CorpusStudio: Surfacing Emergent Patterns In A Corpus Of Prior Work While Writing

Hai Dang*
HCI+AI
University of Bayreuth
Bayreuth, Germany
hai.dang@uni-bayreuth.de

Chelse Swoopes*
SEAS
Harvard University
Cambridge, Massachusetts, USA
cswoopes@g.harvard.edu

Daniel Buschek
Department of Computer Science
University of Bayreuth
Bayreuth, Germany
daniel.buschek@uni-bayreuth.de

Elena L. Glassman
SEAS
Harvard University
Cambridge, Massachusetts, USA
glassman@seas.harvard.edu

A **Ordered distribution** over section title clusters can be expanded on click

B **Text editor** tab key retrieves corpus sentences related to cursor position & local textual context

C **Many retrieved sentences** with automatic mark up + user bookmarks & notes

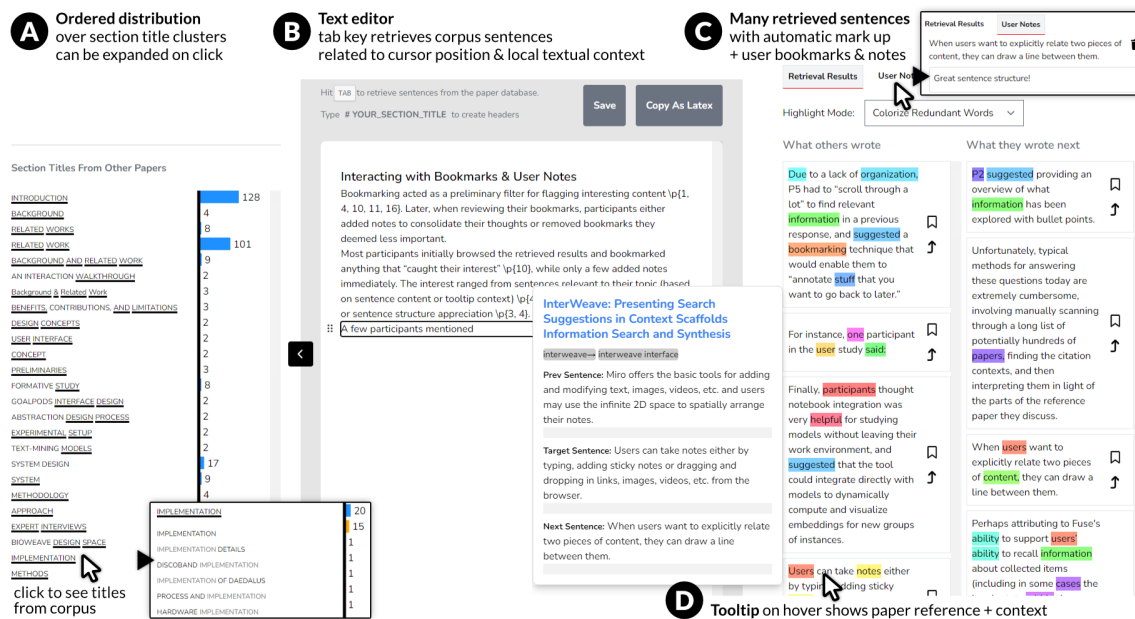**D** **Tooltip** on hover shows paper reference + context

Figure 1: Given a corpus of papers written for the same or similar audiences, e.g., papers previously published at ACM UIST, CorpusStudio supports writers by making visible the writing choices of previous authors in the corpus. To help the writer recognize common and uncommon paper structures, the left sidebar *(A)* shows an ordered distribution of clusters of section titles in the corpus using Positional Diction Clustering [15]. Informed by *(A)*, writers can draft their own outline in the center text editor *(B)*. When fleshing out their outline with prose, to potentially see emergent patterns in previously written papers, writers can press TAB to retrieve analogous sentence examples from the corpus *(C)* based on their cursor's location within their own draft. To reveal emerging patterns, the writer can select different modes of highlighting commonalities and variation across retrieved sentences. The writer can hover over a retrieved sentence *(D)* to see more of the context in which it appeared, and save, annotate, and share retrieved sentence examples that they think fulfill a purpose particularly well or poorly.

*Equal contributions.

Abstract

Many communities, including the scientific community, develop implicit writing norms. Understanding them is crucial for effective communication with that community. Writers gradually develop an implicit understanding of norms by reading papers and receiving feedback on their writing. However, it is difficult to both externalize this knowledge and apply it to one's own writing. We propose two

new writing support concepts that reify document and sentence-level patterns in a given text corpus: (1) an ordered distribution over section titles and (2) given the user's draft and cursor location, many retrieved contextually relevant sentences. Recurring words in the latter are algorithmically highlighted to help users see any emergent norms. Study results (N=16) show that participants revised the structure and content using these concepts, gaining confidence in aligning with or breaking norms after reviewing many examples. These results demonstrate the value of reifying distributions over other authors' writing choices during the writing process.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **Text input**; • **Computing methodologies** → **Natural language processing**.

Keywords

writing assistance, natural language processing, text visualization

1 Introduction

Inexperienced researchers often lack knowledge of the implicit writing conventions and structural expectations of particular communities. While existing writing guides offer general advice, they are often too abstract and high-level, making it challenging for junior researchers to apply this guidance in their own work.

Given that people commonly learn through examples, providing access to relevant examples can help bridge this gap. However, there are obstacles: viewing too few examples may lead to overfitting and unintentional plagiarism, while viewing too many examples—without sufficient cognitive supports within the interface—may overwhelm the writer.

Prior work motivated the design of writing support tools with social cognitive theories and cognitive apprenticeship theory [18, 19] to help writers learn complex writing skills in a professional setting. While this approach allows novices to learn by mimicking model examples, it also requires expert knowledge to create and annotate these examples, which is often time-consuming and not readily available for many communities.

Learning from examples has also been extensively covered in the language learning literature. Non-native English speakers (NNES) may hesitate to choose from AI-suggested content because they struggle to evaluate the content's fluency and naturalness [21, 24, 25]. To address this problem, researchers proposed that novices learn by inspecting sentences from a curated corpus such as the Corpus of Contemporary American English (COCA)[1] or English dictionaries [4] to help students learn the correct use of collocations and phrases [35]. To this end, *LangSmith* proposed a writing support tool for drafting academic NLP papers [20] by generating grammatically correct phrases given the user's current draft along

with prior section and paper titles. In contrast, our work surfaces existing text from a relevant peer-reviewed text corpus. Furthermore, by avoiding automatically generated text that may—without the writer's knowledge—cross the line into plagiarism if used verbatim, the writer only sees previously published example sentences and their provenance, retaining informed control.

To this end, we implemented CorpusStudio, a writing environment with two novel writing support concepts, one at the document level and one at the sentence level, that help writers learn about the writing styles and expectations of particular audiences. These writing supports are designed to expose writers to as many examples as possible to increase the variety of examples shown to the users—with the cognitive support necessary to help them see the emergent patterns that capture otherwise implicit community-specific writing content and style expectations. Specifically, at the document level, CorpusStudio computes and renders an *ordered* distribution [15] over section title sequences in prior papers from the community the writer wishes to write for. This gives the writer a sense of common and uncommon section titles and in what relative order they appear in prior papers (Figure 1A). At the sentence level, CorpusStudio retrieves many example sentences from the corpus given the user's draft-in-progress and their current cursor location within it (Figure 1C). Words across these many example sentences are algorithmically highlighted or de-emphasized to help users see any emergent norms and outliers. We also allow authors to capture their thoughts and deductions from viewing these examples with bookmarks and notes.

We conducted a user study with N=16 participants from the academic community who were actively writing their manuscripts. They used CorpusStudio to draft their paper outline, wrote one to two sections of their manuscript, and answered questionnaires eliciting their overall experience writing within CorpusStudio.

Our results reveal that users felt more confident in their writing when they discovered that other people in their target community have written in a similar way. Inspecting many examples at once was useful for identifying broad structural patterns, especially for sections that typically have more common structure, such as the description of the system implementation details or participants' demographics.

In summary, we contribute the following:

- We designed two novel writing support concepts that make visible emergent patterns at the document-level and sentence-level of a given paper corpus.
- We implemented these writing support concepts within CorpusStudio, an interactive writing environment that helps writers retrieve, inspect, generalize from, and reflect on many examples from their target venue
- We found that users' writing goals guided their perception of relevance and how they adapted content from the corpus.

2 Background and Related Work

In this section we provide a background on writing norms in different communities and how members learn about these norms. We then give an overview of interactive tools that support writing and tools for finding related work.

---

[1]https://www.english-corpora.org – accessed 12.09.2024

2.1 Community Writing Norms

2.1.1 The Importance of Understanding Writing Norms. Flower [12] highlights the difference between *writer-based* and *reader-based* prose, arguing that the former represents the writers' unrefined thoughts while the latter requires their careful *transformation* with the reader goals in mind while ensuring a *shared language* and *shared context*. Prior writing literature emphasizes the importance of writing for an audience [1, 16, 45], but writing for academic or professional communities can be especially difficult because they have implicit writing expectations and failing to address them may confuse its members or waste their time with "unfocused and pointless discussions" [12].

2.1.2 Aspects of Writing Norms. Writing norms are comprised of many aspects, including the language, content, organization of information, and scope of the written document. Although academic venues may provide a high-level overview of the expected content and style of writing such as encouraging authors to write "clearly and concisely, avoid jargon, organize the submission to flow logically and smoothly, provide the right level of detail"[2], it can be challenging for novice writers to apply these suggestions in their own writing, e.g., *What is the expected level of detail for that audience?*

2.1.3 Learning about a Community's Writing Norms. To support writing in an academic context, prior research has proposed the *cognitive apprenticeship* theoretic approach to teach novices complex cognitive tasks by mimicking experienced members of the community [8]. Drawing from these ideas, our work instantiates two strategies of cognitive apprenticeship, namely modeling and reflection.

2.2 Writing with External Text

2.2.1 Writing with Retrieved or Generated Additional Text. Many writing support tools facilitate writing by suggesting content that users did not write themselves. Sentence completions are a frequent form of writing assistance [2, 28]. However, direct content generation can have negative impacts such as diminishing the author's self-expression and ownership over the final text outcome [30]. Related work has also looked into providing writing inspirations [6, 14, 43].

We build on this line of research by surfacing prior text as the user writes new text (e.g., Naeem et al. [37] and *IntroAssist* [18]). A critical difference is the much larger scale of prior text surfaced to the user and post-processed for emergent patterns in our work. For example, *IntroAssist* [18] presents expert-curated email examples to help users draft their help-seeking message in a professional context, while our approach uses computational methods to show many examples drawn from a large text corpus of previously published papers.

2.2.2 Writing with Text from a Corpus. Hui and Sprouse [19] proposed a writing support tool to help college students write in a professional context assisted by *scaffolded annotations*. The authors proposed a *checklist* with multiple annotations that provide cues about the content to include in the professional email. This checklist

is complemented with *model texts* where sections of that text have been annotated with the elements of the checklist. This approach helps users understand the overall structure and expected content while also demonstrating concrete instantiations of the high-level guidelines. Our concepts help users plan the structure and content of their writing by showing many concrete examples of how other members of their target community have written similar texts.

