” meta-comments). The nesting of substructure can have as many levels as desired, leading to the notion of inherited anchors by which a meta-comment (or any annotation) is recursively associated with the anchors of its substructures.⁴ A reply is a special meta-comment that refers to just a single previous annotation.

Another special type of bundle is a **worklist**. An example would be a bundle having the name “Check spelling” with comment text that says “I am not sure how we spell some of the people’s names in our report. Please make sure I have them right.” The recipient list would indicate who is to do the spell-check, and the anchor would indicate all of the places in the document where the names in question appear.

4.3 Linking Bundle Creation to the Co-authoring Process

The spell-checking bundle just described could be created manually, but we envision it being created automatically as a side effect of running a document processor’s spell-checking command. Realizing that the misspelled words are all names of people, a user could indicate that the selected words form a new bundle by clicking on a bundle button in the spell checker’s dialogue box that creates a new bundle whose substructure is the set of edit made by the spelling checker. The bundle would recursively have a multi-location anchor. The user could manually add name and comment attributes by typing text into the appropriate fields in the spell-check dialogue box. Recipients would be selected from a list of

⁴Annotations can only refer to previous annotations, so we do not have to worry about infinite recursion because the substructure relationship is acyclic.

co-authors. In a similar manner a document processor’s “find” command could produce a bundle of all instances of a pattern, and the “replace” command could produce a bundle of all of its replacement edits.

Another candidate for automatic bundle creation as a side effect is the “Track Changes” feature in MS Word. A user should be able to turn on tracking that automatically bundles all new edits (and comments too, if desired) so that at the end of a session there is a ready-made bundle that can be turned into a worklist. The user could then review just the changes from the current session, or highlight the changes for another co-author to review.

The full power of structured annotations lies in the interplay between normal workflow (editing, commenting, and reviewing) and the ability to capture artifacts from that workflow and use them to manage future workflow. In the annotation model proposed by Weng and Gennari [40] users can assign only one pre-defined category, such as “question” or “reply,” to each annotation. Our model allows users to define their own categories by bundling relevant annotations into the substructure of a new annotation whose name attribute identifies the category. Moreover, any annotation could be assigned to multiple categories because the bundling substructure has no restrictions other than the requirement that it be acyclic.

Adding optional user-defined attributes may still be necessary and would be an easy extension to our model. They would be similar to the user-defined attributes available in qualitative and video analysis systems such as Noldus Observer [33], or the common subclassing (derivation) operation in object-oriented languages where new types extend old types by adding additional attributes (with user-defined default values) and behavior.

Chapter 5

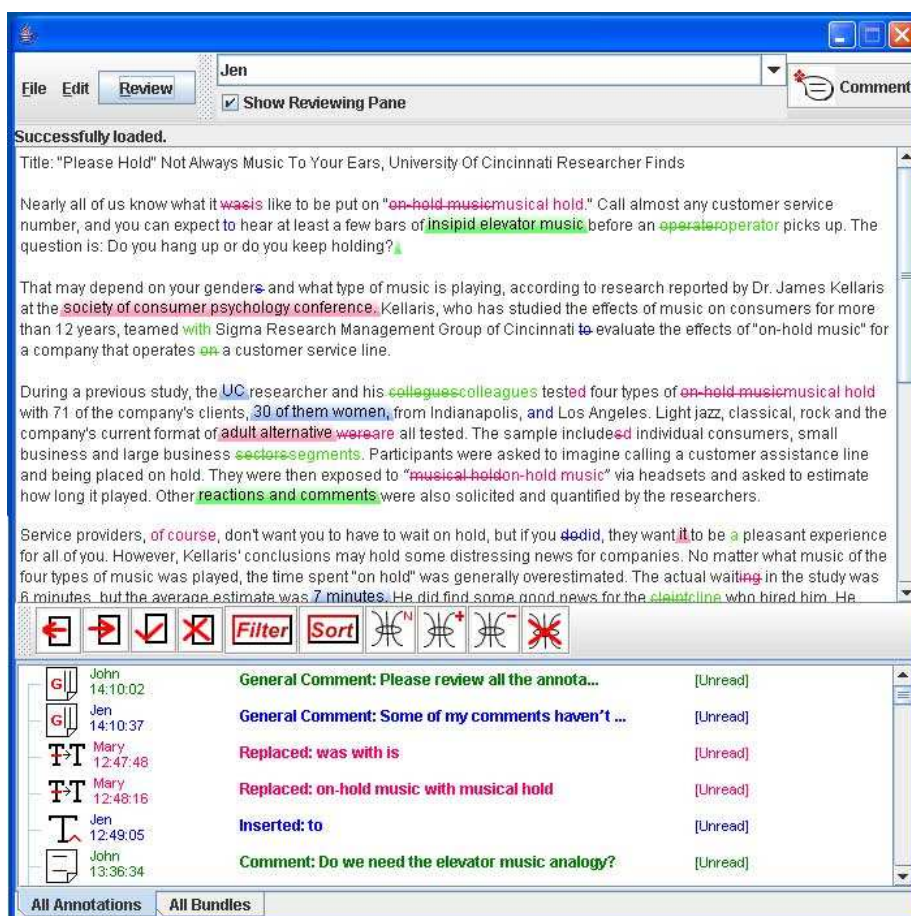
A Prototype of Structured Annotations: The Bundle Editor

In this chapter, we present the Bundle Editor, which is a high-fidelity prototype implementing the structured annotation model described in the previous chapter. We first focus on the major interface components in the Bundle Editor and then describe its functionality. Finally, we step back to briefly describe the iterative design and prototyping, implementation and pilot testing for the system to motivate the design choices that were made.

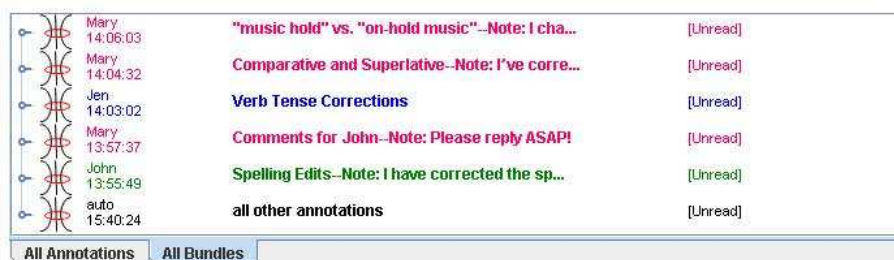
5.1 Major Interface Components

Based on our annotation model, we implemented a high-fidelity prototype called the “Bundle Editor,” which has a number of functions designed to support structured annotations (Figure 5.1). The main component of the system is a two-pane window with an upper **document pane** and a lower **reviewing pane** (similar to MS Word, Acrobat, and XMetal). The document pane displays the annotated document. The reviewing pane is a multi-tabbed pane where each tab pane shows different annotation information. There are two permanent, default tabs at the bottom of the reviewing pane.

The first tab is “All Annotations,” which contains all single annotations (inserts, deletes, replacements, and comments). General comments (i.e., com-



(a)



(b)

Figure 5.1: (a) The Bundle Editor with document and reviewing panes. Annotations are embedded in the document pane and are color coded according to author. The reviewing pane is a multi-tabbed pane where each tab pane shows different annotation information. The first tab in the reviewing pane is “All Annotations”, which contains inserts, deletes, replacements, and comments in the document order. General comments are always placed at the top of the list. (b) The second tab in the reviewing pane, called “All Bundles”, lists all previously created bundles. The last bundle listed in the tab is named “all other annotations”, and is automatically maintained by the system.

ments that pertain to the entire document) appear at the top of the list, and the rest of the annotations appear in the order in which their anchors occur in the document. The second permanent tab is “All Bundles.” It lists all bundles, i.e., the annotations that have substructure. The last bundle listed in this tab is named “all other annotations.” It is maintained automatically by the system and contains all the single annotations that do not belong to any bundle.

5.2 Functional Description of the Bundle Editor

We provide mechanisms for grouping annotations into bundles, annotating previous annotations, filtering to select annotations, and sorting annotations. We describe each of these functionalities, with links to the specific annotation requirements identified in Section 3.1 shown in parentheses. We then present different ways of creating a bundle and describe how to interact with bundles using the Bundle Editor.

5.2.1 Basic Functionality

The Bundle Editor has all of the basic functionality that a typical document editor has, such as insert, delete, and comment (**R1**: basic anchors). It also has specific functions to create a bundle (**R11**: grouping), as shown in Figure 5.2.

Bundles are stored with the document and are linked to various places in the document or to other annotations. Co-authors can add and remove annotations to and from bundles. Any annotation can be in more than one bundle and bundles can be in other bundles. For example, in Figure 5.3, there are two bundles within the “Verb Tense Corrections” bundle. One is called “Jen’s Verb Tense Corrections,” and another is called “Mary’s Verb Tense Corrections.”

Co-authors can annotate a group of annotations by including a note in the appropriate bundle and directing the bundle to a particular set of co-authors (**R8**: specify receiver(s)). For instance, in Figure 5.2, Mary creates the bundle

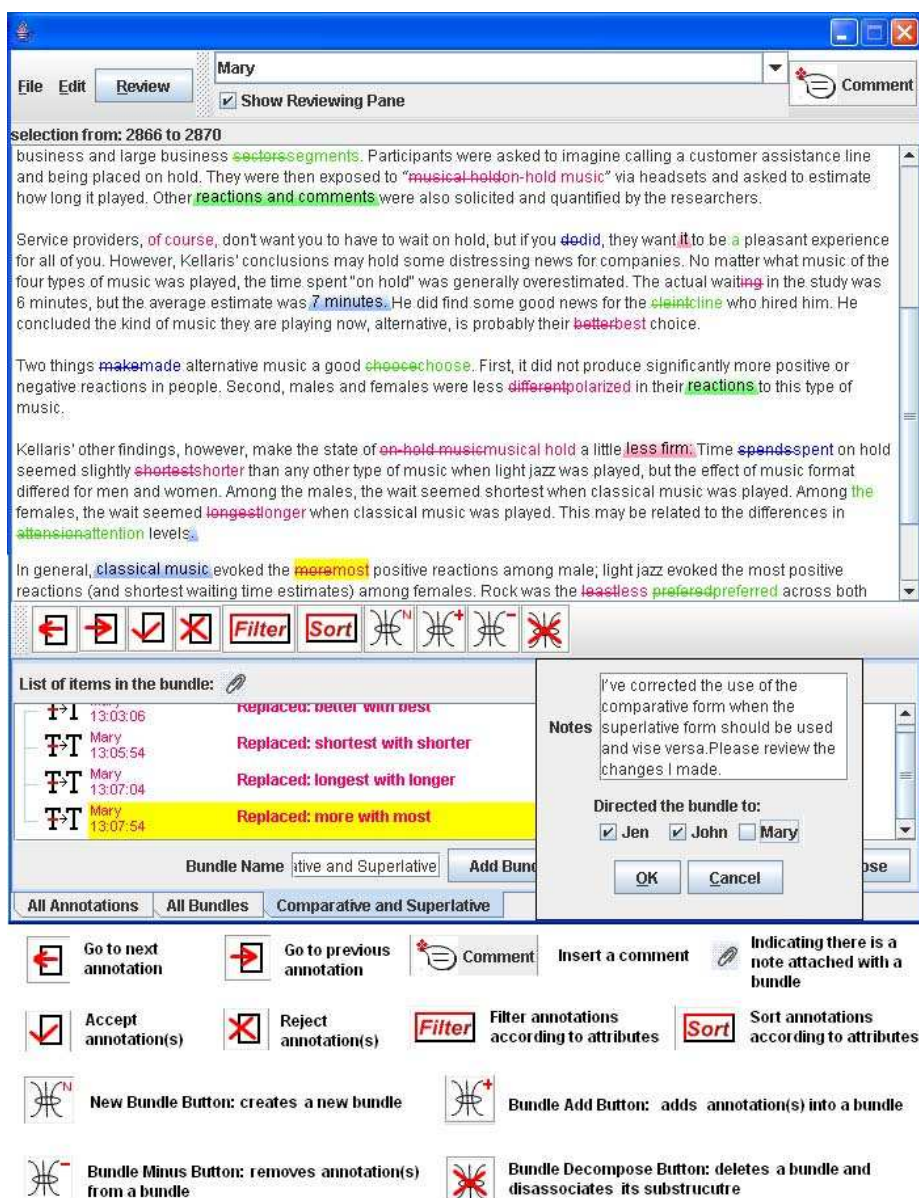


Figure 5.2: A new bundle called “Comparative and Superlative” is being created. Co-author, in this case “Mary,” has added relevant annotations into the bundle using the “Bundle Add” button. She also writes a note with the bundle and specifies the receivers to be Jen and John. A legend for different interface components is also included in the figure. The comment icon allows users to insert a comment with a specified anchor or a general comment. Both “Accept” and “Reject” can be used to accept or reject one or more selected annotations.

