Wikum+: Integrating Discussion and Summarization in Collaborative Writing

by

Sunny Tian

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Author .....
Department of Electrical Engineering and Computer Science
May 12, 2020

Certified by .....
David R. Karger
Professor
Thesis Supervisor

Accepted by .....
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

Wikum+: Integrating Discussion and Summarization in Collaborative Writing

by

Sunny Tian

Submitted to the Department of Electrical Engineering and Computer Science
on May 12, 2020, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

In many instances of collaborative writing, ideation and deliberation about what to write happen in a separate space from the actual document writing. However, having discussion and writing separated may result in a final document that has little connection to the discussion that came before. Furthermore, online users often rely on filtering, voting, and moderation to manage lengthy conversations. In this work, I build upon a hybrid discussion and document-writing tool called Wikum+ to allow groups to mix having discussions and summarizing those discussions in real-time, until the process results in a final document that incorporates and links all discussion points. The system uses collaborative summarization interchanged with deliberation to synthesize the conversation into a meaningful artifact, that can be iterated and improved upon. I conducted a within-subjects user study of 6 small groups where each group used both Wikum+ and a control of Google Docs and a messaging app to collaboratively write proposals. I also conducted a between-subjects user study of 2 larger groups, with the control given only a Google Doc. From analyzing survey and interview results, I found evidence that Wikum+'s integration of discussion and summarization helped users be more organized as well as more inclusive of ideas, leading to a more comprehensive final document. Compared to a control, Wikum+ also allowed for more light-weight coordination and iterative improvements through the incorporation of new ideas.

Thesis Supervisor: David R. Karger
Title: Professor

Acknowledgments

This work would not have been possible without the advice, support, and encouragement of many.

Amy, thank you for teaching me the ins and outs of research, from design to ideation to evaluation. Thank you for encouraging me to submit to conferences and competitions that I never would have even known about, for sending me exciting new literature, for providing guidance on my papers and presentations, and for cheering me on every step of the way. You have been such an amazing mentor and I feel so fortunate to have had the opportunity to work closely with you on Wikum.

David, thank you for your support and guidance throughout my entire time with the Haystack group. It means so much that you take the time to discuss design choices and big-picture goals with me. Your insights always teach me something new and your guidance throughout this work has been incredibly helpful. Thank you for your endless wisdom and kindness throughout this semester.

To the Haystack group, thank you for putting up with my weekly requests to test Wikum's new features, no matter how buggy. Your feedback on the interface and my presentations were so valuable. Thank you for being such a welcoming group and making me always feel so comfortable.

To my friends and the many students who took a part in the user studies of Wikum+, thank you for your participation and willingness to help.

Finally, to my mom, my dad, my brother, and Andy, thank you for all of your love, encouragement, and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cooperation and the sharing of ideas in team production have been shown to have social benefits for groups as well as better and more innovative outcomes [13]. Furthermore, the perceived quality of collaboratively written documents is higher [3], overall productivity increases [10], and interpersonal relationships are enhanced [40]. Today, online discussion platforms provide a liberating medium for groups to collectively interact, ideate, and deliberate across time and distance. However, a major challenge for collective deliberation that is exacerbated in the online setting is refining and combining the mountain of ideas into something meaningful [4]. Large groups tend to develop large amounts of shallow and redundant ideas and have trouble deciding on and selecting a set of good ideas [4, 30]. Moreover, interesting comments and the outcome of deliberations can be buried deep in or scattered throughout online discussion threads, whereas other comments may be controversial, off-topic, or problematic [30]. Users must sift through repetitive and low quality ideas, worsening the creativity of their own contributions and their overall productivity [42].

We introduce the idea of using summaries as a form of synthesis, integrated into the discussion process. Summarization provides a way to synthesize, combine, and collapse part of a discussion into something more concise and conclusive. This reduces efforts of readers and decreases the likelihood of redundant discussion threads, allowing for people to focus on other important unresolved points. The integration of summaries with discussion on a single platform helps readers cross-reference the con-

versation and understand how conclusions were made, only if necessary. By making summary-creation and editing collaborative, we allow for a greater sense of inclusion and a lower chance of bias from select moderators. Additional features such as flags or marking certain comments unsummarized allows minority voices to brings attention to unaddressed issues.

To address document-writing issues that occur with asynchronous collaboration, we use color-coded comments and unread markers to bring more attention to both new and unincorporated comments. This allows for new ideas to hold presence and spark discussion even in seemingly complete documents. As a result, discussions no longer need to be append-only, but can shrink as conversation is synthesized into summaries, and grow as novel discussion springs later on.

By making the summaries collaborative in nature, we aim to reduce the friction of editing summary content created by others. By having contributors working towards the common goal of synthesizing discussion and introducing features to help indicate areas that need review, we hope that users feel more comfortable making updates independently to the summary or document.

Summarization as a form of synthesis is useful because the current tools people use online don't allow for an easy way to synthesize ideas, instead relying more on filtering, voting, or moderation. While these strategies can help reduce the issues at hand, they don't solve them completely. For instance, Redditors may up-vote several top comments that have very similar content, again, leading to redundancy. A controversial sub-discussion topic may have reached a conclusion somewhere in the conversation, but due to the append-only nature of typical conversations, a moderator or reader must wade through the entire sub-topic to find the resolution.

Unfortunately, these strategies can also often lead to biases, as minority ideas are pushed down or filtered out. Moderators can unknowingly select ideas aligning more closely to their opinions, and too much moderation can reduce the overall perceived level of fairness [35]. Furthermore, by only using votes can result in a lack of justification for promoting certain ideas and even if the reasoning lies in the comments, there is no way to directly connect them. Voting systems also have no capability for

more sophisticated *refinement* to improve an idea or *synthesis* of multiple ideas. For instance, some competing ideas may be impossible to combine while others become stronger together.

These issues all make the task of writing up a cohesive plan or document from a discussion more difficult. In addition, the lack of synthesizing capabilities on common online discussion platforms result in users having to move to document-writing software. By migrating away from where they had discussion, users lose the ability to directly reference conversation or easily incorporate what people said. Some software such as Wikipedia have space for discussion but it lives on a separate page and cannot be referenced in the main document [18]. Additionally, collaborative document-writing tools like Google Docs have features for chatting and in-line commenting. However, there is no easy way to connect chat messages to document edits, or to start from a discussion and build to a document section.

When working collaboratively on writing the document, new content is detectable when everyone is working in real-time. But when people must work asynchronously, new changes that are introduced can be difficult to find for returning users, especially when the document becomes lengthy. If a contributor wants to introduce a new idea, they may be wary of incorporating those new concepts without discussing with the others first, and thus opt for not introducing new content at all. Furthermore, because discussions are less likely to happen inside of a document, a contributor who wants to make edits to a portion written by someone else may avoid doing so, not wanting to over-step boundaries without conversing first.

In this thesis, we present Wikum+, a system that integrates collaborative discussion and summarization so that users can collectively create a document synthesized directly from their deliberation in real-time. We allow for users to move seamlessly between the phases of conversation and document-writing, and the ideas formed in the comments maintain their provenance as they are summarized and surfaced into the final document. Furthermore, continued discussion after summarization allows for users to discover topics requiring further deliberation, conversations that lack inclusion, or summaries that are incomplete or biased. By making these comments

discoverable, Wikum+ allows for iterative improvements, ultimately bettering the quality of the final document. A pilot study of our initial interface provided us with insight on usability improvements and preliminary results, allowing us to improve upon both our interface and the design of our study. After upgrading Wikum+ to the system it is today, we tested its effectiveness in achieving our goals by running two user studies. The first study was a within-subjects experiment involving six groups of 5-6 users, with each group using both Wikum+ and a control. Our results show that users found Wikum+ to provide significantly more organization, and that users found it significantly easier to locate both new ideas and topics that still needed work on Wikum+. This allowed for iterative improvements on the summaries and final document. We also found that users generally were more concise on Wikum+ while users wrote more fluidly in the Doc. Our second study evaluated how Wikum+ would perform compared to a control if the groups were larger in a between-subjects experiment. Our results show that Wikum+ users found the process to be significantly more inclusive of their ideas and rated their final document as significantly more comprehensive compared to the control. Overall, our studies provide evidence that Wikum+'s integration of discussion and summarization helped users be more organized as well as more inclusive of ideas in the final document. Wikum+ also allowed for more light-weight coordination and iterative improvements through the incorporation of new ideas.

This thesis details the design and implementation of Wikum+, along with findings from the pilot study and larger user study. Chapter 2 summarizes related work; Chapter 3 details the technical implementation and system design of Wikum+; Chapter 4 describes the pilot study, both user studies conducted, and their results; Chapter 5 discusses the findings and suggestions for other discussion systems while explaining potential future work; and Chapter 6 concludes the thesis with a summary of our contributions.

Chapter 2

Related Work

2.1 Collaborative Decision Making

Online deliberation involves discussions with the use of information and communication technologies. Compared to typical discussion forums, these processes focus more on specific, important topics, with the goal of mutual understanding. As mentioned in the Introduction, one area that online deliberation has shown more presence in is the political sphere. This has allowed citizens of a community or city to discuss issues or objections and propose change for policy makers [27]. In such contentious situations, one might believe that heavy moderation is needed, but research has shown that higher levels of moderation actually negatively impacts the overall perceived fairness of the procedure, validity in claim, as well as policy legitimacy [35].

Moreover, although opinion heterogeneity may be perceived as a potential source of harm for harmonizing opposing voices, exposure to different opinions allows people to better support their own positions, leading to better argumentation [37]. It has also been suggested that high heterogeneity with mixed opinions are important for avoiding positive or negative perception bias in the discussion process [35].

2.2 Collaborative Writing

Collaboration in writing has been shown to help authors pool diverse ideas, provide feedback, and write better texts in terms of task fulfilment, grammatical accuracy, and complexity [46]. Furthermore, authors writing collaboratively generally improve the overall accuracy of their resulting work compared to single-author works [20, 14]. However, trust and willingness to share authorship are crucial for collaborative writing to succeed [45]. In online collaborative environments, writers that share common habits and hold less author-centric perspectives of textual ownership perform better in group writing [17]. Research further suggests that writers involved in collaborative projects may focus more on the product than the process [46]. When writers are pressed for time, this may result in users quickly voting on the best ideas to incorporate and leaving out important minority ideas that were simply voted out.

Today, technologies that allow for more convenient feedback and faster response times can increase motivation and creativity [24]. One of our goals with Wikum+ is to provide an interface that allows for real-time conversation to merge with collaborative writing in a more organized form than traditional free-for-all online document-writing spaces, which are less conducive to discussion-based feedback. We design Wikum+ as a system with similar functionalities to discussion forums, including commenting, tagging, mentioning, and notification abilities. However, Wikum+'s collaborative summarization process ties the interface back to document writing in a more guided fashion, as users are provided with rich text editing tools and have the ability to move text around on the page. This synthesis process occurs on a thread-by-thread basis, gradually building up a group's final document.

2.3 Creativity and Diversity in Ideation

Seeing high quality, inspirational ideas allows individuals to come up with more diverse and creative ideas [42]. Learning about new ideas also helps accelerate the generation of ideas across various categories, improving overall productivity [32]. How-

ever, exposure to repetitive and simple ideas can actually hinder creativity, and when considering discussions at scale, we often find that there are a multitude of these non-productive ideas expressed throughout the conversation [19, 23]. The main problem with this is that people tend to generate or borrow ideas based on the ideas they are exposed to [23, 44, 28]. Thus, if all of the presented ideas take up a very narrow portion of the solution space, this could actually hinder creative ideation rather than help it — a process known as *design fixation* [19].

Therefore, the importance of highlighting and helping users notice useful, diverse, and creative ideas is heightened for discussions at scale, and part of that process includes hiding or minimizing less useful comments. Other existing solutions involve creating visualizations and algorithms to help users recognize the limitations of their generated idea space [15, 8, 36, 42, 43]. Wikum+, on the other hand, provides a tree structure to visualize the commenting structure and flow, as well as a clustering view to help group similar comments together based on their textual content. The tree structure is dynamic, growing and shrinking as the conversation and summarization progresses, instead of existing as a post-discussion analysis tool. Wikum+ also has an option to mark certain comments as unimportant, thus minimizing less useful comments and increasing the likelihood of exposure to more diverse and productive ideas.

2.4 Making Sense of Online Conversation

The most widely used practice for reducing and managing large conversation volume is filtering, either through selected moderators [25] or some form of voting or flagging mechanism [9]. Unfortunately, moderation and filtering sometimes lead to the suppression of opinion minorities and the moderators, themselves, may unintentionally introduce implicit personal biases.

Personalization features can help reduce content load as well, but also risks biases and "filter bubbles" — a situation where only one point of view is represented [33]. While automatic natural language filtering exists, it is typically used for detecting

spam [29] and trolls [6].

Even with filtering and moderation, however, a high quantity of people can simply mean a high quantity of ideas. Thus, Wikum explores using summarization to not only filter out unimportant or repetitive comments, but also synthesize major ideas. Existing automatic summarization techniques focus mostly on extracting important sentences from the text [31], and are primarily made for summarizing long documents, email chains, or dissociated reviews. By using human input, Wikum+ differs by allowing for the possibility of synthesizing points made, such as deciding on the best solution of many based on consensus among users. Furthermore, all automatic summarization systems require massive sets of training data to obtain sufficient accuracy [41], and currently, these do not exist for discussion summaries. Similar human-input based summarization systems exist as discussed below [30, 1], but exist only at the top level of discussion, summarizing the discussion as a whole, and only analyze archived conversations rather than on-going discussions as well.