2.3 Finding Relevant Examples

As previously described, writers learn from inspecting examples, but finding related work with the purpose of writing support is challenging. Search engines such as Google Scholar, the ACM Digital Library, Mendeley, and Semantic Scholar typically surface documents based on topical similarity while writers need specific analogous snippets of text from the manuscript to inform their writing. Current tools also help researchers understand how their own work relates to other work [29] by contextualizing embedded citations [5], drawing connections to follow-on work directly in the paper document [40], and making recommendations based on a paper's metadata such as citations and authorship [7, 23, 39, 42] or its semantic similarity [3, 17, 23]. In contrast, our design focuses on finding related work for the purpose of writing support, i.e., that helps writers identify existing writing conventions from a specific community.

Most interactive research tools either support analyzing related work [26, 29, 33, 38] or writing with computational help [9, 28]. In existing contexts, writers often need to navigate away from their writing environment to search for related work, scroll through the target document, select and process relevant text structures and draw conclusions about the implicit writing conventions encoded in the target document. Performing all of these steps leads to abrupt context switches and breaks the writing flow. Our approach proposes an integrated environment to support writing and searching for model text in the same interface, thereby minimizing the number of context switches and streamlining the writing and research process.

3 Corpus Studio

We designed CorpusStudio to support writers in learning about and applying community-specific writing norms. This is achieved primarily through novel mechanisms for the retrieval and rendering of content from a corpus of existing documents written by and for that community.

3.1 Design Goals

Our overarching design goal is to *empower users with evidence of community-specific emergent norms and outliers* at the document and sentence level as they write their own document for the same audience. In the process of fulfilling this overarching goal, our goal is to reduce, not facilitate, unintentional human- or AI-generated plagiarism.

3.2 Usage Scenario

Alice has been working on a new interactive writing support system and is drafting her manuscript for submission to an HCI venue. She has some prior experience writing academic papers but is unfamiliar

---

with writing papers with a focus on a system contribution. She opens CorpusStudio to learn more about her target venue's writing expectations while working on her draft.

Inspecting the section titles. She looks at the ordered distribution over section titles. As shown in Figure 1A, she expands the cluster named *Implementation*, which typically comes after *System Design* and around the same location as *Methods* (when those sections exist in a given paper). In the expanded cluster, she observes how other authors have phrased this typical section's title, i.e., more commonly *Implementation* and less commonly *Implementation Details* or *Implementation of <System Name>*. Although she likes the idea of adding a system name to the title, she decides to stick with the more generic commonly chosen option for now, and finishes her paper outline of section titles in the text editor beside it.

Inspecting the retrieved sentences. In the text editor, she starts drafting the implementation section immediately under the *Implementation* section title, but she does not know how she should introduce it. She writes a few words and, without moving the cursor, hits the TAB key to retrieve related sentences from other work to get a sense of what (and how) other people wrote the beginnings of their implementation sections. She notices that many authors start with *"We implemented* <name of the system> *to do/test* <some hypothesis>*"* while some also state their design rationale.

Developing a personal style sheet. She bookmarks one of the retrieved sentences to remind herself of this observation and adds the note: "The first paragraph of the implementation is often formulaic and with this recognizable structure." She concludes that it is probably well accepted in the community if she starts her paragraph in a similar way.

Continuing Writing. Alice continues to write down her ideas in the text editor and occasionally retrieves example sentences given where her cursor is when she needs inspiration or wants to confirm how other authors have written their manuscripts.

3.3 System Characteristics

In this work, the evidence of community-specific emergent norms and outliers at both the document and sentence level is provided to writers in the form of *extracted examples* from a corpus of documents written for and by their intended audience. To facilitate the writer's recognition of these emergent norms from examples without facilitating the writer's anchoring on any one example too much, we look to theories of human concept learning:

Variation Theory. Regardless of whether a human or a machine is learning from examples, any learning process can fall victim to overfitting, i.e., learn a rule that does not generalize to the rest of the corpus, when trained on too little data. For example, the Variation Theory of Learning [32] was explicitly designed to minimize human overfitting when learning from examples. With only a few examples or suggestions, previous systems may cause users to anchor [46] on just one of them. By instead showing *many* examples, we hope the user may be (1) less likely to anchor to one of the retrieved examples (or even unintentionally plagiarize from it) and (2) more likely to observe community writing norms and possibilities without overfitting to the retrieved examples.

Structural Alignment Theory. While the many retrieved examples in our work are not algorithmically selected on the basis of Variation Theory, the textual examples are ordered and rendered with visual attributes (e.g., grayed-out repeated words or color-coordinated highlights) designed to support the recognition of emergent commonalities and outliers, based on the human cognitive processes of pattern recognition described by Structural Alignment Theory [13].

3.3.1 Text Editor.
Users outline their paper and draft their prose in the central text editor (Figure 1B). The editor supports two text cell types: section title and paragraph.

Borrowing from the syntax of Markdown editors, the editor allows users to create section titles by pre-pending text with #'s where the number of #'s indicate the level of the section title.

3.3.2 Surfacing Document-level Community-specific Emergent Norms and Outliers.
Inspecting outlines from other papers in the target community can help writers develop their own. To help writers understand how papers are commonly (and uncommonly) structured within their target community at the document level, we show an ordered distribution of section title clusters from the entire corpus immediately to the left of the central text editor (Figure 1A *Section Titles From Other Papers*). This ordered distribution is computed by applying Positional Diction Clustering (PDC) [15] to the ordered lists of top-level section titles extracted from each paper in the corpus.

PDC is a structural mapping engine [11] designed to algorithmically emulate at scale the analogical matching process humans do when comparing two or more structurally similar objects, as described by Structural Alignment Theory [13]. In CorpusStudio, the objects are papers, and the resulting order of section title clusters that PDC generates preserves the section titles' approximate relative positions in their respective papers. For example, in Figure 1A, Introduction is the ubiquitous starting point[3] and, in approximately 80% of papers, it is followed by Related Work. The user can also see outliers like the two papers that have a section titled An Interactive Walkthrough soon after.

Section title clusters are shown with a count and a proportionally long bar indicating the total number of section titles in that cluster found across the entire corpus. The most frequent section title in each cluster is chosen as the cluster name. Clicking on a cluster name expands (or closes) a cluster, showing (or hiding) all its members.

We also provide cognitive support for perceiving the commonalities and variations within each cluster of section titles: The underlined words within each cluster name appear in every section title within it (a strict commonality), serving as information scent about the cluster contents when it is closed. Within each cluster, the PDC algorithm dictates the order of section titles to minimize diction differences between section titles in adjacent rows; words that appeared previously in the ordered list of cluster members are greyed out so that variations might be more visible. This text graying strategy is based on the interleaved view of PDC results in Gero et al. [15].

---

[3]Note that the screenshot in Figure 1 was taken with the system running on a subsample of the corpus, explaining why there are e.g. only 128 introductions, extracted from 128 papers.

3.3.3 Surfacing Sentence-level Community-specific Emergent Norms and Outliers. This writing support is intended to provide insights into both established conventions and areas where variation is common at the sentence level. By retrieving and displaying many relevant sentences simultaneously, users can observe how similar content is expressed across different papers. These distributions can reveal both rigid conventions (narrow distributions) and areas of flexibility (wide distributions). For instance, participant descriptions in user studies often follow stereotypical patterns, while discussion sections typically show more variation to accommodate novel findings.

Spatial Retrieval. At the core of our approach is a novel retrieval mechanism we call *analogous spatial retrieval*, which integrates context, content, and spatial information. Users can request this retrieval of sentences from other papers by hitting the TAB key as they are drafting their text. Given where the user's cursor is placed in the text editor at the moment of retrieval (i.e., the text of the currently focused paragraph cell and the title of the section containing it), we retrieve and display many potentially analogous sentences from similarly titled sections of papers in the corpus (Figure 1C *What others wrote* column) as well as the subsequent sentence in each paper (Figure 1C *What they wrote next* column). See Section 3.4 and Section 4.2 for more spatial retrieval implementation details.

To further help writers "peek into the future," that is, explore what other authors wrote later in a given section, users can insert empty paragraph cells after a non-empty paragraph cell and then trigger sentence retrieval. The number of empty cells corresponds to the offset o (in sentences) to be retrieved from the corpus. Retrieving content with an offset works by first retrieving all sentences that would have been retrieved if the cursor had been in the nearest preceding non-empty paragraph cell but then retrieving and displaying the o-th subsequent sentences.

Scale of Retrieval. Following Gentner's work on analogical learning theory [13], the comparison of many example sentences promotes the extraction of common principles or abstract schemas, which may be more likely to be transferred to future writing tasks. We experimented with displaying 10, 25 and 50 sentences at once and found that 25 sentences provided enough content for writers to explore and potentially recognize common principles and abstract schemas while not overwhelming them.

Sentence Rendering. However, what writers learn from these examples depends not only on their scale and variety, but also on how the writer mentally processes and compares them [13]. To aid in this mental sentence comparison process, the sentence examples can be rendered with algorithmically assigned visual variables (e.g., color-coordinated highlights) designed to support the recognition of emergent commonalities and outliers (and possibly support the mental cross-sentence alignment process necessary for developing abstract sentence schemas). Figure 2 shows a comparison of the two available but optional rendering modes:

- *Distinctly Colorize Recurring Words.* We created a color map based on the twenty most-frequently occurring words across all retrieved sentences and columns (excluding stop-words), then iterated through each retrieved sentence and colorized
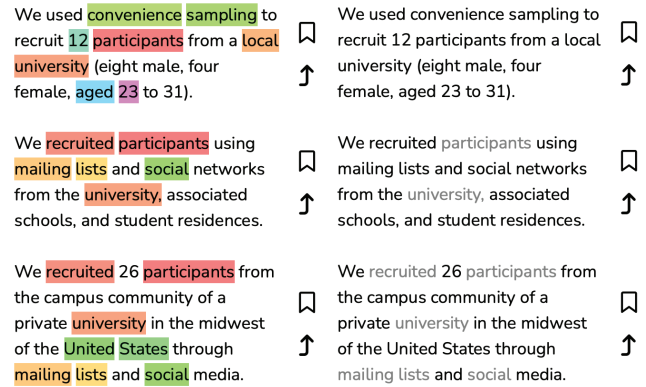

Figure 2: Side-by-side comparison of sentence rendering modes to help the writer identify textual commonalities and variations across retrieved sentences: (left) highlight each commonly recurring word with the same distinct color versus (right) greying out repetitions of words in subsequent sentences
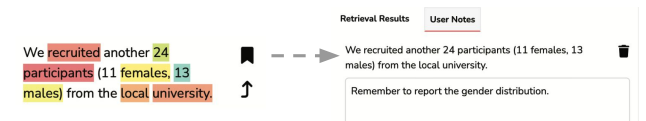

Figure 3: Illustration of the bookmarking feature showing how writers can save sentences and add notes, with the original bookmarked text preserved for context. In this example, the writer added a reminder to note the gender distribution of participants.

the other word occurrences using the color map. This mode is the default.
- *Grey Out Repeated Words.* This mode de-emphasizes previously seen words by iterating through each retrieved sentence and greying out words that have appeared in any of the previous sentences in the result list.