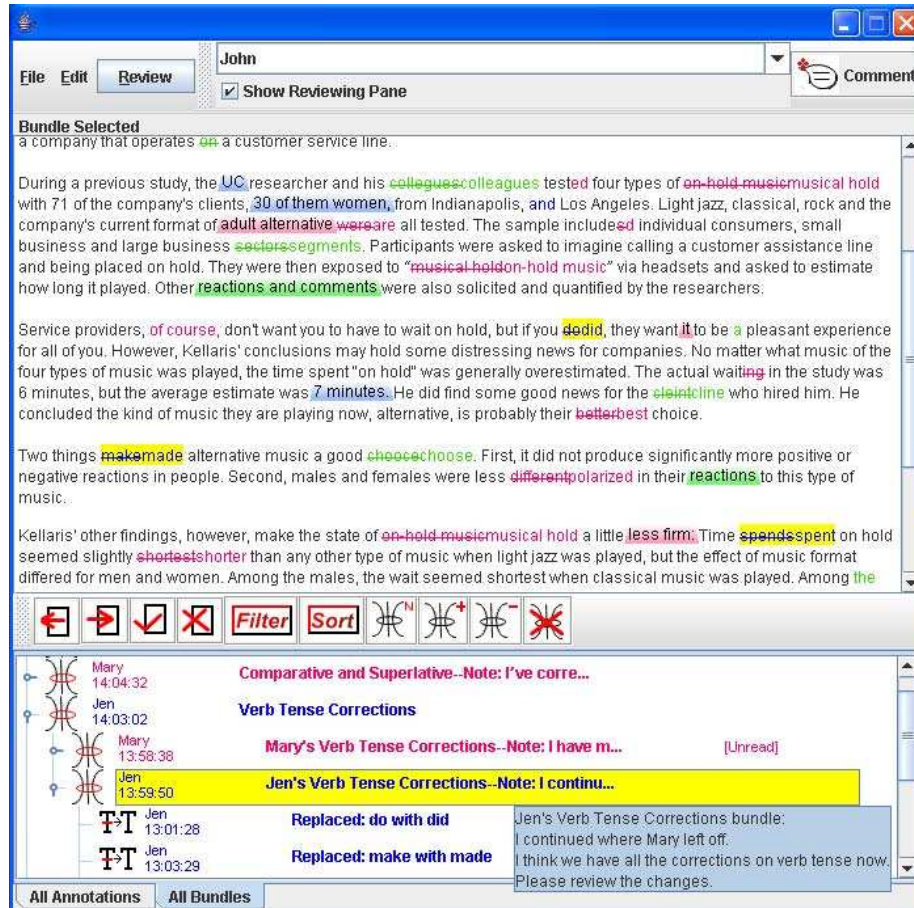


Figure 5.3: The user highlights the “Jen’s Verb Tense Corrections” bundle in the reviewing pane, which highlights all of its sub-annotations’ recursive anchors in the document pane. The “Verb Tense Corrections” bundle contains two sub-bundles. One was created by Jen and one was created by Mary.

“Comparative and Superlative” and attaches a note explaining what changes she made on comparative and superlative forms. The bundle is directed to Jen and John.

The Bundle Editor has functions for replying to annotations, which encourages discussion (**R6**: discussions), and it allows co-authors to make general comments to each other without leaving the document (**R9**: general comments).

The filtering function in the Bundle Editor is more flexible than the filtering functions in existing tools (**R7**: filtering). It allows co-authors to select annotations based on multiple attributes such as “all of Jennifer’s and Brad’s comments.” The filter result is a new bundle that is a subset of the annotations in the current tab in the reviewing pane to which the filter was applied. The result can either replace the bundle in the reviewing pane or appear in a new tab in the reviewing pane.¹ The sort function in the system allows co-authors to sort annotations within a tab of the reviewing pane according to time, location in the document, author, recipient, or any other user-defined or built-in attribute.

Reviewing progress can be tracked by assigning a status to individual annotations (**R5**: status). Depending on the co-authors’ reviewing activities, the system assigns annotation status automatically, so an “unread” annotation becomes “read” by a co-author when it has been selected. Co-authors can always over-ride a system-assigned status by right clicking on the annotation either in the document pane or reviewing pane to set the status. When a bundle’s status is set, users have choice to decide whether the status will propagate to all the annotations in its substructure.

5.2.2 The Four Primary Ways of Creating Bundles

Bundles can be created manually while annotating the document. For example, if Jennifer finds recurring problems in a document, she can create a bundle by explicitly selecting all relevant annotations so she can deal with them as all at

¹For permanent tabs filters always produce their results in a new tab.

once.

Temporary or working bundles are created by filtering and other operations. They can be saved as permanent bundles with a single click. For example, Jennifer might want to look at the comments made by Brad. She can create a working bundle by filtering on “Brad” and “comment” and save the edits as a bundle for later reviewing. Working bundles can also be created by normal editing commands, such as “Find/Replace.” Brad may want to replace all occurrences of “Jennifer” with “Angelina” and then save the results as a bundle so that other co-authors can manipulate all of the annotations in a single operation, such as setting the status to “reject” or changing the replacement field to some other name should he later change his mind.

A bundle is created automatically at the end of every reviewing session. Once Jennifer finishes her session, all of her new annotations from that session form a bundle that other co-authors can review, unless she elects not to save it. This mechanism generalizes the “Track Changes” functionality in current editors and provides a uniform way to capture reviewing history.

A flexible ability to group or bundle annotations is lacking in other co-authoring systems. Bundles provide explicit representations of user-defined workflow, and they integrate normal editing with other annotation activity using a range of implicit to explicit bundle creation.

5.2.3 Working with Bundles

Various techniques help users maintain a mental model of a document and its annotations. In order to capture the structure of annotations, we employ a threaded list of annotations in the reviewing panel (**R6**: discussions). Users can expand or collapse any bundle to view or hide the annotations belonging to it. Once a bundle is expanded (i.e., its substructure is showing), “Next” and “Previous” buttons can be used to traverse the annotations within the bundle. A right-click on any annotation within the document or the reviewing panel gives users the option to view the bundles to which it belongs (**R11**: grouping).

Users can select multiple bundles at a time and perform operations (such as setting the reviewing status) on all of the selected annotations. If a bundle is selected, the anchors for all of its sub-annotations will be highlighted in the document (shown in Figure 5.3). Users can have several bundles active at one time, each in separate tabs of the reviewing pane, and switch between them. Each tab can be sorted according to author, date, document order, or various other attributes. Co-authors can prioritize annotations in a bundle using drag-and-drop techniques (**R10**: prioritization). For example, users can move a bundle up and down in the list of annotations in the reviewing pane. Annotations can also be moved between bundles.

5.3 Iterative Design and Low- to Medium-Fidelity Prototypes

Before we formalized our design in the high-fidelity Bundle Editor prototype, we iterated through a series of paper prototypes and medium-fidelity prototypes using Microsoft PowerPoint. Many design alternatives center around two interface components: the way comments are displayed in the document and the way the single and bundled annotations are organized and displayed in the reviewing pane.

For displaying comments in the document, one design alternative was to use a comment icon similar to the note icon in Acrobat, where the icon is anchored at a user-specified location. However, using icons does not allow us to encode structural information. For example, if a comment explains why a certain edit was made (i.e., a meta-comment), there is no easy way to visually link the comment icon with the edit. Also, the icon design only allows comments to be anchored at a point and not on a range of text. Several rounds of brainstorming and feedback from potential users resulted in a design that displays the comment's location and range as a colored background on the text. If the comment is anchored at a single point then it is displayed as a triangle, otherwise it is

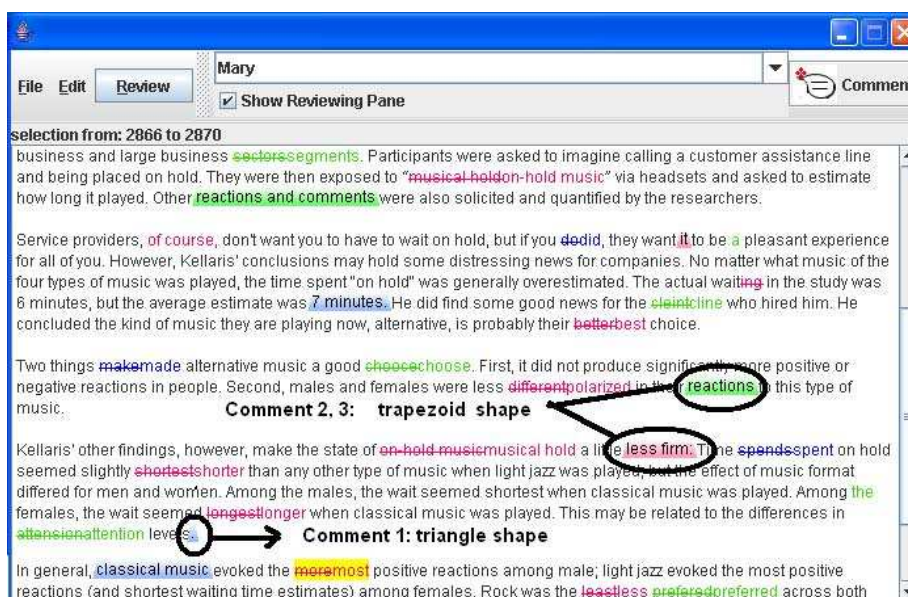


Figure 5.4: Comment's location and range are displayed as a colored background on the commented text. If a comment (e.g., comment 1) is anchored at a single point, then it is displayed as a triangle. If a comment is anchored at a range of text (e.g., comment 2 and 3), it is displayed as a trapezoid.

displayed as a trapezoid over the range of the commented text (Figure 5.4).

In addition to displaying comments in the document, we created a design scheme (Figure 5.5) that allows regular text and annotations to be displayed at the same time. In this scheme, there are three text types: regular text, inserted text, and deleted text. Different text types allow us to show regular document text and insert, delete and replacement annotations. Similar to other reviewing systems such as MS Word 2003, regular document text is shown in black on the default white document background. Insertion is color-coded text according to who made the annotation.² Deletion is color-coded text with a strike through

²Our prototype uses color to code the author attribute. It could be used to code any other attribute by a user-initiated change in the mapping, but overloading color to code multiple attributes simultaneously does not seem wise, based on our limited experience.

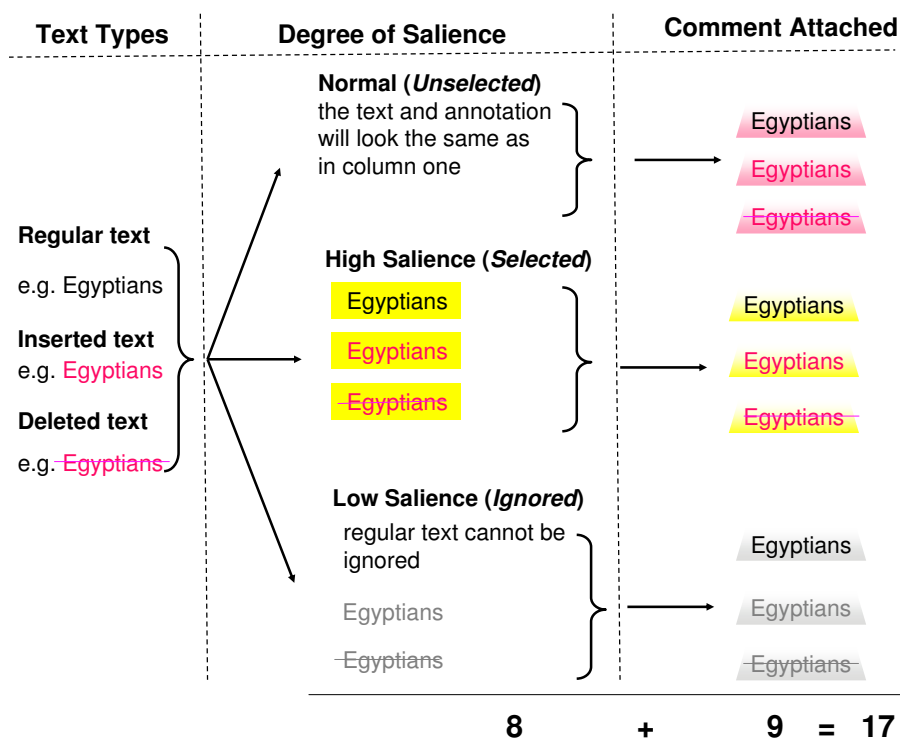


Figure 5.5: The different ways that regular text and annotations can be displayed in the document. Three text types are possible: regular text, inserted text, and deleted text. Each can be in one of three degrees of salience: normal salience (unselected), high salience (selected), and low salience (ignored), except that regular text cannot be in low salience. The result is 8 different display modes that may or may not have an associated comment. Regular text can also be associated with low-salience comments. Therefore, there are 9 different displays with comments attached. In total, there are $8+9=17$ display combinations.

the deleted text. Replacement is displayed as a combination of insertion and deletion.

There are also three degrees of salience for each type of text: normal salience when the text is not selected, high salience when the text is selected, and low salience when inserted text and deleted text are ignored. Low salience is used when users want to focus on just a particular set of annotations (e.g., verb tense edits) and temporarily ignore other annotations. One exception is that regular document text cannot be in low salience since users are not allowed to ignore original document text. Therefore, there are 8 different displays after applying degrees of salience.

The background of selected text will turn to bright yellow³ to catch users' attention, which is common in many other applications. When text is unselected, it resumes its original color-coding (e.g., inserted text is color-coded text on a white background). When inserted or deleted text is ignored, it turns into grey to help users quickly skip it when reviewing the document. These three salience displays are necessary for interacting with the document text and annotations in the document. For example, when annotation "A" has just been created, it will show as being selected in the document. Later, when annotation "B" is created or the cursor has moved to a different location in the document, annotation "A" will show as being unselected. When a user chooses to only show all the annotations belonging to one bundle in the document, all the other annotations are displayed in grey and can be ignored. This helps to reduce visual clutter.

Finally, comments can be associated with each type of text. If a comment is anchored to a range of text, the background of the text will turn into the color assigned to the author of the comment. Since bright yellow is used as a highlighting color for selection, no author can be associated with yellow as his/her reviewing color. Originally, all text is uncommented. Each comment can also be displayed as being normal, high, or low salience depending on user interactions, which results in $3 \times 3 = 9$ different ways to display text with comments.

