Scim: Intelligent Skimming Support for Scientific Papers

Raymond Fok
rayfok@cs.washington.edu
University of Washington
Seattle, Washington, USA

Hita Kambhamettu
hitakam@seas.upenn.edu
University of Pennsylvania
Philadelphia, Pennsylvania, USA

Luca Soldaini
lucas@allenai.org
Allen Institute for AI
Seattle, Washington, USA

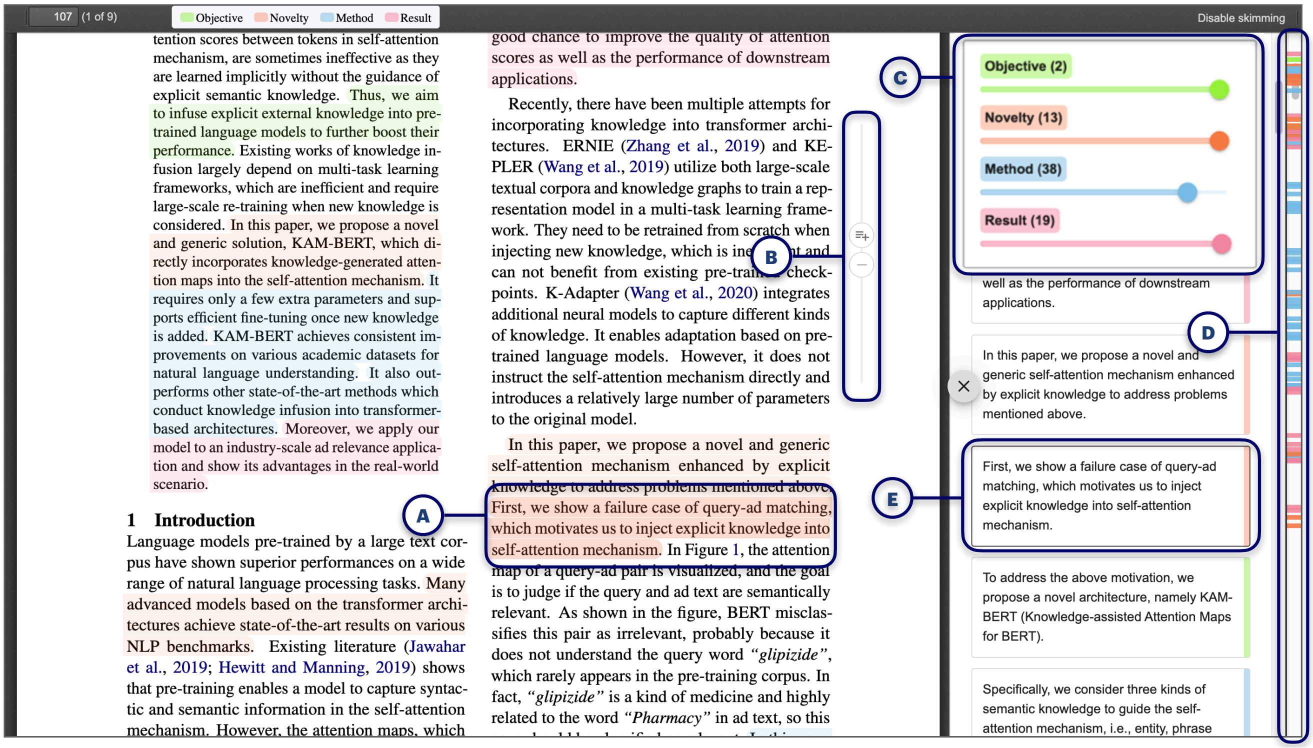
Jonathan Bragg
jbragg@allenai.org
Allen Institute for AI
Seattle, Washington, USA

Kyle Lo
kylel@allenai.org
Allen Institute for AI
Seattle, Washington, USA

Marti A. Hearst
hearst@berkeley.edu
University of Berkeley
Berkeley, California, USA

Andrew Head
head@seas.upenn.edu
University of Pennsylvania
Philadelphia, Pennsylvania, USA

Daniel S. Weld
danw@allenai.org
Allen Institute for AI
Seattle, Washington, USA



tention scores between tokens in self-attention mechanism, are sometimes ineffective as they are learned implicitly without the guidance of explicit semantic knowledge. Thus, we aim to infuse explicit external knowledge into pre-trained language models to further boost their performance. Existing works of knowledge infusion largely depend on multi-task learning frameworks, which are inefficient and require large-scale re-training when new knowledge is considered. In this paper, we propose a novel and generic solution, KAM-BERT, which directly incorporates knowledge-generated attention maps into the self-attention mechanism. It requires only a few extra parameters and supports efficient fine-tuning once new knowledge is added. KAM-BERT achieves consistent improvements on various academic datasets for natural language understanding. It also outperforms other state-of-the-art methods which conduct knowledge infusion into transformer-based architectures. Moreover, we apply our model to an industry-scale ad relevance application and show its advantages in the real-world scenario.

1 Introduction

Language models pre-trained by a large text corpus have shown superior performances on a wide range of natural language processing tasks. Many advanced models based on the transformer architectures achieve state-of-the-art results on various NLP benchmarks. Existing literature (Jawahar et al., 2019; Hewitt and Manning, 2019) shows that pre-training enables a model to capture syntactic and semantic information in the self-attention mechanism. However, the attention maps, which

good chance to improve the quality of attention scores as well as the performance of downstream applications.

Recently, there have been multiple attempts for incorporating knowledge into transformer architectures. ERNIE (Zhang et al., 2019) and KEPLER (Wang et al., 2019) utilize both large-scale textual corpora and knowledge graphs to train a representation model in a multi-task learning framework. They need to be retrained from scratch when injecting new knowledge, which is ine[illegible]nt and can not benefit from existing pre-trai[illegible] checkpoints. K-Adapter (Wang et al., 2020) integrates additional neural models to capture different kinds of knowledge. It enables adaptation based on pre-trained language models. However, it does not instruct the self-attention mechanism directly and introduces a relatively large number of parameters to the original model.

In this paper, we propose a novel and generic self-attention mechanism enhanced by explicit knowledge to address problems mentioned above. First, we show a failure case of query-ad matching, which motivates us to inject explicit knowledge into self-attention mechanism. In Figure 1, the attention map of a query-ad pair is visualized, and the goal is to judge if the query and ad text are semantically relevant. As shown in the figure, BERT misclassifies this pair as irrelevant, probably because it does not understand the query word *"glipizide"*, which rarely appears in the pre-training corpus. In fact, *"glipizide"* is a kind of medicine and highly related to the word *"Pharmacy"* in ad text, so this

well as the performance of downstream applications.

In this paper, we propose a novel and generic self-attention mechanism enhanced by explicit knowledge to address problems mentioned above.

First, we show a failure case of query-ad matching, which motivates us to inject explicit knowledge into self-attention mechanism.

To address the above motivation, we propose a novel architecture, namely KAM-BERT (Knowledge-assisted Attention Maps for BERT).

Specifically, we consider three kinds of semantic knowledge to guide the self-attention mechanism, i.e., entity, phrase

Figure 1: Scim is an intelligent reading interface for skimming scientific articles. To help readers develop a broad overview of content in a paper, Scim intelligently highlights passages (A). The passages are colorized to indicate the rhetorical role of the passage, i.e., whether it describes the research's objectives, novelty, methods, and results. Highlights are distributed throughout the text to support a holistic skim. Readers request additional (or fewer) highlights by using paragraph-local (B) and paper-wide (C) controls. To understand where to find information of a certain kind, readers can glance at highlight markers in the scroll bar (D). Readers can also collect an overview of the paper by reviewing highlighted passages in a sidebar (E).

ABSTRACT

Researchers need to keep up with immense literatures, though it is time-consuming and difficult to do so. In this paper, we investigate the role that intelligent interfaces can play in helping researchers skim papers, that is, rapidly reviewing a paper to attain a cursory understanding of its contents. After conducting formative interviews and a design probe, we suggest that skimming aids should aim to thread the needle of highlighting content that is simultaneously diverse, evenly-distributed, and important. We introduce Scim, a novel intelligent skimming interface that reifies this aim, designed to support the skimming process by highlighting salient paper contents to direct a skimmer's focus. Key to the design is that the highlights are faceted by content type, evenly-distributed across a paper, with a density configurable by readers at both the global and local level. We evaluate Scim with an in-lab usability study and deployment study, revealing how skimming aids can support readers throughout the skimming experience and yielding design considerations and tensions for the design of future intelligent skimming tools.

CCS CONCEPTS

• **Human-centered computing → Interactive systems and tools**; **Empirical studies in HCI**.

KEYWORDS

intelligent reading support, complex texts, skimming, highlights, scientific papers

1 INTRODUCTION

With the rise of knowledge work and a contemporaneous explosion of information, knowledge workers are expected to quickly sift through high volumes of rapidly evolving information. One domain where this trend is particularly pronounced is in scientific research. Scientific researchers spend a tremendous amount of effort staying up to date with the scientific literature. To keep up with the literature, a researcher needs to—with some regularity—forage for relevant articles (hereafter called "papers"), and for any papers they identify, skim the paper, read the paper, and then integrate knowledge gained from reading that paper into their personal records.

Among these stages, skimming is the task that involves reviewing the contents of a paper selectively and rapidly to develop a cursory understanding of a paper's contents. An initial skim of a paper helps a researcher decide whether to read a paper further, and hence skimming is critical to a researcher who has a broad literature to acquaint themselves with. The process of skimming papers is common among researchers [47]. With the shift of scholarly papers to a digital online publications, the practice of skimming has become yet more widespread [34, 55].

While skimming is a pervasive practice, it is not easy. Skimming is a skill that takes time to learn and effectively employ [14, 40, 64]. A key challenge of skimming is that a reader is attempting to understand parts of a paper's contents, while skipping over much of the paper. To skim well thus requires not only quite a bit of attention, but also strategic choice of what to read, where, and when to stop reading a section. A *skimming* session may devolve into a *reading* session should a reader find themselves drawn into the details of a section, a practice we observed in some of our formative observations of researchers (Section 3.2).

In this paper, we investigate how intelligent user interfaces could assist in skimming scientific papers. With widespread use of AIs in scholarly search (cf. [1, 4]) and increased research attention into the design of intelligent scientific reading applications (e.g., [20]), we explore how intelligent tools could help with the task in between reading and search—namely, skimming. As a starting point, we ask how an interface can make judicious use of highlighting of a paper's contents to help readers direct their attention across a sample of passages while they skim.

