



DOI:10.1145/3659096

**Augmenting scholarly documents through AI-powered interactive reading interfaces.****BY KYLE LO, JOSEPH CHEE CHANG, ANDREW HEAD, JONATHAN BRAGG, AMY X. ZHANG, CASSIDY TRIER, MARTI A. HEARST, DANIEL S. WELD, ET AL.**

# The Semantic Reader Project

THE EXPONENTIAL GROWTH in the rate of scientific publication<sup>4</sup> and increasing interdisciplinary nature of scientific progress<sup>27</sup> makes it increasingly hard for scholars to keep up with the latest developments. Academic search engines, such as Google Scholar and Semantic Scholar, help scholars discover research papers. Techniques such as automated summarization help scholars triage research papers.<sup>5</sup> But when it comes to actually *reading* research papers, the process, often based on a static PDF format, has remained largely unchanged for many decades. This is a problem because digesting technical research papers in their conventional formats is difficult.<sup>2</sup>

In contrast, interactive and personalized documents have seen significant adoption in domains outside of academic research. For example, news websites such as the *The New York Times* often present interactive articles with explorable visualizations that allow readers to understand complex data in a personalized way. E-readers, such as the Kindle, provide in situ context to help readers better comprehend complex documents, showing inline term definitions and tracking the occurrence of characters in a long novel. While prior work has envisioned how authoring tools can reduce effort in creating interactive scientific documents,<sup>13</sup> they have not seen widespread adoption. Furthermore, millions of research papers are locked in the rigid and static PDF format, whose low-level syntax makes it extremely difficult for systems to access semantic content, augment interactivity, or even provide basic reading functionality for assistive tools such as screen readers.

Fortunately, recent work on layout-aware document parsing<sup>14,34</sup> and language models for scientific text<sup>3</sup> shows promise for extracting the content of PDF documents and building systems that can better understand their semantics. This raises an excit-

**» key insights**

- The experience of reading information-dense scientific papers has remained unchanged in decades, relying on aging formats with static content and low accessibility.
- Advances in AI and HCI can power intelligent, interactive, and accessible reading interfaces to improve scholarly reading.
- The Semantic Reader project introduces novel user interfaces that augment traditional PDFs to improve reading experiences for scholars, as shown with evaluations both in the lab and the wild.
- We have released an open platform with a public reader tool and software components for the community to experiment with their own AI reading interfaces.



ing challenge: *Can we create intelligent, interactive, and accessible reading interfaces for research papers, even atop existing PDFs?*

To explore this question, we present the **Semantic Reader Project**, a broad collaborative effort across multiple non-profit, industry, and academic institutions to create interactive, intelligent reading interfaces for research papers. This project consists of three pillars: research, product, and open-science resources. On the research front, we are combining artificial intelligence (AI) and human-computer interaction (HCI) research to design, prototype, and evaluate novel, AI-powered interactive reading interfaces that address a variety of user challenges faced by today's scholars. On the product front, we are developing the Semantic Reader (Figure 1),<sup>a</sup> a freely available reading interface that integrates features from our research prototypes as they mature.<sup>b</sup> Finally, we are releasing an open-science research platform<sup>c</sup> with resources that drive both our research and product. The platform brings together open source software,<sup>24</sup> AI models,<sup>5,8,16,34</sup> and open datasets<sup>22,25</sup> to support continued work in this area.

In this article, we summarize our efforts under the *research pillar* of the Semantic Reader Project. We structure our discussion around five high-impact opportunities, each with a dedicated section, to improve the research-paper reading experience:

► **Unlocking Citations for Discovery:** Identifying relevant papers to read is a long-standing challenge for scholars. While exploring citations is a crucial strategy, making sense of the many citations encountered while reading and prioritizing them can be overwhelming. In this section, the article explores ways to visually augment research papers to help readers prioritize their paper exploration while conducting literature reviews.

► **Navigation and Efficient Reading:** The exponential growth of publication makes it difficult for scholars to keep up to date with the literature—

## Can we create intelligent, interactive, and accessible reading interfaces for research papers, even atop existing PDFs?

scholars need to efficiently read many papers while making sure they capture enough details in each. This section explores how support for non-linear reading can help readers consume research papers more efficiently.