2.5 Relevant Systems

Many systems have come into place for online collaborative deliberation. However, most of these systems polarize comments by having a voting system or a methodology to determine a 'winning' side [11, 26, 21, 8]. This not only shuts down or hides valid ideas, but also results in a large number of the participants feeling like they've lost something that could have been collaborative. Furthermore, many platforms allow for little or no discussion due to the structure and purpose of the system [36, 2, 38]. On the other hand, many platforms allow for discussion, but either rely on a single summarizer [5], rely on moderators that select representative comments [7], rely on visualizations to collect ideas [36], or rely on polls and surveys based on the discussion to decide on a specific topic [26].

Some efforts, including the original Wikum interface, have been made to incrementally distill large quantities of messages into more useful, succinct, and durable summaries [30, 1, 47], but these systems focus on stagnated or archived conversations

and do not explore the possibility of continuous and simultaneous conversation in addition to summarization. Wikum+ also differs by allowing users to expand sub-summary discussions and explore the details of subtopics rather than providing a "flat" set of top-level summarizations. By perusing through the topics discussed previously efficiently in Wikum+, a new user can become inspired and add new, innovative ideas, keeping the conversation alive but still focused.

Thus, Wikum+'s novelty exists in its ability to both synthesize conversation and continue discussion, with the objective of improving the quality of the final summarized document, as well as the experiences of the users participating in the deliberation. In the next chapter, we detail the system's implementation and the specific features that significantly distinguish it from the original Wikum.

Chapter 3

System Design and Implementation

3.1 Wikum+

In this work, I present a system that allows for the steps of collaborative ideation, deliberation, synthesis, and document-writing to happen in one space, so that users can move seamlessly from one step to another and back as necessary, and ideas maintain their provenance as they move from original comments all the way to their incorporation into the final document. The system is developed on top of an open source software tool called Wikum [47].

On the original Wikum, users can import a concluded conversation that has happened elsewhere, such as Reddit or Disqus, and collaboratively summarize it in a *recursive* fashion, by summarizing small portions of discussion and then summarizing the summaries, until the entire discussion is summarized into a *summary tree* (Stages (b) through (d) in Figure 3-1). This explorable summary tree allows users to quickly identify and explore the main topics of discussion [47]. Furthermore, the summary tree's default view collapses the contents of full summaries, reducing the number of words shown on the page and shrinking the conversation.

However, the original Wikum assumes that all discussion on a topic is completed before any summarization commences. Users could not add new comments and could only summarize the existing imported discussion. But it is common for collaborations to require multiple stages of deliberation, action, and then reflection and further

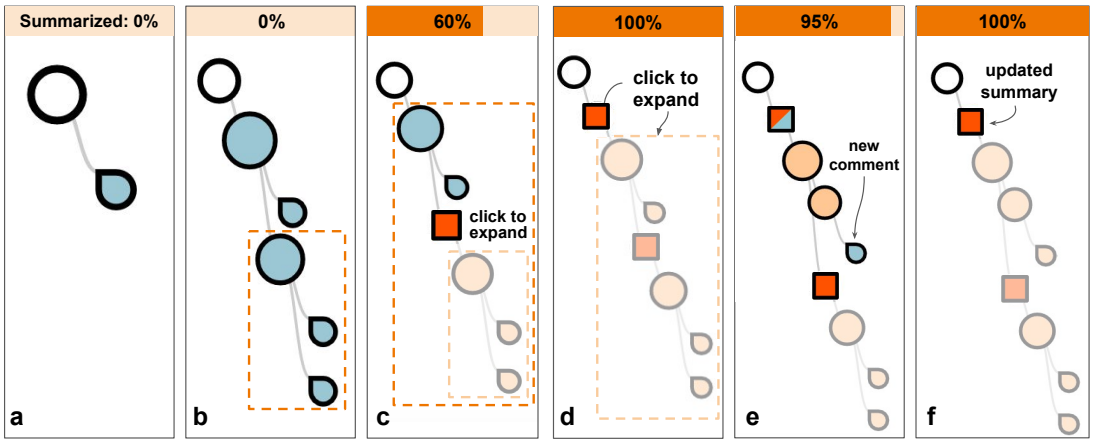

Figure 3-1: Wikum+ discussions grow with discussion and shrink with summarization. By default, only summary nodes and unsummarized comments are shown, unless summaries are clicked-to-expand. A typical workflow is as follows:

a–b Users contribute new comments and replies.

b–c Users can summarize select comments, adding an orange summary node and changing the blue (unsummarized) comments to yellow (summarized).

c–d Summaries can be added at a higher level that include other summaries recursively.

d–e Users can add new comments at any point. Summaries above such comments turn half-blue, half-orange to signify partial summarization of comments below.

e–f Users can edit the partial summary to incorporate new comments.

Stages d through f loop until the summary reaches a stable state.

deliberation of those actions. New ideas and re-considerations may also come at different times. For instance, after ideas have been synthesized into a summary, users may generate more creative ideas and hope to discuss these new ideas before incorporating them. Some users may even want to discuss the quality or content of the summary itself. These new comments are included in the default view of the page, therefore increasing the discussion length and expanding the conversation. We refer to this process of interchanging discussion and summarization as *interleaving summarization*.

In our design of Wikum+, we hoped to improve both the quality of discussion by

encouraging deliberation and inclusiveness, and the quality of the final document by enabling iterative improvements through interleaving summarization. In the following section, I detail the new features I've implemented on Wikum+ to help us achieve our goals.

3.2 The New Features of Wikum+

3.2.1 New Comments.

The first, most crucial feature of Wikum+ introduced is the ability to add new comments. Comments can be added as replies to existing comments, replies to existing summaries, or as brand new comments, starting a new discussion thread. Along with this functionality came the ability to start a new Wikum+ page from scratch, as previous Wikum pages could only be created through the importation of existing discussions from other platforms.

By allowing for replies to summaries, summaries can become outdated, as the summary content may not incorporate the new comments (Figure 3-1 (e)). Furthermore, comments that are considered summarized should be distinguished from comments that are not summarized to indicate which comments still need to be synthesized or integrated.

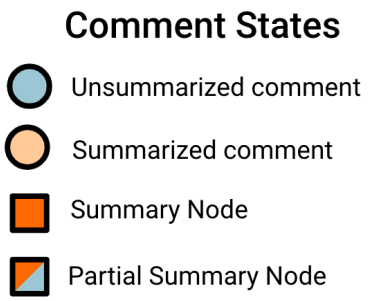

Figure 3-2: Comment States in Wikum+

We use squares to represent summary nodes and circles to represent comments. We use the color blue to represent unsummarized comments and yellow to represent summarized comments. Full summaries are orange squares. Then, summary nodes with unsummarized comments below them are squares colored half-orange and half-blue to symbolize a partial, unfinished summary requiring updating (Figure 3-2).

Summaries in Wikum were already editable, but I added an additional feature

so that users can evaluate summaries to indicate which children comments the summary incorporates. Based on this evaluation, colors of the child comments update accordingly and any positive number of unsummarized children comments leaves the summary node in a partial summary state.

3.2.2 Permissions.

There are instances where we may want to restrict the level of access a user has on a Wikum+ page. Enforcing an initial discussion period of comment-only permissions could increase the quantity of distinct, creative ideas. Enforcing moderator roles with greater edit access may lead to more cohesive final summary documents.

While we considered enforcing a certain minimum of characters or time to pass from the creation of a Wikum+ page before allowing summarization and other editing functionalities, we decided that this minimum varied too much by topic and relied on a variety of factors, such as how controversial or multi-faceted a subject is. Instead we grant the owner or creator of a Wikum+ page authority to control the permissions for their page at the public and private level, much like Google Docs permissions. The owner is always given full edit and comment access. At the public level, owners can set the page permissions to be one of five settings:

1. **Private Access:** Only the owner can access the page.

2. **Publicly Viewable:** No one can edit or comment, but anyone can view the page.

3. **Publicly Editable:** Anyone can view the page and perform editing functionalities including creating or editing summaries, hiding comments or summaries, tagging comments or summaries, evaluating summaries, moving comments around, and marking comments as summarized or unsummarized. Adding new comments is not publicly allowed.

4. **Publicly Commentable:** Anyone can view the page and add comments, but no editing functionalities are publicly allowed.

5. **Publicly Editable and Commentable:** This is the full access setting, combining Publicly Commentable and Publicly Editable allowances.

At the private level, owners can grant specific users one of four settings, each corresponding to its respective public level functionalities:

1. View Access

2. Edit Access

3. Comment Access

4. Edit and Comment Access

These private level permissions are in addition to the public level access settings. For instance, if a page is Publicly Commentable and a user is granted private Edit Access, they will have full edit and comment access to the page.

3.2.3 Outline View.

The original comment and summary visualization existed as a *summary tree*, depicted in circular nodes, as shown in Figure 3-3. While this tree was helpful for gaining insight into the structure of the conversation, the lack of textual context made it more difficult to navigate and find topics of discussion.

As I explain more thoroughly in Section 4.3, some pilot study users expressed frustration in being unable to easily find the comments they had written. Some users said that they had a hard time scrolling through all of the comments on the page to find the different topics discussed. When asked if an textual outline view of the comments, as shown in Figure 3-3, would be helpful in the issues they raised, all of the Wikum+ users agreed.

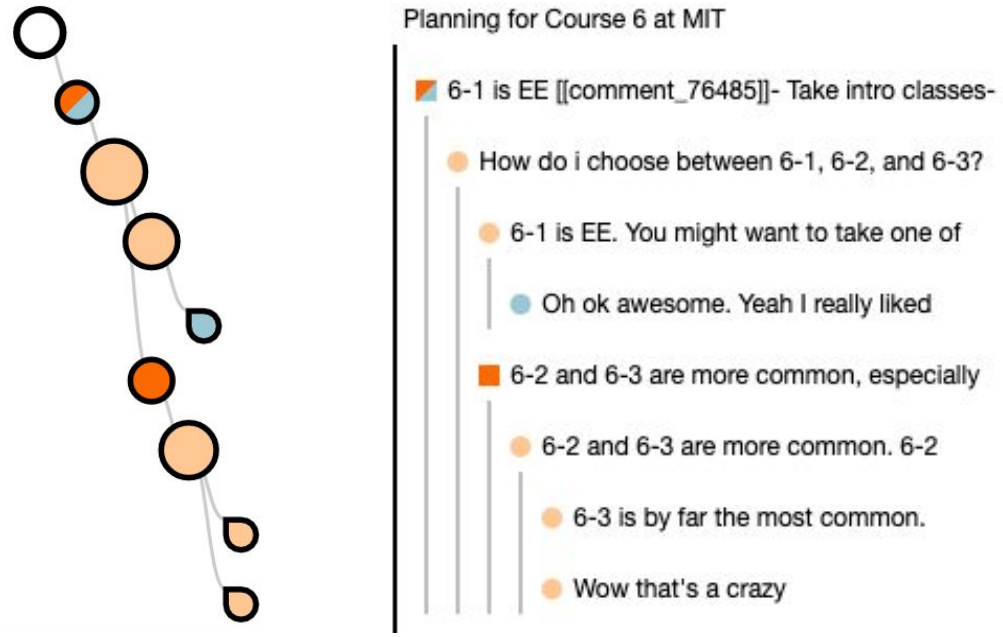

Figure 3-3: The left shows the original summary tree visualization in Wikum. The right shows the new textual outline view visualization in Wikum+. The outline view truncates each comment or summary text to the last word possible, given a max length which is determined by the depth of the comment in the conversation. The text at the top is the title of the Wikum+ page. On page-load the outermost black line is highlighted, showing the entire discussion.

In creating the outline view, we wanted to maintain the threaded tree-like structure from the original visualization. We also wanted to design the outline view so that the dragging and dropping of comments would be easier and more visible. We thus drew ideas from both Google Doc's outline sidebar and Reddit's threaded forum UI to create Wikum+'s outline view (Figure 3-3). While the original summary tree visualization encoded the size of a comment to the size of the circle, we chose to standardize each shape to the same size to reduce bulkiness, especially with the heavier emphasis on the text.