To understand the nuance of what it means to design a usable intelligent highlighting interface for skimming, we conducted formative interviews and observations with researchers, and then preliminary usability studies with prototypes of an intelligent highlighting tool. These initial investigations led to the following learnings (Section 3.2): First, readers desire highlights that cover diverse content, are evenly-distributed across a paper, and mark important paper content. This represents a tension between reader expectations and system implementation, in that it is not always possible to highlight according to passage importance while achieving an even distribution of highlights. Second, readers desire the ability to influence the degree to which a paper is highlighted.

We incorporated these insights into the design of Scim, an intelligent user interface for skimming scientific papers (Figure 1). Scim addresses the above-mentioned tensions in design as follows. First, it identifies important passages by applying a classification model to the text, highlighting those passages which it is most confident contain information of consequence to a reader. Second, it supports skimming diverse kinds of contents by highlighting passages relating to four major kinds of content readers in the formative studies sought: research objectives, novel aspects of the research, methodology, and results. To help readers identify passages containing each of these kinds of content, Scim highlights each kind of content in a distinct color. Third, Scim supports an evenly-distributed skim of a paper, highlighting passages in such a way that most paragraphs contain at least one sentence, with the aim of reducing a reader's concern that they may be missing information of consequence. Finally, Scim lets readers customize the number of highlights in a paper, with controls for configuring highlights across an entire paper, and for requesting additional highlights within individual paragraphs of interest.

This paper concludes with two studies that shed light on how a tool like Scim would affect skimming. We first conducted an in-lab usability study to explore how Scim affects readers' ability to search for specific kinds of information in a paper. We observed a small, significant effect where readers found information in less time with Scim, while reporting approximately the same difficulty in tasks, as with a baseline.

To investigate how Scim could support skimming in more realistic scenarios, we conducted a second, longitudinal diary study spanning two weeks. Twelve NLP researchers were asked to skim one paper a day with Scim for 10 days, where they selected papers of their choice from a collection of 100 recent NAACL papers. Researchers completed daily diary entries and an exit interview. In 70% of diary entries, researchers reported that Scim's highlights helped them skim a paper. The diary entries also brought into focus what circumstances tools like Scim can help readers. For instance, Scim was particularly useful when reading text-dense passages with few visuals, or when reading a paper that falls outside one's area of expertise. Researchers reported that Scim became more usable with time, because they became accustomed to highlights that they had initially found distracting. The study also revealed directions where additional improvements to intelligent skimming support are necessary, such as highlighting passages that provide important background for later highlighted passages, and highlighting in a way that integrates nicely with emphasis authors have already added through boldface font and text formatting.

Taken together, this paper envisions systems that support skimming of scientific papers with intelligent highlighting, reifying the vision in a working reference implementation of Scim, and offering a comprehensive two-study evaluation. In summary, this paper contributes:

- Seven design motivations for intelligent, highlight-based scientific skimming user interfaces, grounded in formative interviews and observations and preliminary usability studies of a prototype tool.
- Scim, an intelligent skimming interface that highlights passages to optimize for importance, diversity, and distribution of content, while affording control over highlighting at both the paper and paragraph level.
- A reference implementation of Scim's end-to-end paper processing backend, including a language model for classifying and highlighting sentences, fine-tuned using a data programming approach, and post-processing heuristics for improving prediction accuracy and achieving well-distributed highlights.
- Insights into the strengths and limitations of Scim based on a controlled in-lab usability study and a two-week diary study.

2 RELATED WORK

2.1 Prior Studies on Skimming

Skimming is widely considered to be a form of rapid reading in which the goal is to get a general idea of the text or visual content, typically accomplished by focusing on information relevant to one's goals and skipping over irrelevant information [39, 47]. Skimming is a particularly necessary and useful skill for scholars who read scientific papers. As the number of published papers continue to increase year over year and technology has caused a gradual move from print towards a digital medium, scholars have adapted by reading more papers while spending less time on each [34, 55].

Prior results from the psychology literature have found that skimming readers are not generally very accurate at selecting goal-relevant information for processing within text, and that physical limitations in the oculomotor system responsible for controlling eye movements largely preclude rapid, accurate placements of eye gaze for extended periods such as when skimming a long document [38, 39]. Beyond limitations in visual acuity, skimming can also be a cognitively demanding task as readers are continually building an ongoing mental model of the text and integrating information across sentences as they read [45, 47, 54].

Other studies suggest that skimming readers may be able to effectively direct attention to more important content, for instance by reading in a satisficing manner [14, 15, 48]. Satisficing is a skim reading strategy in which readers are inherently sensitive to a proxy for information gain. Under this strategy, readers set a information threshold, and if while reading a unit of text they determine that the information gain falls below a designated information threshold, they proceed on to the next unit of text. These studies have found that as a result, readers tend to spend more time at the beginning of paragraphs, toward the top of pages, and at the beginning of documents [14]. We use Scim to study how automated assistance may support skimming by cueing readers towards salient sentences, suggested by an AI system, thereby shifting the initial locus of attention for readers under the satisficing strategy.

One study on skimming for scientific document triage found that readers were hasty and incomplete, and documents were scrolled through quickly with attention paid to highly visual content and section headers [35]. Since information-dense content may be buried within pages of plain text, we see an opportunity for automated assistance in facilitating the discovery of these relevant information units that may otherwise be overlooked. Scientific documents are also laden with visual content, typographical cues (e.g., italicized, bold, or colored text), and structural information. Studies have found that readers draw on document features to support rapid comprehension via these macro- and micro-structures [8, 29, 36] and visual content [24, 64]. Scim's design as an AI-augmented reading interface enables readers to leverage AI assistance while retaining access to a paper's intrinsic visual and structural information.

2.2 Tools for Reading and Skimming

Researchers have long sought to equip readers with tools that support and augment their cognition while reading documents. The nascent days of human-computer interaction saw the introduction of augmented reading interfaces to support the reading process, including fluid documents that provided contextual access to supplemental information between lines of text [9], fluid hypertext [65], visualizations for social annotations within papers [21], and affordances for annotating papers and jumping readers to passages of interest [18, 51]. Since then, several classes of approaches have been proposed to support the various aspects of reading, such as document navigation and comprehension.

2.2.1 *Modified Scrolling Interactions.*

One line of research sought to facilitate the rapid exploration of long documents by modifying the behavior of reading interfaces during scrolling. Applications of content-aware scrolling were used to redefine the presentation order of content within a document [23], provide pseudo-haptic feedback when scrolling past relevant information [26], and dynamically resize document headings within paper thumbnails in a document viewer [6]. Spotlights implemented an attention allocation technique that pinned headings and figures as static overlays to a document as it was continuously scrolled [30].

2.2.2 *Typographical Cueing.*

Another approach involved augmenting reading interfaces with typographical cues (e.g., highlighting).

Studies in cognitive psychology have found that visual cueing mechanisms can be effective in focusing reader attention [10] and improving retention of material [17, 50]. The Semantize system used highlights to visualize sentiment within a document, and underlined words with positive or negative sentiment in different colors [61]. The ScentHighlights system used highlights to identify conceptually relevant text based on a user's query [11]. The HiText technique introduced dynamic graded highlighting of sentences within a document in accordance with their salience [63]. Modern reading interfaces also commonly support readers in marking regions of interest with a document with highlights or free-text annotations. The pervasiveness of highlighting as a technique for drawing readers' attention can be attributed to the von Restorff isolation effect, which states that an item isolated against a homogenous background will be more likely to be attended to and remembered [58]. Studies have since found evidence of this effect on the visual foraging behavior of readers on highlighted documents, finding that highlights attract about half of the total number of fixations within a document, and are often drawn to by readers' eyes [10].

2.2.3 Document Augmentations. Beyond typographical cues, other reading interface augmentations exist to specifically support the reading of scientific papers. For instance, online paper providers like ScienceDirect, PubMed, and Semantic Scholar provide readers with in-context citation information. Experimental systems have linked document text to marks within charts [28] and cells within tables [25], generated on-demand visualizations based on text within the paper [3], augmented static visualizations with animated [19] or interactive [37] overlays, and provided in-context definitions for nonce words [20]. We design Scim with inspiration from many of these prior augmented reading interfaces, augmenting scientific papers with interactive highlights that guide reader attention. Extending prior systems, Scim not only extracts salient sentences, but also classifies each highlight into common classes of information needs for readers.

2.2.4 Summarization. An alternative method to skimming a full paper is to read a shortened representation of the paper's content in the form of a summary. An author-provided summary is de facto included with each paper as an abstract, which researchers often read before continuing to the rest of the paper. Automated summarization has garnered significant interest from the natural language processing community, and extractive and abstractive methods for generating summaries from long-form documents have been developed over the years [2, 43, 52]. Some methods have even been proposed for generating extreme (single sentence) summaries, called TLDRs, from full papers [7].

However, providing only a summary to readers is often unsatisfactory. Despite recent improvements in the quality of generated summaries, they remain error-prone, susceptible to hallucination [66], and are not reliable enough to be used as a standalone replacement for reading the paper itself. Furthermore, summaries do not provide readers with the ability to interact with the full paper. For instance, as readers' goals and interests change while reading a paper, they may wish to explore certain sections in further detail. While traditional summaries cannot support this interaction, augmented reading interfaces naturally retain the context of the paper. We leverage natural language processing techniques to identify salient sentences and classify sentences into rhetorical facets using a pretrained language model, and present the output within a carefully-designed augmented reading interface to support the interactivity and context lacking in standalone summaries.

3 DESIGN MOTIVATIONS

To better understand how to design usable, intelligent skimming interfaces, we undertook an iterative design process. Our process began with interviews and observations of researchers, and continued into an evaluation of an early prototype of Scim (Section 3.1). In this section, we describe our design process. We then distill the lessons learned from our formative research into a set of design motivations (Section 3.2) to guide the design and implementation of intelligent, highlighting-based skimming support tools.

3.1 Design Methodology

Our design process consisted of several stages:

3.1.1 Formative interviews and observations. We conducted formative study sessions with eight researchers to better understand how they skim scientific papers. All researchers belong to the target user group for Scim, and were either graduate students or academic faculty. Researchers were first observed as they skimmed a paper of their choice. Then, the researchers were asked to describe their skimming process, including goals that they have while skimming, strategies that they employ while skimming, and any aspects of skimming that they found difficult or tedious.

3.1.2 Prototype development and evaluation. A prototype of Scim was iteratively designed and developed drawing inspiration from the formative interviews and observations. While many kinds of tools could support skimming, our design exploration focused specifically on skimming aids that would incorporate intelligent highlights.

The prototype was similar to the Scim system described in Section 4, with a few differences. First, the prototype's highlighting policy was different, resulting in fewer highlighted passages, and a less uniform distribution of highlights. Second, the prototype had no paragraph-level or facet-specific controls for the number of highlights, but rather only global-level controls on the number of highlights and switches to turn on or off individual facets.