Re-ranking is another way for users to explore relationships across sentences. When users click on the re-ranking icon (arrow below bookmark icon next to each sentence), the selected sentence moves to the top of its column in Figure 1C. CorpusStudio then sorts the other sentences in ascending cosine-distance from that new top sentence. This triggers the system to recompute the greyed-out words to account for the new sorted column.

3.3.4 User Notes. As shown in Figure 3 and the top right callout in Figure 1C, users can bookmark a sentence and then comment on it in the User Notes section, e.g., with a reminder of why it fulfills its purpose particularly well or poorly. These annotated sentence examples can be exported as a csv file to facilitate sharing observations, preferences, and insights with others.

3.3.5 Contextual Tooltips. Writers may want to know more of the context surrounding a retrieved sentence, e.g., to avoid misinterpreting it. As shown in Figure 1D, users can hover over a retrieved sentence to view a floating window showing the paper's title, the hierarchical section title path, and the sentences immediately before and after. The paper title links to the full paper in the ACM Digital Library. Additionally, if any retrieved sentence includes a citation, users can hover over it to display the full citation information, including the authors and the referenced paper title.

3.3.6 Features for Promoting Learning, not Plagiarism. A combination of features are intended to work together to support users in actively engaging with many examples, rather than anchoring on and copying any one. More specifically, showing multiple examples at once may help novice writers make more informed decisions about plagiarism: when sentences are highly similar across papers, this often indicates accepted community writing norms (like in methodology sections), while sections with more varied writing styles suggest areas where similar phrasing should be avoided. In concert with showing many examples, two additional design choices are intended to further facilitate learning over plagiarism: First, we disable the user's ability to copy and paste retrieved sentences into their own draft. Second, by providing visible affordances for saving and annotating their insights, we may promote deeper reflection on the writing patterns they observe.

3.4 Implementation Details

We implemented our system as a web application using SvelteKit.[4] For sentence retrieval, we used WeaviateDB[5] as our vector database, using its built-in text embedding capabilities and approximate k-Nearest Neighbor (kNN) search implementation [34]. The system was deployed on a workstation with 12 GB of GPU memory and made remotely accessible to study participants.

While determining relevance in retrieval systems is inherently complex and depends on users' dynamic writing objectives, querying based on semantic similarity in text vector space is an established approach in retrieval-based systems [47]. Our initial implementation explored hybrid search techniques, following prior work that used Reciprocal Rank Fusion [41] to combine keyword-based and vector-based retrieval results. However, through initial experiments, we found pure vector-based search provided more contextually relevant matches.

4 Data & Processing

We collected 478 papers from two HCI venues (IUI and UIST) which included the proceedings from 2021 until 2023. These venues typically publish HCI system contributions. We completed the data request process with the ACM.

4.1 Pre-processing Documents

We initially used the NLTK module to split each manuscript into sentences, but we experienced many edge cases where extracted sentences were incomplete. We therefore chose an LLM-based approach instructing the model to separate sentences by inserting

---

[4]https://kit.svelte.dev
[5]https://weaviate.io

[EOS] tokens between sentences. For each extracted sentence, metadata was also saved, e.g., the context shown in the tooltips described in Section 3.3.5.

4.2 Embedding Process

To implement spatial retrieval, we follow an approach resembling *LangSmith* [20]. We prepared our corpus to support this functionality by pre-pending section titles to the extracted sentences before embedding the concatenated sequence using OpenAI's 'text-embedding-small-3' model (Figure 4). This approach enables CorpusStudio to retrieve related sentences based on both the section title and the content of the sentence. We used default parameter values and default un-trimmed vector dimensions.

4.3 Extracting an Ordered Distribution over Section Titles

The sequences of top-level section titles for each paper were extracted and organized into a list of lists using positional diction clustering (PDC) [15]. This is identical to the process that produced the data for the Interleaved PDC view in Gero et al. [15], except instead of processing LLM responses as sequences of sentences, we processed papers as sequences of section titles. We also computed how many papers had each section title in order to populate a corresponding histogram in the CorpusStudio interface.

We experimented with different sample sizes ranging from 200, 500, 1000, to 5000 section titles and found that distinctive distribution characteristics quickly emerged even with smaller sample sizes. For our prototype we decided to render and cluster 1000 section titles to demonstrate both the defining characteristics of the section title distribution while still providing a variety of ways authors have organized and phrased their section titles.

5 User Study

To evaluate our two novel writing support concepts (described at the document level in Section 3.3.2 and at the sentence level in Section 3.3.3), we used CorpusStudio to conduct a controlled within-subjects experiment with participants who were, with one exception, actively working towards writing a manuscript for an HCI venue. Participants completed two writing tasks during the user study: (i) drafting or revising an outline of their paper and (ii) drafting or revising section(s) of their paper. We referred to CorpusStudio as a "set of design features" or "prototype" during our user evaluation of the concepts. We used Google Docs as our baseline interface because it is a commonly used writing environment in this context. We aimed to investigate how participants would use the writing supports to achieve their goals, how the supports would affect their writing process, and how they may help them identify norms of a target community.

5.1 Participants

Sixteen participants were recruited through mailing lists and Slack channels from different universities in the USA, Germany, United Kingdom and South Korea (Table 1). Participants had most commonly authored or co-authored 3-5 academic papers already, but two had authored less and four had authored more than twenty (see Appendix Figure 12 for the entire distribution). The average
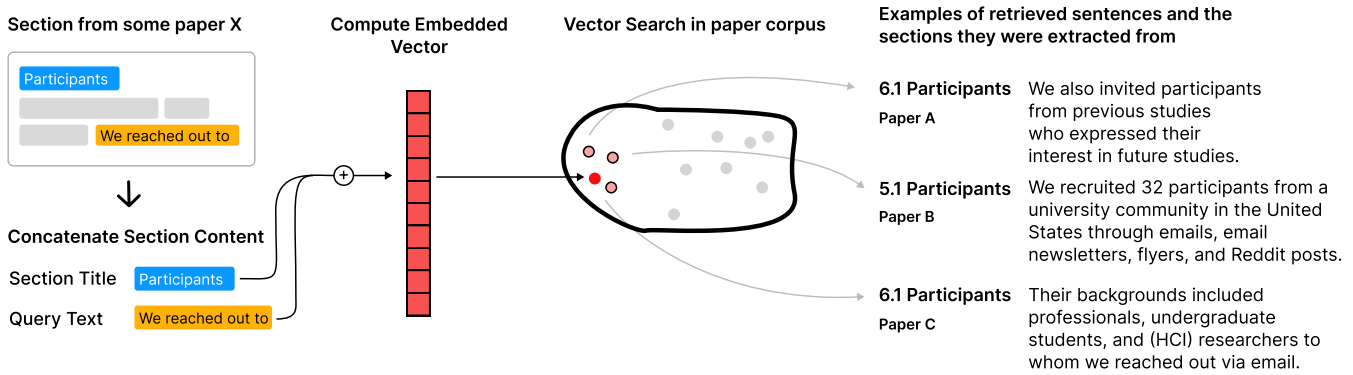
Figure 4: Sentence embedding workflow designed to retrieve analogous content across papers by incorporating section context. The process prepends section titles to sentences before vectorization, enabling the system to find similar content from matching sections (e.g., 'Participants' sections) across different papers, as shown in the three retrieved examples on the right.

Category	Details	
Participants	Total: 16 (8 men, 8 women)	
	Age: 18-24 (4 participants), 25-34 (11), 35-44 (1)	
	Criteria: Over 18 years old, fluent in English, actively writing a manuscript for an HCI venue	
Writing experience	General writing	4.3 years (SD 2.5)
	For target venue	2.6 years (SD 2.8)
Educational background	Undergrad. students	3
	Graduate students	6
	Post-doc	3
	Faculty members	3
	Research associate	1
Target venue	CHI	11
	CHI or NeurIPS	1
	NeurIPS	1
	IUI	2
	Undecided (HCI work)	1

Table 1: Participant information

Need for Cognition (NCS-6) score [31] across all participants was high—4.57 out of 5 (standard deviation: 0.46)—so participants were particularly willing to exert cognitive effort on a given task.

Most participants were targeting HCI venues, and one was targeting a machine learning venue. All participants reported referencing published papers during their writing process: 50% always did, 30% (5/16) often did, and 20% (3/16) sometimes did.

All but one participant had previously used at least one writing support tool. Tools included Grammarly (12), DeepL (3), ChatGPT (4), Google Docs add-ons for reference search (1), Zotero plugins (1), SciSpace (1), and OpenAI's Grammar Checker plug-in (1). Usage frequency varied: 2/16 used them occasionally, 6/16 used them a few times per month, 5/16 a few times a week, and 3/16 once a day or more.

5.2 Study Procedure

All studies were conducted remotely via Zoom and were evenly facilitated between the two lead co-authors. CorpusStudio was accessible as a web application, so participants were not required to install anything. Each study took approximately 60-75 minutes, and participants received $25 (USD) via digital payment (Zelle, Venmo, or PayPal) as compensation for their time.

5.2.1 Consent & Pre-Study Survey. Each participant, after giving informed consent, completed a brief survey about their demographics, experience with scientific communication and writing for a target venue, current strategies for writing, use and experience with writing support tools, and questions relevant to the task that they would be completing. All survey questions are included in Appendix A.

5.2.2 Task Preparation. Participants were informed that CorpusStudio's corpus, from which they would be viewing section titles and retrieved sentences, consisted of papers from two HCI venues, IUI and UIST, between 2021 and 2023. This information was disclosed to ensure participants had a clear understanding of the corpus context, allowing them to engage with the examples accordingly, regardless of their own specific target venue.

Before starting each of the two tasks, participants watched a tutorial video that explained the relevant writing support concept and its supporting features in CorpusStudio. Participants were told that we evaluated the features, not their work.

5.2.3 Task 1: Outline Writing. We chose outlining as our first task because it let us examine how participants structured their work after seeing an ordered distribution over section titles from published papers.

The outlining task had two sub-tasks: drafting section titles for the first and second half of the paper. Participants used either a baseline interface (Google Docs) or the exploratory interface (CorpusStudio) in counterbalanced order. With the baseline, participants relied on their typical resources. With CorpusStudio, they just used

the Figure 1A *Section Titles From Other Papers* feature. Each sub-task had a 5-minute limit. They thought aloud while writing and completed post-task surveys (Appendix A.2, Appendix A.3).

5.2.4 Task 2: Paper Writing. Participants began writing or revising sections of their manuscript for their target venue. They were given the option to bring the outline created in the previous task and any supporting text that they may have already written into this task. Participants were allowed to use any feature of CorpusStudio to help them write. As part of the task, participants were asked to bookmark example sentences that they would want to save and annotate in User Notes for themselves or a mentee. They were given twenty minutes to complete the writing portion and five minutes at the end to revise and update their User Notes; they also were asked to think aloud while completing this task.

Note that, in the tutorial video at the beginning of the task, we included the information about plagiarism and patterns over retrieved sentences described in Section 3.3.6. We also provided the following instruction: "When there is high diversity of sentence content and style, read many sentences before composing your thoughts to avoid mirroring any one sentence directly; this will hopefully help you avoid plagiarism and ensure your work remains authentic."