► **In Situ Explanations for Better Comprehension:** Research papers can be difficult to understand, due to the complexity of their text, scientific terminology, and expectations about readers' prior knowledge. In this section, the article explores how providing definitions, summaries, and auxiliary non-textual explanations can benefit reader comprehension.

► **Bootstrapping Literature Synthesis with Related Work Sections:** The sensemaking process of synthesizing knowledge scattered across many papers is effortful but necessary to produce literature reviews or identify new research opportunities. This section explores how interfaces to support reading across many related work sections can help readers explore different threads of prior research and make connections between many papers.

► **Dynamic Documents for Improved Accessibility:** Static PDFs are an ill-suited format for many reading interfaces. For example, PDFs are notoriously incompatible with screen readers and represent a significant barrier for blind and low-vision readers. Furthermore, an increasing number of scholars access content on mobile devices, on which PDFs of papers are difficult to read. In this section, the article explores methods for converting legacy papers to more accessible representations.

We present research prototypes developed under our project to illustrate how one might apply AI assistance paired with interactive user interface design when tackling these opportunities. We conclude by discussing ongoing research opportunities in both AI and HCI for developing the future of scholarly reading interfaces and provide pointers to our open resources to invite the broader research community to join our effort.

**Unlocking Citations for Discovery**  
Scholars use various methods to dis-

<sup>a</sup> Semantic Reader: <https://bit.ly/3Z6BW3L>

<sup>b</sup> Available for over 7.2M papers as of early August 2024.

<sup>c</sup> <https://openreader.semanticscholar.org>

cover relevant research papers to read, including search engines, word of mouth, and browsing familiar venues. However, once they find one research paper, it is especially common for scholars to use its references and citations to further expand their knowledge of a research area. This behavior, sometimes referred to as *forward/backward chaining* or *foot-note chasing*, is ubiquitous and has been observed across many scholarly disciplines.<sup>30</sup> Supporting this, one popular feature in the Semantic Reader is in situ Paper Cards that pop up when readers click on an inline citation, dramatically reducing the interaction cost caused by jumping back and forth between inline citations and their corresponding references at the end of a research paper (Figure 1). Despite this affordance, during literature reviews, readers may still be overwhelmed trying to make sense of the tens to hundreds of inline citations in each paper.<sup>6,9</sup> Conversely, when reading a given paper, a reader cannot see relevant follow-on research papers that cited the current paper. Interactive reading interfaces can help scholars more effectively explore citations to important relevant work in both these directions.

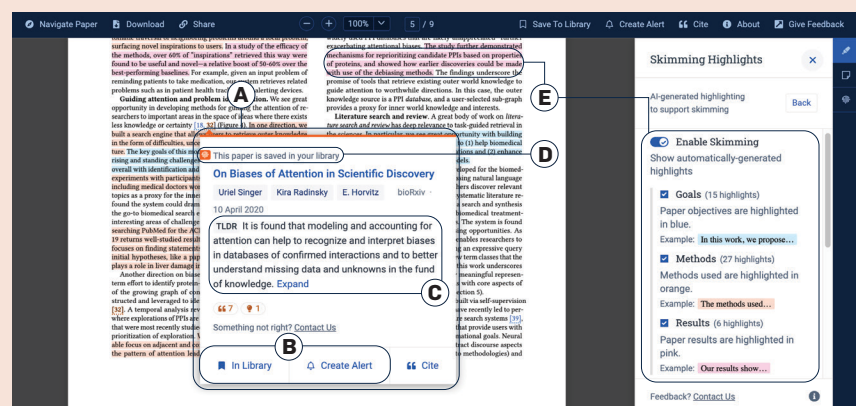
**Making sense of inline citations with personalized context and visual cues.** While most prior work on supporting research paper discovery has focused on developing bespoke interfaces for recommender systems or visualizations based on the citation graph<sup>15</sup> and paper content,<sup>8</sup> research paper discovery via inline citations in a reading interface is important yet under-explored. One study estimates that reading and exploring inline citations accounts for around one in five research paper discoveries during active research.<sup>21</sup> However, while all inline citations are relevant to the current research paper, it is likely that some are more relevant to the current reader than others. For example, a reader reading papers about *aspect extraction of online product reviews* to learn more about *natural language processing techniques* would be less interested in citations to research papers around *e-commerce and marketing*. In addition, citations to the same research papers often have

different surface forms across papers (that is, reference numbers), making it all the more difficult for readers to keep track of all the inline citations they should explore or have already explored during literature reviews.