Two preliminary usability studies were conducted with this prototype. 13 researchers were recruited from university mailing lists, and via direct outreach following purposive and snowball sampling approaches. Study sessions in both studies were 1-hour in length and conducted held on the Zoom video conferencing software. In both studies, participants skimmed papers with Scim for limited amount of time. Then, depending on the study, the participant was assigned one task where they demonstrated their understanding of the paper, for instance by outlining the paper or answering questions about the paper. Following the skimming task, participants were asked to comment on their interactions with Scim during the tasks and what aspects of the system required improvement.

3.1.3 Synthesis. One author conducted a thematic analysis [5, Ch. 5] of data from the formative study and preliminary evaluations. Notes and transcripts from study sessions were reviewed for themes

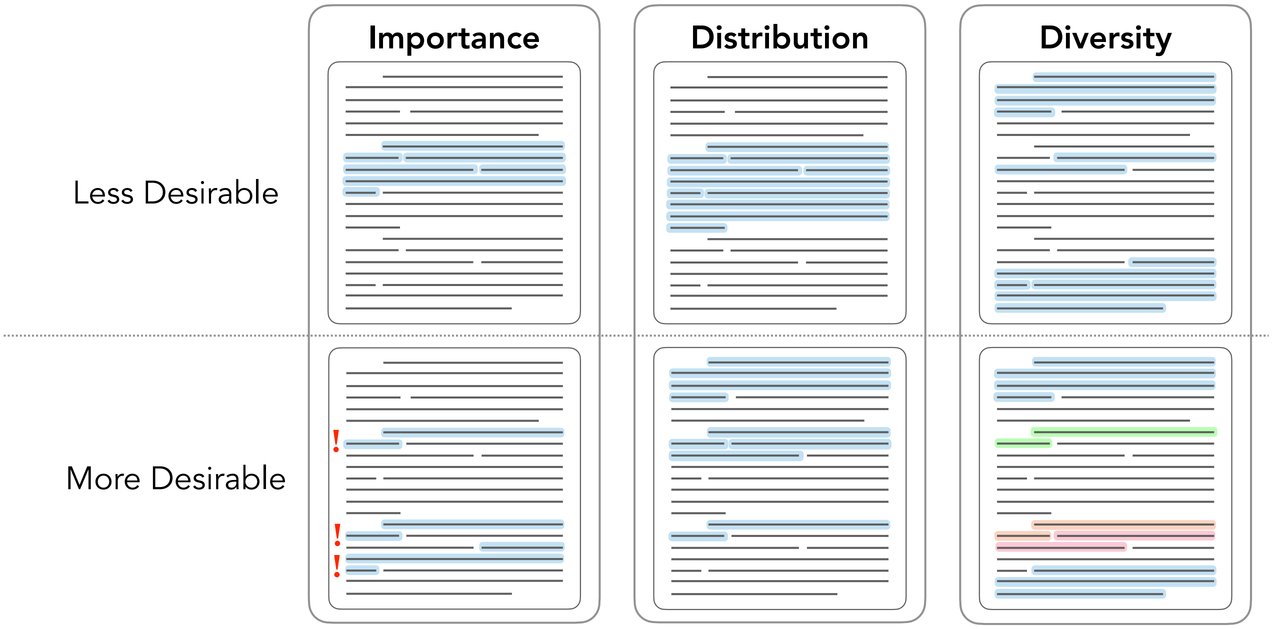


Figure 2: Our formative research revealed that intelligent highlights need to do more than pointing readers to important content. They should also be *well-distributed* throughout a paper (design motivations 3 and 5) and steer readers towards *diverse* content types (design motivation 1).

and supporting evidence. Themes were validated through discussion and review with a second author. Those themes that provided actionable guidance for design are reported in the next section.

3.2 Design Motivations

Here, we introduce seven design motivations for designers and builders of highlighting-based intelligent skimming interfaces that arose from the formative research described above. Each design motivation is listed with a description and supporting evidence from the formative research. When describing supporting evidence, we refer to participants in the formative interviews as F1–8 and participants in the preliminary evaluation as E1–13.

D1. Augment readers' skimming practices. Researchers described myriad strategies that they already used to skim papers. One common strategy was to read the abstract and introduction of a paper. Then, researchers consulted other key material in the paper, including bulleted lists of contributions (F1, F4, F6), summaries of results (F1–3), and conclusions (F1, F3, F6, F7). Consulting these parts of a paper was often regarded as conventional wisdom.

Researchers also employed strategies that were particular to their goals, paper, and level of comfort with the paper. Researchers relied on various visible cues in the text to help them identify important information, including typographical cues (e.g., italics, boldface) (F3, F6), structural cues (e.g., section headers) (F2, F6), visuals (e.g., figures and tables) (F1, F2, F4, F6, F8), and text position (e.g., inspecting the first sentences of paragraphs) (F2, F3, F6). We see these strategies as readers' strengths, and propose that skimming interfaces should let readers leverage these strategies, rather than impeding or replacing them.

D2. Highlight diverse kinds of content. Researchers' skimming goals were diverse. For instance, one researcher sought information about specific techniques introduced in a paper (F1). Other researchers wished to understand a paper's relationship to prior work or their own research, or to discover new research directions (F2–4, F7). Some wished for a high-level understanding suitable for discussing the paper with colleagues (F3, F7). These goals influenced researchers' skimming strategies, leading them to look for answers to specific questions, or for a subset of passages that provided a high-level understanding of the paper's objective, significance, key approaches, and experimental results. Skimming interfaces should support the diversity of readers' goals by supporting review of myriad kinds of paper contents.

D3. Support skimming in the long middle of the paper. Researchers noted that while one recommended strategy for skimming is to read the beginning and ends of paragraphs, important content may reside in the middle of paragraphs. Furthermore, we observed that when participants were asked to skim a paper, often their reading behavior better resembled a complete read of some passages of a paper, leading their skimming session to take quite a bit of time (F1, F3, F5). We propose that skimming tools should help readers identify important passages that conventional strategies do not reach—that is, content in the middle. A skimming tool also may need to provide sufficient highlights of middle material for readers to feel they are not missing important information without a close read.

D4. Minimize distraction. Without careful visual design, an augmented reading tool can occlude text or misdirect readers' attention.

Our prototypes incorporated a variety of text highlighting techniques, including underlines, lowlighting unimportant paper contents (as inspired by the ScholarPhi augmented reading tool [20]), and highlighting text with background color. Underlining was too subtle to consistently catch the reader's eye. Lowlighting some paper contents distracted readers, because it required additional effort for readers to tend to lowlighted content. Highlighting was chosen for its familiar use in PDF patterns, with the colors tuned to distinguish the categories of text, with as little contrast as possible to avoid an unpleasant visual pop-out effect.

Three of our design motivations related particularly to nuance around highlighting a paper's contents:

D5. Supply enough highlights. In our preliminary usability studies, researchers often desired more to see more highlights, after encountering long passages where they expected there was important information but which were not highlighted at all. Some researchers expressed a desire to see highlights distributed more uniformly throughout the paper (as opposed to highlights concentrated primarily in a paper's introduction or conclusion). We suggest the rule of thumb that a highlight should be provided around once per paragraph, and that readers should be able to request additional highlights in particularly dense passages.

D6. Accuracy is key. A side effect of introducing faceted highlights, where highlights were classified according to rhetorical category of the passage rather than being assigned a single color, was that it became obvious to readers when a model made a classification mistake, like labeling a passage about results as instead being about methods. Readers found themselves distracted when the classification of a passage clashed with their expectations of the purpose of the passage, and became more skeptical of the tool's capabilities (E11, E12). If skimming tools provide faceted highlights, it is particularly important to classify these highlights correctly.

D7. Support user customization. Participants desired more control over the amount of highlights shown by the prototype. Many suggested the capability of fine-tuning what was highlighted, either by adaptive personalization of the highlights (i.e., responding to passages that a reader has highlighted themselves or highlights they have deleted) or through manual adjustments (E5, E7, E8, E12).

A final takeaway from our formative research is that participants believed that their comfort using intelligent highlights would change over time, as they became more familiar with the features, the colors associated with the highlights, and the accuracy of the highlights. One participant described this as the issue of "getting used to seeing highlights that aren't my own" (E13). This observation motivated our choice of a longitudinal diary study as one of the summative evaluations of Scim (see Section 7).

4 SCIM

Drawing on our formative research, we designed Scim, an interface that provides intelligent support for skimming scientific papers. In this section, we describe the design of Scim, highlighting how subtle aspects of the system address the design motivations (referred to with abbreviations D1–7) introduced in the prior section.

4.1 Overview

In its envisioned usage, a reader interacts with Scim as a tool that supports and augments their typical skimming process (D1). As they might do without the tool, a reader would start their skim by reviewing a paper's title and abstract. Once they have finished reviewing these materials, a reader then begins their skim with a piecemeal review of the paper.

At this point, what differs about the skimming experience with Scim is that rather than relying on conventional skimming strategies like first sentences of paragraphs or selectively attending to individual sentences, a reader has the option of following Scim's highlights. These highlights extend into the parts of the paragraph that a reader may not notice during a typical skim (D3). Furthermore, the color of the highlights allow a reader to selectively attend to just those sentences that contain information about an aspect of the paper that they care about (D2).

Together, Scim's features provide holistic support for a highlight-driven, configurable skimming experience. Below, we describe each of these features in detail. Readers may also see how these features are used together in the screenshot in Figure 1 and the video figure included in the supplemental material.

4.2 Faceted Highlights

Scim intelligently highlights a paper to direct a reader's attention to a variety of passages that the reader may wish to review during their skim. These highlights support skimming with a combination of three important attributes:

Faceted. Because readers have different goals when skimming, Scim colorizes highlights according to *rhetorical facets* (which we refer to as "facets" below) (D2). The set of facets was selected to broadly encompass a variety of the kinds of information readers described in the formative study. The number of categories was limited to four, in part to minimize the variety of colors in a paper and to promote memorability of each facet, and in part because the selected categories were those that could be reliably detected (see the implementation section).

Numerous schemes exist for sentence-level classification of scientific literature into facets. Coarse-grained schemes classify sentences according to typical section names found in scientific literature [12, 22] and are composed of a small number of facets, while other fine-grained schemes rely on argumentative zones and conceptual structure [31, 32, 56, 57].