5.2.5 Post-Tasks Survey and Interview. To conclude the task and study, participants completed a post-task survey (see Appendix A.4) followed by a brief semi-structured interview (see Appendix A.5).

6 Qualitative Results

The two lead co-authors analyzed the user study transcripts with a *Grounded Theory* approach [22]. They separately collected a list of participants' behaviors, approaches, insights, and suggestions about the interface as well as responses to initial open-ended survey questions and the final semi-structured interview questions. To derive high-level themes, both co-authors jointly organized and clustered the codes through affinity diagramming. Throughout and after this process, co-authors discussed their respective codes, resolved assignment conflicts through discussions, and refined the labeling of the clusters. Once both authors agreed on the clusters, they distributed the themes evenly among the authors who individually identified further evidence for each theme in the study transcriptions. Lastly, all authors reviewed and agreed on the final result report.

Overview. Participants reported that seeing retrieved sentences and section titles helped them overcome the blank page problem and continue drafting. They also found it helpful to view multiple contextually relevant sentences from various papers on a single page within a centralized writing environment, which allowed for easier comparison of structure and patterns. They reported that this layout was more efficient and less overwhelming than traditional methods, such as having multiple papers open to compare texts.

Participants used the retrieved sentences and section titles to gather ideas or mimic good writing, and found the tool helpful for tasks such as drafting introductions, starting sections, describing systems, and beginning paragraphs.

Some participants reported that their understanding of a venue's norms remained the same after utilizing the tool, but they still found

value from other aspects such as terminology, phrasing, structure, and ideas. For example, a participant noted that they were inspired to think more critically about how they wrote rather than just following norms.

6.1 Document-level Writing Support

6.1.1 Existing Practices for Creating Section Titles. Prior to Task 1, participants described their methods for drafting outlines. Many saved a list of papers in Google Docs or a local folder, reviewed the individual works, and aligned their section titles with papers they deemed most relevant to their topic. Some participants followed section titles from their own prior work, while others consulted colleagues familiar with their target venue or reviewed papers from the venue to follow a similar structure.

6.1.2 Ways of Reading and Using the Section Titles From Other Papers Feature. Some participants entered the study with an existing notion of the section titles they needed and used the distribution of section titles as a checklist to validate their choices (P_1, P_3). P_{13} found the feature useful for identifying structure, particularly for new writers. Several participants either chose a section title from the distribution and adapted it to align with their writing goals or modified their pre-existing section title based on inspiration from the distribution. P_7 decided to use the feature to find top-level section titles and then create their own subsection titles.

Participants used section title frequency as an indicator of norm and popularity. P_{10} explained, *"I'm less drawn to section [titles] that only have two or one [paper they show up in] because it doesn't make me feel confident that it's common."*

Post-task, participants described scenarios where the feature could benefit their work. While P_1 did not find it valuable for their future writing, others identified uses such as determining paper structure (P_2, P_3, P_4, P_8, P_9, P_{10}, P_{11}, P_{12}), outlining early drafts (P_3, P_4), writing about unfamiliar subject areas or methods (P_5, P_6), writing for a new venue (P_8), having an *"external"* perspective (P_{11}), learning about research (P_{16}), preparing to break the norm in the research (P_{16}), prototyping an outline rapidly (P_{13}), and navigating when they have *"no idea where to start"* (P_{15}).

6.1.3 User Insights on Variation, Norms, and Breaking Norms. Participants noticed variation across section titles, noting that there were different ways to describe conceptually similar sections. Differences ranged from minor, such as singular versus plural titles (e.g., "Related Work" versus "Related Works") to more significant, like adding new details to a more generic title or the existence of a more distinct title indicating a more unusual paper component. Evaluating the distribution of titles helped some participants see an unexpected range of possible titles and exposed them to new options. For example, P_2 noted, *"The section titles are more diverse than I thought. There are things like, you know, 'expert interviews' which I've never done... There were a few others that I found which are pretty new to me... It definitely broadens my sort of—my horizon."*

Some participants were particularly interested in seeing titles that deviated from the norm. P_3 found it helpful to visualize less common titles, which appeared only once or twice, compared to more popular ones with over a hundred instances. Some participants chose to use or adapt titles from the shown distribution, while

P3
Query Context

Formative Study

3.1 Formative Study

We designed the formative study as shown below.

3 Formative Study

The formative study consisted of a semi-structured interview to investigate current strategies and challenges of obtaining images, and two image generation tasks to explore current strategies and challenges of using text-to-image generation.

3 Formative Study

This study involved a tutorial, a 15-day diary, and a semi-structured interview to explore their daily usage and potential challenges.

3 Formative Study

We then identify user challenges and needs for support in this process via a survey study.

...

P4
Query Context

Challenges

While our system is capable of simulating users in mobility, we acknowledge certain challenges.

3.1 Challenges in existing v-imt systems

Our user evaluation results presented in Section 8 also confirm these challenges well.

7 Conclusion

We will move from simulation into a real-world home testbed environment, which poses its own challenges.

8.3 Limitations and discussion

Through user interviews, we found two major challenges regarding the application of our system in a real environment: mechanical actuation noise and user concerns about collisions.

8.2 Limitations on current design

However, some of the limitations we encountered can be at least partially addressed by improvements to real-time tracking and re-registration on mobile devices.
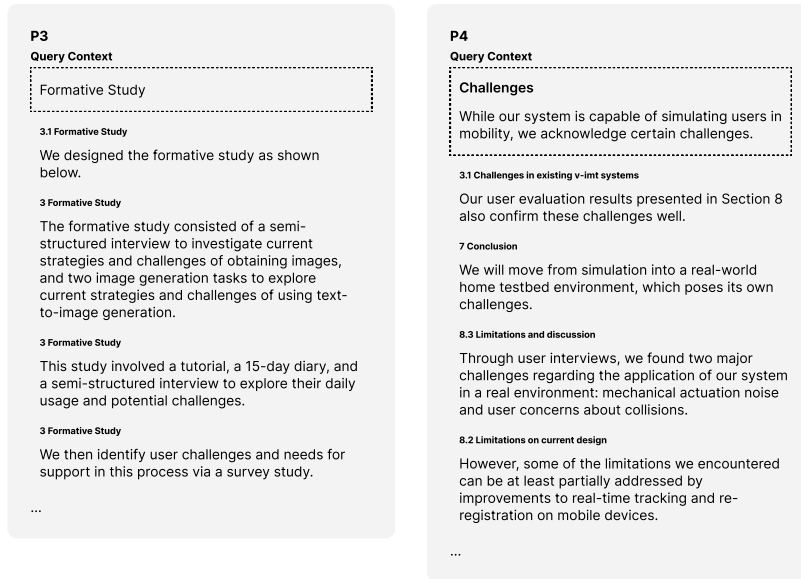
...

Figure 5: Examples of different users' immediate writing contexts when requesting sentence examples, and the first of CorpusStudio's retrieval results. Note that these are representations of system logs, not user interface screenshots. (Left) P3 queried for inspiration on how to introduce a formative study section, retrieving examples of how other papers structured and presented their study introductions. (Right) P4 searched for ways to acknowledge design and study challenges, finding various examples of how other papers discussed limitations and technical constraints.

others chose to intentionally deviate from the perceived norm of the distribution. For example, P_7 decided to continue with their own drafting approach, as their paper was *"more on novel contribution rather than implementation,"* so they believed that frequently used section titles about implementation would not be useful when describing their work. Critically, we observed that participants exercised their agency despite having access to the distribution: If they did not like how a section title was phrased, even if it was frequently used, they would opt for their own desired phrasing.

6.2 Sentence-level Writing Support

6.2.1 Reading and Processing Sentences. We observed that participants read and processed what they had read in two different styles: some did selective reading by skimming across sentences, while others read sentences in their entirety. Regardless of engagement style, participants identified new takeaways. For example, P_2 discovered new ways of structuring sentences through analyzing examples. Some of the (sub)communities represented in the corpus were unfamiliar to P_4 and P_6, and they learned about the habits reflected in that text. P_{14} likes control over the content of her writing, and valued learning about sentence structures that could guide her own composition choices.

Figure 5 shows two examples of system logs (not user interface screenshots) when participants retrieved sentences given a point in their draft. Note that the relevance of these retrieved results varied depending on each user's specific writing goals and context. For example, several participants valued perceived syntax and organizational structures over content (P_1, P_2, P_{12}, P_{10}). P_{11} and P_5 looked for concise sentences while P_{14} sought out more detailed examples.

P_{10} noted similarities in the retrieved sentences and reflected on drafting certain sections of their own.

6.2.2 Aligning, Understanding, and Addressing Terminology. Writing for different venues requires aligning terminology with community expectations to ensure consistency and clarity. Misunderstandings in terminology can make a paper difficult to follow or cause confusion, as P_1 noted about differing definitions of terms like "LLMs" in different venues. P_6 emphasized the importance of using correct terminology, especially in the analysis section, where there are expected norms to follow to avoid confusion. Participants found that they sometimes overlooked writing about important terms, as P_{11} realized when seeing the word "rubric" in the retrieved results, prompting him to address it in his manuscript. However, P_6 cautioned that related terms might appear in different contexts, which requires careful consideration by the reader to avoid misinterpretation.

6.2.3 Finding Related Work. Many participants used the interface as a tool to find related work to cite in their manuscript ($P_8, P_9, P_{10}, P_{11}, P_{15}, P_{16}$). Sometimes they intentionally searched for sources, while other times they came across works that highlighted a necessary concept or a limitation that they had not considered, prompting them to bookmark these sentences for later reference. P_{10} noted that they would use the tool more in the future to retrieve similar papers to cite and P_{11} would use it to find papers that included arguments related to their own work. P_{15} noted that they used the tool primarily for finding related work, not for identifying norms of a venue. In summary, the prototype demonstrated dual utility, serving not only as a writing aid for experienced and novice

writers, but also as a research tool. Participants could develop their outline and manuscripts to align with, or break, expected norms and also discover related work for citation, all within one centralized platform.

6.2.4 Closing Content Gaps.

Writers also used the tool to address content gaps in their drafts. P_{11} mentioned reading a sentence that helped her connect two key aspects of her work, confirming an idea she had previously considered but had not fully decided on. Another sentence prompted her to highlight a strength of her system in her manuscript. The sentences reminded participants of details they had forgotten to include, such as participant compensation (P_{13}). P_{15} used the sentences to identify gaps in their writing, stating, *"It's great to read what others wrote because we perform a similar process, but we did not mention that [...] in our own writing. So it would be great if we can add those."*

6.2.5 Confidence and Perception.

Participants felt more confident in their writing when they discovered that other authors had written about similar ideas or expressed them in a similar style. They also gained confidence in their related works section when they found papers in the retrieved results that aligned with their own literature review (P_8, P_{15}). Two participants, after not finding many related results for their initial text queries, perceived novelty in their writing topics and concluded that the community might not have discussed their topic yet. For example, P_8 noted that his terminology appeared in other contexts but not his specific focus. P_9 saw this as an opportunity to contribute novel insights. Similarly, P_4 was encouraged by the lack of similar work, noting *"I'm also pretty happy that apparently this is a work that might not be yet that present [in] those two communities, or in the HCI community in general. So this is not something that I find bad, but actually something that encourages me, that I'm on the right track."* Note that participants were informed at the beginning of the study that the corpus included papers from two venues over three years, so some relevant work may not have been included.