Intelligent reading interfaces have the potential to address these issues by *augmenting* inline citations with *personalized visual cues and context* to help a reader spot and make sense of relevant work amid a sea of encountered citations. For example, CiteSee<sup>6</sup> highlights inline citations with different colors to signal to a reader whether a paper has been previously

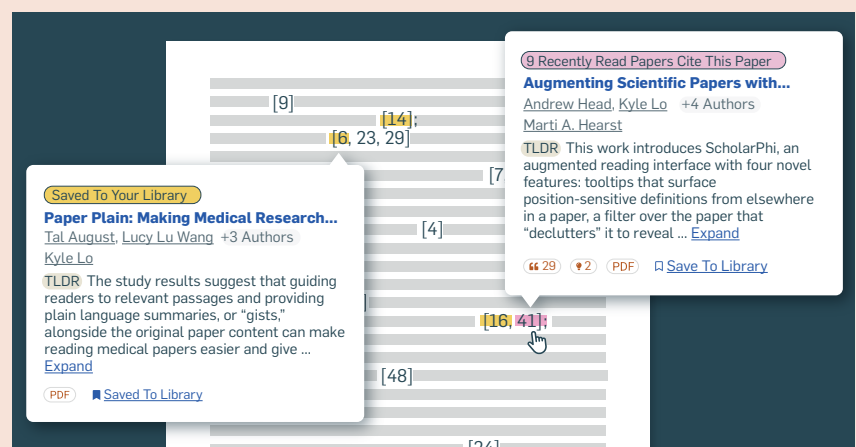
encountered or saved, as well as how it might be relevant to their interests based on their reading history and publication record. CiteSee<sup>6</sup> also imbues the Paper Cards with *personalized context* to explain how cited works relate to the reader, such as citing contexts from other papers that were familiar to the reader (Figure 2). While CiteSee focused on surfacing structured signals (for example, citations between familiar and unfamiliar encountered papers), as the capability of AI methods increases, an exciting future direction is to provide additional personalized explanations

**Figure 1. The Semantic Reader Project consists of research, product, and open-science resources. The Semantic Reader product† is a free interactive interface for research papers. Semantic Reader supports useful augmentations atop the existing PDF—for example, (A) in situ Paper Cards when clicking inline citations and integration with Semantic Scholar, and (B) save to library. We continue to integrate research features into this product as they mature—for example, (C) AI-generated summaries, (D) CiteSee personalized context, and (E) Scim automated highlights.**



† Semantic Reader: <https://www.semanticscholar.org/product/semantic-reader>

**Figure 2. CiteSee<sup>6</sup> highlights citations to familiar papers (for example, recently read or saved in their libraries) as well as unfamiliar papers to help readers avoid overlooking important citations when conducting literature reviews. Clicking on *Expand* surfaces additional context, such as citing sentences from recently read papers.**



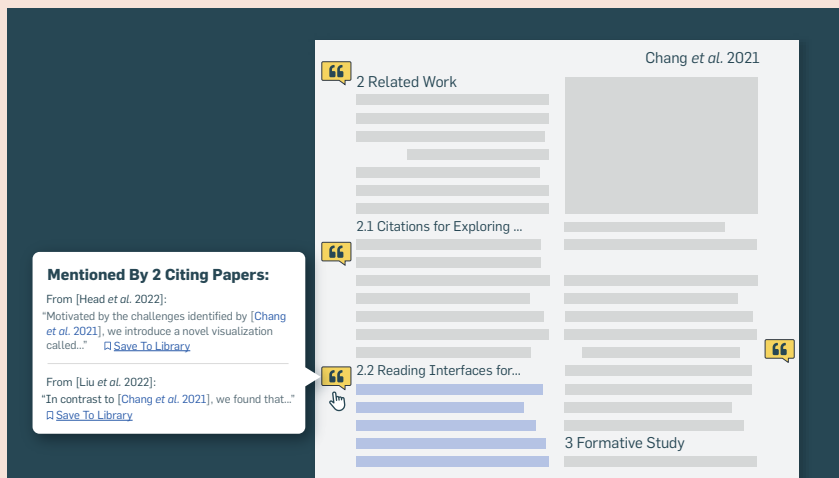
based on the content of papers; for example, generating a description of how an encountered citation may be relevant to one of the reader's publications or how it compares and contrasts with a familiar paper saved in the reader's library.