We formed our own taxonomy of four facets by combining aspects of a coarse-grained schema for classifying scientific abstracts [12] and the NOV_ADV category (i.e., sentences describing the novelty of a paper) from Argumentative Zoning [57]. A separate color is used to denote each facet. The four kinds of facets are: Objective (green), Novelty (orange), Method (blue), and Result (red). Examples of passages of each facet appear in Figure 3.

Low distraction. Text is highlighted using the familiar paradigm of a solid, colored rectangular box behind the text that is commonly used in PDF readers. As described in Section 3.2, this design was selected —rather than underlines or lowlighting unselected passages—to be both noticeable and minimally distracting (D4).

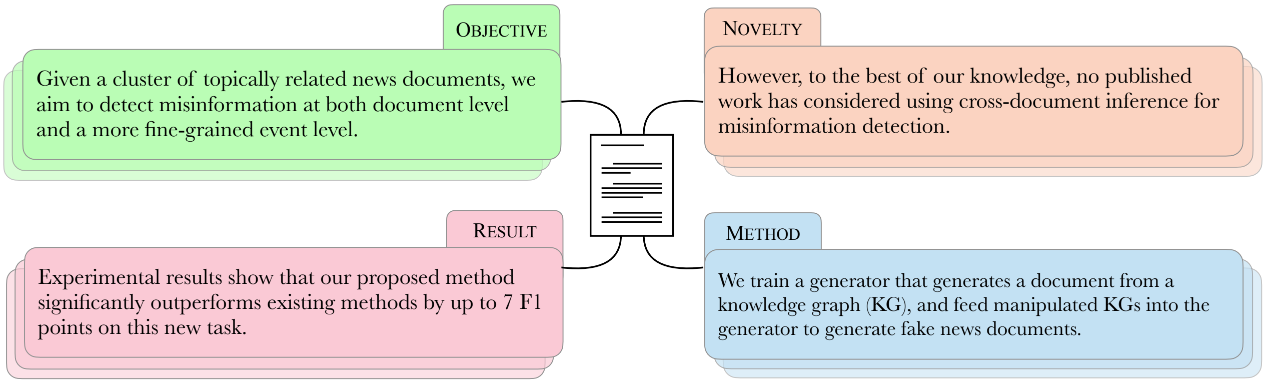

Objective
Given a cluster of topically related news documents, we aim to detect misinformation at both document level and a more fine-grained event level.
Novelty
However, to the best of our knowledge, no published work has considered using cross-document inference for misinformation detection.
Result
Experimental results show that our proposed method significantly outperforms existing methods by up to 7 F1 points on this new task.
Method
We train a generator that generates a document from a knowledge graph (KG), and feed manipulated KGs into the generator to generate fake news documents.

Figure 3: Scim classifies and highlights four facets of information commonly found in papers: Objective, Novelty, Method, and Result. These facets aim to surface specific kinds of paper content that align with common skimming goals identified in formative research, reflecting design guideline *D2*. Above, we show example passages matching each of the four facets. The passages appear in Wu et al.'s scientific paper, "Cross-document Misinformation Detection based on Event Graph Reasoning" [62].

The four facets are assigned the same color in any paper for which Scim is used, with the hope that these colors will become a learned association device for each facet, reducing the distraction that readers may experience when interpreting the highlights during first use (D4). To support readers in first use settings, a legend mapping highlight colors to facet categories appears in the header of the Scim application.

Distributed. To help a reader skim a paper in a way that is both piecemeal yet comprehensive, highlights are generated to be well-distributed across a text. An important goal of the highlighting scheme is to leave no paragraph, or no single section, completely without highlights, anticipating that readers may take this as a cue that they will have to deeply read the section to find information of consequence in that section. As described in the implementation section, heuristics were used to generate highlights approximately evenly throughout the paper, such that most pages had some important highlighted content while not overwhelming the reader in highlights (D5).

4.3 Controls

Readers' goals are diverse, and a single reader's goals may differ across papers, or even evolve over the course of reading a single paper. Because of this, Scim provides two kinds of controls, each of which support a different scenario of when a reader would wish to tailor the of highlighting to their goals (D7):

Paper-level controls. If a reader wishes to perform a more cursory skim of a paper, they can reduce the number of highlights. If they wish to inspect a paper more closely, they can increase the number of highlights. And if a reader does not care to read about a particular kind of content (e.g., if the reader decides they wish to learn about the results of a study but not the study's methodology), the reader can disable individual facets. In Scim's side bar, a reader can access facet-specific sliders that control the number of intelligent highlights for that facet. As a reader adjusts the slider for a facet, they can preview the effect on the highlights on the paper by seeing highlights appear and disappear in the paper, by watching highlight markers appear and disappear in the scrollbar, and by observing the label for the facet change in the sidebar to reflect the total number of highlights (Figure 4, right).

Paragraph-level controls. If a reader encounters a section where they wish for additional highlights (for instance, a dense paragraph of results that has been assigned no highlights by Scim's highlighting algorithm), the easiest way to access highlights is to request them at the paragraph level. When a reader hovers the mouse over a paragraph, a control appears in the margins of the paragraph that allows a reader to request or dismiss highlights from that paragraph, one highlight at a time (Figure 4, left). This feature provides flexibility where the paper-level controls do not, allowing readers to request highlights where they know they need them. With both paper- and paragraph-level controls, additional highlights are incorporated using a sentence prioritization score assigned during the document processing phase, as described in the implementation section.

4.4 Scrollbar annotations

A reader can discover where in a paper they can find information of a certain type by viewing highlight annotations in the scrollbar (Figure 1.D). This feature is inspired by edit wear and read wear affordances [21] and scrollbar marks in integrated development environments (e.g., [41]). When viewed in aggregate, these annotations can be suggestive of the structure of a paper, capable of implying if a paper has a particularly lengthy methods or results section. In addition, these marks provide feedback to a reader to help them adjust the number of highlights using global controls until the number and distribution of highlights in the scrollbar appears as they intend.

4.5 List of highlighted passages

Prior to skimming a paper, a reader can review key passages of a paper at a glance by opening up a list of all highlighted passages in the side bar (Figure 1.E). This list dynamically updates as readers adjust the controls for the number of highlights. Highlights in the

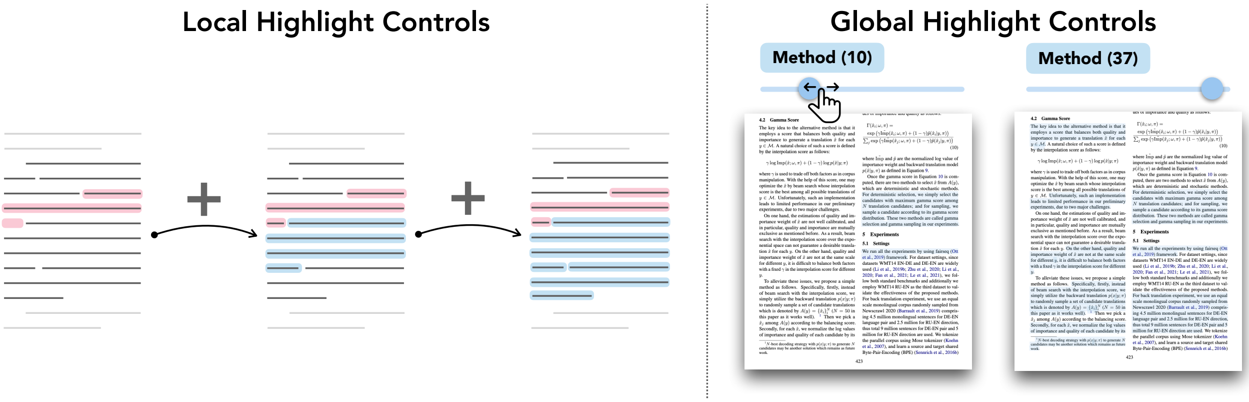


Figure 4: Scim allows readers to adjust the level of automated guidance they receive while skimming via local (paragraph-level) and global (document-level) highlight density controls.

browser are ordered by their position within the paper, and grouped by paper section (Figure 1.E). A colored indicator is also displayed to the right of each highlight in the browser, a subtle cue for its classified facet. Readers desiring more context for a highlighted passage can click that passage in the list, after which Scim scrolls the paper to that passage's position in the paper, a feature we refer to as *context linking*.

5 IMPLEMENTATION

Scim includes an end-to-end document processing pipeline that leverages a pretrained language model fine-tuned via weak supervision to identify and classify salient sentences within papers (Figure 5). Below, we describe the implementation of our pipeline that enables Scim to be deployed at scale over the domain of scientific literature.

5.1 Paper Component Extraction

We used the open-source Multimodal Document Analysis (MMDA) library [16] to process all textual tokens, mathematical symbols, section headers, and metadata within each paper (as PDF documents). We performed sentence segmentation and merged token and row bounding boxes to form sentence bounding boxes. We also labeled each sentence with its section header and paragraph index to support subsequent heuristics for sentence prioritization.

5.2 Sentence Classification

To classify sentences into facets, we adapted the sequential sentence classification architecture from [12] and substituted the base BERT model with a pretrained MiniLM model [59, 60]. The model is designed to incorporate the surrounding context—up to a combined sequence length of 512 words or 10 sentences—when classifying a target sentence. We first fine-tuned the model with train splits from the CSAbstruct dataset [12], a corpora of abstracts from computer science papers with manually-curated "gold" labels. Since the dataset only contained sentences from paper abstracts, we found the model insufficient for classifying sentences within full papers, and we sought to further fine-tune the model.

5.2.1 Data Programming. However, creating manually-curated datasets of "gold" facet labels for sentences from full papers is prohibitively expensive and time-consuming, potentially requiring hundreds of hours from domain experts annotating scientific literature. Instead, we used a data programming approach and weak supervision to further fine-tune our model. In weak supervision, we assume access to large unlabeled dataset and one or more weak supervision sources (e.g., heuristics encapsulating domain expertise, crowdsourcing, or knowledge bases), which are used to generate noisy and potentially conflicting labels for the dataset. Weak supervision offers a simple and model-agnostic way for us to incorporate domain expertise into our model, without the need for comprehensive manual annotation of a dataset. While a naive aggregation of these weak supervision sources could themselves be sufficient as a standalone sentence classifier, we sought to generalize beyond the coverage of these precise but incomplete labeling functions. We therefore employed a data programming paradigm to further unify and de-noise our weak supervision sources, creating a weakly labeled training set of sentences for downstream fine-tuning.