³If all colors are reserved for coding other attributes, selection could use reverse video or other monochromatic cues.

Altogether, there are $8+9=17$ different ways that regular text and anchoring text can be displayed. Figure 5.5 shows a summary of all 17 display combinations. Evaluating the effectiveness of these displays remains for future research.

The design of the multi-tabbed reviewing pane also went through iterative design. In our original design, we had one permanent tab which displayed all the single and bundled annotations. However, we found that this made the annotation display too cluttered and that it required frequent use of filtering to select either single or bundled annotations. We then decided to use the two permanent tabs so the user could use the “All Bundles” tab as an initial guide when starting to review annotations. They could then refer to the “All Annotations” tab to see the complete list of document-embedded annotations and general comments.

5.4 Implementation

The Bundle Editor is implemented using Java Swing 1.5.0. The most important underlying component of the editor is the Annotation class. The Annotation class encapsulates all the annotation features, in particular the attributes and structure of an annotation. It also encodes various operations that can be performed on annotations, such as adding/removing annotations to/from sub-structures. The two main interface components are the document pane and the reviewing pane, which govern different annotation displays.

5.5 Pilot Testing

Prior to running our formal evaluation on the Bundle Editor, we performed two iterations of pilot testing with a total of 10 participants. All participants were Computer Science graduate students who volunteered to participate. The first pilot test, with 6 participants, was used primarily to test the robustness of the system. Observations from the test session and user feedback led to improved interface functionality. For example, we modified the system so that users could

not open the same bundle multiple times or create two bundles with the same name. We also made small changes to the interface design such as changing the highlighting color in the reviewing pane to match the yellow highlighting in the document and modifying button designs and their tooltip text to provide a better mental model to users.

The second pilot test, with 4 participants, evaluated the full experimental protocol, described in the next chapter. Pilot testing improved the clarity of questions in the questionnaire and the training session was significantly modified to include more information in a training video to illustrate how to use the Bundle Editor, and to provide participants with more hands-on task practice rather than simply having them watch the experimenter to demonstrate how to complete tasks.

Chapter 6

Evaluation of Structured Annotations

The goal of our research is to use structured annotations to support collaborative writing workflow, specifically during the collaborative reviewing stage. We conducted a usability study to determine whether structured annotations can reduce an individual co-author's workload and improve their document reviewing quality. Recall that our focus is on documents embedded with a large number of annotations, which is very common in collaborative writing. It may not be necessary to use structured annotations where there are only a few annotations in the document.

The usability study focused on the workflow experienced from the annotation receiver's perspective, rather than from the annotation creator's perspective. This is because we believe the power of bundles is best demonstrated at the annotation reviewing stage. The usability study documented here represents a first step in the evaluation of bundles. We described different ways of creating bundles in the previous chapter. In section 6.3.6, we present an initial argument for the cost and benefit tradeoff of creating bundles, which explains why we believe that users will be willing to create bundles. A usability study evaluating bundle creation remains as future work and is discussed in Chapter 7.

Our usability study consisted of a single experiment. In this chapter, we first describe the experimental methodology, including the two annotation systems used in the experiment, tasks, measures, design, participants, the procedure we followed, and the two major hypotheses. Quantitative and qualitative results

are then reported followed by a discussion of the implications of the experiment.

6.1 Methodology

Our experiment compared two annotation systems: the Bundle System, which supports structured annotations, and the Simple System, which supports edits, deletes, replacements, and comments. The Simple System was intended to be representative of current co-authoring systems such as MS Word 2003. Participants were asked to assume the role of co-author for two documents and to review annotations related to the documents made previously by other co-authors. Participants were instructed that reviewing annotations meant accepting the annotations they agreed with and rejecting the others, according to a prescribed task.

Each participant saw both systems, with a different document for each system. The two documents used were chosen from ScienceDaily [38]. The first (docB, 528 words, 7 paragraphs) is about the growth of black holes, while the second (docM, 535 words, 7 paragraphs) is about customer reaction to “on-hold music” when calling a customer service phone line. The two documents have an almost identical level of reading difficulty, as determined by the Kincaid Readability Test and the Flesch Reading Ease Score [16] (Flesch Reading Ease: 52.2 for docB, 52.1 for docM; Flesch-Kincaid Grade Level: 10.9 for docB, 10.2 for docM). A third document was used during two practice sessions. Because the document was common to all experimental configurations, we were not concerned with its similarity to the other two documents. All of the documents described above are included in Appendix B.

6.1.1 Two Systems

Both the Simple System and the Bundle System were created by modifying our Bundle Editor (Figure 5.2), so that they differed only in their annotation functions. Because we were evaluating bundles from the annotation receiver’s

perspective, we disabled and removed the following functions from the Bundle Editor to create the two systems:

- Functions that allow subjects to create/decompose bundles manually. For example, the “New Bundle” and “Bundle Decompose” buttons were removed from the Bundle Editor.
- Bundle manipulation functions such as adding and removing basic annotations from a bundle were removed. We also disabled the function that adds bundles to an existing bundle.
- Editing and annotating functions such as “cut,” “copy,” “paste,” “comment,” and “find/replace” within the Bundle Editor were disabled because in the experiment, participants only needed to review existing annotations in the document, not create new ones.
- Meta-controls of the Bundle Editor such as “open,” “save,” and “exit” were disabled because the system automatically generated these events in the study.

Although the modified Bundle System (see Figure 6.1) was simpler than the Bundle Editor, it contained all the functionality required for the tasks used in the experiment. More importantly, it was sufficient to evaluate our hypotheses.

For the Simple System (see Figure 6.2), we also removed the multi-tabbed reviewing pane and replaced it with a single-pane reviewing pane, which contains the list of document-embedded annotations. The bundled annotations that could not be included in the Simple System were displayed in a separate simulated email window, beside the system interface. Table 6.1 summarizes the differences between the two systems.

The experiment was conducted on a single Linux machine running SUSE 9.0 with a Pentium 4 CPU and 512 MB of RAM. The software for both of the experimental systems was written in Java 1.5.1.

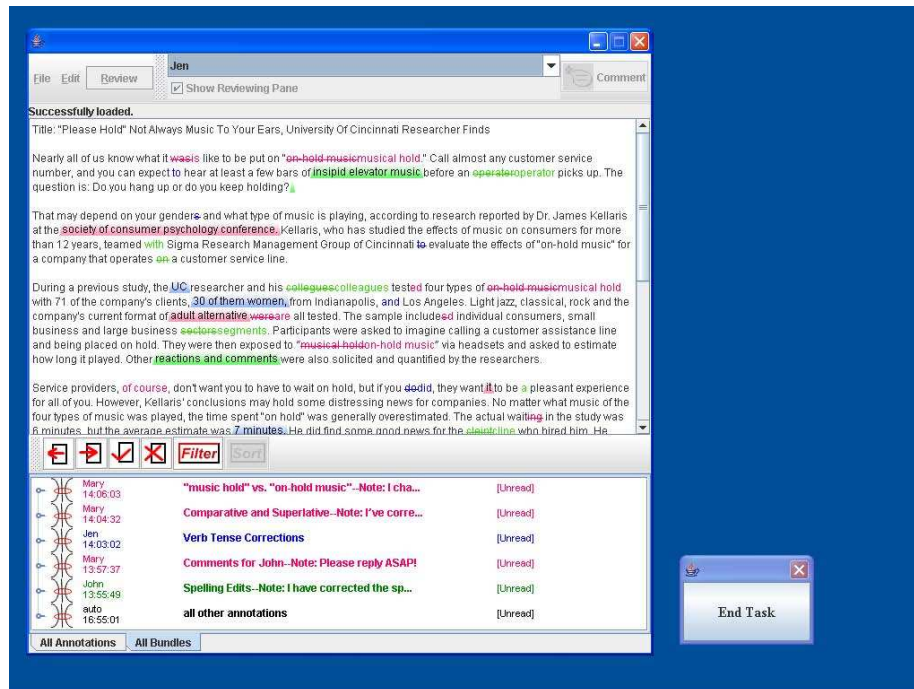


Figure 6.1: The Bundle System used in the usability study. It was created by modifying the Bundle Editor. The “New Bundle”, “Bundle Add”, “Bundle Minus”, and “Bundle Decompose” buttons were removed. Participants were not able to create new annotations or manually exit the system. There was a task control button (e.g., “End Task” button shown in the figure) located at the bottom right of the screen for participants to start or end a task during the experiment.

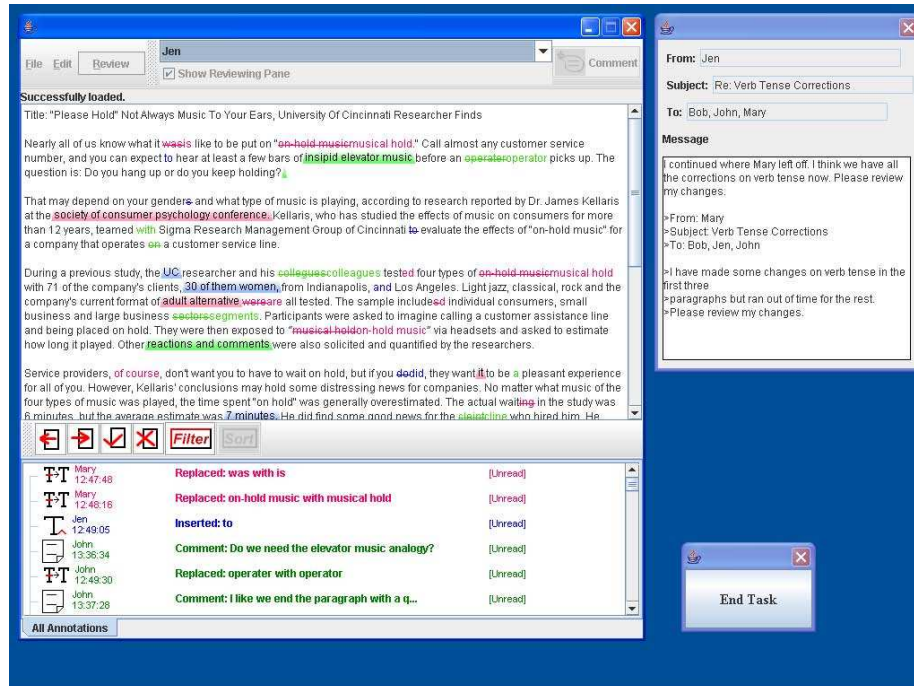


Figure 6.2: The Simple System used in the usability study. Compared to the Bundle System's interface, the Simple System only had a single-pane reviewing pane to display document-embedded annotations. General comments and higher-level annotations were displayed in a separate simulated email window, to the right of the system interface. Similar to the Bundle System, there was a task control button for the Simple System. The two systems were otherwise the same.

	Bundle System	Simple System
Interface components	document panel, multi-tabbed reviewing panel	document panel, single pane reviewing panel
Basic annotations (excluding general comments)	embedded in the document and listed or grouped in the reviewing panel	embedded in the document and listed in the reviewing panel
General comments	listed at the top of the “All Annotations” tab in the reviewing panel	shown in the simulated email window
Group of related annotations	listed in the “All Bundles” tab in the reviewing panel	shown in the simulated email window
Filtering functions	AND, OR filtering on all or a subset of annotations	OR filtering on all annotations

Table 6.1: Comparison of the Bundle System and the Simple System

6.1.2 Tasks

There were six representative tasks to complete for each document, which were designed to gauge the strengths and weaknesses of the Bundle System. The annotations for all tasks were present from the outset. We controlled for the number, type, and authorship of annotations in the documents: 52 basic annotations (8 insertions, 5 deletions, 25 replacements, and 14 comments); Jennifer, John, and Mary made 15, 15, and 25 annotations, respectively. In addition, we controlled for reviewing difficulty with respect to the amount of context participants needed to review in order to accept/reject an annotation; 36 annotations could be processed by reading a single sentence, 10 annotations required reading two sentences, and 6 required reading a full paragraph. Both documents with annotations embedded are included in Appendix B. Among the 52 annotations, 32 were related to the selected tasks, while 20 served as “distractors” as the participants were performing their tasks.

Task #5

----- Task Background -----

In this document, “**musical hold**” is used to describe the stage of holding with music playing in the background, and “**on-hold music**” is used to describe the specific type of music discussed in the document. The study described in the document is trying to evaluate the effects of “on-hold music” on customers when they are put on “musical hold” after calling a company’s service line.

During Mary’s previous reviewing session, she discovered there are some misuses of the two phrases, so she first ran “Find/Replace” function to find all the incorrect uses of “musical hold” and replaced it with “on-hold music.” Then, she ran “Find/Replace” again to find all the incorrect uses of “on-hold music” and replaced them with “musical hold.”

----- Task Instructions -----

Review the annotations regarding “musical hold” and “on-hold music.” **Accept** the ones that you agree with and **reject** the ones that you disagree with.

Click “**Start Task**” when you are ready to start.

Figure 6.3: Task 5 instruction in docM. Task background explains how the phrases “musical hold” and “on-hold music” should be used in the document followed by the specific instructions for the task.

In this section, we describe each task in terms of instructions, presentation, relevant annotations, and expectations. For each task, a task instruction screen was shown first. Some tasks also had task background to inform or refresh participants on basic English grammar or specific words used in the document, as shown in Figure 6.3. For each document, the same task instructions were given for both the Simple System and the Bundle System. Because the documents differed in content, some tasks were adjusted slightly to fit the document content, but always with the goal of making them equivalent. The full set of tasks instructions shown to the participants is included as Appendix C.