6.2.6 Sentence Rendering Modes.

We analyzed participants' perception of the cross-sentence structure mapping engine support features and how these features helped participants process the retrieved sentences. While initially overwhelming for some, participants found *distinctly colorized recurring words* valuable. It helped participants scan text efficiently. P_7 explained, *"I can just look [for the] participants in that color rather than spending my whole time reading over every block."* P_4 noted, *"I do like the color version the most while it could be perceived as a bit overwhelming... the color does help me find the relevant words."* P_6 and P_9 emphasized the ability to read and skim through the text faster. The feature served multiple purposes: P_8 found it helpful for providing structure, while P_{13} used it to guide their attention to content (*"I just looked at where's like the least or the lowest redundancy"*).

There were more mixed comments on and less use of *greying out repeated words*. P_8 found it counterproductive: *"I can't think of a reason that I would want to have those grayed out... These were sort of the key pieces of the sentence that I was trying to match."* P_{12} did not even notice the feature. In contrast, P_4 and P_3 preferred it. P_3 reported that their eyes were still drawn to greyed-out words,

suggesting the words' relevance mattered more than their visual appearance.

6.3 Bookmark and User Notes

Participants used bookmarking and note-taking to manage content during their writing process ($P_1, P_4, P_{10}, P_{11}, P_{16}$). Initially, they bookmarked content that *"caught their interest"* (P_{10}), focusing on topic relevance (P_4, P_8, P_{10}, P_{16}) and writing style (P_3, P_4). When reviewing these bookmarked sentences, participants would either enhance them with notes or unbookmark them if they seemed no longer sufficiently relevant.

Participants used notes on bookmarked sentences in several ways. Some analyzed writing techniques, with P_{11} examining content focus and P_4 studying structural elements. Notes also served as a teaching tool (P_6, P_8), with P_6 particularly valuing sentence-specific annotations for student guidance: *"If I were planning to share this with someone else I think that'd be really useful, especially to my students [...] Being able to specifically reference the part of that paper, I think is useful."* However, not all participants found notes helpful—P_7 preferred editing drafts directly. Overall, this dual system of bookmarks and notes supported both content curation and reflection, while potentially enabling knowledge transfer to more junior community members.

6.4 Tooltips

Tooltips provided useful additional context (P_3, P_9, P_{13}, P_{12}), enabling participants to contextualize retrieved sentences at the paper, section, and paragraph levels. The main critique was the additional text, which required extra effort to read.

At the paper level, the paper title offered clues about the overall topic, helping participants determine if a sentence was discussed in a similar context. For example, P_9 commented that he *"ended up using that tooltip quite a lot. [...] because it was quite nice to directly make a connection from the sentence to where it came from."*

When looking at the section level, participants found value in reviewing the section title path. P_3 observed how other authors structured the formative study, introducing *"the design"* and *"the steps in the study"* before detailing the procedure. This led her to consider splitting her own formative study into subsections and adopting a less abrupt transition to better inform the system design.

At the paragraph level, participants reported that the sentences could sometimes be ambiguous when interpreted in isolation, so they read the surrounding sentences in the tooltip for clearer context. P_{12} used the tooltip *"to see what they're talking about before and after."*

While the tooltips were helpful for many, three participants used it less frequently (P_2, P_7, P_{14}). P_7 commented, *"It just requires a lot of time to read it all [...] so I ended up not using it as much."* P_2 found that *"the two columns"* on the right hand side (i.e., *What others wrote*, and *What they wrote next*) provided enough context without needing the tooltip. Meanwhile, P_{14}, writing for a different venue, felt that HCI-focused contextualization was not relevant to her work.

6.5 Use of Features, Especially Across Writing Stages

Overall, participants found the design features useful both early in the writing process and during iterative revisions at later stages. Section title clusters were helpful for creating an initial outline. Three participants preferred to use it in early drafting because inspecting examples of other outlines helped them plan their writing for venues or contributions that they were unfamiliar with. Commenting on the integration of retrieved sentences, P_1 mentioned that he would use them in the initial phase to establish terminology and sentence structure but not later, as the style, structure, and outline would already be set. Similarly, P_{10} preferred not to be influenced by other papers' writing styles once his initial draft was more complete. P_{11} added that section title clusters and retrieved sentences are most helpful early on, when there are *"fewer restrictions"* on content and style.

Once the style, terminology, and structure were set, e.g., when participants had written a complete paragraph, some participants shifted focus to refining their own writing rather than inspecting others' work—but not all. Two participants (P_8, P_{11}) mentioned that they would use section title clusters in later stages to revise their outline, either to *"check the framing of the paper"* or adjust subsection titles. P_3 suggested using the feature in different iterations of the paper to revise sections already written, explaining that she would *"highlight that section and compare it"*, via sentence retrieval, to similar work in the venue. Likewise, P_4 noted that she would revisit the section titles, as the outline evolves during the writing process and may no longer align with the original version.

Some participants developed strategies to manage the complexity of all the features available to them. Often, they primarily focused on the *What others wrote* column, consulting the *What they wrote next* column and contextual tooltips only when needed for deeper understanding. P_2 observed: *"I didn't look at the 'What they wrote next' column that much... I also didn't look at the tooltips a lot, because I don't know. I was too absorbed by the 1st column—what others wrote. And because I, at a given position or location in the paper... I was just trying to focus on what to write then."* This selective attention may have helped them balance the information these features provided with the inherent cognitive load of lots of text that cognitive supports like special text rendering might not completely mitigate.

6.6 Efficacy of Anti-plagiarism Features

We implemented measures intended to promote learning and prevent plagiarism. As expected, while participants drew inspiration from the retrieved sentences, none showed an intent to plagiarize. Instead, they emphasized the importance of (i) modifying any examples to avoid plagiarism and align with their own ideas and (ii) saving work of interest to reference in their own paper, suggesting a strong awareness of the need to draft an original manuscript and cite properly.

7 Quantitative Results

We analyzed participants' responses to survey questions using a non-parametric Mann-Whitney U test. For Task 1, i.e., outline writing, we analyzed participants' responses comparing the two sub-tasks (first half of the outline vs. second half of the outline) and two

interface conditions (using a regular text editor without document-level support vs. with section title rendering in CorpusStudio). For Task 2, the prose writing task, we evaluated participants' reported frequency of using CorpusStudio features and their perceived usefulness.

7.1 Task 1: Outline Writing

7.1.1 Usefulness of Drafting Tools. Participants rated the usefulness of drafting their outline in a blank document (baseline) and drafting it in CorpusStudio (Figure 6). Results indicated a statistically significant preference for CorpusStudio (p = 0.003). As discussed in Section 6.1, participants valued aspects such as creating a better structure for their outline, learning how to write for a new venue or topic, and confirming their existing titles.

7.1.2 Ease of Drafting Section Title Outlines and Venue Alignment. For both conditions, participants answered a series of questions focusing on how easily they drafted section titles that were relevant to their work, aligned with the style of the target venue, adhered to norms, included all desired aspects, and identified when others had broken norms. The average responses for each condition are shown in Figure 7. A Mann-Whitney U test revealed no statistically significant differences between the interface conditions for any of these five Likert items.

7.1.3 Differences Between Tasks. Between the sub-tasks, i.e., drafting the first half of the document compared to the second half, there was a statistically significant difference (p=0.014) in participants' responses to the question, "It was easy to create the section title structure that my target venue expects." This suggests potential learning effects as participants found it easier to create the section title structure in the second half. Participants could have adapted their writing strategies from drafting the first half to the second half. We found no significant differences in the other questions within this category (see Appendix A.3.2 for the full questions).

We also evaluated whether different interface conditions impacted the ease of drafting section titles and aligning them with the target venue's expectations for both the first and second halves of the outline. Although no significant differences were found, our qualitative findings suggest that participants had greater confidence in their writing when using CorpusStudio due to the availability of examples, which helped reduce uncertainty.

7.1.4 Participant Perceptions of Ease and Overwhelm. Fourteen of the 16 participants (87.5%) reported that writing the section titles in CorpusStudio with *Section Titles of Other Papers* made the task easier (via alleviating the blank page problem and having examples at hand in a centralized writing environment), while the remaining two participants reported that writing in the baseline interface was easier.

Nine of the 16 participants (56%) concluded that writing section titles in CorpusStudio felt more overwhelming because they had to mentally process the distribution of section titles—yet most acknowledged that, despite the initial challenge, the process was useful for generating ideas. The remaining seven participants (44%) reported that completing the task in the baseline interface felt more overwhelming because they had to begin from an empty page and
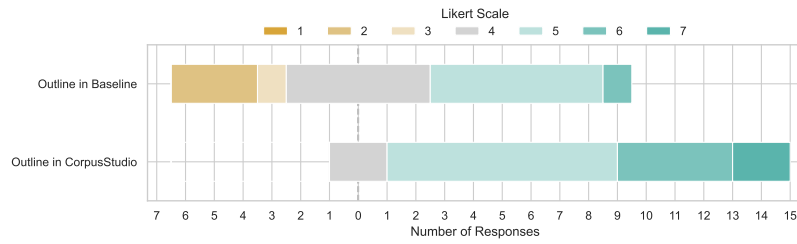
Figure 6: Usefulness of drafting an outline in the baseline condition compared to in CorpusStudio
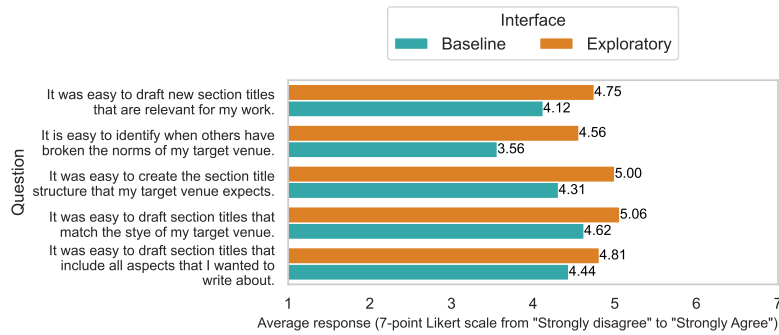

Figure 7: Ease of drafting an outline in the baseline condition compared to in CorpusStudio.

they felt uncertainty about how to name and structure the section titles.

7.1.5 Identifying and Breaking Norms. We evaluated whether there was any significant difference in participants' ability to identify norms of their target venue when reading past papers (prior to the study or during the study in the baseline condition) in comparison to reviewing the distribution over section titles. The results showed no significant difference. Similarly, for identifying when it is appropriate to break norms, no significant differences were found (see Appendix A.4.3 for the full questions). However, qualitative data highlighted that with CorpusStudio, participants more frequently reflected on breaking norms intentionally, using the example distributions and their own writing to justify deviations. While these reflections did not result in significant quantitative differences, they suggest a deeper engagement with norms during the task.

7.2 Task 2: Writing a Section of a Manuscript

Based on the NASA-TLX questionnaire (Figure 8), participants found the writing task mentally demanding, which balanced with their perception of how hard they had to work. They still felt successful in completing the task.