### Surfacing incoming citations to enable awareness of follow-on work.

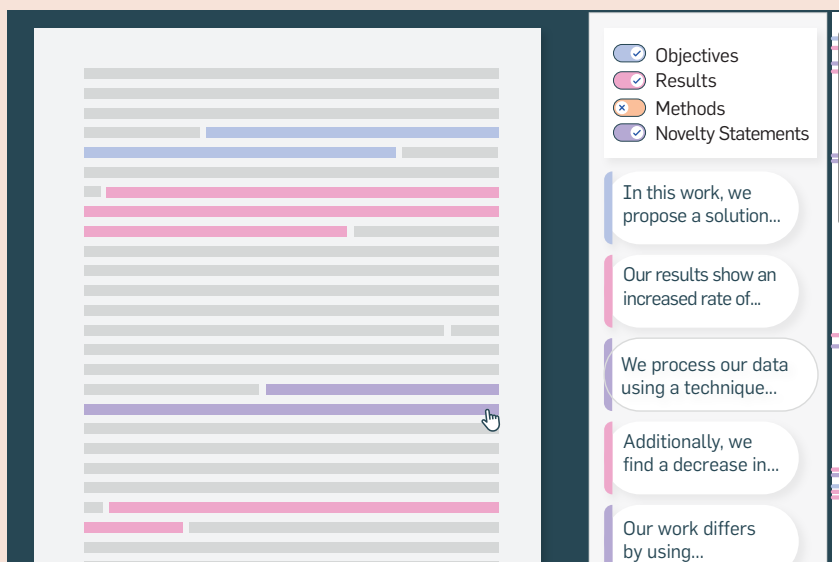
While augmenting inline citations can help readers better prioritize the most relevant prior work,<sup>6</sup> many relevant research papers are not cited in the first place—in particular, follow-

on work published after the current paper is not cited. To address this, reading interfaces could bring additional citations into the current paper as annotations, so that readers can become aware of relevant work not cited in the current paper. For example, taking inspiration from social document annotation systems,<sup>40</sup> CiteRead<sup>33</sup> creates margin notes in the current paper with citation sentences, or *citations*, from citing papers published afterward, as a form of commentary on the current paper (Figure 3).

**Figure 3.** CiteRead<sup>33</sup> finds subsequently published citing research papers, extracts the citation context, and localizes it to relevant parts of the current research paper as margin notes. This allows readers to become aware of important follow-on work and explore them in situ.



**Figure 4.** The Scim<sup>10</sup> interface guides reader attention using color highlights corresponding to discourse facets. A sidebar allows users to toggle facets on/off. Clicking a color-coded snippet scrolls the reader to the relevant passage.



To produce these annotations automatically, reading interfaces could leverage AI to determine which papers have the most relevant commentary to surface, so as to avoid overwhelming the reader with excessive annotation. CiteRead, for example, determines relevant work using a trained model that derives signal from citational discourse and textual similarity—that is, from scientific paper embeddings.<sup>8</sup> Reading interfaces should also determine how to best localize margin annotation to a specific position in the paper, a particularly challenging task because citations do not typically reference specific locations or passages in a cited paper. To address this, CiteRead determines when it is feasible to localize to particular spans of text in the paper being read, and when to fall back to coarser document units (for example, sections). While CiteRead focuses on a paper's citations as relevant commentary, another exciting direction is to augment the current paper with passages from relevant follow-on work that may fail to cite the current paper or even with broader AI-generated commentary.