Task 1: Location Pointers.**Instructions.**

- *docB*: review annotations on quantifying words (e.g., at least, at most).
- *docM*: review annotations on comparative and superlative forms of adjectives.

Presentation.

- *Bundle System*: a bundle with a note attached containing all relevant annotations.
- *Simple System*: an email message containing location pointers for relevant annotations.

Relevant annotations. 5 task-relevant annotations from 1 co-author distributed in each document. 3 were designed to be accepted and 2 were designed to be rejected according to the document context.

Expectations. Better performance for both speed and accuracy in the Bundle System.

Task 2: Localized Annotations.

Instructions. Review all annotations in a specified paragraph.

Presentation.

- *Bundle System*: a general comment describes which paragraph to review. No relevant bundle created.
- *Simple System*: an email message describes which paragraph to review.

Relevant annotations. 5 task-relevant localized annotations from multiple co-authors. 4 were designed to be accepted and 1 was designed to be rejected according to the document context.

Expectations. Similar performance in both systems.

Task 3: Spelling Edits.

Instructions. Review spelling edits in the document.

Presentation.

- *Bundle System:* a bundle with a note attached containing all relevant annotations.
- *Simple System:* an email message describing relevant annotations.

Relevant annotations. 6 task-relevant annotations from 1 co-author distributed in the document. 4 were designed to be accepted and 2 were designed to be rejected according to the document context.

Expectations. Better performance for speed in the Bundle System. Similar performance for accuracy in both systems because spelling edits are easy to review.

Task 4: Multiple Co-authors Annotations.

Instructions. Review all verb tense edits in the document.

Presentation.

- *Bundle System:* a bundle with two bundles (created by 2 co-authors) in its substructure. Each sub-bundle contains task-relevant annotations and comments.
- *Simple System:* two email messages (from 2 co-authors) are shown (one is a reply to the other) describing the relevant annotations.

Relevant annotations. 8 task-relevant annotations from 2 co-authors (4 annotations from each co-author) distributed in the document. 6 were designed to be accepted and 2 were designed to be rejected according to the document context.

Expectations. Better performance for both speed and accuracy in the Bundle System.

Task 5: Global Replacements.**Instructions.**

- *docB*: review all the replacements between “grow” and “growth.”
- *docM*: review all the replacements between “on-hold music” and “musical hold.”

Presentation. A writing tip explaining how to use each word in the document was provided before participants started the task.

- *Bundle System*: a bundle with a note attached containing relevant annotations.
- *Simple System*: an email message describing the relevant annotations.

Relevant annotations. 5 task-relevant annotations from 1 co-author distributed in the document. 3 were designed to be accepted and 2 were designed to be rejected according to the document context.

Expectations. Better performance for speed in the Bundle System. Similar performance for accuracy in both systems because these replacement edits are easy to identify.

Task 6: Unaddressed Comments.

Instructions. Review a co-author’s comments that have not been accepted or rejected.

Presentation.

- *Bundle System*: a general comment describes which co-author’s comments to review. No relevant bundle created.
- *Simple System*: an email message describes which co-author’s comments to review.

Relevant annotations. 3 task-relevant comments from 1 co-author distributed in the document. 2 were designed to be accepted and 1 was designed to be rejected according to the document context.

Expectations. Filtering functions are likely to be used in both systems. Better performance for both speed and accuracy in the Bundle System because of multi-attribute filtering.

Each task discussed above was representative of tasks we saw in our field investigation, where authors connected higher-level communication in email with lower-level document-embedded annotations. For example, task 1,3,5 all represent the tasks of reviewing summaries of edits. Task 2 represents reviewing general comment. Task 4 represents reviewing to-do items, and task 6 represents reviewing comments-on-comments. The task difficulty was primarily for the user to find/navigate to the right set of annotations to review, which was our main focus. Some individual annotations in our study were a bit cosmetic (e.g., word replacements, spelling edits) because subjects' understanding of the document was limited (they were not authors). This minimized individual differences in reviewing skills and comprehension by making it straightforward to decide accept/reject, so we could measure time spent navigating to the relevant changes. This is discussed in the next section.

6.1.3 Measures

Our main dependent variables were speed and accuracy. Speed consisted of total completion time per task, which was the aggregate of navigation time and decision time. Navigation time was calculated by adding three types of time segments: initial navigation time, between selection navigation time, and final navigation time. See Figure 6.4 for details on how these time segments were measured. Decision time was calculated by adding the time segments between selecting an annotation to accepting or rejecting the annotation. During each task, a participant could be either navigating in the annotated document or deciding whether to accept or reject a particular annotation.

Accuracy was assessed with three measures: the number of task-relevant annotations reviewed (accepted/rejected), the number of task-relevant annotations reviewed correctly, and the number of non-task-relevant annotations

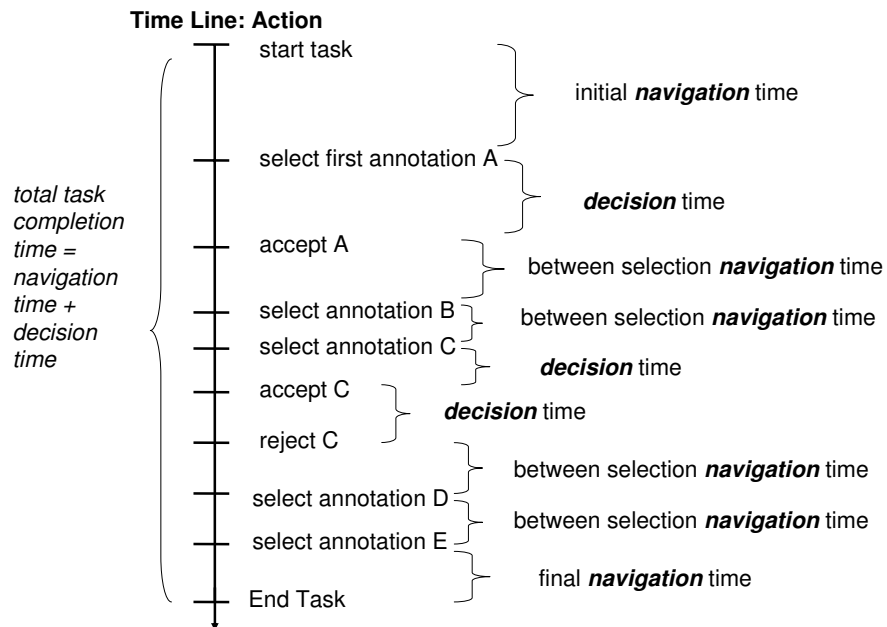


Figure 6.4: An example of how the navigation and decision time were measured. Total task completion time is the sum of navigation time and decision time. Navigation time is composed of initial navigation time, between annotation navigation time, and final navigation time. All time from selecting an annotation to accepting or rejecting the same annotation is measured as decision time.

reviewed. We also recorded the number of times the filtering function was used. Self-reported measures captured through questionnaires included ease of finding annotations, ease of completing tasks, confidence in performing task, ease of use, ease of learning, and overall system preference.

6.1.4 Experimental Design

The experiment was a within-subjects 2x6 (system type x task) factorial design. Document type was a within-subjects control variable, and both order of presentation for system and order of presentation for the document were between-subject controls. A within-subjects design was chosen for its increased power and because it allowed us to collect comparative comments on the two systems. To minimize learning effects, we counterbalanced the order of presentation for both system type and document, resulting in four configurations. The tasks were presented in the same order to each participant.

6.1.5 Participants

A total of 20 people (8 females) participated. They were undergraduate and graduate students recruited through online mailing lists and newsgroups. They were paid \$20 for their time. All spoke English as their native language. Seventeen used a word processor (mainly Microsoft Word) every 2-3 days, and 3 did so once a week. All felt very confident about using their word processor, although 5 had never used any annotation functions. They had all been involved in collaborative authoring: 6 participants fewer than 5 times, 7 participants between 5 and 10 times, and 7 participants more than 10 times.

6.1.6 Procedure

The experiment was designed for a single two-hour session. A questionnaire was administered to obtain information on past computer and writing experience. Participants were then shown a training video on general concepts such as collaborative authoring and how to use the first system, followed by a practice

session of six reviewing tasks using the first system. For each task, a participant first read the task instruction screen then clicked on the “Start Task” button. The system loaded and the data logging and timing functions started. After the participant finished a task, s/he clicked “End Task” and the next task instruction appeared. The practice tasks were similar to the experimental tasks described previously, but in a different order and on a practice document different than either of the test documents.

Participants were next asked to read the original version of the task document (i.e., with no annotations), after which they had to perform the six tasks in the order they were given. A second questionnaire was administered to collect feedback on the first system. Participants were given a 5-minute break and then were shown a video on how to use the second system, followed by six practice tasks using the same practice document then the six experiment tasks for the second document. A final questionnaire solicited feedback on the second system and asked the participants to directly compare the two systems. A short de-briefing was conducted with some of the participants based on their questionnaire data.

6.1.7 Hypotheses

Our main hypotheses were as follows:

H1. The Bundle System will reduce the time participants spend navigating to relevant annotations. Some tasks (as identified above) will be more affected than others.

H2. Participants will perform more accurately in the Bundle System than the Simple System. Some tasks (as identified above) will be more affected than others.

6.2 Results

Here we report on both the quantitative data captured through software logging as well as the self-reported data from our questionnaires.

Before testing our hypotheses, we checked to make sure that there was no effect of document. Investigation of an interaction effect between document and task on total time ($F(4,64) = 4.706$, $p = .002$, $\eta^2 = .227$) revealed that task 1 was more difficult in docB than in docM. Our goal had been to create two documents that were as equal in difficulty as possible, and so we removed task 1 from our remaining analysis and focused exclusively on tasks 2 through 6.¹

To test our hypotheses we ran 2 systems x 2 order of systems x 2 order of documents x 5 tasks ANOVAs for our speed and accuracy measures. System and tasks were within-subjects factors, and orders of system and document presentation were both between-subjects factors. For our secondary analysis, a series of two-tailed t-tests were used to investigate performance differences between the two systems for each of the tasks. Along with statistical significance, we report partial eta-squared (η^2), a measure of effect size, which is often more informative than statistical significance in applied human-computer interaction research [21]. This value is usually interpreted as .01 being a small effect size, .06 a medium effect size, and .14 a large effect size [11].

6.2.1 Testing Hypotheses

Total navigation time (across all 5 tasks) was significantly less in the Bundle System ($p < .001$). Participant's decision time, however, was not impacted by the two systems ($p = .336$). The large navigation time effect was sufficient to influence the total completion time, which was also significantly lower in the Bundle System ($p < .001$). The means are given in Table 6.2. As

¹This is a potential confound because the six tasks were always done in the same order. Thus subjects might have experienced an asymmetric transfer effect from task 1 to the other tasks, but we think this was minimal at best because the six tasks were relatively independent of each other.

Speed	Mean (in sec)		F	Sig.	η^2
	Bundle	Simple			
<i>Navigation</i>	39.3	58.3	40.1	< 0.001	0.715
<i>Decision</i>	60.8	64.5	0.98	0.336	0.058
<i>Completion</i>	100.2	122.8	22.9	< 0.001	0.589

Table 6.2: Speed measures across five tasks. Df = (1,16). N=20.

hypothesized in H1, and as Figure 6.5 shows, some tasks required less navigation time than others. There was an interaction between task and system, ($F(4, 64) = 16.09, p < .001, \eta^2 = .354$).

T-tests revealed that tasks 3, 4, and 5 were all significantly faster in the Bundle System (all $df = 19, p < .001$). There were no differences detected for tasks 2 and 6. Consistent with hypothesis H2, accuracy was also significantly better with the Bundle System. Across all 5 tasks, participants reviewed more task-relevant annotations ($p < .001$), they correctly processed more task-relevant annotations ($p = .018$), and they made fewer identification errors, meaning they reviewed fewer non-task-relevant annotations ($p < .001$) in the bundle condition. Means for these errors are shown in Table 6.3.

There was an interaction between task and the number of non-task-relevant annotations reviewed ($F(1, 16) = 21.93, p < .001, \eta^2 = .578$), prompting us to investigate which tasks were affected differently by the two systems. A series of five two-tailed t-tests showed that there were significantly more non-task-relevant annotations reviewed in the Simple System for task 4 and task 6 (both $df = 19, p < 0.001$). These differences are apparent in Figure 6.6.

6.2.2 Other Effects

In addition to the main effect of system type, we also found a main effect of task across all measures. This was expected because we designed each task to match a particular type of annotation activity; some activities are inherently more difficult and time consuming than others.

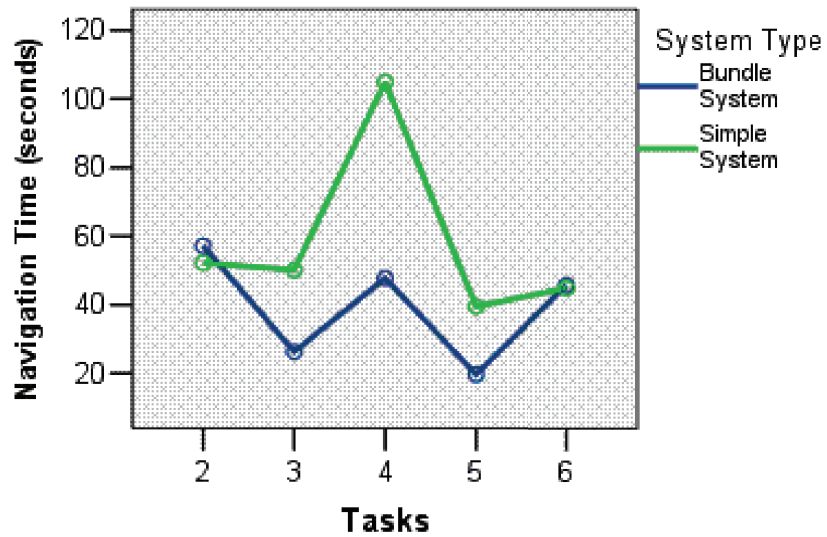


Figure 6.5: Line graph for mean navigation times per task in the two systems. N=20.