Participants reported on how often they used each feature (Figure 9), with the counts reflecting total reported usage across all participants while completing the writing task. The retrieved sentences in the *What others wrote* column was most frequently used, followed by the *colorizing redundant words* mode of rendering sentences, which participants reported using for reasons including

finding relevant aspects, remaining focused, reading faster, and skimming text (Section 6.2.6). They also frequently used the *tooltips*, noting that they were useful for contextualizing retrieved sentences and identifying their source (Section 6.4). *Greying out repeated words* was least frequently used. Some participants did not find it useful to de-emphasize words, while others did not notice the feature at all (Section 6.2.6). There was an overall positive sentiment on the usefulness of all other features (Figure 10).

Eight of 16 participants (50%) answered that they had a clearer sense of when it was appropriate to break the writing expectations of the target venue after interacting with CorpusStudio while four did not think that their understanding had changed and four gave a neutral response (Figure 11). From the study, we observed that at least three people commented on deviations between their prior assumptions about the structure or content and the retrieved content (Section 6.1).

All but two participants found that evaluating many sentences from previous papers gave them a better understanding of the typical writing expectations of the target venue (Figure 11). This aligns with the qualitative results where participants commented positively about the usefulness of inspecting retrieved sentences, which gave them more confidence in their writing.

8 Discussion

Our analysis revealed several ways that writers used CorpusStudio to support their academic writing process. Here, we reflect on our design and results on conceptual and empirical levels.
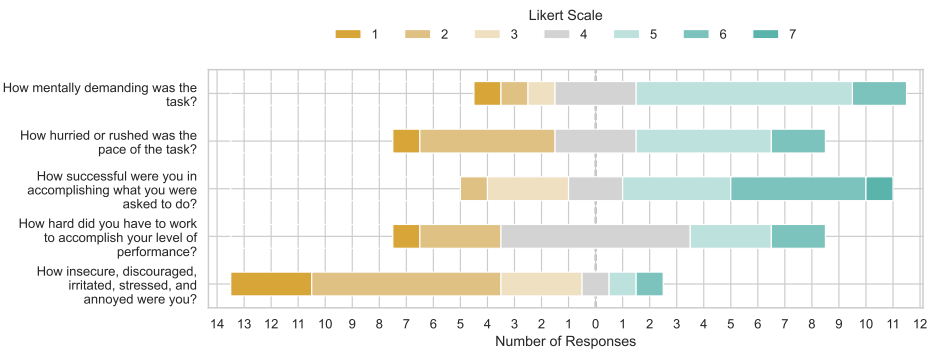
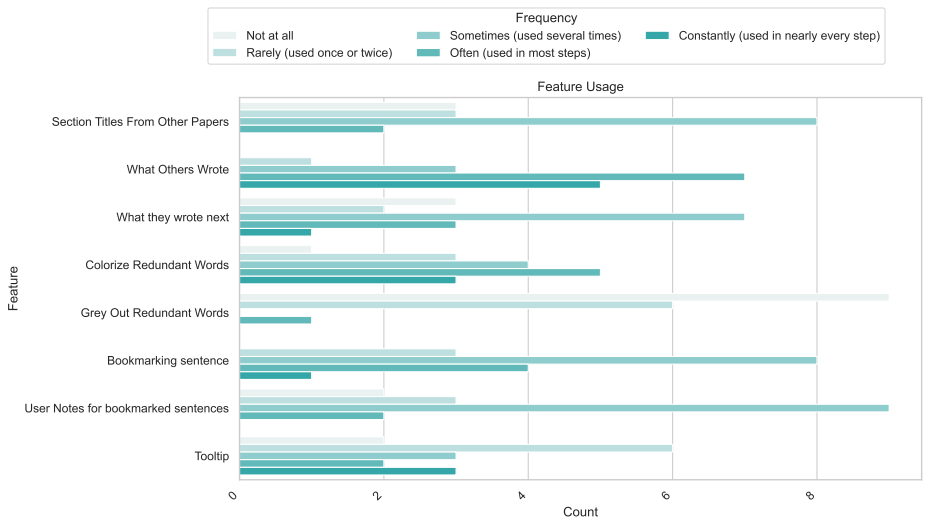**Figure 8: Responses to the NASA-TLX questionnaire during Task 2: writing a section of a manuscript**

Figure 9: Participant-reported frequency of using each feature while completing the writing task
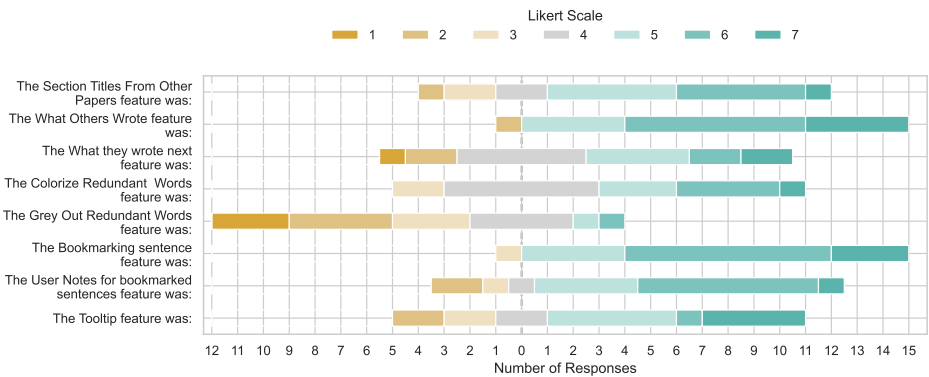

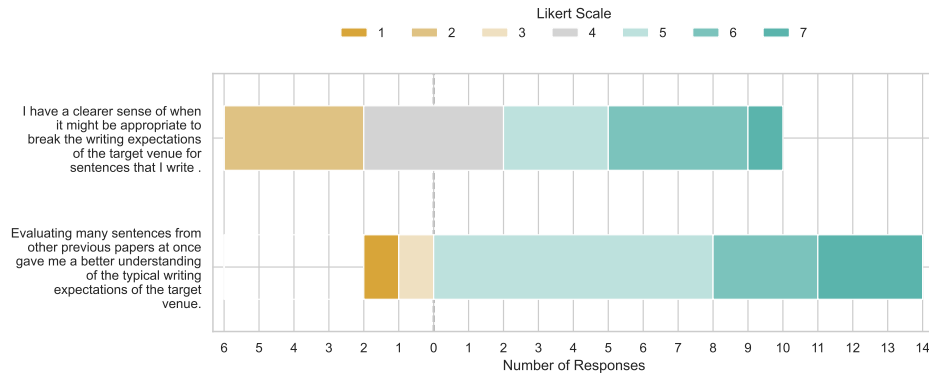**Figure 10: Usefulness of each feature during the drafting process**

Figure 11: Perceptions on understanding and breaking typical writing expectations of the target venue after using CorpusStudio

8.1 Relevance of Retrieved Sentences

We found that relevance of retrieved sentences was tightly coupled with users' writing intentions, as they found different uses and rationales for integrating the retrieved results into their writing process.

8.1.1 Users' Diverse Applications of Retrieved Sentences. While CorpusStudio uses text embeddings to capture semantic and syntactic relationships [36], our study revealed that relevance was also determined by participants' writing goals rather than semantic similarity alone. For instance, semantically similar sentences were sometimes less useful than structurally similar ones when writers sought to understand the organization of the paper.

Participants sought inspiration at multiple levels of writing: from paper organization through section titles to section-level transitions and standard academic phrases. The writers shaped these results through queries using their draft text itself, similar to patterns observed in interactions with AI in writing [10]. This challenges fundamental information retrieval assumptions about query-document similarity as an (only) indicator of relevance. While it provided a useful starting point, our findings suggest that relevance of retrieved text in writing support systems is not fixed, but dynamically constructed based on writers' immediate goals.

8.1.2 Interpreting Retrieved Results: User Behaviors and Dataset Limitations. Our study revealed how retrieved examples influenced participants' writing in unexpected ways. When participants were looking for something specific and did not find it in the retrieved results, this led to two notable behaviors: First, participants often rephrased their text to get different results, blurring the line between writing and querying. Second, some interpreted a lack of results as evidence of their work's novelty, a problematic conclusion given CorpusStudio's limited scope (two HCI venues: IUI and UIST from 2021-2023). These findings suggest that writing support systems should actively communicate their scope and limitations, for instance by suggesting broader venue searches when topic-specific queries yield no results, helping writers make informed decisions about how to interpret search results.

8.2 Designing a System with Many Retrieved Examples

While many writing support systems prioritize simplicity through aggregated results, we deliberately presented many examples (N=25) to increase the variety of examples users saw. To help users process the larger scale of information, we provided sentence rendering modes that modulate visual variables of the text in a way intended to support users in recognizing cross-sentence relationships. While it provided support, it did not eliminate all the additional cognitive demands of the extra information.

The study results suggest that showing multiple examples at once is valuable despite its cognitive demands: seeing variations in how similar ideas were expressed helped participants distinguish between rigid conventions and flexible patterns in academic writing. This suggests a trade-off between cognitive load and learning potential, a consideration for future writing support systems that aim to support learning rather than just text production.

Given that readers have different visual preferences, we provided more than one method for cognitive support. We encourage designers who wish to show multiple parallel examples to also provide alternate ways to help users visually identify patterns.

8.3 Designing Writing Support for and with Community Integration, not AI Integration

Our concept provides relevant text snippets to support writing but it does not use any AI-generated text, in contrast to recent trends in writing assistant systems [27]. Instead of training AI on a corpus, we surface examples from the corpus directly. In our case, these examples can be assumed to meet the community's quality standards, because they are evaluated by its members beforehand, that is, the published papers in the corpus went through peer review.

We argue that this approach values and supports communities. For example, it might aid communities in discovering their own community-level patterns of expression and in "onboarding" new community members. Supporting this, our design includes functionality for taking one's own notes, which might later contribute to exchanges between community members about writing norms, i.e., beyond individual interactions with our system.

This approach may also make a community's past work more discoverable to its members. Our retrieval method is based on the local text in the draft, and not on other metrics, such as paper citation counts or author metrics. In this way, our concept might contribute to discovering less well known but relevant work that the writer would not have noticed otherwise. This could be examined in a dedicated future study, and the UI design could be developed further in this direction.

8.4 Putting Metacognitive Strategies "Ready to Hand"

Tankelevitch et al. [44] recently motivated designing for metacognitive challenges in interaction. One emerging approach here is to do so by embedding specific thinking strategies directly into the interface. For instance, Dang et al. [9] designed for the strategy of "reverse outlining" by providing a sidebar of paragraph-wise summaries, updated live. Similarly, we designed for learning from examples, with respective material provided in sidebars directly within the writing environment. Our design also supports the externalization of users' self-reflection (e.g., what does this mean for my text?) in User Notes. As our results show, participants found this helpful and made use of it in diverse ways.

Tankelevitch et al. [44] also suggests "support[ing] users in increasing their self-awareness and adjusting their confidence." We observed that, by becoming aware of what authors in a community wrote, many participants' confidence levels increased with respect to writing for that same community.