Just as in the original Wikum, full summaries collapse the content they encapsulate on page-load and the words shown on a Wikum+ page, by default, contain unsummarized content, summaries, and partial summary threads. Therefore, portions of the discussion that have reached a conclusion can be summarized and condensed, shrinking the conversation. As shown in Figure 3-3, users see the top-level discussion

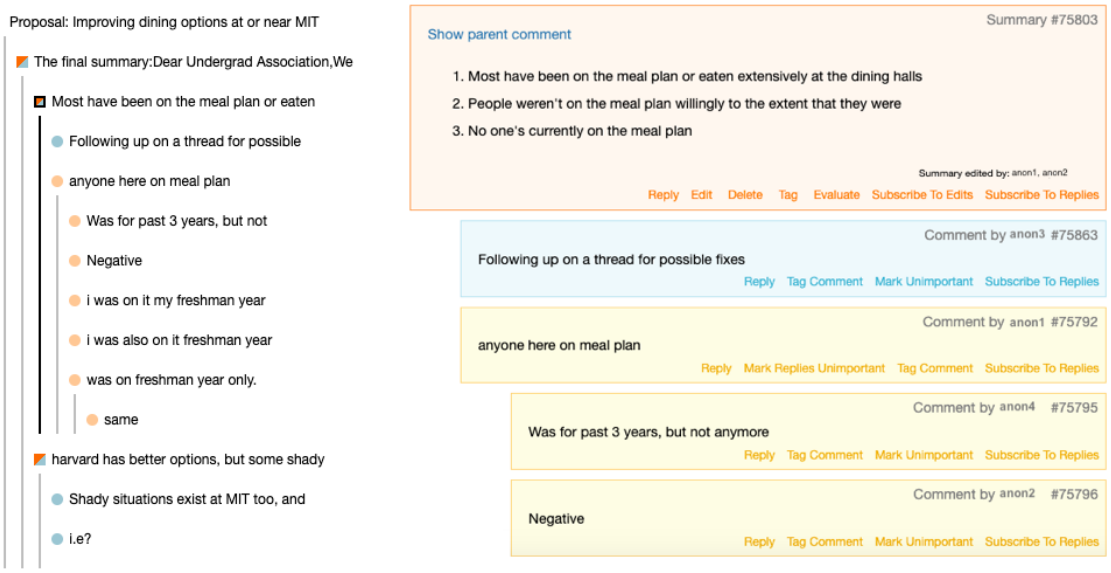

Figure 3-4: Comment threads can be selected, showing only the comments or summaries in that thread on the right. Selected comments and threads in the outline view are highlighted in black.

or outermost thread by default, indicated by the black outermost highlighted line in the outline view. However, if a user selects any specific comment or summary, its corresponding thread will be highlighted (Figure 3-4) and only that threads comments will be visible on the right. This view is also achievable by clicking on the thread line itself. Then, clicking any surrounding whitespace or the title text will show the default, full view again. Just as summaries can be expanded by default, any thread in the outline can be collapsed or expanded by clicking the particular thread twice.

The outline view acts not only as a condensed version of the comments, it also functions as a navigation tool for exploring the conversation. By hovering over any comment in the outline view, the right hand side will scroll to the corresponding comment if it is in the selected thread. On the other hand, a user can quickly understand the context of a comment by hovering over a comment on the right, which will scroll the outline view to the corresponding element on the left, allowing the user to scan surrounding comments efficiently.

In some cases, it makes sense to group separate threads or select related comments to synthesize into a single summary. To do this in the outline view, users can hold

the "command" key (on Macs), or the "Ctrl" key (on Windows and other PCs), while selecting the comments or summaries to group together. Previously this Summarize Selected feature was done by clicking, dragging, and selecting a group of comments on the left. Due to the introduction of drag-and-drop (described below), this dragging method of selection may unintentionally drag and move a thread instead of selecting a set of comments. In the future, we may bring back the drag-to-select feature by distinguishing these two functionalities with an additional key-press. For instance, perhaps drag-and-drop should only occur if the Shift key is held down while drag-to-select can happen without any key-press.

3.2.4 Read Markers.

Consider a synchronous setting where one user is working in a discussion thread near the top of the conversation, while another user is working in a discussion thread near the bottom. The users may want to see what other comments had been added while they were busy with their respective parts. Consider an asynchronous setting where a user has finished adding their thoughts to a lengthy discussion for the time being and leaves the page for a day. When they come back to the page, some other users may have left new comments deep in a discussion thread. In either scenario, an indicator for which comments the user has and hasn't seen would benefit the user greatly. Users could scroll up or down the page and find the unread indicators and read the new comments, or if they left the page, could come back to the page and find the deeply embedded comments, simply by scanning for the unread indicators.

In designing these unread indicators, we borrowed ideas from emails, where unread emails have their preview text bolded. Similarly, we bold the text of outline view comments that have not yet been read by a user. Just as email preview text becomes unbolded upon click, clicking on any comment's outline view text will also unbold that text, marking it as "read." Furthermore, if any comment box on the right has been in view for 5 seconds, or if a user hovers over the comment box for one second,

it is also marked as read and unbolded in the outline view. Finally hovering over any outline view text will mark a comment as read after one second, if the corresponding comment box is in view.

Upon a user's first time visiting a page, we automatically mark all comments as read so that no comment is in bold. Therefore, the only comments that are marked in bold for a user are comments that have been added after their initial visit. This prevents a user from being overwhelmed by the unread markers upon visiting existing Wikum+ pages, while still serving their intended purpose of highlighting new comments added.

3.2.5 Drag-and-Drop: Giving Order.

One of our primary goals in creating Wikum+ is to improve the quality and experience of collaborative writing. Since Wikum was originally built for imported discussions, the default order of the comments was simply their existing order in the source they originated from. However, with the introduction of new comments and to give Wikum+ the same functionalities that documents provide, we allow for dragging and dropping through the outline view for a persistent, saved ordering of comments and summaries (Figure 3-5). In particular, a discussion in the reply thread of a post may head off in a direction that is more suitable in a separate thread or under a different, existing thread. Users would require the ability to move threads around in a persistent manner.

Sorting (by import order, date created, number of replies and text length) remains a functionality, but these are only temporary front-end sorts. The default saved view is the persistent drag-and-drop ordering, and when the page is sorted in a different order, the drag-and-drop feature is turned off. Note that when a comment or summary is selected and dragged, its reply thread will also be dragged with it. This is because we assume that replies are relevant to a particular comment or summary. Therefore, if a Comment object belongs elsewhere, its replies should likely follow. Users are,

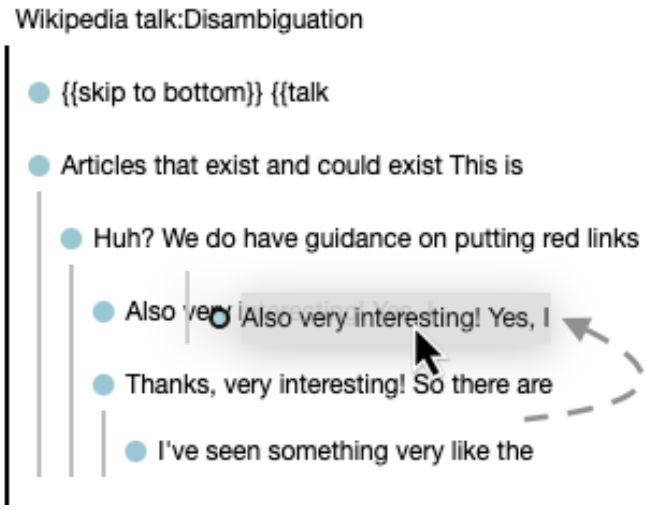

Figure 3-5: Drag-and-drop in the outline view

however, given the option to drag child Comment objects out of a reply thread or any Comment object into a reply thread. For a Comment object with no children, dragging another Comment object under it will create a new reply thread.

I chose to implement this feature through doubly linked lists, with each Comment object (including both comments an summaries) pointing to its parent, its first child, its last child, its previous sibling, and its next sibling. The Comment objects at the top level point to no parent and the Article object points to the first Comment object, as well as the last Comment object. When a Comment object moves, its own pointers, its old siblings' pointers, its new siblings' pointers, potentially its old parent's first and last child pointers, and potentially its new parent's first and last child pointers must update. When a summarized comment thread under some summary node is moved to a different parent, the new parent may not be a summary. Even if the new parent is a summary, its content may not incorporate the contents of the moved comment thread. Therefore, the comments in the thread that are not contained in a sub-summary become unsummarized. Additionally, if the new parent was a full summary node, it becomes a partial summary (Figure 3-6).

Many back-end functions, including adding new comments, summarizing a comment and its replies, summarizing selected comments, and deleting a comment or a group of comments, all require updating the pointers of surrounding comments and

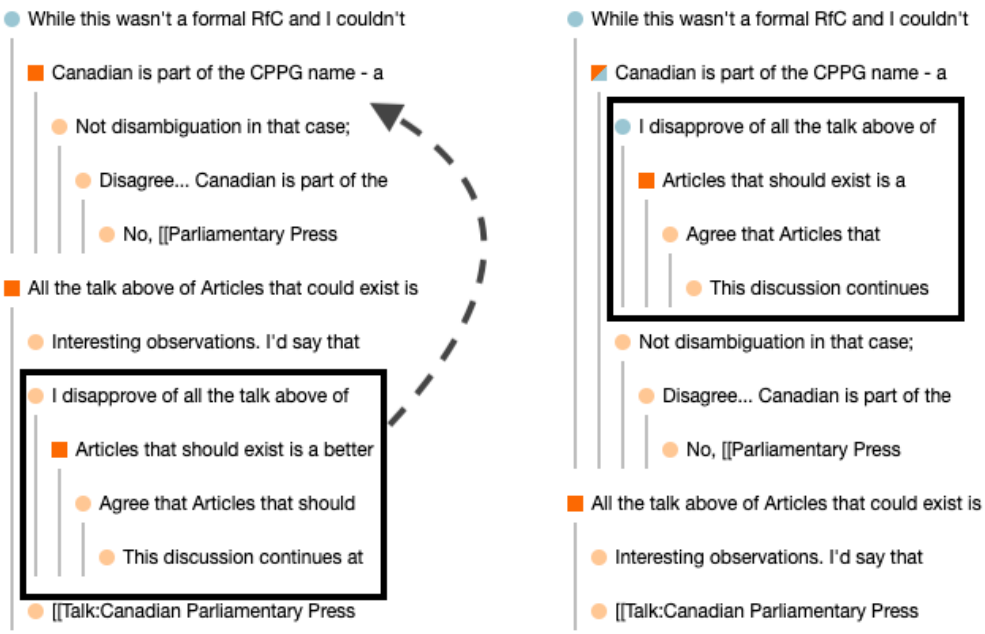

Figure 3-6: When a comment thread is dragged to a new parent, we check to see if the nearest summary above has changed. In this case it has changed from "All the talk..." to "Canadian is part...", so we automatically mark the moved comments that are not contained in sub-summaries as unsummarized (blue). There is only one comment of this type here: "I disapprove of...". The nearest summary above the new location of the dragged comments (i.e., "Canadian is part...") then becomes a partial summary.

summaries. These pointers must be maintained so that they can be traversed when displaying the Wikum+ page.

An additional complication that arises with large discussions is pagination. For efficiency and readability reasons, we truncate a single page to have at most 30 top-level comments. Thus a long discussion with many top-level comments will span multiple pages, which can be accessed with the "Next" and "Previous" buttons. I take care not to rely on the front-end for obtaining pointer values, as sibling pointers might not always exist on the page. For instance, if the Wikum+ has multiple pages, the last Comment object on the first page will not have its next sibling displayed because that Comment object is on the next page. Thus, pointers are obtained almost entirely from the back-end.

3.2.6 Real-time Updates and Locking.

Wikum originally only allowed for asynchronous collaboration in creating summaries. With the introduction of discussion in Wikum+, we found that synchronous updates were necessary for real-time collaboration. Due to Wikum's Django back-end framework, we decided to implement real-time updates with Django Channels. Extra implementation details are found in Section 3.3.

There were two primary race conditions we faced with real-time updates. One issue was that multiple users might try to summarize overlapping sets of comments or edit the same summary. To prevent this from happening, I introduced real-time locks that prevent another user from trying to create a new summary from comments currently in use and disables the Edit Summary button for others when a user is updating a summary.

It is possible that a user begins to edit or create a summary but then forgets to finish, leaving the page idle for 15 minutes or longer. Another possibility is that a user exits the page, either voluntarily or involuntarily. In each of these scenarios, the lock will be released and a different user can acquire the summary lock. Due to efficiency issues, a future goal of Wikum+ is to allow for synchronous editing of the same summary, similar to that of Google Docs.

The second issue was that multiple users may try and drag the same comment, one user may try and drag one comment while another user drags a parent comment, or one user may try and drag one comment while another user drags a child comment simultaneously. Due to the doubly linked-list back-end pointers discussed above, this race condition could result in problematic linking of comments such as cycles (e.g., Comment A's parent pointer pointing to Comment B, whose parent pointer points back to Comment A). Under the assumption that a single drag-and-drop action should take no longer than a couple of seconds and that the action would not have frequent use, we created a drag-and-drop lock for the entire set of comments so that only one user can move any comment at a given time. The moment a user releases or drops the comment, another user can acquire the drag-and-drop lock and begin to move

comments.

3.2.7 Rich Text Editing.

In Wikum+, discussion happens in comments, but document writing happens primarily in summaries. I integrate rich text editing into summary nodes to match document writing more closely, allowing for a greater range of writing functionality.

3.2.8 Notifications and Comment Links.

One of the most requested features from the pilot study (Section 4.3) was for Wikum+ to have notifications. Some users wanted to be able to subscribe to updates on summaries, while others wanted the ability to mention another user in a comment.

All notifications on Wikum+ occur through the email associated with each user's account. Notifications occur in six cases:

1. **Reply in Thread:** when another user replies to a comment in your comment's reply thread
2. **Subscribed Reply:** when another user replies directly to a comment or summary you are subscribed to
3. **Summary Edit:** when another user edits a summary you contributed to
4. **Subscribed Edit:** when another user edits a summary you are subscribed to
5. **Summarized Comment:** when another user summarizes your comment or includes your comment in a summary
6. **Mention User:** when another user mentions you in a comment

To use the mention feature, a user can type the @ symbol in a comment text box, causing a list of users, with a matching prefix, to drop-down next to the user's cursor

and populate the list. When the user submits the comment, the mentioned user will receive a notification, along with a link to the that comment in their email. Thus, each comment has an associated URL link; upon load, the page will scroll to place that comment in view. Additionally, the page URL will update with the comment link if a user clicks on any comment reference found in the summaries.