Accuracy	Mean (# of annos)		F	Sig.	η^2
	Bundle	Simple			
<i>Task-relevant annotations reviewed</i>	5.25	5.01	19.53	< 0.001	0.550
<i>Task-relevant annotations reviewed Correctly</i>	4.84	4.61	59.02	0.018	0.306
<i>Non-task-relevant annotations reviewed (errors)</i>	0.05	0.65	7.05	< 0.001	0.787

Table 6.3: Accuracy measures across five tasks. Df = (1,16). N=20.

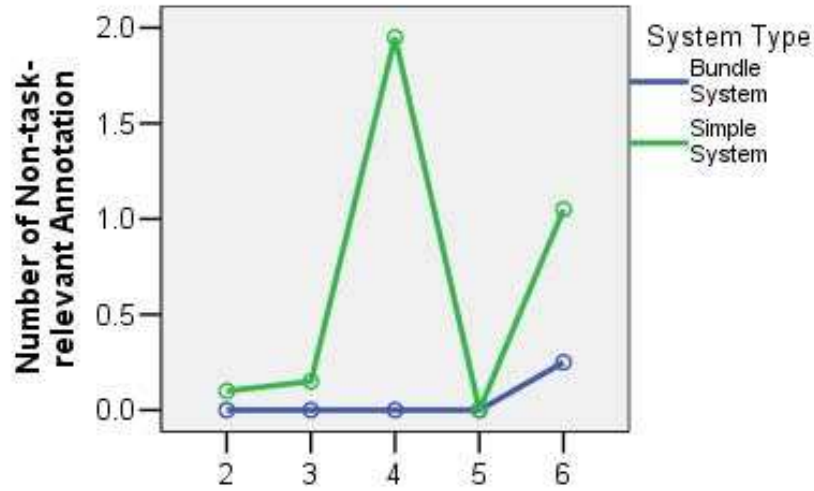


Figure 6.6: Line graph for mean number of non-task-relevant annotations reviewed. $N=20$.

We found a number of multi-way interactions involving task, and system and document presentation orders. Systematic investigation of each of the interactions revealed no clear interpretation of the interactions.

Not surprisingly, participants used the filtering functions more in the Simple System than in the Bundle System ($F(1, 16) = 39.42, p < 0.001, \eta^2 = 0.711$).

6.2.3 Self-reported Measures

We ran the Wilcoxon Signed-Rank Test on the questionnaire data. Consistent with our navigation and accuracy findings, analysis of the self-reported measures showed that with the Bundle System participants found it easier to find annotations ($p = 0.002$), easier to complete tasks ($p = 0.012$), and were more confident in their answers ($p = 0.014$). They also had an overall preference for the Bundle System ($p = 0.003$). But there was no significant difference in the ease of learning ($p = 0.667$) or ease of use ($p = 0.26$) between the two systems. When asked which of the two systems they would prefer to continue using, 18

out of the 20 participants (90%) chose the Bundle System.

6.2.4 Other Feedback

We asked our participants how they currently review documents with their co-authors. Among the 20 participants, the most popular reviewing method is writing email messages to co-authors (18/20) that include suggested changes and comments about the document. The next most popular methods are directly editing the document using a word processor (16/20), and printing out the document and marking up using a pen (15/20). These are followed by using annotation functions in existing word processors such as “Track Changes” in MS Word 2003 (12/20) and using online newsgroups like Yahoo Groups (10/20). Participants usually use multiple reviewing methods (e.g., direct editing + email). The reviewing methods and their frequencies of use are shown in Figure 6.7.

After using each of the systems, participants were also asked to estimate whether they spent more time finding annotations of interest or deciding whether to accept or reject annotations. They could also indicate that they spent roughly the same amount of time on the two activities. As we expected, participants felt they spent more time deciding whether to accept or reject annotations (13/20) in the Bundle System than finding annotations of interests (3/20). The remaining 4 participants felt they spent roughly the same amount of time on the two activities. In the Simple System, opinions were almost evenly split (8 chose finding annotations, 5 chose accepting/rejecting annotations, and 6 chose roughly the same). The results are summarized in the Figure 6.8.

Participants provided free-form comments at the end of the questionnaire about what they liked and disliked about each system. For the Simple System, although not actually integrated with the system, most participants indicated that they liked the email window, which provided them with more information to complete tasks. Interestingly, many participants who used the Simple System first indicated they liked the filtering function; however, of those participants who had first been exposed to the Bundle System, almost all disliked the

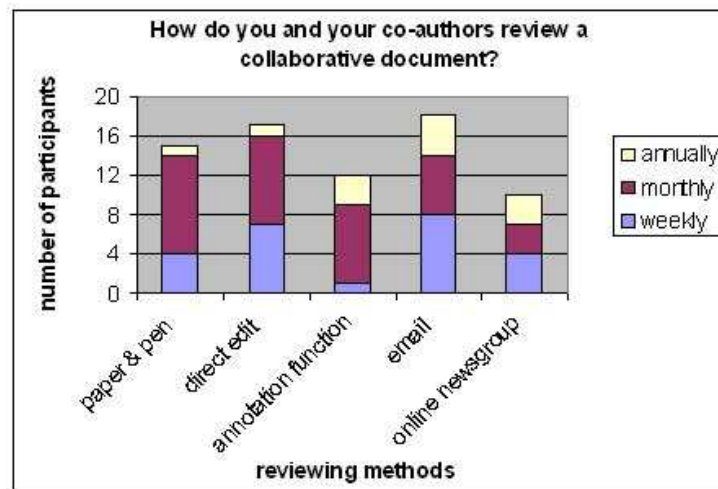


Figure 6.7: Comparing different collaborative reviewing methods and their frequency of use. $N = 20$.

comparatively limited filtering functions in the Simple System. For the Bundle System, participants noted the time saved using bundles and were surprised by how easy it was to learn to use bundles. They also liked the flexible filtering provided in the Bundle System. One suggestion for improvement in the Bundle System was to increase the size of the reviewing pane. Participants felt the current reviewing pane was small and required too much scrolling.

6.2.5 Summary of Results

To summarize, the Bundle System allowed participants to navigate among annotations significantly faster for tasks 3, 4 and 5. Participants were also significantly more accurate with the Bundle System; for example, they reviewed significantly fewer non-task-relevant annotations for tasks 4 and 6. Overall, 90% of participants preferred the Bundle System.

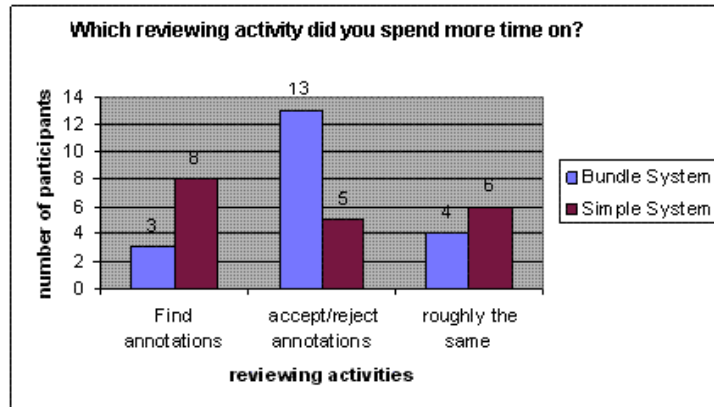


Figure 6.8: Participants' perceived length for reviewing activities.

6.3 Discussions

There are a number of interesting findings from the study, which we discuss in the remainder of this chapter.

6.3.1 Bundle Concept Is Intuitive

All participants developed the strategy of using a bundle list as their guide for completing tasks. They searched first for an existing bundle related to the current task description before directly searching for annotations in the document. Based on their interaction sequences with the prototype and their feedback, it was clear that the bundle concept, and its fit within the task workflow, was intuitive.

6.3.2 Bundles Reduce Navigation Time

Once participants found a relevant bundle, locating each annotation in the document was a single click away. By contrast, in the Simple System, most of the navigation time was spent searching through the document for the next relevant annotation, which was time consuming. Bundling reduced the navigation time

for tasks 3, 4, and 5. All three tasks had annotations distributed throughout the document that were not amenable to basic filtering. For task 6, filtering was a good strategy in both systems. Even though the Bundle System had the advantage of filtering on both the comment and author attribute, it was easy in the Simple System to filter on author and then identify the comments. So it was not surprising that task 6 did not show a difference. As one would hope, there was no difference in navigation time for tasks that were localized within the document (task 2).

6.3.3 Bundles Improve Accuracy

Once the correct bundle was found, users were guaranteed to find the task-relevant set of annotations. This minimized the number of extra annotations reviewed and allowed users to concentrate on reviewing the actual annotations. The biggest difference was found in task 4, where 39 extra annotations were reviewed across all participants in the Simple System, and none extra were reviewed in the Bundle System. The cause of this was users mistakenly identifying annotations as verb tense changes; for example, in docB replacing “grow” with “growth” was treated as a verb tense change. This was quite surprising, given that all our participants were native English speakers. But it shows that bundling can overcome even basic misunderstandings of the English language.

6.3.4 Users Group Annotations

Participants filtered significantly more often in the Simple System than in the Bundle System. They did so to reduce the number of annotations under consideration for a task. Participants were effectively creating their own temporary task-based annotation groups. Not only might there be a cost to having the reviewer do the grouping (see discussion on cost below), but current systems do not allow users to store filter results for subsequent usage. Bundling supports the easy creation and reuse of annotation groups formed through filtering.

6.3.5 Scalability of Bundles

Our target context for bundles is sophisticated documents that are heavily annotated. We chose simpler documents for our experiment in order to keep the tasks manageable. We speculate, however, that a comparison between the Simple System and the Bundle System for sophisticated documents would be even more dramatic. As a document increases in length, causing relevant annotations to be spread further apart, navigation time will increase without bundles.

6.3.6 Cost/Benefit Tradeoff

Our experiment only evaluated the annotation reviewing stage of authoring. Bundles shift some of the effort that is traditionally spent on annotation reviewing to annotation creation. At first glance this might appear to be a zero sum game, and that effort is only being shifted within the authoring workflow. We argue that authors are currently communicating a large amount of information through email, and that manually creating bundles should be more efficient than incurring overhead through the inefficiencies of email. Automatically generated bundles should clearly be faster than email communication. A tradeoff to explore, however, will be between the value of bundles and the increased overall complexity they bring to the annotation system. Evaluating bundle creation, and the impact of bundles on the complete co-authoring workflow, is an obvious next step in our work.

6.3.7 Bundles Provide a More Pleasant User Experience

When participants were asked which system they preferred, 90% stated that it was the Bundle System. The elements of the Simple System they liked the most were the email message and filtering function. We note that the experimental design provided a single email message per task, with clear instructions, which underestimates the workload in real situations when users need to locate the relevant email, and possibly an entire email thread describing the task. The two participants who favored the Simple System were both experienced Microsoft

Word users, but neither had used the annotation functions. They were excited by the functionality in the Simple System, and they found the Bundle System to be complex and confusing. However, they both recognized the potential advantages of bundles and thought that after becoming accustomed to basic annotation functions, they might desire more complex ones.

Chapter 7

Conclusion and Future Work

7.1 Conclusions

In this thesis, we have presented a structured annotation model, which includes annotation groups called bundles. Bundles are designed to improve co-authoring workflow by fully integrating annotations (both basic and higher-level annotations) with the document. We have implemented a preliminary prototype called the Bundle Editor and compared it to a system that offers only basic annotation functions. Our study focused on annotation reviewing and showed that structured annotations can reduce the time it takes to navigate between task-relevant annotations and can improve reviewing accuracy.

7.2 Future Work

Ultimately, we would like to work towards a lightweight and robust annotation tool that can be integrated into existing word processors (e.g., Microsoft Word) or online reviewing systems (e.g., XMetal Reviewer) to support co-authoring. We summarize some of the potential future work below.

Evaluation of Bundle Creation

Now that there are confirmed benefits at the reviewing stage, our next step will be to investigate the usability of bundle creation and, more generally, how

bundles support the full co-authoring workflow. We mentioned in section 5.2.2 that there are four ways to create bundles. More research is needed to explore the practical situations for using each bundle creation method as well as more accurate estimates of the effort required.

Supporting Version Control and Synchronous Co-authoring

Broader issues of how bundles can support collaborative writing in general still remain. These include investigating how bundles might be extended to support version control and synchronous co-authoring, which are both classic problems in the collaborative writing literature.

Our initial intuition is that bundles provide more organized annotations, which can reduce the workload co-authors experience during each reviewing and editing cycle. Thus, co-authors can complete the document with fewer reviewing cycles, resulting in fewer versions. Although our structured annotation model does not directly target the problem of version control, we hypothesize that it can minimize it. More research is required to measure the effect of bundles on version control.

In our research, we investigated asynchronous collaboration because it is more common during the reviewing and editing stages of collaborative writing. However, structured annotations may also be extended to synchronous co-authoring environments. For example, in a synchronous setting, co-authors could choose to send a bundle back and forth through instant messaging, or bundles could update themselves automatically for all co-authors to see real-time changes. Synchronous conversations between co-authors could be saved as bundles for later retrieval by other co-authors. Many potential synchronous uses for bundles remain to be explored.