To create an unlabeled dataset, we extracted full paper sentences from the proceedings of NAACL 2018, 2019, and 2021, and ACL 2020, 2021, and 2022. In total, the dataset consisted of 3,051 papers with 606,400 unlabeled sentences. We then created weak supervision sources from heuristic rules and keyword matches to provide noisy facet labels for sentences in the dataset. For example, one rule-based supervision source detected sentence salience based on the presence of author intent via keywords such as "we", "our", or "this paper" and their aliases. Other supervision sources relied on keyword matches to perform facet labeling. For example, sentences were weakly labeled as Novelty if any relevant keywords (e.g., "novel", "propose", or "differ" and their aliases) could be found. We used Snorkel [46] to unify these weak supervision sources and output a dataset of weakly labeled sentences.

The dataset created through the data programming approach was further improved by weakly labeling additional negative sentences from full papers in the CSAbstruct. Briefly, we used the `all-mpnet-base-v2` model [53] from the Sentence Transformers library [49] to score how similar sentences were in the full text

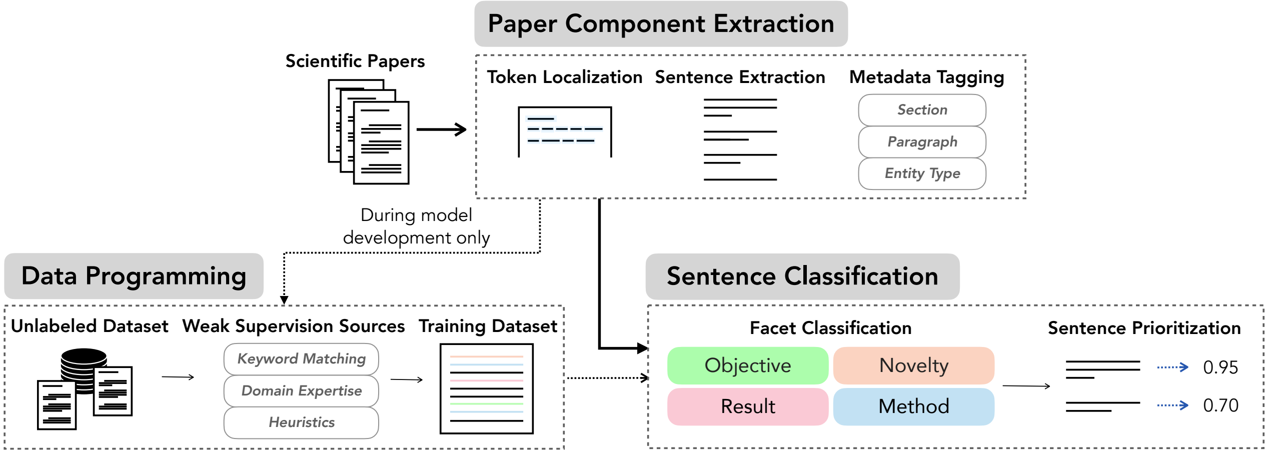


Figure 5: Overview of the implementation of Scim's paper processing pipeline. Scim processes scientific papers (as PDF documents) into localized sentences, and classifies these sentences into four rhetorical facets. Facet classification is performed with a large language model fine-tuned via a data programming approach. Highlights extracted by the model are visualized within Scim's user interface.

to the abstract; we labeled the most dissimilar[1] sentences to the abstract as not relevant for any facet. Model fine-tuning was done on an NVIDIA A6000 GPU, using 0.1 dropout rate and Adam optimizer [27] over 5 epochs, and $5 \cdot 10^{-5}$ learning rate. All parameters were determined empirically on the CSAbstruct validation split.

5.2.2 Evaluation. We conducted a preliminary evaluation of our sentence classification model over a set of 20 NLP papers. We recruited annotators from Upwork, a crowd-work site for hiring freelancers. All hired annotators were required to have experience with NLP and scientific writing. Annotators were asked to identify significant, complete sentences within each of the 20 papers. We provided annotators with detailed instructions, a snippet of which is shown below:

> *You have been hired by a scientific communication magazine (for example, Communications of the ACM or MIT Technology Review) to help create abridged versions of scientific papers. These are shorter versions of full academic papers that are often read by researchers from the same academic field who might not have time to read full-length manuscript. Your boss, the editor, has given you a target length of 200 words per paper page, with some wiggle room. You have to decide which sentences from the manuscript are significant, and thus should be kept while fitting within the target length. Anything you do not select will be mercilessly thrown away by the ruthless editor.*

Each paper took up to 30 minutes to annotate, and was annotated by three Upworkers using PAWLS [44]. Sentences selected by at least two of the three annotators were collected to form a test set of significant sentences. On this test set, our classification model achieved an F1 score of 0.533, compared to an annotator-annotator F1 score of 0.725 (which we consider as a gold-standard, i.e., a performance ceiling, since there is inherent variability in which sentences annotators believe are significant for skimming). We note that our goal with this preliminary evaluation was not to necessarily to progress the state-of-the-art, but rather to verify that the model was sufficient for enabling Scim's faceted highlights.

5.3 Sentence Prioritization

Scim's user interface selected sentences to be highlighted based on the model's facet label and probability score, along with other heuristics. One heuristic enforced consistency between facet labels and sentences within known sections (e.g., Method within a Methods section or Novelty within a Related Work section). Another heuristic encouraged a more uniform distribution of highlights throughout a paper, prioritizing sentences within paragraphs that did not already contain other highlighted sentences.

5.4 User Interface Implementation

The interface retains text markup that may aid readers in skimming, such as hyperlinks, interactive citations, bold and italicized text, and other visual cues provided by the authors. Scim is implemented as a web application built atop the PDF rendering platform `pdf.js` [42], and reduces adoption friction by adapting design patterns from existing document viewers and integrated development environments.

6 STUDY 1: IN-LAB USABILITY STUDY

We first conducted an in-lab usability study to assess how Scim affects readers' ability to search for specific kinds of information in a paper. Participants in the study were asked to complete a series of short tasks using both Scim and a normal document reader. Our usability study sought to answer the following two research questions:

[1] We empirically determined a cosine similarity threshold of 0.25.

RQ1. *Does Scim enable readers to skim papers more quickly?*

RQ2. *How does Scim affect the self-reported difficulty and ability to identify relevant information after a skim?*

6.1 Study Design

6.1.1 Participants. We recruited 19 participants (8 male, 10 female, 1 non-binary) for our study. We also conducted pilot studies with three additional participants to test and refine the study design; these individuals were not included in the main analysis. Our sole inclusion criteria was that the participants had some experience reading NLP papers, as they would be required to do so during the study. Participants received $25 USD compensation for their time. Participants ranged from 21 to 30 years of age, and included 11 doctoral students, 5 master's students, 2 software engineers, and 1 industry researcher. Participants self-reported an average of 3.78 (on a 5-point Likert scale) for comfort with reading NLP papers, suggesting participants overall were familiar with the type of literature used in the study. Participants were recruited through university-affiliated mailing lists and Slack channels, as well as from the pool of participants not selected for the diary study. We obtained Internal Review Board (IRB) approval from all involved institutions prior to conducting our study.

6.1.2 Procedure. Participants first provided consent and then were led through a tutorial of Scim's features. Our study used a within-subjects design, and consisted of three tasks, each with two sub-tasks, one for each of the two reading interface conditions—Scim and a normal document reader. We designed our study to be completed in under one hour to limit participant fatigue. The studies were conducted remotely via Zoom, an online video conferencing platform. In order to minimize biases, we counterbalanced the order of the reading interfaces and papers used in each task. Below, we describe the three tasks.

- *Task 1*: Participants skimmed a paper and identified a passage in the paper that described a key feature (e.g., dataset creation or evaluation) of the paper. They had no time limit to complete these tasks and informed the researcher upon completion. We intended for this task to be used only to familiarize participants with the two different interfaces, so we did include any measures from this task in our analysis.
- *Task 2*: Participants skimmed a paper and answered two multiple-choice questions based on information found in the paper. Answers to these questions *could* be found in text highlighted by Scim. We believed that the main points highlighted by Scim should be easier for participants to locate, and this task was designed to test that hypothesis.
- *Task 3*: Participants skimmed a paper and answered two multiple-choice questions based on information found in the paper. Answers to these questions *could not* be found in text highlighted by Scim. In contrast to Task 2, we anticipated Scim might not help as much in finding information that was outside of the highlights, and this task was designed to test that hypothesis.

Participants skimmed a different paper for each of the sub-tasks. The six papers for these tasks were selected from the proceedings of NAACL 2022, corresponding to the following types: (1) technical papers introducing new datasets or metrics, (2) exploratory papers investigating the effectiveness of current tools and proposing new design guidelines, and (3) technical papers proposing novel language models for specific applications. In Tasks 2 and 3, the multiple-choice questions focused on various aspects of a paper a reader might be interested in while skimming, such as the paper's evaluation metrics or the motivation behind a proposed method. For these questions, participants were given multiple attempts and asked to skim the paper until they located the correct answer, since we were interested in measuring the time it took to correctly answer a question using Scim, and control for participants who might otherwise guess.

For each question, we measured the following quantitative metrics:

- *Time* — The number of seconds taken by the participant to answer the question, from when the paper was first opened to when a final, correct answer choice was selected.
- *Accuracy* — A binary variable indicating whether the participant's first response to the question was correct.
- *Ease* — A seven-point Likert scale variable indicating the participant's self-assessment of the following prompt: "I found the task difficult."

6.1.3 Analysis. We compared readers' time, ease, accuracy, and subjective ratings for perceived difficulty using linear mixed-effects models [33] with reading interface as a fixed effect, task and question number as nested fixed effects, and participant as a random effect. We first conducted F-tests for any significant difference across the system variants, and then we conducted post-hoc t-tests for differences in the estimated fixed-effects between Scim and a normal document reader.

6.2 Results

Participants answered questions more quickly with Scim (M=94.3s, SD=74.9s) than with a normal document reader (M=117.7s, SD=76.4s), a statistically significant difference ($p < .05$). This difference was more pronounced in questions where the correct answer was located within one of the highlighted sentences suggested by Scim (Task 2), and there was no significant difference for questions where the correct answer was not located within one of the highlights (Task 3). There was no significant difference in participants' reported difficulty answering questions with Scim compared to a normal document reader. There was also no significant difference in participants' accuracy in answering comprehension questions with Scim (M=0.80, SD=0.40) compared to a normal document reader (M=0.76, SD=0.43).

After completing the tasks, some participants noted how Scim had a learning curve, and that continued usage may be needed to better understand how to effectively use Scim's highlights and facet color associations. These results are confirmatory of observations from our formative usability studies, and suggest that a longitudinal evaluation could provide more nuanced insights into the advantages and limitations of Scim.