### Navigation and Efficient Reading

In pursuit of efficiency, scholars often read papers non-linearly. For example, they might jump forward and read only the most relevant parts of a paper, return to a previously read passage to recall some information, or even switch to another paper to look-up specific information needed to understand the current paper. While jumping can help scholars focus their reading to sections of interest, it can also be disorienting due to constant context-switching. Non-linear navigation can be especially burdensome when the reader is interested in a particular *type* of information (for example, skimming a paper for the main results) but does not know precisely where to find it within the paper. Interactive reading interfaces can help readers navigate efficiently through a paper toward high-value, relevant information.


**Guided skimming with faceted highlights.** Scholarly reading is often a *sensemaking* process involving interleaved *foraging* for relevant passages and *comprehension* of found




information to integrate it into one's existing knowledge. Readers spend a significant amount of time foraging for relevant passages of papers, especially when they skim. A time-pressed reader might attempt to learn more from a paper in less time by skimming its abstract, section headers, text styling, and/or visuals to identify its most significant information. That said, skimming in this way may not connect a reader with the breadth of significant ideas in a paper.

Reading interfaces can help readers encounter more important ideas in a paper, in less time. In the Scim<sup>10</sup> project, we augmented the reading application in a way that helped expose readers to the breadth of important ideas in a paper with faceted (that is, multi-category) highlights. Scim uses a tuned language model with custom post-processing to identify a set of highlights for passages meriting reader attention. These highlights are approximately evenly distributed throughout a paper to encourage examination of major ideas not just in the front matter of the paper, but in all sections. The highlights are tuned to be sparse enough that they can be rapidly reviewed, and dense enough so as to avoid the perception of a tool that “missed” a passage. Furthermore, they are faceted, representing four kinds of major paper findings—research objectives, novel aspects of the research, methodological aspects, and results. Finally, the highlights are controllable: Readers can tune the skimming experience using controls that alter the density of highlights in the paper as a whole, or within individual passages. This project represents a vision of AI as a helper in foraging for relevant information in a paper. Tools like Scim could be yet more powerful with improved AIs for identifying relevant passages, personalization to the information needs of individual readers, and the ability to point out not just significant textual content, but also significant visual content (for example, important aspects of figures and tables).

**Reader-sourced hyperlinks for low-vision navigation support.** The task of navigating between sections and retrieving content can be par-



**As the reliability of AI improves, interfaces can employ it as a means to suggest or even fully automate scalable link creation.**



ticularly challenging for blind and low-vision readers due to limitations in auditory information access or small viewports under high magnification.<sup>36</sup> A small viewport can make navigation difficult and necessitate scrolling,<sup>31</sup> which is problematic when the reader needs to jump back and forth to understand the content. In-document hyperlinks, like those in tables of content or inline references, can help but are typically unidirectional and do not exist (for example, jumping between a results section and relevant experimental design). Most existing tools do not address such challenges associated with low-vision and magnification. Reading interfaces might minimize scrolling requirements for low-vision readers by augmenting the paper with new hyperlinks. For example, Ocean<sup>31</sup> provides bi-directional hyperlinks that enable navigating to and from associated content without disrupting the viewport. These links allow for easy revisiting of portions of the paper with tabbed reading. AI automation may play an important role in scalable creation of links to power such interfaces, though naive application may not yield desired results. For example, Ocean found in an exploratory field-deployment study with mixed-ability groups of low-vision and sighted readers that readers placed greater value in links curated by teammates over crowd or machine generations. In lieu of automation, Ocean includes an authoring interface that allows readers to create and share paper links during reading, thereby enabling readers to build shared interpretations of a paper through collaboration. Of course, as the reliability of AI improves, interfaces can employ it as a means to suggest or even fully automate scalable link creation.

### **In Situ Explanations for Better Comprehension**

One foundational promise of intelligent reading aids has been to help readers better understand a document by extending their cognition. Existing tools have approached this by surfacing relevant information in situ through on-demand tooltips. A classical example is the embedded

term *gloss*—an extension to a reading interface that shows a reader an explanation of a phrase when they click it. Glosses appeared in early research interfaces for reading hypertext<sup>39</sup> and have since become part of widely used reading interfaces including Wikipedia and Kindle. Well-executed glosses have been shown to reduce the time it takes readers to find answers to questions involving the understanding of terminology.<sup>12</sup> In this section, we consider the various forms these *in situ explanations* might take atop research papers, ranging from familiar aids like definitions of terms and symbols, to more novel augmentations, such as plain-language summaries of paper passages and alternative forms of expression beyond text (for example, embedded video).