Enhancing Annotation Structures in the Document

Another area for further research is the display of bundled annotations in the document. The reviewing pane clearly shows which bundles have been created

and their substructures. However, when the user selects (i.e., highlights) an annotation in the document pane, there is no visual cue suggesting which bundle or bundles the annotation belongs to or whether there are other similar annotations (i.e., belonging to the same bundle) nearby in the document. Currently, users need to explicitly right-click on the annotation and choose the option to view the bundle(s) it belongs to. Visualization techniques need to be applied carefully to display the relationships among annotations in the document while at the same time not overloading the document pane.

Including Free Text Search

In our usability study, some participants stated a need for a free text search on annotations in the Bundle Editor. They pointed out that when there are a large number of annotations including bundles, it would be easier if they could type the bundle name into a search box and locate the bundle quickly in the reviewing pane. This shows another advantage of having structured annotations: if annotations were distributed in both the document and in emails, it would be nearly impossible to implement one search mechanism to search on two different applications. Structured annotations make it possible to conduct a single search on all document-related annotations (i.e., single or bundled annotations).

Structured Annotations for Rich Annotation Types

Another future research area is how to apply structured annotations on rich annotation types such as image, audio, and video annotations. The attributes we defined in the annotation model (see section 4.1) will need to be modified accordingly. For example, the note attribute might be an image or audio file instead of just a simple string of text.

Bibliography

- [1] Allen, R. (2005). Workflow: An introduction. Workflow Management Coalition. <http://www.wfmc.org/information/Workflow-An-Introduction.pdf>.
- [2] Baecker, R. M., Nastos, D., Posner, I. R., and Mawby, K. L. (1993). The user-centered iterative design of collaborative writing software. In *CHI '93: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 399–405, New York, NY, USA. ACM Press.
- [3] Brush, A. B. (2002). *Annotating Digital Documents for Asynchronous Collaboration*. PhD thesis, University of Washington.
- [4] Brush, A. J. and Borning, A. (2003). 'today' messages: lightweight group awareness via email. In *CHI '03: extended abstracts on Human factors in computing systems*, pages 920–921, New York, NY, USA. ACM Press.
- [5] Brush, A. J. B., Barger, D., Grudin, J., and Gupta, A. (2002). Notification for shared annotation of digital documents. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 89–96, New York, NY, USA. ACM Press.
- [6] Brush, A. J. B., Barger, D., Gupta, A., and Cadiz, J. J. (2001). Robust annotation positioning in digital documents. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 285–292, New York, NY, USA. ACM Press.
- [7] Cadiz, J. J., Gupta, A., and Grudin, J. (2000). Using web annotations for asynchronous collaboration around documents. In *CSCW '00: Proceedings*

-
- of the 2000 ACM conference on Computer supported cooperative work, pages 309–318, New York, NY, USA. ACM Press.
- [8] Catlin, T., Bush, P., and Yankelovich, N. (1989). Internote: extending a hypermedia framework to support annotative collaboration. In *HYPERTEXT '89: Proceedings of the second annual ACM conference on Hypertext*, pages 365–378, New York, NY, USA. ACM Press.
- [9] Chandler, H. E. (2001). The complexity of online groups: a case study of asynchronous collaboration. *ACM Journal of Computer Documentation*, 25(1):17–24.
- [10] Churchill, E. F., Trevor, J., Bly, S., Nelson, L., and Cubranic, D. (2000). Anchored conversations: chatting in the context of a document. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 454–461, New York, NY, USA. ACM Press.
- [11] Cohen, A. L., Cash, D., Muller, M. J., and Culberson, C. (1999). Writing apart and designing together. In *CHI '99: CHI '99 extended abstracts on Human factors in computing systems*, pages 198–199, New York, NY, USA. ACM Press.
- [12] Dourish, P. (1996). Consistency guarantees: exploiting application semantics for consistency management in a collaboration toolkit. In *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 268–277, New York, NY, USA. ACM Press.
- [13] Dourish, P. and Bellotti, V. (1992). Awareness and coordination in shared workspaces. In *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 107–114, New York, NY, USA. ACM Press.
- [14] Ede, L. and Lunsford, A. (1990). *Singular Texts/Plural Authors: Perspectives on Collaborative Writing*. Carbondale: Southern Illinois UP.

-
- [15] Fish, R. S., Kraut, R. E., and Leland, M. D. P. (1988). Quilt: a collaborative tool for cooperative writing. In *Conference Sponsored by ACM SIGOIS and IEEECS TC-OA on Office information systems*, pages 30–37, New York, NY, USA. ACM Press.
- [16] Flesch-Kincaid Readability Test. http://en.wikipedia.org/wiki/Flesch-Kincaid_Readability_Test.
- [17] Huang, E. M. and Mynatt, E. D. (2003). Semi-public displays for small, co-located groups. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 49–56, New York, NY, USA. ACM Press.
- [18] Jackson, L. S. and Grossman, E. (1999). Integration of synchronous and asynchronous collaboration activities. *ACM Computing Surveys*, 31(2):12.
- [19] Jaeger, T. and Prakash, A. (1996). Requirements of role-based access control for collaborative systems. In *RBAC '95: Proceedings of the first ACM Workshop on Role-based access control*, page 16, New York, NY, USA. ACM Press.
- [20] Kahan, J. and Koivunen, M.-R. (2001). Annotea: an open RDF infrastructure for shared web annotations. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 623–632, New York, NY, USA. ACM Press.
- [21] Landauer, T. (1997). *Handbook of Human-Computer Interaction*, chapter 9: Behavioral research methods in human-computer interaction., pages 203–227. Amsterdam:Elsevier Science B.V.
- [22] Margolis, M. and Resnick, D. Third voice: *Vox Populi Vox Dei?* <http://firstmonday.org/issues/issue4.10/margolis/index.html>.
- [23] Marshall, C. C. (1997). Annotation: from paper books to the digital library. In *DL '97: Proceedings of the second ACM international conference on Digital libraries*, pages 131–140, New York, NY, USA. ACM Press.

-
- [24] Mendoza-Chapa, S., Salcedo, M. R., and Oktaba, H. (2000). Group awareness in collaborative writing systems. In *Proceedings of the 6th International workshop on groupware*, pages 112–118.
- [25] Miller, E. (1998). An introduction to the resource description framework. <http://www.dlib.org/dlib/may98/miller/05miller.html>.
- [26] Munson, J. P. and Dewan, P. (1994). A flexible object merging framework. In *CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 231–242, New York, NY, USA. ACM Press.
- [27] Neuwirth, C. M. (2000). Computer support for collaborative writing: a human-computer interaction perspective. In *Third annual collaborative editing workshop*.
- [28] Neuwirth, C. M., Chandhok, R., Charney, D., Wojahn, P., and Kim, L. (1994a). Distributed collaborative writing: a comparison of spoken and written modalities for reviewing and revising documents. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 51–57, New York, NY, USA. ACM Press.
- [29] Neuwirth, C. M., Chandhok, R., Kaufer, D. S., Erion, P., Morris, J., and Miller, D. (1992). Flexible diff-ing in a collaborative writing system. In *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 147–154, New York, NY, USA. ACM Press.
- [30] Neuwirth, C. M., Kaufer, D. S., Chandhok, R., and Morris, J. H. (1990). Issues in the design of computer support for co-authoring and commenting. In *CSCW '90: Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, pages 183–195, New York, NY, USA. ACM Press.
- [31] Neuwirth, C. M., Kaufer, D. S., Chandhok, R., and Morris, J. H. (1994b). Computer support for distributed collaborative writing: defining parameters of interaction. In *CSCW '94: Proceedings of the 1994 ACM conference on*


-
- Computer supported cooperative work*, pages 145–152, New York, NY, USA. ACM Press.
- [32] Noel, S. and Robert, J.-M. (2004). Empirical study on collaborative writing: What do co-authors do, use, and like? *Computer supported cooperative work*, 13(1):63–89.
- [33] Noldus Observer. <http://www.noldus.com/site/doc200401012>.
- [34] Ovsiannikov, I. A., Arbib, M. A., and McNeill, T. H. (1999). Annotation technology. *Int. J. Hum.-Comput. Stud.*, 50(4):329–362.
- [35] Phelps, L. (1997). Active documentation: wizards as a medium for meeting user needs. In *SIGDOC '97: Proceedings of the 15th annual international conference on Computer documentation*, pages 207–210, New York, NY, USA. ACM Press.
- [36] Qixing Zheng, J. M. and Booth, K. (2006). Co-authoring with structured annotations. In *CHI '06: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA. ACM Press.
- [37] Rhyne, J. R. and Wolf, C. G. (1992). Tools for supporting the collaborative process. In *UIST '92: Proceedings of the 5th annual ACM symposium on User interface software and technology*, pages 161–170, New York, NY, USA. ACM Press.
- [38] ScienceDaily. <http://www.sciencedaily.com/>.
- [39] Sharples, M., Goodlet, J., Beck, E., Wood, C., Easterbrook, S., and Plowman, L. (1993). *Research issues in the study of computer supported collaborative writing*, chapter 2, pages 9–28. Springer-Verlag.
- [40] Weng, C. and Gennari, J. (2004). Asynchronous collaborative writing through annotations, note. In *Proceedings of ACM Conference on Computer supported cooperative work (CSCW'04)*, pages 564–573, Chicago, IL.

-
- [41] Whitehead, E. J. J. (2001). WebDAV and deltaV: collaborative authoring, versioning, and configuration management for the web. In *HYPertext '01: Proceedings of the twelfth ACM conference on Hypertext and hypermedia*, pages 259–260, New York, NY, USA. ACM Press.
- [42] Wojahn, P. G., Neuwirth, C. M., and Bullock, B. (1998). Effects of interfaces for annotation on communication in a collaborative task. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 456–463, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.
- [43] Woods, D. D., Patterson, E. S., Roth, E. M., and Christoffersen, K. (2002). Can we ever escape from data overload? A cognitive systems diagnosis. In *Cognition, technology and work*, volume 4, pages 22–36.
- [44] XMetal 4.6 Reviewer. http://www.xmetal.com/en.us/products/xmetal_reviewer/index.x.

Appendix A

Usability Study

Questionnaire

	Using Structured Annotations in Collaborative Writing Study Questionnaire Form							
Instructions								
Please try to respond to all of the items listed below. For those items that are not applicable, specify N/A .								
Part 1: Past Computer and Writing Experience <i>(To be completed before the study)</i>								
<p>1. Which word processor do you currently use for writing documents (e.g. essays, reports, letters, conference papers, journal articles, etc.)?</p> <p>2. How often do you use the word processor?</p> <p> <input type="checkbox"/> Once a month <input type="checkbox"/> Once a week <input type="checkbox"/> Every 2-3 days <input type="checkbox"/> Every day </p> <p>3. How confident do you feel about using the word processor?</p> <table border="1" data-bbox="295 981 906 1014"> <tr> <td></td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td></td> </tr> </table> <p>Not at all confident <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> Very confident</p> <p>4. Do you use the annotation functions in the word processor? (e.g. Track Changes and Commenting functions in Microsoft Word)</p> <p> <input type="checkbox"/> Yes. </p> <p> <input type="checkbox"/> No, please specify why: _____ </p> <p>5. Have you previously written or reviewed documents with other people?</p> <p> <input type="checkbox"/> None <input type="checkbox"/> less 5 times <input type="checkbox"/> between 5~10 times <input type="checkbox"/> more than 10 times </p> <p>Word processor used in collaborative writing: _____</p>			1	2	3	4	5	
	1	2	3	4	5			
<p style="text-align: center;">(Continue on the next page)</p>								

5. How do you and your co-authors review a collaborative document? (Check all the items that apply.)

Reviewing Method	Frequency of Use
<input type="checkbox"/> Print out the document, mark on the document using a pen, and then hand back the marked document to co-author(s).	<input type="checkbox"/> weekly <input type="checkbox"/> monthly <input type="checkbox"/> annually
<input type="checkbox"/> Directly edit on the document using a word processor.	<input type="checkbox"/> weekly <input type="checkbox"/> monthly <input type="checkbox"/> annually
<input type="checkbox"/> Use the annotation function in the word processor to edit the document and add comments.	<input type="checkbox"/> weekly <input type="checkbox"/> monthly <input type="checkbox"/> annually
<input type="checkbox"/> Write an email message that includes suggested changes and comments about the document to other co-authors.	<input type="checkbox"/> weekly <input type="checkbox"/> monthly <input type="checkbox"/> annually
<input type="checkbox"/> Use online communication groupware (e.g., Yahoo! Groups) to discuss about the changes to the document.	<input type="checkbox"/> weekly <input type="checkbox"/> monthly <input type="checkbox"/> annually
<input type="checkbox"/> Other. Please specify: _____	<input type="checkbox"/> weekly <input type="checkbox"/> monthly <input type="checkbox"/> annually

Part 2: (To be completed after completing tasks using the **first** system)

		1	2	3	4	5	6	7	
1. It was easy to learn to use this system.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
2. Navigating through annotations using this system was easy.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
3. Completing the given tasks using the system was easy.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
4. Finding annotations of interest using this system was easy.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
6. Overall, I was satisfied with how easy it was to use this system.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
7. I was confident about my answers to the tasks.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
8. I would like to use this system for my co-writing activities.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
9. I enjoyed using this system.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree

Questions:

1. What was the most difficult task for you to complete using the system?

2. Overall, on which of the following two activities do you feel you spent the most time?

- ☐ Finding the annotations of interest in the system
- ☐ Determining whether to accept or reject annotations
- ☐ I spent roughly the same amount of time on the above two activities

3. What particular aspect(s) of this system did you **like**?

4. What particular aspect(s) of this system did you **dislike**?

Part 3: (To be completed after completing tasks using the **second** system)

		1	2	3	4	5	6	7	
1. It was easy to learn to use this system.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
2. Navigating through annotations using this system was easy.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
3. Completing the given tasks using the system was easy.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
4. Finding annotations of interest using this system was easy.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
6. Overall, I was satisfied with how easy it was to use this system.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
7. I was confident about my answers to the tasks.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
8. I would like to use this system for my co-writing activities.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree
9. I enjoyed using this system.	strongly disagree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	strongly agree

Questions:

1. What was the most difficult task for you to complete using the system?

2. Overall, on which of the following two activities do you feel you spent the most time?

☐ Finding the annotations of interest in the system

☐ Determining whether to accept or reject annotations

☐ I spent roughly the same time on the above two activities

3. What particular aspect(s) of this system did you **like**?

4. What particular aspect(s) of this system did you **dislike**?

5. If you could choose only one of the systems to continue using, which would it be?

☐ First System

☐ Second System

6. Any additional comments?

Appendix B

Usability Study Documents

B. 1 Original black hole document (docB) with no annotations

B. 2 Black hole document with annotations

B. 3 Original music and consumer document (docM) with no annotations

B. 4 Music and consumer document with annotations

B. 5 Weather and mood document (practice document) with annotations

Title: NASA Observatory Confirms Black Hole Limits

The very largest black holes reach a certain point and then growth no more. That's according to the survey of black holes made by NASA's Chandra X-ray Observatory. Scientists also discover that previously hidden black holes are well below their weight limit.