7 STUDY 2: LONGITUDINAL DIARY STUDY

Participants in our usability studies noted how novel reading interfaces such as Scim—which interact with existing, traditionally

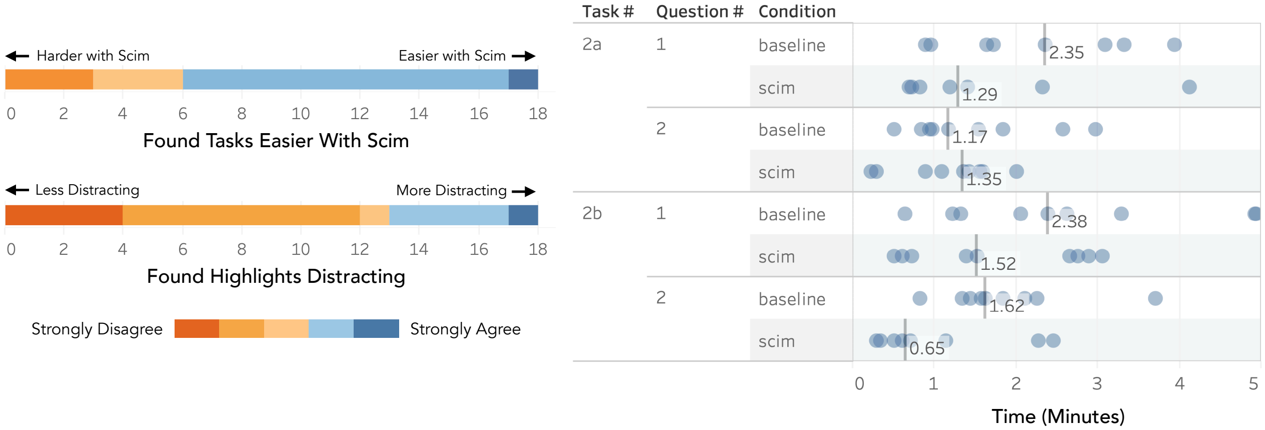


Figure 6: Results from the in-lab usability study. Left: Participants' subjective ease of completing tasks with Scim versus a normal document reader, and whether they found Scim's highlights distracting. Right: Timing results for Task 2. Vertical bars indicate median of responses.

manual processes—may have a learning curve requiring repeated usage to evaluate holistically. As a result, we conducted a two-week long diary study to understand how readers might use Scim within their daily skimming workflow. In contrast to an in-lab usability study, a diary study enables participants to engage with Scim for more realistic skims and demonstrates the feasibility of our paper processing approach at scale. It allows participants to retain their agency in choosing papers they want to read and when, and also mitigates any potential observational biases caused by study facilitators. A diary study also provides valuable longitudinal data that allows to study how various usage patterns may evolve over continued usage of Scim. We designed our diary study and qualitative analyses to explore the following research questions:

RQ1. *How do researchers make use of Scim as they skim papers?*
RQ2. *What value does a tool like Scim provide as researchers skim papers?*
RQ3. *In what circumstances do researchers find Scim useful?*
RQ4. *What are the limitations of Scim's approach to intelligent highlighting?*
RQ5. *What additional features should future intelligent skimming tools provide?*

7.1 Study Design

7.1.1 Participants. We recruited participants via university-affiliated mailing lists and Slack channels, and the authors' social media (Twitter). We required participants to have some prior experience in reading or writing research papers, and preferred those who had experience reading papers in the field of natural language processing (NLP). A total of 12 participants were recruited for the study (6 male, 6 female). All but one of the participants were current doctoral students, and one was a master's student. None of the participants also participated in our in-lab usability study. All participants who completed the study received $100 USD compensation for their time.

7.1.2 Reading Materials. While Scim's paper processing pipeline could feasibly be invoked at run-time to extract faceted highlights, for this study we wanted to eliminate any computational overhead participants could encounter when initially loading a paper with Scim. To still ensure that participants had an expansive selection of papers to read for the study, we preprocessed all papers from the proceedings of NAACL 2022. We chose these papers for several reasons: (1) Scim was fine-tuned on datasets primarily containing sentences from NLP papers, (2) all participants reported familiarity with reading NLP papers, and (3) participants may be more motivated if the study also allowed them to catch up on recent papers relevant to their own research. Specific papers not in this set but requested by participants during the study were also processed and made available within Scim on-demand.

7.1.3 Procedure. The diary study consisted of three parts for each participant: an introductory session, a two-week long observational period, and an exit interview. In the introductory session, participants were led through a tutorial of Scim's features, and given a few minutes to explore and ask any questions about the interface. Each participant was also given an online diary (a Google Docs document shared between each participant and the authors) to record their experience after each day's reading session with Scim. For each day throughout the observational period, participants were asked to skim at least one paper for 5 to 10 minutes and complete one entry in their online diary. Participants skimmed papers using a normal document reader for the first day to get familiarized with the diary study, before completing a total of nine skimming sessions using Scim and nine diary entries on subsequent days in the two-week long period. The short diary entries were designed to be completed immediately following each skimming session and elicit participants' feedback on how their skimming experience was influenced by Scim and what could have been improved. Each entry consisted of the following prompts:

(1) Which papers did you skim today, and how long did you spend skimming each one?

(2) What highlights (if any) drew your attention to something you might have missed without the highlights?
(3) Did highlights help you skim this paper? Explain.
(4) List one or more ways the system could have helped you better skim this paper.

After participants completed their diary entries, we conducted exit interviews to discuss their overall experience using Scim and feedback from their diary entries. We obtained Internal Review Board (IRB) approval from all involved institutions prior to conducting our study.

7.1.4 Analysis. Two authors conducted a thematic analysis on the diary entries and transcripts from exit interviews, following a qualitative approach described in [13]. One author extracted relevant utterances from participants, iteratively developing and refining a codebook and emergent themes, while another author verified these themes. 177 responses to questions from participants' diary entries were analyzed. We also instrumented and analyzed behavioral logs detailing interactions with Scim for each participant.

In the results section, we refer to participants with the pseudonyms P1–P12. The utterances presented below were edited to elide identifying information while preserving their meaning.

7.2 Results

In this section, we describe the results of the diary study as they relate to our five research questions.

7.2.1 How researchers make use of Scim as they skim papers (RQ1). All participants used Scim's highlight browser to view a condensed view of highlights more than once throughout their skims. Most participants used Scim's global highlight controls once or twice to configure an acceptable density of highlights (which then persisted throughout their skimming sessions), while a few adjusted the highlights multiple times for the different papers they skimmed. A detailed listing of per-participant usage frequencies appears in Table 1.

Many participants were satisfied with the default density of highlights (P2, P6, P7). All participants adjusted the number of highlights with paper-level controls at least once. The exit interviews confirmed that participants tended to tune the level of highlights to their preferred level on the first day, and then continued to use that level of highlights in subsequent skimming sessions. Some participants desired highlight controls with coarser granularity, e.g., providing two modes, one for the most important highlights, and another supporting a deeper skim (P2).

7.2.2 The value of Scim as a skimming aid (RQ2). **Helping readers attain a high-level understanding of papers.** Participants reported that Scim helped them attain a high-level understanding of the papers they were skimming (P5, P6, P9, P10). For these participants, Scim helped them identify key concepts and review the main ideas of papers, as one participant described:

> *For both papers the highlights showed important contributions and what the paper does. They were helpful for me to get a "gist" of the paper beyond what was in the abstract.* (P5)

Highlights helped participants review both the paper as a whole, as well as particular aspects of a paper that a participant wanted to understand. For instance, two participants noted how the highlights helped them to understand the results of the paper more quickly (P1, P2).

Drawing attention to information that would have been skipped. Scim drew readers' to interesting details in sections of papers that might have otherwise been skipped over in a typical skim (P1, P4, P11):

> *The ethical conduct and related sections were highlighted well which I would have usually skipped while skimming the paper, but this time that came to my attention.* (P1)

One participant described this as "slowing down," rather than speeding up, to skim with greater care:

> *This was a paper that is very light on methods and most content is about results, which I tend to skim over. So the highlights helped me slow down and slightly more carefully read a few places.* (P4)

Helping readers skim papers in a single pass. Some participants reported that Scim reduced the need to skim a paper in multiple passes. For these participants, a typical skimming approach without Scim consisted of first skimming the paper to identify relevant passages, and then re-reading passages of interest in greater detail (P5, P8). With Scim, they instead skimmed the paper a single time:

> *With highlights, I usually spend more time reading and understanding the highlighted content and skimming the other content. Without the highlight, I need to scan the entire content first, identify the critical points and then understand it. The highlights save me time in skimming the whole paper.* (P8)

> *They helped me scan the paper in order in one go and get relevant information. Otherwise I often have to come back to sections or search for specific details.* (P5)

7.2.3 Circumstances in which Scim is useful (RQ3). In 74 of 105 (70.4%) diary responses to the question, "Did highlights help you skim this paper?" participants replied in the affirmative. There were a handful of circumstances in which participants reported Scim was particularly useful:

Skimming dense texts. Scim could be particularly useful for passages and papers that were dense with text. Participants described how intelligent highlights made long, visualization-bare passages more approachable:

> *I largely relied on the highlights for skimming Section 3 and onwards, especially since they were text-dense passages with hardly any visual support (e.g., in the form of figures)* (P3)

> *The highlights were good in the experiment setup section. This paper was a human study and therefore there were many details on how the study was setup and what different conditions were considered. The details were dense so I might have skipped it if not for the highlights.* (P5)

Table 1: Behavioral log data from participants in the diary study. The number of times each participant used each of Scim's features (excluding faceted highlights) while skimming papers over a two-week deployment period.

Feature	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Highlight Browser	5	10	8	10	9	4	12	3	5	20	6	19
Global Highlight Controls	1	0	3	0	2	16	3	3	0	22	9	4
Local Highlight Controls	3	2	1	8	0	12	0	0	0	16	3	34
Context Linking	0	1	0	3	30	2	0	0	0	6	0	8

Participants also noted that they found Scim helpful not just for dense textual passages, but also for papers that were text-heavy as a whole, such as survey papers (P5, P11).

Skimming papers from unfamiliar domains. Intelligent highlights were seen as useful for supporting readers in seeking out important information in papers that were about a topic that they did not typically read about:

> *For me it was also generally useful for reading papers that were a little out of my comfort zone. Papers that were very focused on machine learning… In that case the highlighting helped me focus on, read, and conceptualize better certain parts of the methodology in order to better understand the conclusions.* (P10)

> *The highlights identify the critical steps in the method and facilitate me to focus on the principal modification of the model. This is especially useful when I am unfamiliar with the topic and don't know how to skim this paper.* (P8)

Skimming with low engagement. One participant noted that Scim helped them review the contents of a paper when they were not motivated to do a thorough read of this paper, remarking that "I ended up being not particularly interested in this paper, so I just looked at the highlights for a summary" (P4).