8.5 Limitations and Future Work

Our results reveal how writers interact with and perceive examples retrieved from a community-specific document corpus. However, readers should be cautious when generalizing these findings beyond the scope of our study. All participants had a background in HCI, and the corpus was drawn from two HCI venues, focusing primarily on system contribution papers. Additionally, the writing task was based on an academic writing scenario. Future research should explore other writing contexts where communities play a role, such as fiction, poetry, or argumentative writing. In our user study, participants were given 20-25 minutes to draft their text and engage with the various system features. While this offered valuable insights into their interactions and perceptions of the writing environment, it would be worthwhile to explore how prolonged use of these features might influence users' writing habits and style.

9 Conclusion

CorpusStudio is a writing support system that integrates example retrieval—at both the document level and sentence level—directly into the drafting process. Writers can compare their text against a curated corpus from their target academic community, with the system presenting many real examples in parallel. This tight coupling between writing and retrieval enables authors to identify established writing patterns, understand community conventions, and validate their writing choices through direct comparison with published work. Unlike systems that generate or remix text, CorpusStudio deliberately preserves and attributes original manuscripts, using

search based on text embeddings that incorporate spatial information and context to surface relevant examples in an accessible format. Thereby we support academic writing while maintaining transparency and acknowledging existing scholarship. Our work aims to advance an interaction model where writing support tools help authors overcome writing barriers while respecting and learning from the contributions of others in their field.

References

[1] Karim Benharrak, Tim Zindulka, Florian Lehmann, Hendrik Heuer, and Daniel Buschek. 2023. Writer-Defined AI Personas for On-Demand Feedback Generation. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2023). https://api.semanticscholar.org/CorpusID:261945156

[2] Advait Bhat, Saaket Agashe, Parth Oberoi, Niharika Mohile, Ravi Jangir, and Anirudha Joshi. 2023. Interacting with Next-Phrase Suggestions: How Suggestion Systems Aid and Influence the Cognitive Processes of Writing. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 436–452. doi:10.1145/3581641.3584060

[3] Joel Chan, Joseph Chee Chang, Tom Hope, Dafna Shahaf, and Aniket Kittur. 2018. SOLVENT. *Proceedings of the ACM on Human-Computer Interaction* 2 (2018), 1 – 21. https://api.semanticscholar.org/CorpusID:53236685

[4] Jim Chang and Jason J. S. Chang. 2015. WriteAhead2: Mining Lexical Grammar Patterns for Assisted Writing. In *North American Chapter of the Association for Computational Linguistics*. https://api.semanticscholar.org/CorpusID:1529175

[5] Joseph Chee Chang, Amy X Zhang, Jonathan Bragg, Andrew Head, Kyle Lo, Doug Downey, and Daniel S Weld. 2023. Citesee: Augmenting citations in scientific papers with persistent and personalized historical context. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[6] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://api.semanticscholar.org/CorpusID:247625751

[7] Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S. Weld. 2020. SPECTER: Document-level Representation Learning using Citation-informed Transformers. *ArXiv* abs/2004.07180 (2020). https://api.semanticscholar.org/CorpusID:215768677

[8] Allan M. Collins, John Seely Brown, and Susan E. Newman. 1988. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. https://api.semanticscholar.org/CorpusID:63647121

[9] Hai Dang, Karim Benharrak, Florian Lehmann, and Daniel Buschek. 2022. Beyond Text Generation: Supporting Writers with Continuous Automatic Text Summaries. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 98, 13 pages. doi:10.1145/3526113.3545672

[10] Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. 2023. Choice Over Control: How Users Write with Large Language Models using Diegetic and Non-Diegetic Prompting. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://api.semanticscholar.org/CorpusID:257365446

[11] Brian Falkenhainer, Kenneth D Forbus, and Dedre Gentner. 1989. The structure-mapping engine: Algorithm and examples. *Artificial intelligence* 41, 1 (1989), 1–63.

[12] Linda S. Flower. 1979. Writer-Based Prose: A Cognitive Basis for Problems in Writing. *College English* (1979). https://api.semanticscholar.org/CorpusID:142218225

[13] Dedre Gentner and Christian Hoyos. 2017. Analogy and abstraction. *Topics in cognitive science* 9, 3 (2017), 672–693.

[14] K. Gero, Vivian Liu, and Lydia B. Chilton. 2021. Sparks: Inspiration for Science Writing using Language Models. *Proceedings of the 2022 ACM Designing Interactive Systems Conference* (2021). https://api.semanticscholar.org/CorpusID:239009871

[15] Katy Ilonka Gero, Chelse Swoopes, Ziwei Gu, Jonathan K Kummerfeld, and Elena L Glassman. 2024. Supporting Sensemaking of Large Language Model Outputs at Scale. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.

[16] Susanne Göpferich. 2009. Comprehensibility assessment using the Karlsruhe Comprehensibility Concept 1. https://api.semanticscholar.org/CorpusID:203672248

[17] Tom Hope, Joel Chan, Aniket Kittur, and Dafna Shahaf. 2017. Accelerating Innovation Through Analogy Mining. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2017). https://api.semanticscholar.org/CorpusID:11811861

[18] Julie Hui, Darren Gergle, and Elizabeth Gerber. 2018. IntroAssist: A Tool to Support Writing Introductory Help Requests. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018). https://api.semanticscholar.org/CorpusID:5052431

[19] Julie Hui and Michelle L. Sprouse. 2023. Lettersmith: Scaffolding Written Professional Communication Among College Students. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023). https://api.semanticscholar.org/CorpusID:258216852

[20] Takumi Ito, Tatsuki Kuribayashi, Masatoshi Hidaka, Jun Suzuki, and Kentaro Inui. 2020. Langsmith: An Interactive Academic Text Revision System. In *Conference on Empirical Methods in Natural Language Processing*. https://api.semanticscholar.org/CorpusID:222272178

[21] Takumi Ito, Naomi Yamashita, Tatsuki Kuribayashi, Masatoshi Hidaka, Jun Suzuki, Ge Gao, Jacky Jamieson, and Kentaro Inui. 2023. Use of an AI-powered Rewriting Support Software in Context with Other Tools: A Study of Non-Native English Speakers. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (2023). https://api.semanticscholar.org/CorpusID:264350505

[22] Pamela Jordan. 1994. Basics of qualitative research: Grounded theory procedures and techniques. *Journal of Pediatric Nursing* 9 (1994), 205–206. https://api.semanticscholar.org/CorpusID:79244797

[23] Hyeonsu B Kang, Xin Qian, Tom Hope, Dafna Shahaf, Joel Chan, and Aniket Kittur. 2022. Augmenting Scientific Creativity with an Analogical Search Engine. *ACM Transactions on Computer-Human Interaction* 29 (2022), 1 – 36. https://api.semanticscholar.org/CorpusID:249209576

[24] Yewon Kim, Thanh-Long V. Le, Donghwi Kim, Mina Lee, and Sung-Ju Lee. 2024. How Non-native English Speakers Use, Assess, and Select AI-Generated Paraphrases with Information Aids. *ArXiv* abs/2405.07475 (2024). https://api.semanticscholar.org/CorpusID:269757551

[25] Yewon Kim, Mina Lee, Donghwi Kim, and Sung-Ju Lee. 2023. Towards Explainable AI Writing Assistants for Non-native English Speakers. *ArXiv* abs/2304.02625 (2023). https://api.semanticscholar.org/CorpusID:257952442

[26] Bongshin Lee, Mary Czerwinski, George G. Robertson, and Benjamin B. Bederson. 2005. Understanding research trends in conferences using paperLens. *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (2005). https://api.semanticscholar.org/CorpusID:1452553

[27] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Antonette Shibani, Disha Shrivastava, Lila Shroff, Agnia Sergeyuk, Jessi Stark, Sarah Sterman, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia Ha Rim Rho, Zejiang Shen, and Pao Siangliulue. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 1054, 35 pages. doi:10.1145/3613904.3642697

[28] Mina Lee, Percy Liang, and Qian Yang. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022). https://api.semanticscholar.org/CorpusID:246016439

[29] Yoonjoo Lee, Hyeonsu B Kang, Matt Latzke, Juho Kim, Jonathan Bragg, Joseph Chee Chang, and Pao Siangliulue. 2024. PaperWeaver: Enriching Topical Paper Alerts by Contextualizing Recommended Papers with User-collected Papers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 19, 19 pages. doi:10.1145/3613904.3642196

[30] Zhuoyan Li, Chen Liang, Jing Peng, and Ming Yin. 2024. The Value, Benefits, and Concerns of Generative AI-Powered Assistance in Writing. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2024). https://api.semanticscholar.org/CorpusID:268532053

[31] Gabriel Lins de Holanda Coelho, Paul HP Hanel, and Lukas J. Wolf. 2020. The very efficient assessment of need for cognition: Developing a six-item version. *Assessment* 27, 8 (2020), 1870–1885.

[32] Mun Ling Lo. 2012. Variation Theory and the Improvement of Teaching and Learning. https://api.semanticscholar.org/CorpusID:59529353

[33] Jock D. Mackinlay, Ramana Rao, and Stuart K. Card. 1995. An organic user interface for searching citation links. In *International Conference on Human Factors in Computing Systems*. https://api.semanticscholar.org/CorpusID:18174079

[34] Yu A Malkov and Dmitry A Yashunin. 2018. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE transactions on pattern analysis and machine intelligence* 42, 4 (2018), 824–836.

[35] Deena Mohammad Mansour. 2017. Using COCA to Foster Students' Use of English Collocations in Academic Writing. https://api.semanticscholar.org/CorpusID:55705638

[36] Niklas Muennighoff, Nouamane Tazi, Loic Magne, and Nils Reimers. 2023. MTEB: Massive Text Embedding Benchmark. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, Andreas Vlachos and Isabelle Augenstein (Eds.). Association for Computational Linguistics, Dubrovnik, Croatia, 2014–2037. doi:10.18653/v1/2023.eacl-main.148

[37] M. Asif Naeem, I Wayan Sathya Linggawa, Aftab Ahmad Mughal, Christof Lutteroth, and Gerald Weber. 2018. A Smart Email Client Prototype for Effective Reuse of Past Replies. *IEEE Access* 6 (2018), 69453–69471. https://api.semanticscholar.org/CorpusID:54449952

[38] Antoine Ponsard, Francisco Escalona, and Tamara Munzner. 2016. PaperQuest: A Visualization Tool to Support Literature Review. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (2016). https://api.semanticscholar.org/CorpusID:7511242

[39] Jason Portenoy, Marissa Radensky, Jevin D. West, Eric Horvitz, Daniel S. Weld, and Tom Hope. 2021. Bursting Scientific Filter Bubbles: Boosting Innovation via Novel Author Discovery. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2021). https://api.semanticscholar.org/CorpusID:237485597

[40] Napol Rachatasumrit, Jonathan Bragg, Amy X Zhang, and Daniel S Weld. 2022. Citeread: Integrating localized citation contexts into scientific paper reading. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 707–719.