The new results corroborate recent theoretical work about how black holes and galaxies grow. The biggest black holes, those with at least 100 million times heavier the mass of the sun, eat voraciously during the early universe. Nearly all of them ran out of "food" billions of years ago and went onto a forced starvation diet. On the other hand, black holes 10 to 100 million solar masses follow a more controlled eating plan. Because they took smaller portions of their meals of gas and dust, they continue growth as of today.

"Our data show some super massive black holes seem to binge, while others prefer to graze," said Amy Barger of University of Wisconsin and University of Hawaii. Barger is the lead author of the paper describes the results in the latest issue of the astronomical journal. "We understand better than ever how super massive black holes grow."

One revelation is that there is a strong association between the growth of the black holes and the birth of stars. Previously, astronomers do careful studies of the birth of stars in galaxies but didn't know as much about the black holes at their centers. "These galaxies lose material into their central black holes at the same time they make their stars," Barger said. Therefore, whatever mechanism governs star formation in galaxies also governs black hole growth.

Astronomers have made an accurate census of both the biggest, black holes in the distance, and the relatively smaller, clamer ones closer by Earth. Now, for the first time, the ones in between the two extreme have been properly counted. Co-author Richard Mushotzky of NASA's Goddard Space Flight Center, Greenbelt, Md. said that they needed to have an accurate head count over time of all growth black holes if they ever hoped to understand black holes' habits.

This study relies on the X-ray images obtained, the Chandra Deep Fields North and South, plus a key wider-area survey of an area called the "Lockman Hole." The distances to the X-ray sources were determined by optical spectroscopic follow-up at the Keck 15-meter telescope on Mauna Kea in Hawaii, and show the black holes range from a billion to 12 billion light-years away. The very long-exposure images are crucial to keep observing black holes within a billion light-years away and find the black holes that otherwise would go unnoticed.

Chandra found many of the black holes smaller than about 100 million suns are buried under large amounts of dust. This prevents detection of the optical light from the heated material near the black holes. The X-rays were more energetic and able to dig through this dust and gas. However, the black holes show little sign of being obscured by dust or gas. In a form of weight self-control, powerful winds generated by the black hole's feeding frenzy may have cleared out the remaining dust and gas.

Title: NASA Observatory Confirms Black Hole Limits

The very largest black holes reach a certain point and then ~~growth~~grow no more. That's according to ~~the survey~~of back holes made ~~by~~with NASA's Chandra X-ray ~~Observatory~~Observatory. Scientists also discovered that previously ~~hidden~~black holes are well below their weight limit.

~~The~~These new results corroborate recent ~~theoretical~~theoretical work about how black holes and galaxies ~~grow~~growth. The biggest black holes, those with ~~at least~~at most 100 million times ~~heavier~~the mass of the sun, ~~eat~~ate voraciously during the early universe. Nearly all of them ran out of "food" ~~billions~~of years ago and went onto a forced starvation diet. On the other hand, black holes ~~approximately~~ 10 to 100 million solar masses followed a more controlled eating plan. Because they took smaller portions of their meals of gas and dust, they continued ~~growth~~growing as of today.

"Our data show some super massive black holes seem to binge, while others prefer to graze," said Amy Barger of ~~the~~University of Wisconsin ~~and~~with University of Hawaii. Barger is the lead author ~~e~~for the paper ~~describes~~describing the results in the latest issue of the ~~astronomical journal~~. "We understand better than ever how super massive black holes grow."

One revelation is that there was a strong ~~association~~connection between the growth of ~~the~~black holes and the birth of stars. Previously, astronomers ~~do~~had done careful studies of the ~~birth of stars~~in galaxies but didn't know as much about the black holes at their centers. "These galaxies lose material into ~~their~~central black holes at the same time they make their stars," Barger said. Therefore, whatever mechanism ~~governs~~given star formation in galaxies also governs black hole ~~growth~~growing.

Astronomers have made an ~~accurate census~~of both the biggest, ~~active~~black holes in the distance, and the relatively ~~smaller~~small, ~~claim~~erclaimers ones closer ~~by~~to Earth. Now, ~~for the first time~~, the ones in between ~~the two extreme~~ have been properly counted. Co-author ~~Richard Mushotzky~~ of NASA's Goddard Space Flight Center, Greenbelt, Md. ~~said~~says that they needed to have an accurate head count over time of all ~~growth~~growing black holes if they ever hoped to understand black holes' habits.

This study relied ~~ed~~ on the ~~X-ray images~~ever obtained, the Chandra Deep Fields North and South, plus a key wider-area survey of an area called the "Lockman Hole." The distances to the X-ray sources were determined by optical spectroscopic follow-up at the Keck ~~15-meter~~telescope on Mauna Kea in Hawaii, and show the black holes range from ~~less than~~ a billion to 12 billion light-years away. The very ~~long-exposure images~~are ~~crucial~~crucial to keep observing black holes within a billion light-years away and find the black holes that otherwise would go unnoticed.

Chandra found many of the black holes smaller than about 100 million suns are buried under large amounts of ~~dust~~. This prevents detection of the ~~optical light~~from the heated material near the black holes. The X-rays ~~were~~are more ~~energetic~~energetic and able to ~~dig~~burrow through this dust and gas. However, the ~~larger~~black holes show little sign of being obscured by dust or gas. In a form of weight self-control, powerful winds generated by the black hole's feeding frenzy may have cleared out the remaining dust and gas.

Figure B.2: Black hole document with annotations.

Comments in docB (listed in document order):

Jen: I think it's the best survey to date.

John: Would readers understand what hidden black holes are?

Jen: I think it should be "millions."

Mary: We need to capitalize the first character for each word in the journal name.

John: Precisely, it should be the birthrate of stars.

John: I don't think "their" has a clear reference.

Mary: Shall we add a reference here for readers to follow?

Jen: I don't think it's the first time.

Mary: I don't think it's clear that Richard is the co-author of which paper.

Mary: They are the deepest X-Ray images ever obtained, right, John?

Jen: I think the telescope is only 10-meter long.

John: Do we need to explain more about how the long-exposure images are obtained?

Jen: We also need to include gas.

Mary: John, do we need to explain what optical light is?

Title: "Please Hold" Not Always Music To Your Ears, University Of Cincinnati Researcher Finds

Nearly all of us know what it was like to be put on "on-hold music." Call almost any customer service number, and you can expect hear at least a few bars of insipid elevator music before an operator picks up. The question is: Do you hang up or do you keep holding? That may depend on your genders and what type of music is playing, according to research reported by Dr. marketing James Kellaris at the society of consumer psychology conference. Kellaris, who has studied the effects of music on consumers for more than 12 years, teamed Sigma Research Management Group of Cincinnati to evaluate the effects of "on-hold music" for a company that operates on a customer service line.

The UC researcher and his colleagues tested four types of on-hold music with 71 of the company's clients, 30 of them women, from Indianapolis, Los Angeles. Light jazz, classical, rock and the company's current format of adult alternative were all tested. The sample include individual consumers, small business and large business segments. Participants were asked to imagine calling a customer assistance line and being placed on hold. They were then exposed to "musical hold" via headsets and asked to estimate how long it played. Other reactions and comments were also solicited and quatiified by the researchers.

Service providers, don't want you to have to wait on hold, but if you do, they want it to be pleasant experience for all of you. But Kellaris' conclusions may hold some distressing news for companies. No matter what music is played, the time spent "on hold" was generally overestimated. The actual waiting in the study was 6 minutes, but the average estimate was 7 minutes.

He did find some good news for the cleint who hired him. "The kind of music they're playing now, alternative, is probably their better choice. Two things made it a good chooce. First, it did not produce significantly more positive or negatives reactions in people. Second, males and females were less different in their reactions to this type of music."

Kellaris' other findings, however, make the state of on-hold music a little less firm: Time spend on hold seemed slightly shorter when light jazz was played, but the effect of music format differed for men and women. Among the males, the wait seemed shorter when classical music was played. Among the females, the wait seemed longest when classical music was played. This may be related to the differences in attention levels.

In general, classical music evoked the more positive reactions among males; light jazz evoked the most positive reactions (and shortest waiting time estimates) among females. Rock is the least preferred across both gender groups and produce the longest waiting time estimates. "The rock music's driving beat kind of aggravates people calling a customer assistance line with a problem," said Kellaris. "The more positive the reaction to the music, the shorter the waiting timeis seemed to be. "So maybe time does tend to fly when you're having fun, even if you're on musical hold," Kellaris quipped.

Title: "Please Hold" Not Always Music To Your Ears, University Of Cincinnati Researcher Finds

Nearly all of us know what it ~~was~~ like to be put on "~~on-hold music~~ musical hold." Call almost any customer service number, and you can expect to hear at least a few bars of ~~insipid elevator music~~ before an ~~operator~~ operator picks up. The question is: Do you hang up or do you keep holding?

That may depend on your ~~genders~~ and what type of music is playing, according to research reported by Dr. James Kellaris at the ~~society of consumer psychology conference~~. Kellaris, who has studied the effects of music on consumers for more than 12 years, teamed ~~with~~ Sigma Research Management Group of Cincinnati ~~to~~ evaluate the effects of "on-hold music" for a company that operates ~~on~~ a customer service line.

During a previous study, the ~~UC~~ researcher and his ~~colleagues~~ colleagues tested four types of ~~on-hold music~~ musical hold with 71 of the company's clients, ~~30 of them women~~, from Indianapolis, and Los Angeles. Light jazz, classical, rock and the company's current format of ~~adult alternative~~ ~~were~~ all tested. The sample included ~~ed~~ individual consumers, small business and large business ~~sectors~~ segments. Participants were asked to imagine calling a customer assistance line and being placed on hold. They were then exposed to "~~musical hold~~ on-hold music" via headsets and asked to estimate how long it played. Other ~~reactions and comments~~ were also solicited and quantified by the researchers.

Service providers, ~~of course~~, don't want you to have to wait on hold, but if you ~~did~~ did, they want ~~it~~ to be a pleasant experience for all of you. However, Kellaris' conclusions may hold some distressing news for companies. No matter what music of the four types of music was played, the time spent "on hold" was generally overestimated. The actual waiting in the study was 6 minutes, but the average estimate was ~~7 minutes~~. He did find some good news for the ~~client~~ line who hired him. He concluded the kind of music they are playing now, alternative, is probably their ~~better~~ best choice.

Two things ~~make~~ made alternative music a good ~~choice~~ choose. First, it did not produce significantly more positive or negative reactions in people. Second, males and females were less ~~different~~ polarized in their ~~reactions~~ to this type of music.

Kellaris' other findings, however, make the state of ~~on-hold music~~ musical hold a little ~~less firm~~. Time ~~spend~~ spent on hold seemed slightly ~~shortest~~ shorter than any other type of music when light jazz was played, but the effect of music format differed for men and women. Among the males, the wait seemed shortest when classical music was played. Among ~~the~~ females, the wait seemed ~~longest~~ longer when classical music was played. This may be related to the differences in ~~attention~~ attention levels.

In general, ~~classical music~~ evoked the ~~more~~ most positive reactions among male; light jazz evoked the most positive reactions (and shortest waiting time estimates) among females. Rock was the ~~least~~ less preferred across both gender groups and produced the longest waiting time estimates. "The ~~rock music's~~ driving beat kind of aggravates people calling a customer assistance line with a problem," said Kellaris. He concluded that the more positive the reaction to the music, the shorter the waiting time ~~is~~ seemed to be. So maybe time does tend to fly when you're having fun, even if you're on ~~musical hold~~ on-hold music and waiting for customer service.

Figure B.4: Music and consumer document with annotations.

Comments in docM (listed in document order):

John: Do we need the elevator music analogy?

John: I like we end the paragraph with a question.

Mary: Capitalize the first character for each word in the conference name.

Jen: I think UC refers to University of Cincinnati.

Jen: I thought 30 of them are men.

Mary: John, I think adult alternative means a mix of contemporary styles.

John: Do we need to include more details about the study?

Jen: I don't think "it" has a clear reference.

John: I think it should be 7 minutes and 6 seconds.