7.2.4 Limitations of Scim's approach to intelligent highlighting (RQ4).

Missing context. Participants frequently reported that they lacked the context to fully understand the highlighted passages they encountered when skimming with Scim (P1–3, P7-8, P11). While Scim was designed with the expectation that readers would simply look to the surrounding text for context, in practice, it could be disruptive for readers to seek this context. Necessary context may appear just before the highlighted passage in the paragraph, though in some cases it appeared in prior sections. One participant described the challenge as follows:

> *When reading the highlights, the context is often missing. Sometimes it is just in the lines before and after, but sometimes we need to find it which then makes reading difficult as there is now more back and forth instead of a linear reading.* (P7)

Integration with existing visual cues. Participants noted inconsistencies between the visual cues that authors introduced into their papers, and the highlights introduced by Scim (P2, P4, P5). For instance, authors may have bolded text that described key results, or emphasized contributions by placing them in a list. Scim does not take these author-provided cues into account, and often missed these passages.

Other issues in understanding paper contents. Participants noted several other cases where Scim's intelligent highlights behaved in unexpected ways. In some cases, Scim highlighted only one contribution in a list of bulleted contributions, when readers believed it should have highlighted all of them (P1, P7). Participants also found Scim often had unpredictable behavior when highlighting passages that contained dense math notation (P1, P6, P11), and wished that highlights extended into visual artifacts like tables and figures (P2, P5, P7, P12).

7.2.5 Directions for developing future intelligent skimming tools (RQ5).

Highlighting-based support like that provided by Scim represents just one way that tools could support skimming. Participants indicated several other ways in which future tools could have supported their skimming experiences:

Abstractive summarization. In some cases, Scim's highlights provided readers with more detail than readers wished to see, particularly if a reader desired only a very high-level understanding of the material (P6, P8). Several participants suggested that abstractive summarization of papers' contents could help reduce the effort to understand dense sections of papers (P1–2, P7–P8, P12):

> *I think the best way to summarize this results section (and probably others) is not in terms of highlights but maybe abstractive summarization with a bit of info potentially pulled from tables/graphs/figures/examples.* (P12)

Enhanced navigation support. One reader suggested that a paper's introductory material, such as an abstract, could serve as an index into related highlights in the rest of the paper (P2):

> *Sometimes, the abstract, intro, figures, and tables are key to skimming. I wonder if there is a way to link sentences in abstract/intro to related highlights in the rest of the paper.* (P2)

A related idea was to provide an index into a paper that supported navigation to passages that answered important questions about the paper:

> *A summary of the paper in QA format would be perfect. Example questions that I am interested in or try to extract when I skim paper are: What are the research questions? What are the novelties/contributions of this study? What data/model/evaluation methods do they use? What are the main results? What are the limitations?* (P8)

7.2.6 *Further validation of Scim's design.*

Participants' experiences also served to validate of several aspects of Scim's design:

Augmenting existing skimming strategies. Participants interleaved conventional skimming techniques with usage of Scim. To provide one example, P9 navigated through the main sections of a paper as they might in a typical skim, and used Scim's highlights to identify important information within those specific sections (P9). As described in Section 7.2.2, Scim drew readers attention to passages they might have otherwise skipped, and gave them license to skip passages they might have read. The information that Scim provided could be complementary to other visual cues: one participant described skimming by looking at both the section headers and the highlights (P4).

Previewing paper contents in the list of highlighted passages. While the predominant method of interaction with Scim was to view highlights within the paper, several participants described using the list of highlighted passages in the side bar to support navigation or help them gain a rapid understanding of paper contents (P7, P9, P10). One participant described referring to this list as an "extractive summary" (P2). Another participant described the list of highlights as a "better way to skim" in comparison to reading the paper with highlights, which at the time of their diary entry, they found made the paper "difficult to read" (P7).

Accustomization to Scim with repeated use. Over time, participants found themselves more comfortable with Scim, becoming accustomed to features that were sometimes found to be distracting during initial use:

> *On the first day I was not that trusting of the tool, so I tried to skim the paper by my own way… After ten days I trusted the highlight more, and I tried to rely on the highlights first to identify a sentence, to identify the key important information.* (P8)
>
> *I feel like I just got more used to the highlights. ... When I would see an objective highlight, I would trust it. I found the results highlights to be very helpful, so I would immediately focus on those. I would open the side panel right away instead of waiting during the end of the paper. I just got used to the tool, and I learned how to use it fast, depending on the paper and what I wanted to get from the paper.* (P10)

This suggests that for many users, the benefits of a tool like Scim may depend on continued use over several sessions.

8 DISCUSSION

8.1 Summary of Results

The in-lab usability study revealed that Scim reduced the amount of time that complete short information seeking tasks in scientific papers, albeit without significantly changing the difficulty of the tasks. In a diary study, 12 researchers made us of Scim for 10 days each, with most researchers invoking its features for browsing lists of highlights and adjusting highlights. Scim helped researchers develop a high-level understanding of papers, while helping them determine which passages to skim or to skip. Scim was seen as particularly useful for skimming dense texts and papers from unfamiliar domains. The studies also revealed opportunities for improvements in skimming aids, including awareness of context for understanding highlighted passages and integrating with existing visual cues.

8.2 Future Work

The design process of Scim and our studies present exciting opportunities for future research in AI-assisted augmented reading interfaces. We discuss a few of these below.

8.2.1 Model Enhancements. The effectiveness of Scim, like that of many other AI-infused user interfaces, is tightly coupled with the performance of the underlying AI models that enable its features. Beyond the facet classification approach we describe in this paper, highlights could instead be extracted with long-form summarization models tuned with heuristics based on our highlight-relevant design guidelines. Features such as a paper's hierarchical structure, author-cued content, and visual content could also be integrated into models to improve the quality and trustworthiness of the recommended paper content. Improvements to the performance of visual PDF processing modules—which all of the prior natural language techniques depend upon—would also significantly improve the end user experience within a reading interface like Scim; for instance, current PDF processing errors result in content such as footnotes, section headers, tables, or figures being concatenated with paper sentences before being passed to other downstream classification models.

8.2.2 Social Highlights. Our studies suggested one reason readers may be hesitant to adopt an augmented reading interface like Scim is distrust of the system's ability to provide the most relevant highlights. Some mentioned potentially greater trust in highlights created by other people (e.g., fellow researchers), and combining social and AI-powered highlights raises interesting design challenges for augmented reading. We note platforms such as Medium show "popular highlights," which suggests the potential for social highlights in reading tools for scientific literature as well.

8.2.3 Personalized Skimming Aids. As readers continue to interact with these augmented reading interfaces, we envision an opportunity for these AI-powered systems to adaptively learn from repeated reader interactions, perhaps providing personalized and proactive reading support to help mitigate some of the undesirable cognitive overhead introduced by these systems. They could also be tailored to readers' individual reading behaviors and tendencies, for instance by modeling users' characteristics of reading behavior, such as their experience reading papers within a particular field, their typical information needs during reading, or their goals for reading a particular paper.

8.3 Limitations

The results of our diary study should be considered amidst its limitations. The study was conducted with a small group of twelve researchers, mostly doctoral students, focused exclusively on NLP papers, and ended after two weeks. The behaviors we observed may not generalize to more senior researchers, additional academic disciplines, and longer periods of use. As interfaces like Scim are developed further, they should be evaluated on many disciplines, and the affordances and frequency of highlights may need further fine-tuning.

9 CONCLUSION

Our formative research yielded seven guidelines to motivate the design of intelligent tools for skimming scholarly papers. We instantiate these motivations in Scim, an intelligent skimming interface which supports skimming with faceted, evenly-distributed, minimally intrusive, configurable highlights. An in-lab usability study showed that participants found information in papers more quickly with Scim than with a baseline. In a longitudinal diary study, 12 participants used Scim daily for two weeks, and reported that Scim supports rapid, high-level skimming of papers. Scim was found to be particularly useful for dense textual passages and papers from unfamiliar domains. Together, these studies suggest the promise of intelligent skimming tools for supporting researchers in skimming scholarly literature.

ACKNOWLEDGMENTS

This project is supported in part by NSF Grant OIA-2033558, NSF RAPID award 2040196, ONR grant N00014-21-1-2707, and the Allen Institute for Artificial Intelligence (AI2).

REFERENCES

[1] Waleed Ammar, Dirk Groeneveld, Chandra Bhagavatula, Iz Beltagy, Miles Crawford, Doug Downey, Jason Dunkelberger, Ahmed Elgohary, Sergey Feldman, Vu Ha, Rodney Kinney, Sebastian Kohlmeier, Kyle Lo, Tyler Murray, Hsu-Han Ooi, Matthew Peters, Joanna Power, Sam Skjonsberg, Lucy Wang, Chris Wilhelm, Zheng Yuan, Madeleine van Zuylen, and Oren Etzioni. 2018. Construction of the Literature Graph in Semantic Scholar. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*. Association for Computational Linguistics, New Orleans - Louisiana, 84–91.

[2] Abdelkrime Aries, Djamel Eddine Zegour, and Walid-Khaled Hidouci. 2019. Automatic text summarization: What has been done and what has to be done. *CoRR* abs/1904.00688 (2019).

[3] Sriram Karthik Badam, Zhicheng Liu, and Niklas Elmqvist. 2019. Elastic Documents: Coupling Text and Tables through Contextual Visualizations for Enhanced Document Reading. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 661–671.

[4] Joeran Beel and Bela Gipp. 2009. Google Scholar's Ranking Algorithm : An Introductory Overview. In *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, Vol. 1. "International Society for Scientometrics and Informetrics", "Rio de Janeiro, Brazil", 230–2041.

[5] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI research: Going behind the scenes. *Synthesis lectures on human-centered informatics* 9, 1 (2016), 1–115.

[6] George Buchanan and Tom Owen. 2008. Improving skim reading for document triage. In *Proceedings of the second international symposium on Information interaction in context - IIiX '08*. ACM Press, London, United Kingdom, 83.

[7] Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel Weld. 2020. TLDR: Extreme Summarization of Scientific Documents. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 4766–4777.

[8] Patricia L. Carrell. 1985. Facilitating ESL Reading by Teaching Text Structure. *TESOL Quarterly* 19, 4 (1985), 727–752.

[9] Bay-Wei Chang, Jock D. Mackinlay, Polle T. Zellweger, and Takeo Igarashi. 1998. A Negotiation Architecture for Fluid Documents. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, California, USA) *(UIST '98)*. Association for Computing Machinery, New York, NY, USA, 123–132.