Some users may not want email notifications for all of the six cases above. A user can select which notifications types they would like receive in the settings page. However, users can still see an entire list of notifications on the notifications page, found by clicking the bell in the menu bar. These notifications are organized by unread notifications, shown first, followed by read notifications, shown after. Both of these groups are sorted in reverse chronological order. As soon as the user loads the notification page, we mark all notifications as read in the back-end, so that the next time the user visits the page, they will only see new updates.

3.3 Implementation

Wikum+ is a Django web application and its data is stored in a MySQL database. The front-end is written in JavaScript, HTML, and CSS. Due to the existing back-end framework of Wikum+, we use Django Channels (the Django integration layer), Daphne (the HTTP and Websocket termination server), asgiref (the base ASGI library), and channels_redis (the Redis channel layer backend) to handle synchronous functions within a Wikum+ page [12].

In January 2020, Python 2 was deprecated and the latest version of Django Channels required Python 3. So Wikum+ upgraded Wikum's back-end, which was written in Python 2.7, to Python 3.7.

The original D3 summary tree visualization in Wikum [47] was replaced with an outline view in Wikum+, which uses the library SortableJS (Figure 3-3). The rich text editing feature of summary creation and editing makes use of TinyMCE (Subsection 3.2.7). The email notifications and notification settings in Wikum+ are accomplished with Pinax Notifications and Pinax Templates. Finally, we use jquery-textcomplete

for auto-completion of @-user-mentions (Subsection 3.2.8).

Chapter 4

Evaluation

4.1 System Goals and Research Questions

Through Wikum+, there were a number of a research questions that we wanted to explore. We designed Wikum+ to help us gain insight into how users might use such a tool and what they might need in the system. Below, I detail each research question that led us to implement Wikum into what it is today. In this chapter, we primarily focus on evaluating the first two research questions, although our results provide implications about the other two as well.

How does integrating discussion alongside summarization affect the quality of collaboration and the quality of the final document? Our goal was that through Wikum+'s ability to incorporate discussion alongside document writing, more ideas and perspectives would surface instead of being lost or ignored in existing discussion systems. In this way, we hoped that redundancy would decrease, feelings of inclusiveness in the process would increase, and the quality of the document would also increase.

We have already seen that in reading a discussion, Wikum's summaries helps users more efficiently find and explore main topics [47]. Our goal with Wikum+ is to allow for summarization while conversing in a modular and incremental way, encouraging more organization. We hoped that this would reduce users' cognitive load, increase efficiency, produce more creative ideas, and improve the document outcome.

We considered allowing for voting on the interface, but chose to leave it out so that users would instead focus on using summarization as the primary form of synthesis. Our goals for the design choice of interleaving summarization are: (1) to encourage minority voices to be heard, and (2) encourage users to have a conversation on *why* an idea should or should not be incorporated, rather than simply up-voting ideas that they agree with with no explanation. We hope that in this way, more voices, especially those in a dissenting minority, can be addressed through actual deliberation.

What are the trade-offs of enforcing greater structure (e.g., phases) in the discussion versus summarization process? Since Wikum+ introduces a new way to discuss online, it was not clear how the natural progression of the conversation would flow. Would users discuss as much as possible and then summarize everything afterwards? Would users summarize portions of the discussion as they conversed? What portions would users choose to summarize first and how might that relate with their level of consensus? Quantity over quality has been shown to be critical in brainstorming good ideas [34]. We therefore hoped to analyze the effects of enforcing a certain amount of discussion time before allowing for summarization. As there can be multiple rounds of discussion, summarization, and more discussion, we wondered if multiple, iterative periods of enforced discussion followed by summarization could improve the discussion's overall creativity and increase the diversity of ideas.

On the other hand, some topics could be very controversial and enforcing too long of a discussion period could result in parts of the conversation drawn out, potentially reducing overall productivity. By allowing summarization of a sub-discussion at any time in the conversation process in a seamless way, we hypothesize that synthesis and consensus may both occur faster, potentially improving efficiency.

How do people self-organize into parts of the discussion? While asynchronous discussion platforms such as forums and wikis allow for greater freedom in navigating through the conversation and adding comments in various threads, typical conversations on synchronous platforms such as chat rooms happen linearly, as new messages appear at the bottom. This is likely due to the lack of organization in synchronous discussion which can result in less discoverable new comments if they

appear out of order. Wikum+ allows for threaded discussions, many layers deep, similar to forums and wikis. As a result, we hypothesized that users will self-organize by interest, and discussions will likely not occur linearly. However, by making Wikum+ both threaded and synchronous, we needed to compensate by highlighting new comments so that users would not be overwhelmed by comments appearing all over the discussion page.

What are the trade-offs of enforcing roles in the discussion versus summarization process? Some current online systems rely on moderators to ensure the organization and progression within a discussion [5, 39, 7]. We imagined that granting some users greater edit access might achieve a similar effect, where these users are dedicated members who are given extra permissions to assemble the comments into summaries, flag parts of the discussion requiring extra work, and hide irrelevant or inappropriate comments. By allowing comment-only or full-edit roles in Wikum+, we hoped to explore the impact of enforcing roles within a conversation, particularly in how well-written and cohesive the final summary document is and how levels of inclusion may be impacted. We hypothesize that granting some users full-edit permissions might allow for a better quality final document. Fewer users creating the summaries could result in a document that flows better from beginning to end. However, the same issues that occur with moderators on other platforms may arise with these enforced roles on Wikum+ [35]. Users with full-edit roles may introduce biases and leave out ideas that others may deem important, reducing the perception of fairness. Instead, allowing more people to participate in synthesizing the discussion may increase feelings of inclusiveness.

4.2 Evaluation

I conducted one pilot study in Fall 2019 with 12 MIT undergraduate students and, after iterating on Wikum+'s functionalities, conducted two full user studies in Spring 2020, one with 35 MIT undergraduate students and the other with 29 MIT undergraduate students. In Section 4.3, I briefly outline the differences between the version

of Wikum+ tested in the Fall and the current deployed version, the methodology and results of the pilot study, and the motivation for Wikum+'s changes since then. In Section 4.4, I detail the the two full user studies, including their methodologies, data analyses, and results.

4.3 The Pilot

In the pilot study, the only new features implemented from Section 3.2 were New Comments, Permissions, Real-time Updates and Locking, and Rich Text Editing. The Wikum+ at the time also had the node-view visualization rather than the outline view. The other features were implemented later, partially inspired from this study, as explained in Subsection 4.3.3.

4.3.1 Methodology

The pilot study I conducted had two groups, one using Wikum+ and one using a Google Doc and Facebook Messenger as a control. The study had two parts, including a 30-minute synchronous portion, and a 3-day asynchronous portion. In both the short and long portions, users in the Wikum+ condition (4 males, 2 females; mean age 21) could only communicate through Wikum+ and users in the control group (2 males, 4 females; mean age 21.5) could only communicate over Facebook Messenger and within a shared Google Doc. The Wikum+ group received an additional 15-minute training and a Wikum+ guide to learn the tools and interface that Wikum+ provides. Users were MIT undergraduates and each received $10 for participating.

Users were given the task of writing an email to their campus administration containing a one-page proposal for improving on-campus dining options. The proposal required 5 components to encourage deliberation:

1. Why current dining options are problematic
2. A comprehensive list of dining solutions
3. A cohesive and convincing argument of best 1-2 options from list

4. A trade-off analysis of the selected options

5. A proposed timeline on how to implement options by 2020

Users were instructed to have their final proposal contained on the first page of the Doc or in the top level Wikum+ summaries. At the end of the synchronous session, each group's goal was to get through as many components as possible. Then, in the following three days of asynchronous work, the groups finished up the proposal.

Code	**Definition**
Independent	Independent writing; division of roles
Chaotic	Feelings of confusion or disorganization
Inclusive	Comments addressed; proposal content inclusive of ideas
Comprehensive	Covers significant content; detailed
Organized	Structured proposal-making process
Clear State	Can determine new or unaddressed content
New Content	Added new content to discussion
Avoidance	Avoid editing others' work

Table 4.1: Interview Code Definitions for Open Coding in Tables 4.3 and 4.4

I conducted post-study interviews after both portions, asking users about the collaboration quality and use of their tools. I conducted iterative open coding of the interviews to achieve the 8 codes defined in Table 4.1, using grounded theory qualitative techniques. I then counted their frequency during the interviews, tallied in Table 4.3 and 4.4. Users also filled out a post-study survey, with Likert questions from 1–7. Metrics with average differences of at least 1 are shown in Table 4.2.

4.3.2 Results

Figure 4-2 shows how the two groups worked over time. The Wikum+ users started by discussing in a non-linear manner, jumping between existing comment threads and

starting new threads. During the synchronous portion, they formed small summaries for stagnant, unchanging threads as early as 2 and 6 minutes in. As summaries formed and collapsed nested comments, the number of words shown by default decreased (Figure 4-2).

Users also used tags (Figure 4-1) to group ideas that helped when skimming the page. With Wikum+'s recursive summary structure, users combined several sub-summaries together to form the final proposal by the end of the 30 minutes, though they had not yet completed 2 task items. During the asynchronous portion, Wikum+ users added new threads on the remaining points, growing the default number of words shown, and eventually summarized them and incorporated them into the top-level summary, thus shrinking the default number of words shown. Users also flagged summaries as incomplete to point others to work on them (Figure 4-1). In general, users kept their summaries brief and made use of bullet points in the final proposal. As seen in Figure 4-2, the Wikum+ final summary ended up being about 1000 words shorter than the control.

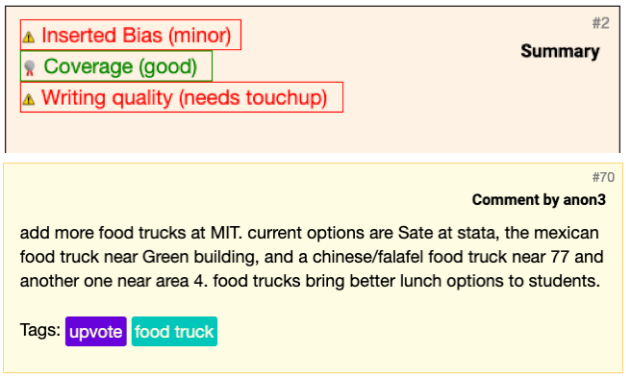

Figure 4-1: Users added flags to evaluate summaries and point out areas that needed work. Users also used tags to help with skimming and grouping.

In the control condition, users discussed in a linear fashion, largely staying on a particular topic and conversing in the Google Doc. A user explained that "*having two platforms dilutes information,*" so the group used the Google Doc for discussion and Messenger for meta decisions (such as pacing) or moderation. During the synchronous portion, the control group, like the Wikum+ group, did not have time to discuss the last 2-3 task items. However, at the 16 minute mark, the group hopped to Messenger to decide on a +1 voting mechanism to determine which discussion ideas to include into the proposal. The users divided up the five task items, each claiming one, and

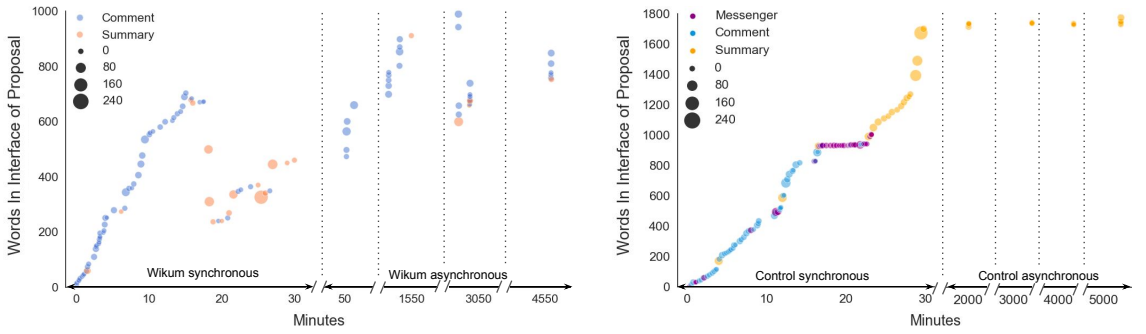

Figure 4-2: Plot of user activity in the two conditions over time. Each node is a user's comment or edit to a summary/proposal, with the size equal to the number of words. In the Wikum+ group, I distinguish between comments and summaries, while in the control, users had discussions both in Google Docs and Messenger, while wrote up their proposal in Google Docs. For the y-axis, I count the number of words present in the default Wikum+ view, including all top-level summaries and unsummarized comments (with summarized comments collapsed), while in the control, I count the number of words in the Google Doc. Thus, orange summary points in Wikum+ result in y-value dips as a summary replaces comment text. Dotted lines indicate a break in time where there was no activity.

independently wrote a paragraph each. Unlike the Wikum+ group, the control group wrote up the remaining task items without discussing first. By the end of the 30 minutes, all 6 users reported that they had not read the entire proposal. During the asynchronous portion, all members of the control group were unclear where changes in the proposal had been made since the synchronous portion. Users had no further discussion and avoided editing others' work, focusing on clarifying existing ideas or fixing grammatical errors.

	Wikum	**Control**
Inclusivity	5.5	4.5
Effort	3.2	4.8
Comprehensive	3.2	5.2

Table 4.2: Post-study survey metrics with average score differences ≥ 1. The Wikum+ group (W) felt a greater sense of inclusion and required less effort, but the control (C) felt that they were more comprehensive.