Mary: "Polarized" is a better word choice here.

Jen: I don't understand what we mean by "less firm."

Jen: I think musical preference is another factor.

Jen: I think it should be light jazz.

Mary: We meant adult alternative here, right, John?

Title: Warm Weather Boosts Mood, Broadens The Mind

The positive impact warm, sunny weather can have on mental health and mood are real if you get out for at least 30 minutes in the warm weather, according to ~~the~~new University of Michigan (U-M) research. Taking a trip to someplace ~~hot~~warm in the middle of winter or lingering outside when spring arrives can be especially ~~good~~beneficial, with pleasant weather improving mood, memory and broadening cognitive style (openness to new information and creative thoughts) as time spent outside increased, researchers ~~find~~found previously. However, they also found hotter weather during the summer, lowered mood levels and the effect of pleasant weather was far less noticeable ~~in other seasons~~.

"Being outside in pleasant weather really offers a way to re-set your mind-set," says Matthew Keller, the U-M post-doctoral researcher who ~~lead~~sled the psychology study. Everyone thinks weather affects mood but ~~the tests~~ of this theory in 2000 found no relationship, so we went back and found there are two important variables: how much time you spend ~~inside~~outside and what the season is. If you go from winter to spring and spend enough time outside, there's a noticeable change." The findings, completed by a team that also included U-M social psychology professors Barbara Frederickson and Oscar Ybarra, will be ~~soon~~ published in an upcoming issue of the Journal Psychological Science.

A set of ~~three studies~~ involved more than 600 ~~participants~~participants. In one study conducted during last spring in Ann Arbor, participants who ~~were~~are randomly assigned to be outdoors during ~~warm~~hot and sunny days showed improved mood and memory compared to participants who were outside when the weather was not ~~that~~ pleasant and compared to participants who spent the time inside.

The impact of weather on moods and cognition has been difficult to demonstrate because of people in industrialized countries, on average, spend 93 percent of their time indoors, making them largely ~~unrelated~~disconnected from the impact of changing weather outside.

Keller himself experienced the ~~phenomenon~~phenomenon and was inspired to conduct the research after leaving Michigan for a mid-winter trip to Mexico, quickly being reminded that "life is full of color and life is fun. It's easy to forget that during the wintertime." He notes that most people ~~feel~~felt less blue in the winter and better in the spring, with the most extreme example being among those who suffer from Seasonal Affective Disorder (SAD), a recurrent depression with typical onset during the fall or winter. Previous research has found a connection between ~~hot~~warm weather and violent behavior, and sunnier weather improving stock performance.

The researchers also found the optimal temperature for mood for most Americans is 72 degrees, about room temperature, with mood decreasing if temperatures became significantly higher or lower. There were regional differences, however, with mood peaking at 65 degrees in Michigan and 86 degrees in considerably warmer Texas.

For weather to improve mood, subjects needed to spend ~~30 minutes~~ outside in hot, sunny weather. Contrary to ~~their~~there initial expectations, researchers found that spending time indoors when the weather outside was pleasant actually decreased mood and narrowed cognitive style. They suspected ~~this~~ was perhaps because people resent being cooped-up indoors when weather becomes better, or perhaps because of improved weather can make normal ~~activities~~ feel boring or irritating.

The researchers note that it should not be surprising that weather and seasons ~~effect~~ human behavior, given that humans have evolved with seasonal and weather changes since the dawn of the species. Calling for further research into the subject, the researchers offer a straight forward prescription: "If you wish to reap the psychological benefits of good springtime weather, go outside."

Figure B.5: Weather and mood document (practice document) with annotations.

Comments in the practice document (listed in document order):

John: I think “beneficial” suits better.

Jen: I don’t think it’s clear here which seasons we are referring to.

Jen: We should add “Dr.” in front of his name.

Mary: Since we did find out the tests in 2000 are the biggest tests, I think we should indicate that here.

Jen: We need to include where the participants are from.

John: We need to add “at least” before 30 minutes to be consistent with the first sentence.

Mary: We need to enhance these are indoor activities.

Jen: We should use “affect” here.

Appendix C

Usability Study Task Sets

Before starting to review each document, participants were asked to imagine themselves in the following scenario:

Scenario: Imagine you are Bob, and you are one of the co-authors for the above document. Other co-authors are Jen, John, and Mary. This is your first time reviewing the document after the first draft of the document has been completed. Other co-authors have already reviewed the document and made annotations. Please complete the following tasks.

Note: In the following tasks, you will be asked to review groups of annotations. “Review” means to accept the annotations that you think are correct and reject the ones that you think are wrong either according to **English grammar** and/or **document context**. Some annotations may already been accepted or rejected by other co-authors.

Task instructions are assumed to be the same in both systems unless indicated otherwise.

C.1 Task Instructions in Black Hole Document

Task 1

Task Instructions

One of the co-authors has made annotations regarding quantifying words. Review these annotations. Accept the ones that you agree with and reject the ones

that you disagree with.

Click “Start Task” when you are ready to start.

Task 2 in the Simple System

Task Instructions

You have received an email from John. Review the annotations mentioned in John’s email. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 2 in the Bundle System

Task Instructions

You have received a general comment from John. Review the annotations mentioned in John’s general comment. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 3

Task Background

John ran the system spell checker during his turn reviewing. He corrected the spelling of some words according to the spell checker’s suggestions. The changes he made are embedded as edits in the document.

Task Instructions

Review the spelling edits John made. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 4

Task Instructions

Three of the co-authors have made annotations regarding verb tense in the document. Review these annotations. Accept the ones that you agree with and

reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 5

Task Background

“Grow” (and its different verb forms according to verb tenses) is a verb, and it should be used when we talk about an action. However, “growth” is noun, and it should be used when we talk about the process of growing.

During Mary’s previous reviewing session, she discovered there are some misuses of the two words, so she first ran “Find/Replace” function to find all the incorrect use of “grow” and replaced it with “growth.” Then, she ran “Find/Replace” again to find all the incorrect use of “growth” and replaced them with “grow.”

Task Instructions

Review the annotations regarding “grow” and “growth.” Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Before Task 6

You will need the following list of facts about the document to complete the next task. Please read them carefully before you proceed.

Facts:

1. The study of counting black holes relied on the deepest X-ray images ever obtained.
2. The distances to the X-ray sources were determined by optical spectroscopic follow-up at the Keck 10-meter telescope on Mauna Kea in Hawaii.
3. The biggest black holes ran out of “food” billions of years ago.
4. Whatever mechanism governs star formation in galaxies also governs black hole growth.

5. Black holes approximately 10 to 100 million solar masses took smaller portions of their meals of gas and dust.
6. It is the first time that we are able to count the black holes between the biggest, active black holes in distance and smaller, calmer ones closer to Earth.
7. NASA's Chandra X-ray Observatory made the best survey to date of black holes.
8. Previously, astronomers had done careful studies of the birthrate of stars in galaxies.
9. Dr. Amy Barger is from the University of Wisconsin and University of Hawaii.
10. Many of the black holes smaller than about 100 million suns are buried under large amounts of dust and gas.

Task 6 in the Simple System*Task Instructions*

Make sure you have a paper copy of the fact sheet before you start this task.

You have received an email from Jen. Review the annotations mentioned in Jen's email. Accept the ones that you agree with and reject the ones that you disagree with.

Click "Start Task" when you are ready to start.

Task 6 in the Bundle System*Task Instructions*

Make sure you have a paper copy of the fact sheet before you start this task.

You have received a general comment from Jen. Review the annotations mentioned in Jen's general comment. Accept the ones that you agree with and reject the ones that you disagree with.

Click "Start Task" when you are ready to start.

C.2 Task Instructions Music and Consumer Document

Task 1

Task Background

Adjectives can express degrees of modification by using their degree of comparison forms, namely positive, comparative, and superlative forms. Here are some simple examples:

Positive, Comparative, Superlative

Short, Shorter, Shortest

Long, Longer, Longest

Little, Less, Least

Much, More, Most

Good, Better, Best

The comparative form of an adjective is usually used when we compare two objects. The superlative form is used when we compare three or more objects.

Example:

Jerry is shorter than Tom. - *Comparing two objects*

Jerry is the shortest among all three cartoon characters. - *Comparing three objects*

Task Instructions

One of the co-authors has made annotations regarding adjectives' comparative and superlative forms. Review these annotations. Accept the ones that you agree with and reject the ones that you disagree with.

Click "Start Task" when you are ready to start.

Task 2 in the Simple System

Task Instructions

You have received an email from John. Review the annotations mentioned in John's email. Accept the ones that you agree with and reject the ones that you

disagree with.

Click “Start Task” when you are ready to start.

Task 2 in the Bundle System

Task Instructions

You have received a general comment from John. Review the annotations mentioned in John’s general comment. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 3

Task Background

John ran the system spell checker during his turn reviewing. He corrected the spelling of some words according to the spell checker’s suggestions. The changes he made are embedded as edits in the document.

Task Instructions

Review the spelling edits John made. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 4

Task Instructions

Three of the co-authors have made annotations regarding verb tense in the document. Review these annotations. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 5

Task Background

In this document, “musical hold” is used to describe the stage of holding with music playing in the background, and “on-hold music” is used to describe the

specific type of music discussed in the document. The study described in the document is trying to evaluate the effects of “on-hold music” on customers when they are put on “musical hold” after calling a company’s service line.

During Mary’s previous reviewing session, she discovered there are some misuses of the two phrases, so she first ran “Find/Replace” function to find all the incorrect use of “musical hold” and replaced it with “on-hold music.” Then, she ran “Find/Replace” again to find all the incorrect use of “on-hold music” and replaced them with “musical hold.”

Task Instructions

Review the annotations regarding “musical hold” and “on-hold music.” Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Before Task 6

You will need the following list of facts about the document to complete the next task. Please read them carefully before you proceed.

Facts:

1. Among the females, the wait seemed longest when classical music was played.
2. The effect of music format differed for men and women may be related to the differences in attention levels and musical preferences.
3. There are 30 female clients in the study described in paragraph
4. Males and females were less polarized in their reactions to alternative music.
5. Adult alternative is a mix of contemporary styles.
6. The actual wait in the study was 6 minutes, but the average estimate was 7 minutes and 6 seconds.
7. James Kellaris is from University of Cincinnati.

8. The rock music's driving beat kind of aggravates people calling a customer assistance line with a problem.
9. In the study, participants were asked to imagine calling a customer assistance line and being placed on hold.
10. Classical music evoked the most positive reactions among males.

Task 6 in the Simple System*Task Instructions*

Make sure you have a paper copy of the fact sheet before you start this task.

You have received an email from Jen. Review the annotations mentioned in Jen's email. Accept the ones that you agree with and reject the ones that you disagree with.

Click "Start Task" when you are ready to start.

Task 6 in the Bundle System*Task Instructions*

Make sure you have a paper copy of the fact sheet before you start this task.

You have received a general comment from Jen. Review the annotations mentioned in Jen's general comment. Accept the ones that you agree with and reject the ones that you disagree with.

Click "Start Task" when you are ready to start.

C.3 Task Instructions in the Practice Document

Task 1*Task Background*

Acronym is a word formed from the initial letters of a series of words.

Example: IEEE is an acronym for *I*nstitute of *E*lectrical and *E*lectronics *E*ngineers.

Task Instructions

One of the co-authors has made annotations regarding the use of acronyms. Review these annotations. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 2 in the Simple System

Task Instructions

You have received an email from John. Review the annotations mentioned in John’s email. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 2 in the Bundle System

Task Instructions

You have received a general comment from John. Review the annotations mentioned in John’s general comment. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 3

Task Background

John ran the system spell checker during his turn reviewing. He corrected the spelling of some words according to the spell checker’s suggestions. The changes he made are embedded as edits in the document.

Task Instructions

Review the spelling edits John made. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 4

Task Instructions

Three of the co-authors have made annotations regarding verb tense in the document. Review these annotations. Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Task 5

Task Background

In the document, “warm” should be used to describe pleasant weather, whereas “hot” should be used to describe the unpleasant summer weather.

During Mary’s previous reviewing session, she discovered there are some misuses of the two words, so she first ran “Find/Replace” function to find all the incorrect use of “warm” and replaced it with “hot.” Then, she ran “Find/Replace” again to find all the incorrect use of “hot” and replaced them with “warm.”

Task Instructions

Review the annotations regarding “warm” and “hot.” Accept the ones that you agree with and reject the ones that you disagree with.

Click “Start Task” when you are ready to start.

Before Task 6

You will need the following list of facts about the document to complete the next task. Please read them carefully before you proceed.

Facts:

1. Matthew Keller is a post-doctoral researcher at University of Michigan.
2. A set of three studies conducted by Keller and his colleagues involved more than 600 participants from throughout the United States.
3. For weather to improve mood, subjects needed to spend at least 30 minutes outside in warm, sunny weather.
4. The researchers note that it should not be surprising that weather and seasons affect human behavior, given that humans have evolved with seasonal

and weather changes since the dawn of the species.

5. The tests that were conducted in 2000 on whether weather affects mood are the biggest tests so far in the theory.

Task 6 in the Simple System

Task Instructions

Make sure you have a paper copy of the fact sheet before you start this task.

You have received an email from Jen. Review the annotations mentioned in Jen's email. Accept the ones that you agree with and reject the ones that you disagree with.

Click "Start Task" when you are ready to start.

Task 6 in the Bundle System

Task Instructions

Make sure you have a paper copy of the fact sheet before you start this task.

You have received a general comment from Jen. Review the annotations mentioned in Jen's general comment. Accept the ones that you agree with and reject the ones that you disagree with.

Click "Start Task" when you are ready to start.