[10] Ed Chi, Michelle Gumbrecht, and Lichan Hong. 2007. *Visual Foraging of Highlighted Text: An Eye-Tracking Study*. Vol. 4552. Springer Berlin Heidelberg, Berlin, Heidelberg.

[11] Ed H. Chi, Lichan Hong, Michelle Gumbrecht, and Stuart K. Card. 2005. ScentHighlights: highlighting conceptually-related sentences during reading. In *Proceedings of the 10th international conference on Intelligent user interfaces - IUI '05*. ACM Press, San Diego, California, USA, 272.

[12] Arman Cohan, Iz Beltagy, Daniel King, Bhavana Dalvi, and Daniel S. Weld. 2019. Pretrained Language Models for Sequential Sentence Classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Online, 3691–3697.

[13] John W Creswell and Cheryl N Poth. 2016. *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications, United States of America.

[14] Geoffrey Duggan and Stephen Payne. 2009. Text Skimming: The Process and Effectiveness of Foraging Through Text Under Time Pressure. *Journal of experimental psychology. Applied* 15 (09 2009), 228–42.

[15] Geoffrey B. Duggan and Stephen J. Payne. 2011. Skim reading by satisficing: evidence from eye tracking. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 1141.

[16] Allen Institute for Artifical Intelligence. 2022. MMDA - Multimodal Document Analysis. https://github.com/allenai/mmda

[17] Robert L. Fowler and Anne S. Barker. 1974. Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology* 59, 3 (1974), 358–364.

[18] Jamey Graham. 1999. The reader's helper: a personalized document reading environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*. ACM Press, Pittsburgh, Pennsylvania, United States, 481–488.

[19] Tovi Grossman, Fanny Chevalier, and Rubaiat Habib Kazi. 2015. Your Paper is Dead!: Bringing Life to Research Articles with Animated Figures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 461–475.

[20] Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, and Marti A. Hearst. 2021. Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 413, 18 pages.

[21] William C. Hill, James D. Hollan, Dave Wroblewski, and Tim McCandless. 1992. Edit Wear and Read Wear. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 3–9.

[22] Kenji Hirohata, Naoaki Okazaki, Sophia Ananiadou, and Mitsuru Ishizuka. 2008. Identifying Sections in Scientific Abstracts using Conditional Random Fields. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*. Association for Computational Linguistics, Hyderabad, India.

[23] Edward W. Ishak and Steven K. Feiner. 2006. Content-aware scrolling. In *Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST '06*. ACM Press, Montreux, Switzerland, 155.

[24] S. Keshav. 2007. How to read a paper. *Comput. Commun. Rev.* 37 (2007), 83–84.

[25] Dae Hyun Kim, Enamul Hoque, Juho Kim, and Maneesh Agrawala. 2018. Facilitating Document Reading by Linking Text and Tables. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, Berlin Germany, 423–434.

[26] Juho Kim, Amy X. Zhang, Jihee Kim, Robert C. Miller, and Krzysztof Z. Gajos. 2014. Content-aware kinetic scrolling for supporting web page navigation. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, Honolulu Hawaii USA, 123–127.

[27] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.).

[28] Nicholas Kong, Marti A. Hearst, and Maneesh Agrawala. 2014. Extracting references between text and charts via crowdsourcing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 31–40.

[29] Natasha Lacroix. 1999. Macrostructure construction and organization in the processing of multiple text passages. *Instructional Science* 27, 3/4 (1999), 221–233.

[30] Byungjoo Lee, Olli Savisaari, and Antti Oulasvirta. 2016. Spotlights: Attention-Optimized Highlights for Skim Reading. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 5203–5214.

[31] Maria Liakata, Shyamasree Saha, Simon Dobnik, Colin Batchelor, and Dietrich Rebholz-Schuhmann. 2012. Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics* 28, 7 (April 2012), 991–1000.

[32] Maria Liakata, Simone Teufel, Advaith Siddharthan, and Colin Batchelor. 2010. Corpora for the Conceptualisation and Zoning of Scientific Papers. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), Valletta, Malta.

[33] Mary J. Lindstrom and Douglas M. Bates. 1990. Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics* 46, 3 (1990), 673–687.

[34] Ziming Liu. 2005. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation* 61, 6 (Jan. 2005), 700–712. Publisher: Emerald Group Publishing Limited.

[35] Fernando Loizides and George Buchanan. 2009. An Empirical Study of User Navigation during Document Triage. In *Research and Advanced Technology for Digital Libraries*. Vol. 5714. Springer Berlin Heidelberg, Berlin, Heidelberg, 138–149.

[36] Tonja Machulla, Mauro Avila, Pawel Wozniak, Dillon Montag, and Albrecht Schmidt. 2018. Skim-Reading Strategies in Sighted and Visually-Impaired Individuals: A Comparative Study. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference* (Corfu, Greece) *(PETRA '18)*. Association for Computing Machinery, New York, NY, USA, 170–177.

[37] Damien Masson, Sylvain Malacria, Edward Lank, and Géry Casiez. 2020. Chameleon: Bringing Interactivity to Static Digital Documents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13.

[38] Michael E. Masson. 1982. Cognitive processes in skimming stories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8, 5 (1982), 400–417.

[39] Michael E. J. Masson. 1983. Conceptual processing of text during skimming and rapid sequential reading. *Memory & Cognition* 11, 3 (May 1983), 262–274.

[40] Martha J. Maxwell. 1972. Skimming and Scanning Improvement: The Needs, Assumptions and Knowledge Base. *Journal of Reading Behavior* 5, 1 (March 1972), 47–59. Publisher: SAGE Publications.

[41] Microsoft. 2022. Scroll bar map mode and bar mode - Visual Studio (Windows). https://docs.microsoft.com/en-us/visualstudio/ide/how-to-track-your-code-by-customizing-the-scrollbar

[42] Mozilla. 2022. PDF.js. https://github.com/mozilla/pdf.js

[43] Ani Nenkova and Kathleen McKeown. 2012. A Survey of Text Summarization Techniques. In *Mining Text Data*, Charu C. Aggarwal and ChengXiang Zhai (Eds.). Springer US, Boston, MA, 43–76.

[44] Mark Neumann, Zejiang Shen, and Sam Skjonsberg. 2021. PAWLS: PDF Annotation With Labels and Structure. arXiv:2101.10281 [cs.CL]

[45] David N. Rapp and Paul van den Broek. 2005. Dynamic text comprehension: An integrative view of reading. *Current Directions in Psychological Science* 14, 5 (2005), 276–279.

[46] Alexander Ratner, Christopher De Sa, Sen Wu, Daniel Selsam, and Christopher Ré. 2016. Data Programming: Creating Large Training Sets, Quickly. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (Barcelona, Spain) *(NIPS'16)*. Curran Associates Inc., Red Hook, NY, USA, 3574–3582.

[47] Keith Rayner, Elizabeth R. Schotter, Michael E. J. Masson, Mary C. Potter, and Rebecca Treiman. 2016. So Much to Read, So Little Time: How Do We Read, and Can Speed Reading Help? *Psychological Science in the Public Interest* 17, 1 (May 2016), 4–34.

[48] William R. Reader and Stephen J. Payne. 2007. Allocating Time across Multiple Texts: Sampling and Satisficing. *Hum.-Comput. Interact.* 22, 3 (aug 2007), 263–298.

[49] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://arxiv.org/abs/1908.10084

[50] John P. Rickards. 1980. Notetaking, Underlining, Inserted Questions, and Organizers in Text: Research Conclusions and Educational Implications. *Educational Technology* 20, 6 (1980), 5–11.

[51] Bill N. Schilit, Gene Golovchinsky, and Morgan N. Price. 1998. Beyond Paper: Supporting Active Reading with Free Form Digital Ink Annotations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI '98)*. ACM Press/Addison-Wesley Publishing Co., USA, 249–256.

[52] Athar Sefid and C. Lee Giles. 2022. SciBERTSUM: Extractive Summarization for Scientific Documents. *CoRR* abs/2201.08495 (2022).

[53] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. MPNet: Masked and Permuted Pre-training for Language Understanding. *arXiv preprint arXiv:2004.09297* (2020).

[54] Craig S. Tashman and W. Keith Edwards. 2011. Active Reading and Its Discontents: The Situations, Problems and Ideas of Readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2927–2936.

[55] Carol Tenopir, Donald King, Sheri Edwards, and Lei Wu. 2009. Electronic Journals and Changes in Scholarly Article Seeking and Reading Patterns. *Carol Tenopir* 61 (01 2009).

[56] Simone Teufel and Marc Moens. 2002. Articles Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status. *Computational Linguistics* 28, 4 (2002), 409–445.

[57] Simone Teufel, Advaith Siddharthan, and Colin Batchelor. 2009. Towards Domain-Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Singapore, 1493–1502.

[58] Hedwig von Restorff. 1933. Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung* 18, 1 (Dec. 1933), 299–342.

[59] Wenhui Wang, Hangbo Bao, Shaohan Huang, Li Dong, and Furu Wei. 2021. MiniLMv2: Multi-Head Self-Attention Relation Distillation for Compressing Pretrained Transformers. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Association for Computational Linguistics, Online, 2140–2151. https://doi.org/10.18653/v1/2021.findings-acl.188

[60] Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. MINILM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS'20)*. Curran Associates Inc., Red Hook, NY, USA, Article 485, 13 pages.

[61] Alan J. Wecker, Joel Lanir, Osnat Mokryn, Einat Minkov, and Tsvi Kuflik. 2014. Semantize: visualizing the sentiment of individual document. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces - AVI '14*. ACM Press, Como, Italy, 385–386.

[62] Xueqing Wu, Kung-Hsiang Huang, Yi Fung, and Heng Ji. 2022. Cross-document Misinformation Detection based on Event Graph Reasoning. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Seattle, United States, 543–558.

[63] Qian Yang, Gerard de Melo, Yong Cheng, and Sen Wang. 2017. HiText: Text Reading with Dynamic Salience Marking. In *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*. ACM Press, Perth, Australia, 311–319.

[64] Ji Soo Yi. 2014. QnDReview: read 100 CHI papers in 7 hours. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 805–814.

[65] Polle T. Zellweger, Bay-Wei Chang, and Jock D. Mackinlay. 1998. Fluid Links for Informed and Incremental Link Transitions. In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia: Links, Objects, Time and Space* (Pittsburgh, Pennsylvania, USA) *(HYPERTEXT '98)*. Association for Computing Machinery, New York, NY, USA, 50–57.

[66] Zheng Zhao, Shay B. Cohen, and Bonnie Webber. 2020. Reducing Quantity Hallucinations in Abstractive Summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 2237–2249.