Users of Wikum+ felt that the process was more inclusive. From the surveys and interviews, Wikum+ users reported feeling a greater sense of inclusion when reaching a consensus on which ideas to include (*Inclusivity* in Table 4.2 – Wikum+: 5.5, Control: 4.5; *Inclusive* code in Table 4.3 – Wikum+: 5, Control: 1). Unlike the control group, Wikum+ users did not write ideas in the final proposal without first collectively discussing (*Independent* code in Table 4.3 – Wikum+: 1, Control: 5), allowing for greater collective ownership of the summaries and final proposal. This led Wikum+ users to feel that "*the summaries were very collaborative,*" so that they were "*comfortable changing the proposal because it wasn't just [their own] ideas.*" In contrast, the +1 voting mechanism of the control left some users feeling as though their ideas were unaddressed or left out.

In addition, due to the control group's decision to divvy up task items and work independently, users did not feel comfortable discussing or editing other users' sections once they had been written. In the end, over 53.3%, consisting of the last 2 bullet points, of their proposal was written up without any discussion of the contents.

	Wikum	**Control**
Independent	1	5
Chaotic	6	6
Inclusive	5	1
Comprehensive	3	5
Organized	4	0

Table 4.3: Open coding results from interviews after the synchronous portion.

Wikum+ allowed for more light-weight coordination and reduced user effort. One potential reason for higher levels of inclusion is the ability in Wikum+ to coordinate work in a light-weight way. Wikum+ users explained that it was easy to pinpoint unresolved, unsummarized blue comments or flagged summaries, encouraging them to go in and summarize or edit summaries without having to ask anyone

explicit permission. The half-blue/half-orange summary nodes also helped Wikum+ users see when new or unresolved comments appeared (*Clear State* code in Table 4.4 – Wikum+: 6, Control: 0). In contrast, in the control condition, users described having trouble knowing what had been done and what still needed work. This resulted in much less activity during the asynchronous portion when coordination is harder.

	Wikum	**Control**
Clear State	6	0
New Content	6	0
Avoidance	1	3

Table 4.4: Open coding results from interviews after the asynchronous portion.

Another benefit of self-organization is that it makes the synthesis process easier. By summarizing early on and throughout the process, the Wikum+ group felt more organized (*Organized* code in Table 4.3 – Wikum+: 4, Control: 0), despite the matching levels of chaos initially (Table 4.3). As a result, it reduced the Wikum+ group's overall efforts when it came time to write a final proposal based on all the discussion (*Effort*–Wikum+: 3.2, Control: 4.8 in Table 4.2). However, it is possible this also resulted in a reduced level of comprehensiveness (*Comprehensiveness* – Wikum+: 3.2, Control: 5.2 in Table 4.2) in the final proposal. As Wikum+ users started to summarize ideas early in the process, this may have cut down on the time spent ideating.

Wikum+ allowed for iterative improvements and incorporation of new ideas. Another potential reason for higher levels of inclusion is the support in Wikum+ for multiple rounds of discussion and summarization. In the asynchronous portion, though both groups started with a draft proposal, the Wikum+ users added new ideas and discussion that were then incorporated into the proposal (*New Content* code in Table 4.4 – Wikum+: 6, Control: 0). In contrast, the control group had only one round of discussion and summarization. Once the initial draft was created, users avoided making any large changes in the Doc (*Avoidance* code in Table 4.4 –

Wikum+: 1, Control: 3) or starting new discussion. This may be because Wikum+ users were used to mixing discussion and summarization from early on, while the control users felt they had graduated to the "writing" phase of the collaboration.

4.3.3 Pilot Study Discussion and Design Implications

Results from the pilot study suggest that the design of Wikum+ helps promote inclusiveness and self-organization compared to a control, with differences magnified in asynchronous settings. Features like flags, tags (Figure 4-1), and blue markers of new/un-summarized versus old/summarized content could help in lightweight coordination and making sense of messy threads. The pilot also showed that Wikum+ users could interweave discussion and summarization of different strands of ideas, suggesting that mandating designated discussion and summarization modes or only having one round of each may be limiting. Wikum+ could require an extended initial ideation step before permitting any summarization to encourage a sufficient quantity of initial ideas without prematurely converging.

Users in the Wikum+ group found that they had a hard time finding threads that they had originally contributed to. While the recursive node-view visualization represented the structure of the discussion, many noted that it was not conducive to finding topics they were looking for. Instead, a recursive outline view with topic sentences or the first few words of the comment would help users navigate the conversation, find new topics, and keep track of where they were, while preserving the nested shape of the discussion.

Finally, many users said that they would have appreciated notifications if some users replied directly to them in the Wikum, particularly in the asynchronous portion of the study. One user mentioned that tagging other users would have also helped drive the conversation forward, especially because users felt that they jumped around a lot in the discussion. Thus, we implemented tagging and notifying other users to help bring them back and continue discussions that they had previously left.

4.4 User Studies

After iterating on the features of Wikum+ to include all of the new features described in Section 3.2, we ran two new user studies. The first user study we performed was within-subjects and consisted of six groups of 5-6 users each. The second user study we performed was between-subjects and consisted of one large group working solely in Wikum+ and a separate large group working solely in Google Docs.

Through university-affiliated mailing lists, we recruited participants who were asked to fill out an interest survey and were filtered based on the requirement that users must be MIT undergraduate students taking online classes. This requirement had to do with the particular tasks chosen for both studies.

4.4.1 User Study 1: Within-Subjects, Multiple Small Groups

Experiment Design

111 people filled out the interest survey and the first 40 eligible people were chosen as participants. These participants were split into six groups (four groups of 7 people each, two groups of 6 people each). By the end of the study, out of the original 40 participants, 35 (Mean age 20.2, 8 Male, 27 Female) had completed the task and the post-study surveys, and the six groups ended up consisting of five groups of 6 users each and one group of 5 users. Each user that completed the study was compensated $35.

The study had two rounds so that each user could use both Wikum+ and the control (Google Docs and Slack). In the first round, every group was given the task of writing a one-page proposal to the administration on how they would like MIT to operate if classes continued to be online in the Fall semester. The proposal for Topic 1 required 5 components to encourage deliberation:

1. How grades should be, including the group's opinions on the pros and cons of a mandatory Pass or No Record grading scheme

2. Whether tuition should or should not be reduced and why

3. How much homework workload should be given

4. How midterms and finals should be structured

5. How student resources should be provided

In the second round, every group was given Topic 2, the Campus Dining task from the Pilot Study (Subsection 4.3.1). We randomly assigned participants to one of the six groups, with half of the groups in each round on Wikum+ and the other half in the control, with order counterbalanced (Table 4.5). Thus at the end of the study, the six groups produced 6 Wikum+ proposals and 6 Google Doc proposals. We chose this experiment design so that we could both compare the experiences, participation levels, and outcomes of Wikum+ versus Google Doc and Slack from each user, controlling for both the topic of discussion across interfaces and the order of the condition across groups.

	Round 1: Days 1-3, Topic 1: Online Classes	**Round 2:** Days 4-6, Topic 2: Campus Dining
Wikum+	G1, G3, G5	G2, G4, G6
Control	G2, G4, G6	G1, G3, G5

Table 4.5: Task and order assignments for groups.

Procedure

I video-called with each group at the start of both rounds and had each group's members meet through a 5-minute icebreaker. During the Wikum+ group calls, I gave a 15-minute tutorial of Wikum+, had all of the users sign up for Wikum+ accounts, and added them, by group, to private-access Wikum+ pages where each group would hold their discussion and create their proposal. The proposals for these groups were to be written in the top level summary nodes in Wikum+. The control groups were each added to their own private-access Google Doc and a private Slack channel. The control groups were allowed to discuss in either their assigned Slack

channel or their Google Doc, but their proposals were to be written in the first page of the Doc.

To spur initial conversation, each group spent 5 minutes conversing and ideating synchronously on their given platform(s) for the given topic of the round. Then, users had 3 days to continue discussion and collectively create an email proposal in their interface, thereby completing the task. However, after the first round was over, we found that one control group had not written a proposal in their Doc. Therefore, after both rounds had completed, we gave each group one extra day to finish up either rounds' proposals. After completing each condition, users filled out surveys on their perceived task load [16], their experience, and the document outcome. After both tasks were completed, users also answered some open-ended questions comparing the two experiences.

Results

Figure 4-3 shows the activity in each group over time in both the Wikum+ condition (left) and the control condition (right). Likely due to the nature of the first task topic, engagement in the interface was generally higher in Round 1 of the study. Furthermore, 3 users had dropped out of the study before or during the the second round, one from each of Group 1, Group 2, and Group 6, further lowering contributions in the second rounds of those groups.

Groups 1 and 3, both who had the control condition second, chose not to use Slack at all, working entirely in their Google Docs. Group 6 users, on the other hand, worked solely in Slack for the given time period, still maintaining high participation levels. The group did not manage to collect their ideas and work on the Google Doc email proposal at all in the 3 days given. The users in the group explained that their ideas and discussion points were scattered and disorganized in the Slack channel. As one user put it, "there was an initial activation barrier to overcome" in collecting the thoughts and moving them over to the Doc that no one took initiative for. In the final extra day given to finish up either proposal, one user in the group went back to the Doc and wrote up a bullet-point summary of their Slack discussion. Although

feasible at the scale of the user study, this would not scale to a larger Slack discussion.

Generally, in the Wikum+ condition, the default (words in summaries and unsummarized comments) number of words grew and shrank as users contributed new ideas, summarized sub-threads, continued discussing or critiquing summaries, and eventually created a final proposal (Figure 4-3). On the other hand, in the control condition, Groups 2-6 all generally conversed for some time before building a final summary based on their conversation, continuously growing the discussion size. Group 1 was the only group that started removing relatively large amounts of text as the proposal started to form, drawing ideas and organizational skills from their previous round in Wikum+.

First round Wikum+ users reused strategies and tools that helped them in Wikum+ when creating the Doc proposal in round 2. As we saw with Figure 4-3, users in Group 1 drew inspiration from the practices that were beneficial in Wikum+ when working in the control condition during round 2. In particular, both Groups 1 and 3 used bullet points to create threaded discussions in the Google Doc, simulating the nested discussion tree in Wikum+. A Group 1 user explained that they were "immediately overwhelmed by the amount of text" in the Doc, and some users in the group started deleting discussion text and replacing it with summarizing text. As one user started summarizing and working on the Doc proposal, they also removed entire portions of the discussion that they had summarized into the proposal (Table 4.8, G1, Avg Words Removed). Another user avoided deleting the discussion text but striked out text that they had incorporated. While deleting or striking out text could reduce clutter and may help users see the remaining discussion that still needs to be included in the proposal, neither allow for a direct link pointing from the proposal text to the discussion that generated and inspired it.

While Group 5 also had the control condition for the second round, unlike Groups 1 and 3, they still decided to discuss in Slack. Right past the 3000 minute mark in Figure 4-3, the large purple point represents one user who took the time to go through all of their group's Slack conversation, summarize each Slack reply thread into a bullet-point list, and send it to the group. Afterwards, other users built off of

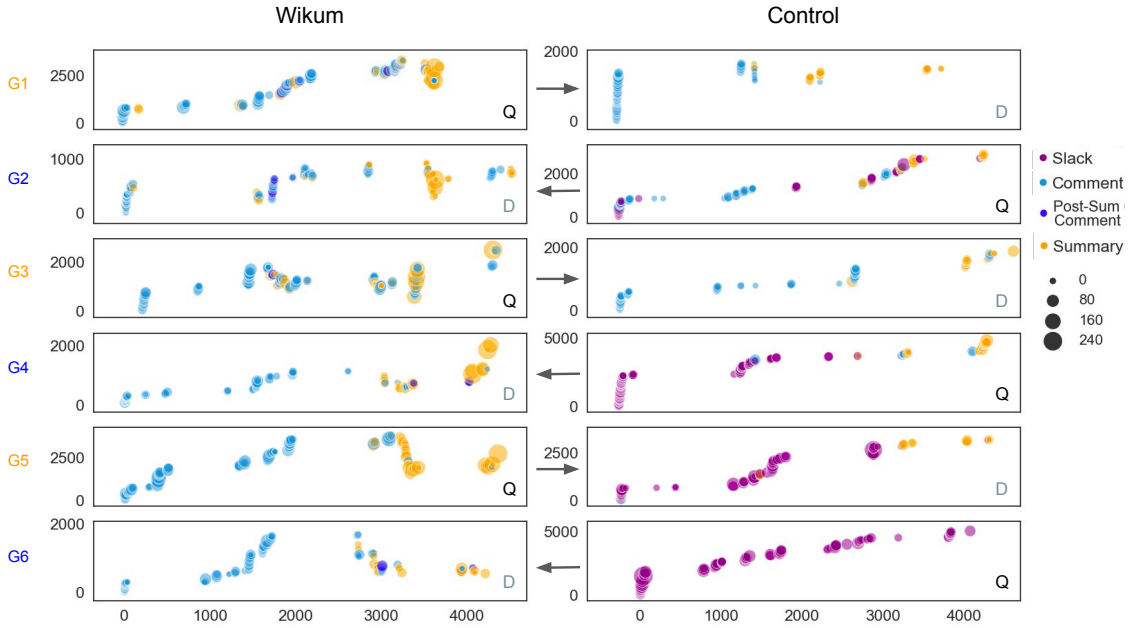

Figure 4-3: Default words shown over time (minutes) for each condition (columns) by group (rows). Each node's size represents the net number of words a user made through an edit, with a minimum set to 0. The x-axis is in minutes and the y-axis is the number of words shown by default. In Wikum+, this includes all top-level summaries and unsummarized comments (because summarized comments collapsed), while in the control, the y-axis counts the total number of words in both Slack and the Google Doc. Comments in the control refer to Doc discussion and comments while summaries refer to summary text created in the discussion as well as proposal text. Each graph is labeled with the task topic, either Q (Quarantine Online classes) or D (Dining options), and arrows point to which condition happened second. Note that three users dropped out before or during the second round, one from each of G1, G2, and G6, and levels of participation generally dropped from Round 1 to 2.

this Slack message to write up the email proposal in the Doc. Each bullet point in the list summarized one discussion thread in the Slack, much like a summary would in Wikum+, but just like in the Doc from Groups 1 and 3, this lacks a direct reference to the conversation thread that sourced the summary. By having summarization collaborative in Wikum+, it also helps scale this synthesizing approach to larger conversations.