[41] Zackary Rackauckas. 2024. Rag-Fusion: A New Take on Retrieval Augmented Generation. *International Journal on Natural Language Computing* 13, 1 (Feb. 2024), 37–47. doi:10.5121/ijnlc.2024.13103

[42] Amanpreet Singh, Mike D'Arcy, Arman Cohan, Doug Downey, and Sergey Feldman. 2022. SciRepEval: A Multi-Format Benchmark for Scientific Document Representations. In *Conference on Empirical Methods in Natural Language Processing*. https://api.semanticscholar.org/CorpusID:254018137

[43] Nikhil Singh, Guillermo Bernal, Daria Savchenko, and Elena L. Glassman. 2023. Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 68 (sep 2023), 57 pages. doi:10.1145/3511599

[44] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 680, 24 pages. doi:10.1145/3613904.3642902

[45] Reinhard Tausch. 2002. Sich verständlich ausdrücken. https://api.semanticscholar.org/CorpusID:176068335

[46] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science* 185, 4157 (1974), 1124–1131.

[47] Ryan Yen and Jian Zhao. 2024. Memolet: Reifying the Reuse of User-AI Conversational Memories. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) (*UIST '24*). Association for Computing Machinery, New York, NY, USA, Article 58, 22 pages. doi:10.1145/3654777.3676388

A Surveys

The surveys we used in the study.

A.1 Pre-study Survey

A.1.1 Demographics.

(1) What is your participant ID?
(Given to participant by interviewee)
(2) What is your age?
○ 18-24 ○ 25-34 ○ 35-44 ○ 45-54 ○ 54+ ○ Prefer not to disclose

(3) What is your gender?
  ○ Woman ○ Man ○ Non-binary ○ Prefer not to disclose ○ Other:
(4) What is your profession?
  ○ Undergraduate student ○ Masters student ○ PhD student/candidate
  ○ Faculty ○ Other:

A.1.2 Need for Cognition (NCS-6) scale. The following questions are on a 7 point likert scale from "not at all like me" to "very much like me."

(1) I would prefer complex to simple problems.
(2) I like to have the responsibility of handling a situation that requires a lot of thinking.
(3) Thinking is not my idea of fun.
(4) I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.
(5) I really enjoy a task that involves coming up with new solutions to problems.
(6) I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

A.1.3 Writing Experience and Approach.

(1) What is your field of study?
  (Open-ended)
(2) What is the topic for the paper you are working on today?
  (Open-ended)
(3) What is your target venue for this writing?
  (Open-ended)
(4) How many years of experience do you have with academic writing? Use a numeric answer.
  (Open-ended)
(5) How many years of experience do you have with writing for your selected target venue? Use a numeric answer.
  (Open-ended)
(6) How familiar are you with the process of writing for your selected target venue?
(7) How many academic papers have you authored or co-authored? Use a numeric answer.
  (Open-ended)
(8) How confident are you in identifying the writing style and content expected by a particular academic venue?
  (Answered on a 7-point likert scale from "Not confident at all" to "Extremely confident")
(9) Have you used any writing support tools (e.g., Grammarly, Ref-n-Write, SciSpace) in the past?
  ○ Yes ○ No
(10) What writing support tools have you used in the past?
  (Open-ended)
(11) How often do you use writing support tools?
  ○ Never ○ I've used them a few times, but not regularly ○ A few time a month ○ A few times a week ○ Once a day or more
(12) What strategies do you typically use when drafting a section of an academic paper?
  (Open-ended)

(13) What strategies do you typically use when drafting a section of a paper for your selected target venue?
  (Open-ended)
(14) How often do you refer to published papers while writing?
  ○ Never ○ Rarely ○ Sometimes ○ Often ○ Always
(15) What challenges do you usually face when writing for an academic venue?
  (Open-ended)

A.2 Task 1 Outlining Survey - Page without section titles

A.2.1 Participant and Task Identification.

(1) What is your participant ID?
  (Given to participant by interviewee)
(2) What outlining task did you just complete?
  ○ First half of manuscript ○ Second half of manuscript

A.2.2 Interface Comparison. Participants answered these questions after outlining the second half of the manuscript.

(1) Which interface made the task easier?
  ○ Page with section titles ○ Page without section titles
(2) Which interface felt more overwhelming?
  ○ Page with section titles ○ Page without section titles

A.2.3 Task Completion. The following questions are on a 7 point likert scale from "Strongly disagree" (1) to "Strongly agree" (7):

(1) It was easy to draft section titles that include all aspects that I wanted to write about.
(2) It was easy to draft section titles that match the stye of my target venue.
(3) It was easy to create the section title structure that my target venue expects.
(4) It is easy to identify when others have broken the norms of my target venue.
(5) It was easy to draft new section titles that relevant for my work.

A.2.4 Writing Expectations.

(1) What is your current strategy for comparing paper outline from related papers
  (Open-ended)
(2) When I looked at previous papers from my target venue in the past, evaluating many section titles from those papers gave me a better understanding of the typical writing expectations of the target venue.
  7 point likert scale from "Strongly disagree" (1) to "Strongly agree" (7)
(3) Based on when I looked at previous papers from my target venue in the past, I have a clearer sense of when it might be appropriate to break the writing expectations of that target venue for section titles that I write.
  7 point likert scale from "Strongly disagree" (1) to "Strongly agree" (7)

A.2.5 Feature Usefulness. The following question is on a 7 point likert scale from "Not useful at all" (1) to "Very useful" (7):

(1) Writing in the document was:

A.3 Task 1- Page with section titles

A.3.1 Participant and Task Identification.

(1) What is your participant ID?
(Given to participant by interviewee)
(2) What outlining task did you just complete?
○ First half of manuscript ○ Second half of manuscript

A.3.2 Task Completion.
The following questions are on a 7 point likert scale from "Strongly disagree" (1) to "Strongly agree" (7):

(1) It was easy to draft section titles that include all aspects that I wanted to write about.
(2) It was easy to draft section titles that match the style of my target venue.
(3) It was easy to create the section title structure that my target venue expects.
(4) It is easy to identify when others have broken the norms of my target venue.
(5) It was easy to draft new section titles that relevant for my work.

A.3.3 Writing expectations.
The following questions are on a 7 point likert scale from "Strongly disagree" (1) to "Strongly agree" (7):

(1) Evaluating many section titles from other papers at once gave me a better understanding of the typical writing expectations of the target venue.
(2) After evaluating many section titles from other papers at once, I have a clearer sense of when it might be appropriate to break the writing expectations of the target venue for section titles that I write.

A.3.4 Feature Usefulness.
The following question is on a 7 point likert scale from "Not useful at all" (1) to "Very useful" (7):

(1) The Section Titles From Other Papers feature was:

A.4 Task 2 - Post-writing Survey

A.4.1 Participant identification.

(1) What is your participant ID?

A.4.2 NASA-TLX.
The following questions from the *NASA-TLX)* scale are on a 7 point likert scale from "Very low" (1) to "Very high":

(1) How mentally demanding was the task?
(2) How hurried or rushed was the pace of the task?
(3) How successful were you in accomplishing what you were asked to do?
(4) How hard did you have to work to accomplish your level of performance?
(5) How insecure, discouraged, irritated, stressed, and annoyed were you?

A.4.3 Writing expectations.
The following question is on a 7 point likert scale from "Strongly disagree" (1) to "Strongly agree" (7):

(1) Evaluating many sentences from other previous papers at once gave me a better understanding of the typical writing expectations of the target venue.

(2) I have a clearer sense of when it might be appropriate to break the writing expectations of the target venue for sentences that I write.

A.4.4 Feature Evaluation.
For each feature, participants ranked the usefulness on a 7-point likert scale from "Not useful at all" (1) to "Very useful" (7); we included a screenshot of the respective feature for participants to reference when answering the question. They then reported their frequency of using the feature for completing the task with the following options: Not at all, Rarely (used once or twice), Sometimes (used several times), Often (used in most steps), and Constantly (used in nearly every step). Lastly, they answered an open-ended question on scenarios for which the feature would be most valuable in their future writing tasks. The questions for each feature are listed below.

(1) The Section Titles From Other Papers feature was:
(2) How frequently did you use the Section Titles From Other Papers feature while completing the task?
(3) In what scenarios do you see the Section Titles From Other Papers feature being most valuable in your future writing tasks?
(4) The What Others Wrote feature was:
(5) How frequently did you use the What Others Wrote feature while completing the task?
(6) In what scenarios do you see the What Others Wrote feature being most valuable in your future writing tasks?
(7) The What they wrote next feature was:
(8) How frequently did you use the What they wrote next feature while completing the task?
(9) In what scenarios do you see the What they wrote next feature being most valuable in your future writing tasks?
(10) The Colorize Redundant Words feature was:
(11) How frequently did you use the Colorize Redundant Words feature while completing the task?
(12) In what scenarios do you see the Colorize Redundant Words feature being most valuable in your future writing tasks?
(13) The Grey Out Redundant Words feature was:
(14) How frequently did you use the Grey Out Redundant Words feature while completing the task?
(15) In what scenarios do you see the Grey Out Redundant Words feature being most valuable in your future writing tasks?
(16) The Bookmarking sentence feature was: How frequently did you use the Bookmarking sentence feature while completing the task?
(17) In what scenarios do you see Bookmarking sentence feature being most valuable in your future writing tasks?
(18) The User Notes for bookmarked sentences feature was:
(19) How frequently did you use the User Notes for bookmarked sentences feature while completing the task?
(20) In what scenarios do you see User Notes for bookmarked sentences feature being most valuable in your future writing tasks?
(21) The Tooltip feature was: How frequently did you use the Tooltip feature while completing the task?
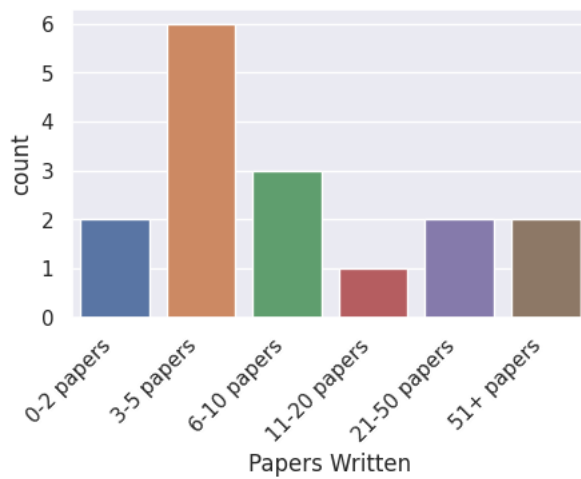(22) In what scenarios do you see Tooltip feature being most valuable in your future writing tasks?

Figure 12: "How many academic papers have you authored or co-authored?"

A.5 Semi-structured Interview

After participants completed both tasks and all surveys, we asked participants the following questions:

(1) How was your experience using these features?

(2) What was your approach when writing with the different features?

(3) Now that you have a better understanding of the different features, when and where would you see the features have the greatest impact on your writing process?

(4) Did you face any challenges while writing this section?

(5) Imagine you didn't have this tool, and think back to your experiences today. What approach would you have taken if you wanted to accomplish the same thing, but didn't have this interface?

(6) For each named feature (Section Title Explorer, Colorize similar words, Grey Out Redundant Words, Tooltip, Bookmarking sentences, User Notes)

(7) What did you like about this feature?

(8) What did you not like about this feature?

(9) What do you wish it could do?

(10) Would you use this feature in the future? If so, how and for what purpose? If not, why not?

(11) Has your understanding of the writing norms of your target community changed? And if so how?

(12) Outside of the discussed features, are there any recommendations you'd make for improving the tool?

(13) Is there anything that I did not ask about that you want to share?

A.6 Additional Figures