On Wikum+, users made use of comment-links, leading to specific conversations from their summaries, thereby keeping the original thought process and deliberation intact while creating a reference that allows us to see *why* some portion of the

summary might have been incorporated.

Figure 4-4: Histograms of all metric values (out of 7) from the conditional post-study surveys. Metrics that resulted in significantly different means from paired sample t-tests are boxed.

Users in Wikum+ found that the writing process was more organized than in the control. From the post-study surveys, we found that to many users, Wikum+'s recursive summarization process provided their group structure, allowing for noticeable, gradual progress. A paired sample t-test yielded significant evidence that Wikum+ is more organized ($t = 3.23$, $p = 0.003$, Figure 4-4) and in the comparison survey, 85.7% of users felt that Wikum+ provides greater organization as a system (Table 4.6). One user explained that the interface of Wikum+ "helped coagulate the comments so that none are lost in the cracks... As you summarize, the comments are displayed beside the summary, so you can ensure you are touching on all topics discussed." Another user "liked building up the summaries in small contributions", making it "easier when bringing [the discussion]... into the final document."

Users in Wikum+ found it easier to locate both new ideas and topics that still needed work, allowing for iterative improvements on summaries. Many users in the post-study survey pointed out that a number of Wikum+ features

	Preferred System	More Organized	More Inclusive, Cooperative	More Successful Proposal
Wikum+	74.3%	85.7%	65.7%	54.3%
Doc & Slack	25.7%	14.3%	17.1%	34.3%
Equal	0	0	17.1%	11.4%

Table 4.6: Qualitative Comparison Survey Results

helped them find areas in the conversation to jump into. When users came back to the Wikum+ page after a period of time, bold unread markers and blue comments under summaries helped them find new ideas introduced while they were gone. Users also reported that summary flags, the summarization progress bar, blue unsummarized comments, and partial summary nodes, helped them easily pinpoint areas that still needed work. From the survey metric results, we found that Wikum+ helped users significantly in discovering newly introduced ideas ($t = 4.30$, $p = 0.00013$, Figure 4-4) and in locating topics that had not been included or still required further deliberation ($t = 2.92$, $p = 0.0063$, Figure 4-4).

Although in both conditions, users expressed that they felt rushed to create the proposal near the end of the 3 days, in many groups, the discussion still continued to grow as content was being summarized, although generally more so in the Wikum condition, as shown in the blue nodes of Figure 4-3. In Wikum+, 5 out of the 6 groups added new ideas as replies within existing summaries – or post-summary comments – that were all addressed or incorporated either through edits or evaluations of the summary (Figure 4.7). In 3 out of 6 groups, these post-summary comments contended with or made suggestions to the summaries they were created under, resulting in improvements and edits to all of those summaries, incorporating the new comments. Other post-summary comments either continued part of a discussion already incorporated in the summary, or voiced agreement to the summary they replied to. These post-summary comments distinguish Wikum+ from the Doc because they allow for iterative improvements, enabled by the interface's ability to help users find specific areas in a potentially lengthy conversation that still require further discussion

or improvements. In the control condition, only one user, who worked in the Doc for round 2 after working in Wikum+ for round 1, made a single comment directly on the summary content, reminding the others on the task objectives. This led to some restructuring of the proposal, and the comment was marked as resolved. However, in general, users in the Doc had trouble finding newly added content or areas that still needed work. One user explained "re-reading over the entire discussion was the only way to determine new content... [and] what had or hadn't already been added," as there was almost "nothing to help track unincorporated points of discussion." A handful of users resorted to using the edit history to find recent changes. One user said: "I really had to think about how each of the new updates would contribute to the overall task – it was a little overwhelming."

	Total Post Summary Comments	**Improving The Summary**	**Agree With Summary**	**Response to Comment**
G1	8	4	0	4
G2	10	6	1	3
G3	2	0	0	2
G4	3	0	1	2
G5	0	0	0	0
G6	4	4	0	0

Table 4.7: Open Coding of Post-Summary Comments in Wikum+, or comments created as replies within existing summaries. Comments categorized under "Improving The Summary" contended with or made suggestions to the summary, and were *all* later incorporated into the summary they were created under. Comments under "Agree With Summary" approved of the summary created, and comments under "Response to Comment" continued a discussion within the comments under the summary.

Although users wrote more summary text in Wikum+, when it came to writing the final document, users in Wikum+ took a more concise, by-topic approach, while users in the Doc wrote more fluidly. On average, users wrote around 254 more words in Wikum+ than in the control, with approximately 212 of those words on summarizing content (Table 4.8). Paired sample t-test analyses

comparing the number of words added in the Wikum+ condition against the number of words added in the control, by type of content, yields significant differences, providing evidence that a greater quantity of contributions occurred in Wikum+ (number of words: $t = 2.31$, $p = 0.027$), particularly when comparing summary contributions (number of summary words: $t = 3.15$, $p = 0.0034$).

However, on average, Doc proposals had about 38 more words than Wikum+ proposals (Table 4.8), and were written with, what users deemed, a higher quality of writing ($t = -2.26$, $p = 0.03$, Figure 4-4). This suggests that a significant portion of the summary contributions in Wikum+ were made prior to creating the proposal, more gradually synthesizing the discussion along the way. We found that the proposals in Wikum+, written in top level summary nodes, were organized by the summaries created for each topic thread. About half of the Wikum+ proposals also maintained some bullet-point structures found from the groups' prior summaries, making them more concise. In contrast, users had only the conversation to go off of in most groups during the control condition, the exception being Group 5, where one user synthesized the discussion into a summary before the group wrote the proposal. It is possible that this general lack of pre-existing summaries in the control resulted in more fluidly written Doc proposals, and that the succinct style of summaries in Wikum+ influenced the writing style of the following proposals.

Overall feedback on the experience. 65.7% of users felt that Wikum+ was more inclusive of ideas and allowed for greater levels of cooperation, the remaining proportion split between the control experience and finding the two equally inclusive (Table 4.6). A user explained that in Wikum+, "being able to summarize helped people interact with others' ideas and not just focus on their own," broadening perspectives, while still "[being] able to add your own ideas into the summaries." This made "the summary mechanic feel like a very collaborative activity."

Furthermore, 54.3 % of users found the Wikum+ proposal more successful compared to the 34.3% of users who found the Doc proposal more successful (Table 4.6). The remaining users found the two proposals about equally successful. From analyzing the open-ended responses given, users generally expressed that their familiarity

		Avg Words Added	Avg Words Removed	Avg Summary Words Added	Avg Days Active	Words In Proposal
G1	W	1116.72	0.14	421.86	2.43	717
	C	360.33	109	78	2.17	360
	W-C	756.3	-108.9	343.9	0.26	357
G2	W	466	29.83	148.67	2.83	229
	C	414	0	92.57	2.83	656
	W-C	52	29.83	56.1	0	-427
G3	W	980.5	8.67	342.5	3.33	363
	C	316.33	1.67	80	2.5	469
	W-C	664.17	7	262.5	0.83	-106
G4	W	568.33	13.83	338	2.67	758
	C	804	0	149.83	2.83	891
	W-C	-235.67	13.83	188.17	-0.17	-133
G5	W	1145.5	3.33	455.5	3.5	657
	C	559.83	1.33	102.67	3.17	607
	W-C	585.67	2	352.83	0.33	50
G6	W	452.67	11.5	92.83	3.5	205
	C	750.43	0	25	2.83	175
	W-C	-297.76	11.5	67.83	0.67	30
Avg	W	788.29	11.22	299.89	3.04	488.17
	C	534.15	18.67	88.012	2.72	526.33
	W-C	254.13	-7.45	211.88	0.32	-38.17

Table 4.8: Average contribution levels by group in Wikum+ (W), the control (C), and their difference (W-C). Note that 3 users dropped out before or during Round 2 and that Round 1 received more engagement.

and interest in the task given influenced their success rating. Users felt that the topic on "Classes in Quarantine was more pressing" and "felt more strongly about the topic," resulting in "more ideas, more thoughts, and more solutions." Another user felt that people were more "strong-minded about dining options on campus, and normally have a lot to complain about in that realm," leading to generally "more complaints than solutions" in the second task. On the other hand, some users felt that "having the experience of going through the first task helped [their group] with the second task – [they] kept the conversation more focused and relevant to what would ultimately be in the final summary."

Finally, in the comparison survey, 74.3% of users preferred Wikum+ as a system overall. Users participated somewhat more consistently in Wikum+ on a day-to-day basis, with each user's number of active, contributing days approximately 0.321 days higher, on average, in Wikum+ than in the control (Table 4.8).

4.4.2 User Study 2: Between-Subjects, Two Large Groups

In this study, we were interested in seeing how larger group sizes would affect the results of and experiences on Wikum+ compared to a control. Leveraging the number of participants who signed up, we created one of each group at random and ran a between-subjects experiment.

Methodology

Every student who filled out the interest survey for User Study 1, but was not selected, was emailed with the opportunity to participate in User Study 2. An additional interest form was sent out to campus mailing lists to gather more interest as well. 47 people signed up in total that had not participated in Study 1. All were selected as participants and were randomly assigned into two groups: one for Wikum+ and one for a Google Doc. 20 people of the Wikum+ group completed the initial step of signing up for an account and 20 people of the control group completed the initial step of giving me a G-mail account for access to the private Doc. 15 users contributed

to Wikum+ and 17 user contributed to the Doc at least once during the course of the study. After the 5 days, all Wikum+ participants and 14 Doc participants completed the post-study survey. The results below are based on what took place during the study with all 32 participants (Mean age 20.0, 8 Male, 32 Female), as well as the experiences of the 29 participants who completed the survey (Mean age 20.0, 6 Male, 23 Female). Each user that completed the study and survey was compensated $15.

In this study, the control group was only given a Google Doc, but could communicate within the Doc using any of its features (comments, chat, typing directly in the Doc). The Quarantine Online Classes task from Study 1 (Subsubsection 4.4.1) was given to both the Wikum+ group and the Google Doc group due to its higher levels of engagement previously. I created an 8-minute video tutorial and an online written tutorial for the Wikum+ users to review before starting on their task. Each user had approximately 5.5 days to work on their given interface and worked asynchronously throughout the study. We staggered sending out the initial instruction emails to users over the course of 6 hours with the intention to also stagger the starting times of the users. This way, some users would arrive after a bit of discussion had already occurred, allowing us to gain feedback from users who had varying degrees of how much they needed to catch up on. The Wikum+ group was instructed to write their email proposal in a top-level summary node and the control group was instructed to write their email proposal in the first few pages of the Doc. After the 5.5 days, each user filled out a post-study survey based on their condition, similar to the surveys from Study 1.

Results

In this study, the control group made heavy use of the commenting functionality in Google Docs, allowing for the group to have additional discussions and find areas that still needed work. Doc comments also allow for email notifications, so users could be notified when new comments were added. Due to the asynchronous nature of the study, participants did not use the Doc chat feature.

In Figure 4-5, we can see a significant contrast between the events of each condi-

tion. After around two days of discussion, Wikum+ users began summarizing rapidly, synthesizing and condensing many of the page's threads and sub-threads. However, discussion continued and some users voiced disagreement with some of the points made in the summaries, resulting in both edits to the existing summaries and new summaries created, incorporating the divergent ideas. In the final day, Wikum+ users primarily edited and combined existing summaries into a final proposal. In comparison, Doc users spent a large proportion of their time in disagreement, trying to determine which opinions to include in the final proposal. In the end, the group did not finish writing their proposal, as some of the points from the Task were unaddressed.

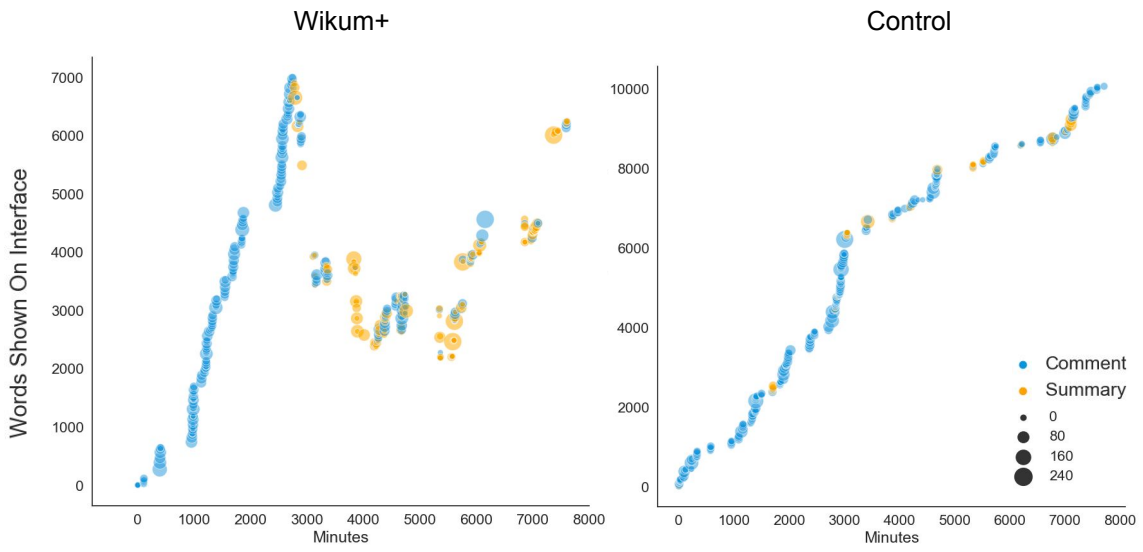

Figure 4-5: Plot of user activity in the two conditions over time. Each node is a user's comment or edit to a summary/proposal, with the size equal to the number of words. The y-axis represents the number of words shown by default on each interface. In Wikum+, this value includes all top-level summaries and unsummarized comments (because summarized comments are collapsed), while in the control, the y-axis counts the number of words in the Google Doc.

Just as in Study 1, the style of writing in the Wikum+ proposal was more concise and consisted of many bulleted lists, likely drawing inspiration directly from the summaries created for sub-threads. The Doc proposal, on the other hand, was more essay-like, written in full paragraphs, typically with one or two users responsible for

each paragraph. However, in this study, Wikum+ users found the quality of writing in their proposal higher than Doc users found theirs (average Quality of Writing in Proposal metric – W: 5.3, C: 4.6), though not by a significant amount ($F = 2.70$, $p = 0.11$).

Although both conditions had comments disagreeing with summary content, Doc users decided to vote for a majority voice, while Wikum+ users iterated on the summaries to include multiple opinions. The topic of mandatory versus optional Pass or No Record grades in the Task caused divisions in both groups, but each group handled this disagreement differently. In the Doc, a couple of users took charge in setting up a system where each user can summarize their viewpoints in the proposal. Then, users can sign off on which summary they are in favor of, only keeping the majority-vote summary in the final proposal.

As discussed in System Goals (Section 4.1), we purposefully avoided adding any voting functionality in Wikum+, and in this study, we avoided moderator roles. One of our goals for this was to promote minority voices and encourage conversation, instead of voting, to reach consensus. Users in the Wikum+ condition decided that instead of choosing to only express the majority opinion, they would update summaries to include the other opinions, allowing for all voices to be heard.

Furthermore, summaries in Wikum+ were in the form of bullet-point lists, making appends and modifications easy to do, as opposed to trying to find the right words to fit into an existing paragraph written by someone else. A user in the control group said that they "felt really uncomfortable editing others' choice of words since that has a direct impact on changing how they initially delivered that thought."

	Words Added		Words Removed		Summary Words Added		Days Active		Proposal Words
	σ	μ	σ	μ	σ	μ	σ	μ	Count
W	1151.7	800.2	28.3	51.6	436.9	601.5	3.73	1.22	1403
C	608.5	353.1	14.4	50.3	54.1	90.3	3.35	1.27	960

Table 4.9: Contribution levels (mean and standard deviation) in both the Wikum+ (W) and the control (C) conditions.

The Wikum+ group's approach of presenting multiple options led to more discussion content in Wikum+ (Table 4.9, Average Words Added – Wikum: 1151.7, Control: 608.5), more summary content in Wikum+ as the discussion ideas were all synthesized (Table 4.9, Average Summary Words Added – Wikum: 436.9, Control: 54.1), and a greater quantity of summary edits in Wikum+, allowing for improvements and multiple iterations of summaries (Figure 4-6 and Figure 4-5).

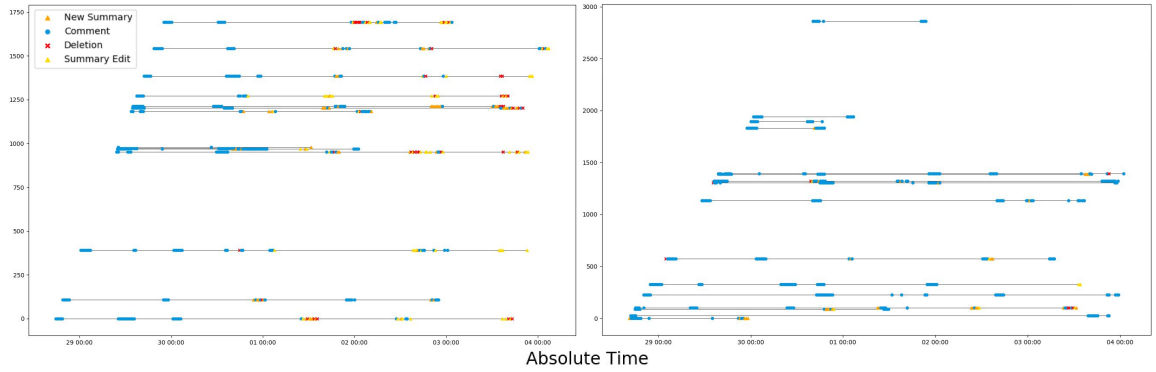

Figure 4-6: Plot of user contributions in the two conditions over time. Each line represents one user and markers symbolize the type of contribution made. The x-axis shows absolute time while the y-axis shows the start times for each user, in minutes from the first contribution made. There were two outlier users, one from each group, who both started after Day 3 of the study and made ≤ 5 total contributions that are omitted for better visual spacing on the graphs.

Users in Wikum+ presented the final proposal as a full list of options, whereas Doc users went with a majority vote, leading to a lowered sense of inclusiveness. Two distinct summaries were created in the Doc for the first Task point, one in favor of a mandatory Pass or No Record (P/NR) grading scheme and the other in favor of making this grading system opt-in by student. However, a majority of users agreed with the mandatory P/NR idea, resulting in more users contributing to writing that summary. Meanwhile, only 3 users voiced that they preferred the optional P/NR idea, resulting in an unfinished alternative summary, as fewer users worked on it. After users voted on which summary was preferred, one user took the initiative to move the minority summary out of the proposal and to the bottom of the Doc.

One user who ranked levels of inclusiveness low in the post-study survey said: "I

felt like I was in a dissenting minority. On the topic of grades, the majority group didn't take into account the full extent of pros/cons. Generally, I felt like the doc was written from a perspective that very much was not mine, so it was hard to write much." As a result, some users in the control said that they stuck with "edit[ing] things related to grammar, sentence structure, and clarity," instead of contributing to writing up an opinion they didn't agree with.

Running a one-way ANOVA test on post-study survey metrics showed a significant difference between levels of inclusion among users in each group ($F = 6.12$, $p = 0.019$), with Wikum+ users feeling more included than Doc users did (Figure 4-7). One Wikum+ user said that "every comment had responses to it, either agreeing or adding on points, or bringing another viewpoint. They were all discussed, including my comments, so I felt included."

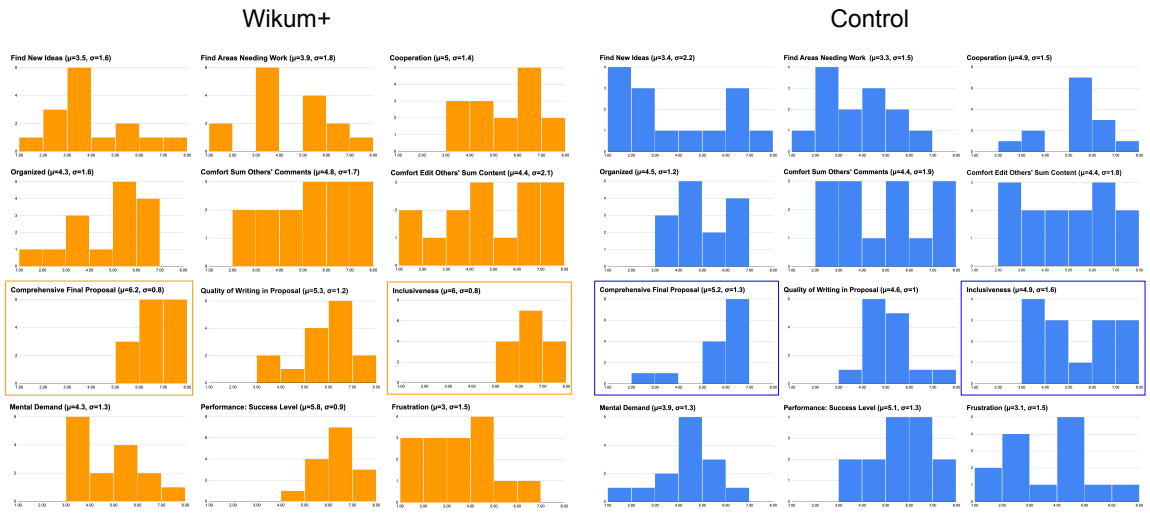

Figure 4-7: Histograms showing contrasting metric values (out of 7) from the large group conditional post-study surveys. Metrics that resulted in significantly different means from one-way ANOVA tests are boxed.

With every side's opinion expressed in Wikum+, more users could collaboratively summarize the discussion, leading to a greater sense of comprehensiveness. Users in the control whose opinions were voted out had a harder time contributing to the final proposal. One user said: "The document wasn't finished... there seemed to be a lot of unresolved comments where people voiced disagreement with parts of the first half. Towards the end, though, it seemed mostly

like a one to two person effort, so I think this is where the quality of writing falls off...the critiques were not acted upon." With too much time spent on the grading aspect of the Task, coupled with minority voices losing engagement and critiques being unaddressed, the control group's final proposal lacked completeness and the variety of ideas that were deliberated in conversation.

Compared to the control group, Wikum+ users felt that their proposal was significantly more comprehensive in content ($F = 6.61$, $p = 0.016$, Figure 4-7). The Wikum+ group made use of flags marking certain summaries as incomplete or biased to improve upon summaries cooperatively. Although the results were not statistically significant, participants in Wikum+ indicated a higher level of comfort on summarizing other users' comments, on average, than those in the control (Figure 4-7 – Wikum+ average comfort: 4.8; control average comfort: 4.4). One participant said that "users took it upon themselves to try and present information in equal light. Later, edits were committed to try and clear up the record and preserve a neutral voice." Another user said that "adding comment references and the color scheme of unsummarized comments played a huge role in making sure someone reads everything in the sub-tree when they're summarizing a thread." As a result, Wikum+ users incorporated a diverse set of ideas and opinions into the sub-tree summaries, which were then translated into the final document.

Chapter 5

Discussion

5.1 Creating Meaningful Artifacts From Conversation Through Collaborative Summarization

In our work, we introduce collaborative summarization as a way to synthesize a conversation into a valuable final document, while still encouraging continued discussion and improvements on both the document and the incremental summaries made along the way. We believe that this gradual summarization process can be applied to many different discussion platforms that exist already today. For instance, Reddit users on a job-searching forum could work together to both discuss and summarize the lessons learned and advice given by the many users who share their diverse experiences. Newcomers to the page can quickly gain the insights from the current artifact, contribute to existing discussions in the forum, or help build the artifact from new comments created.

Current methods for handling lengthy online discussions include filtering, voting, and moderation, all of which can introduce biases or leave out comments without a discussion as to why they were not included. Summarization on Wikum+ distinguishes itself as a solution by making the synthesis process more collaborative and encouraging group deliberation in an organized fashion. While the option for moderators exists, it is not required; and although the majority idea may still surface in

the end, anyone can trace back to the conversation on why that idea was selected.

5.2 Greater Structure in Deliberation

With a free-form discussion, such as one that might occur on Google Docs, the need for structure will grow with the size of the conversation, particularly for the purpose of creating an artifact. In the control groups of our studies, we saw participants use threaded bullet points, Doc commenting functionalities, and by far the most prevalent, voting for the majority option. Usually, these solutions are introduced by a couple of users who want to bring more organization to the conversation, and default to the familiarity of votes or tools that the interface provides. But as we saw with one group in User Study 1, a lack of someone in the group who is willing to introduce structure or improve organization can lead to a conversation that dies off or fails to reach its end goal.

If instead, a method for organizing the discussion is set in place before it occurs or – better yet – already exists inherently on the interface, as in Wikum+, this could reduce users' efforts during the process and help them in creating the final artifact without defaulting back to voting. As one user put it, on Wikum+, "everything was already naturally organized into threads" and "already had a discussion and summarization method built in." This "reduced our efforts because the document-building process happened gradually,...allow[ing] the final summary to be compiled quite easily." In our within-subjects experiment in User Study 1, we also saw that many users learned the threaded-discussion and gradual summarization process of Wikum+ and applied the same strategies later on when working in the control condition. One user said the their group "learned to break up [the] different sections of the report and work on those sections and then eventually take those comments, summarize them, and put them all back together into one essay." Wikum+, therefore, not only provides an interface for integrating discussion and summarization to create a consequential document, it also introduces a method for organizing conversations on *any* online discussion platform to achieve a similar result.

5.3 Signaling In Non Linear Discussions

Chat interfaces have the advantage of always showing the newest content at the bottom, so users know where to find the most recent messages. However, as one user explained, "sometimes, messages just get lost in the chat...it felt like we could only talk about one topic at a time, and if we moved on, we couldn't go back." Beyond the limited linear nature of chats, users who are not interested in or less familiar with the current issue discussed have a harder time participating in the conversation. Threaded discussions instead allow for multiple discussions to occur at the same time. Unfortunately, this often means that new content becomes hard to find, as replies can sprout from anywhere on the page.

Symbols on the page can help signal important information such as where new comments have been made or areas that still need improvement, allowing for more productive discussions. In the case of creating a final document artifact, these signals can also allow users to add on to existing discussions, raise new issues, and make continuous improvements to the document. In Wikum+, we use blue comments, partial summaries, and flags to help users find areas needing work and bold font to help users find new, unread comments and summaries. As one user puts it, "in Wikum+, it felt more like multiple conversations were happening in parallel, but in a way that I could follow multiple at a time." They added that in Wikum+, "it was easier to visualize the different comments at once under each sub-topic...which made it easier to track points that were already addressed. It was also easier to transition from comments to small summaries and work into a bigger summary."

5.4 Greater Inclusion Fosters More Participation

As we saw in User Study 2, when participants began disagreeing on one of the Task points, they inadvertently cut short their chance for a more full-fledged ideation period to suggest and deliberate alternative solutions. This led to the minority voice being drowned out and lower participation levels from that group in forming the final

proposal. Instead, the Wikum+ group acknowledged both original ideas that caused a division in the control by updating summaries to include both ideas as options. Then, users felt more included and generated a much broader range of solutions that were all discussed, with many surfacing as options in the final proposal. Furthermore, in Wikum+, most summaries were created a list of bullet points, allowing for quick edits. In the Doc, full paragraph summaries made it harder for users to jump in and make changes beyond grammar and clarification, not willing to break the flow of the sentences. More generally, synthesis of a discussion through summarization may allow for more ideas to be incorporated as options due to the neutral nature of summaries. This may allow for a greater sense of inclusion, which can result in increased collaboration in generating more ideas and in working towards the final artifact.

5.5 Wikum+

We created a system integrating collaborative real-time discussion and summarization, allowing for the creation of a document synthesized directly from the generated ideas and deliberation. Through the process of interleaving summarization, users are able to continue discussion even after summaries have been made in a discoverable way, allowing for iterative improvements to be made on summaries that users deem incomplete, biased, or poorly-written. We provide users with situational awareness through read-markers to help them find newly added ideas, tags to help them visually group related comments, flags and partial-summaries to help them find areas needing further work, an outline view to help them find various topic threads in the conversation, and notifications to help them follow-up on parts of the discussion. The threaded, flexible structure of the interface, coupled with recursive summarization, allows users to stay organized while gradually working towards the final document.

In our evaluation, we found that our design choices succeeded in helping users locate both new ideas and topics requiring work on Wikum+. We also found that the workflow of Wikum+ was adaptable to other interfaces and kept users organized

throughout their discussion and document-creation process. Our results showed that interleaving summarization allowed for iterative improvements on the summaries and the final document. Furthermore, with larger groups and an increased diversity in ideas, we provide evidence that users on Wikum+ felt a greater sense of inclusion as they managed to incorporate conflicting ideas into more neutral and more comprehensive summaries.

5.6 Future Work

5.6.1 Integration of a Document & Collaborative Editing

In our studies, we found that users wrote in a more flowing manner in the Doc than in the Wikum+ summaries, despite being given the same Task prompt. We hypothesize that users are more accustomed to summaries being concise and list-like, therefore leaving their final proposal written in that style. To formalize the final artifact in Wikum+ and potentially help users create a more cohesive document from their top-level summaries, we would like to integrate an actual document into the interface that users can switch their view to. This document could consist of exactly the contents of the top level summaries or perhaps a single final top-level summary. Users would be able to toggle which version they see, but updating the summary in the original view would automatically update the final document view as well, and vice-versa. This would allow users to focus on writing the document directly when they reach that stage of the process, while still being able to trace back to the original discussion that resulted in the document's ideas.

We currently use locks to prevent users from editing the same summary at once, which may result in race conditions. However, we saw, especially during the longer synchronous portion of the pilot study, that a lack of real-time multi-user edits prevented the group from working on the final proposal together. Furthermore, with the integration of a document, the need for real-time collaborative summary editing increases. Overall, implementing this feature into Wikum+ allow for greater efficiency

and simulate collaborative document writing even more closely.

5.6.2 Wikum+ for a Greater Range of Conversation Types

In all of the studies we conducted, users were asked to create a proposal after discussing a topic on how to improve an existing system. While the topic on online classes in quarantine resulted in some levels of disagreement, we would like to explore how Wikum+ performs for more controversial topics. In particular, we are interested in seeing how summarization and deliberation, instead of voting or moderation, affects how quickly a group reaches consensus, or perhaps dissensus. Given that artifact-creation from conversation is a primary advantage of Wikum+, we hope to also study how more casual conversation topics or advice-type threads might find the interface useful for collecting thoughts.

Many platforms like Kialo [21], ConsiderIt [8], and Deliberatorium [22] structure and organize their argumentation process in an interaction much richer than voting, but doesn't allow for the free-from discussion that Wikum+ provides. It is possible that on Wikum+, users engaging in civic discourse could begin forming similar pros-and-cons lists as enforced by ConsiderIt. Wikum+ users engaging in a debate could use tags to color-code their replies as supporting evidence or refutations. We hope to explore how different types of conversations on Wikum+ may naturally organize themselves into productive strategies for reaching resolution or creating valuable documents.

5.6.3 Measuring the Effects of Greater Enforcements on Time and Roles

In our studies, we primarily analyzed our first research question: "How does integrating discussion alongside summarization affect the quality of collaboration and the quality of the final document?" In other research questions, we wanted to see how enforcing greater structure in the discussion versus summarization process and enforcing specific roles for users might affect the experiences of the users and the

outcome of the final document. We hope to conduct further studies to explore the effects of varying these different factors.

Chapter 6

Conclusion

On today's online discussion platforms, it is too easy for conversations to grow out of hand with the current tools we use. Existing solutions such as filtering, voting, or moderation can suppress minority ideas or reveal biases. Also there's no way to synthesize the ideas so users have to move to a separate document writing software, meaning that users can't refer to the conversation in their document.

In this work, we introduce Wikum+, a system we developed for creating something meaningful from conversations by integrating discussion and synthesis on the same platform. This allows ideas to maintain their provenance as they move from original comments into the final document. We ran one pilot study, allowing us to improve upon Wikum+'s features. We then ran one within-subjects study consisting of 6 smaller groups of 5-6 users and one between-subjects study consisting of 2 larger groups of 15-17 users. Overall, these studies suggest that Wikum+ helps promote inclusiveness, organization, and light-weight coordination compared to a control. It also suggests that the collaborative summarization on Wikum+ allows for more iterative improvements to documents and promotes the incorporation of new ideas. We hope that Wikum+ can inspire new methods and interleaving summarization on existing conversation platforms for synthesizing discussions in an organized, collaborative manner, creating valuable artifacts across the Internet's plethora of discussions.

Bibliography

[1] Mark S Ackerman, Anne Swenson, Stephen Cotterill, and Kurtis DeMaagd. I-diag: from community discussion to knowledge distillation. In *Communities and technologies*, pages 307–325. Springer, 2003.

[2] Arguman. Arguman - argument analysis platform, 2020. `https://en.arguman.org/`, Last accessed on 2020-04-20.

[3] Eevi E Beck. A survey of experiences of collaborative writing. In *Computer supported collaborative writing*, pages 87–112. Springer, 1993.

[4] Osvald M Bjelland and Robert Chapman Wood. An inside view of ibm's' innovation jam'. *MIT Sloan management review*, 50(1):32, 2008.

[5] bluenove. Assembl, 2020. `https://bluenove.com/en/offers/assembl/`, Last accessed on 2020-04-20.

[6] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. Antisocial behavior in online discussion communities. In *Ninth International AAAI Conference on Web and Social Media*, 2015.

[7] Civilized Discourse Construction Kit, Inc. Discourse - civilized discussion, 2020. `https://www.discourse.org/`, Last accessed on 2020-04-20.

[8] Consider.it. Consider.it, 2020. `https://consider.it/`, Last accessed on 2020-04-20.

[9] Kate Crawford and Tarleton Gillespie. What is a flag for? social media reporting tools and the vocabulary of complaint. *New Media & Society*, 18(3):410–428, 2016.

[10] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. Social coding in github: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 1277–1286. ACM, 2012.

[11] Democracia en Red. Democracyos - change the tool, 2017. `http://democracyos.org/`, Last accessed on 2020-04-20.

[12] Django. Django channels, 2020. `https://channels.readthedocs.io/en/latest/`, Last accessed on 2020-04-20.

[13] Steven Dow, Julie Fortuna, Dan Schwartz, Beth Altringer, Daniel Schwartz, and Scott Klemmer. Prototyping dynamics: Sharing multiple designs improves exploration, group rapport, and results. pages 2807–2816, 05 2011.

[14] Idoia Elola. Collaborative writing: Fostering foreign language and writing conventions development. *Language Learning & Technology*, 14(3):51–71, 2010.

[15] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1175–1184. ACM, 2010.

[16] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[17] Rik Hunter. Erasing "property lines": A collaborative notion of authorship and textual ownership on a fan wiki. *Computers and Composition*, 28(1):40–56, 2011.

[18] Jane Im, Amy X Zhang, Christopher J Schilling, and David Karger. Deliberation and resolution on wikipedia: A case study of requests for comments. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):74, 2018.

[19] David G Jansson and Steven M Smith. Design fixation. *Design studies*, 12(1):3–11, 1991.

[20] Greg Kessler, Dawn Bikowski, and Jordan Boggs. Collaborative writing among second language learners in academic web-based projects. *Language Learning & Technology*, 16(1):91–109, 2012.

[21] Kialo. Kialo, 2020. `https://www.kialo.com/`, Last accessed on 2020-04-20.

[22] Mark Klein. How to harvest collective wisdom on complex problems: An introduction to the mit deliberatorium. *Center for Collective Intelligence working paper*, 2011.

[23] Nicholas W Kohn and Steven M Smith. Collaborative fixation: Effects of others' ideas on brainstorming. *Applied Cognitive Psychology*, 25(3):359–371, 2011.

[24] FS Lam and Martha C Pennington. The computer vs. the pen: A comparative study of word processing in a hong kong secondary classroom. *Computer Assisted Language Learning*, 8(1):75–92, 1995.

[25] Cliff Lampe and Paul Resnick. Slash (dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 543–550. ACM, 2004.

[26] Loomio Cooperative Limited. Loomio - make decisions without meetings, 2020. https://www.loomio.org/, Last accessed on 2020-04-20.

[27] Madrid City Council. Decide madrid, 2020. https://decide.madrid.es/, Last accessed on 2020-04-20.

[28] Richard L Marsh, Joshua D Landau, and Jason L Hicks. How examples may (and may not) constrain creativity. *Memory & cognition*, 24(5):669–680, 1996.

[29] Gilad Mishne, David Carmel, Ronny Lempel, et al. Blocking blog spam with language model disagreement. In *AIRWeb*, volume 5, pages 1–6, 2005.

[30] Kevin K Nam and Mark S Ackerman. Arkose: reusing informal information from online discussions. In *Proceedings of the 2007 international ACM conference on Supporting group work*, pages 137–146. ACM, 2007.

[31] Ani Nenkova and Kathleen McKeown. A survey of text summarization techniques. In *Mining text data*, pages 43–76. Springer, 2012.

[32] Bernard A Nijstad, Wolfgang Stroebe, and Hein FM Lodewijkx. Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of experimental social psychology*, 38(6):535–544, 2002.

[33] Eli Pariser. *The filter bubble: What the Internet is hiding from you.* Penguin UK, 2011.

[34] Paul B Paulus, Nicholas W Kohn, and Lauren E Arditti. Effects of quantity and quality instructions on brainstorming. *The Journal of Creative Behavior*, 45(1):38–46, 2011.

[35] Simon T Perrault and Weiyu Zhang. Effects of moderation and opinion heterogeneity on attitude towards the online deliberation experience. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 17. ACM, 2019.

[36] Polis Technology Inc. Polis, 2020. https://pol.is/home, Last accessed on 2020-04-20.

[37] Vincent Price, Joseph N Cappella, and Lilach Nir. Does disagreement contribute to more deliberative opinion? *Political communication*, 19(1):95–112, 2002.

[38] Quora. Quora, 2020. https://quora.com/, Last accessed on 2020-04-20.

[39] reddit Inc. Reddit, 2020. https://reddit.com/, Last accessed on 2020-04-20.

[40] Rodney P Rice and JT Huguley. Describing collaborative forms: A profile of the team-writing process. *IEEE transactions on professional communication*, 37(3):163–170, 1994.

[41] Alexander M Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*, 2015.

[42] Pao Siangliulue, Kenneth C Arnold, Krzysztof Z Gajos, and Steven P Dow. Toward collaborative ideation at scale: Leveraging ideas from others to generate more creative and diverse ideas. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 937–945. ACM, 2015.

[43] Pao Siangliulue, Joel Chan, Steven P Dow, and Krzysztof Z Gajos. Ideahound: improving large-scale collaborative ideation with crowd-powered real-time semantic modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 609–624. ACM, 2016.

[44] Steven M Smith, Thomas B Ward, and Jay S Schumacher. Constraining effects of examples in a creative generation task. *Memory & cognition*, 21(6):837–845, 1993.

[45] Candace Spigelman. *Across property lines: Textual ownership in writing groups*. SIU Press, 2000.

[46] Neomy Storch. Collaborative writing: Product, process, and students' reflections. *Journal of second language writing*, 14(3):153–173, 2005.

[47] Amy X Zhang, Lea Verou, and David Karger. Wikum: Bridging discussion forums and wikis using recursive summarization. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 2082–2096. ACM, 2017.