**Understanding terms and symbols with on-demand, in situ definitions.** Understanding a paper requires understanding the vocabulary it uses, including acronyms, symbols, and invented terms. However, this is by no means an easy task; a typical paper may contain dozens of such terms. Ideally, readers would be able to summon informative definitions with little effort. In the context of scientific papers, familiar gloss designs work less well at helping readers understand terms. One reason is that terms can have multiple senses in a single paper, so a gloss would ideally need to select the sense that matches the context. Another reason is that a reader may rely on usages of a term, and not just its definitions, to understand its meaning. This is especially the case when terms lack explicit definitions.

Intelligent reading interfaces might bring about more effective glosses for scientific papers by providing context-relevant explanations that provide access to the sum of information about a term. ScholarPhi,<sup>12</sup> for example, favors definitions that appear just before a term's usage when a term has multiple definitions. Furthermore, ScholarPhi's glosses consolidate information of all kinds, providing access to all definitions, descriptions, and in-context usages in a compact widget. Finally, to define complex mathematical formulas, ScholarPhi presents its explanations with high economy, show-



**Intelligent reading interfaces might bring about more effective glosses for scientific papers by providing context-relevant explanations that provide access to the sum of information about a term.**



ing definitions for all symbols (and nested sub-symbols) at once, automatically placed adjacent to them in the formula's margins (see Figure 5). Additional advances in AI could lead to even better experiences for understanding terms. For instance, AI could be used to generate definitions of terms and symbols even when no explicit definitions are supplied in the text.<sup>26</sup> Furthermore, AI could be used to identify and prioritize which terms to define for users, and how to explain them in approachable terms, should they have an accurate model of what the reader knows.

**Simplifying complex papers with passage-level plain-language summaries.** Helping a reader understand individual terms and phrases addresses only part of the problem. Papers often contain passages so dense and complex that individual definitions are not enough to help someone read.<sup>11</sup> Can AI be brought into reading applications to make such passages understandable? One solution is to incorporate techniques for plain-language summarization to remove jargon and make passages more approachable. Yet, while modern AI can produce plain-language summaries of long texts, it is not clear how these can enhance the reading experience of the original document. For example, simply attaching a plain-language summary to a paper does not help as papers are read non-linearly; such summaries can even detract from reading the original paper.

Intelligent reading interfaces can grant readers access to plain-language summarization for any passage when and where they need them. For example, in Paper Plain,<sup>1</sup> when a reader encounters a difficult section, clicking on a button adjacent to the section header brings up a generated summary of that section in the paper margin (see Figure 6). To help readers who are so overwhelmed that they do not even know where to begin reading, Paper Plain greets them with a sidebar of questions (for example, *What did the paper find?* or *What were the limitations?*), links to automatically extracted answering passages, and language-model-generated plain-language summaries of those answering passages. Taken together, these

passage-level summaries provide an “index” into the paper’s text, helping readers understand the “gist” of complex passages. For a tool like Paper Plain to see widespread use, several issues in AI need to be addressed. First, how can simplifications be generated without hallucinations, so that a reader can be confident that a simplification accurately reflects the findings of the underlying paper? This is particularly important for domains such as biomedicine, where an inaccurate simplification could lead to a reader making dangerous health decisions. Second, how can AI be optimized to support rapid, on-demand simplification of any passage, large or small, that a reader selects?

**Fusing papers and presentation videos to create engaging multimodal experiences.** Sometimes, the best explanation of an idea is non-textual. For example, an algorithm might be better explained through an animation, and a user interface might be better showcased through a screen recording, as opposed to the prose of a paper. Instead of consuming the two formats independently, could interactive reading interfaces offer readers access to these alternative, more powerful descriptive forms as they read? One approach is for reading interfaces to align external media with the paper text and allow reader traversal in one medium to automatically trigger traversal in the other. For example, in Papeo,<sup>20</sup> paper passages are linked to excerpts of talk videos, and readers skimming through the paper will see corresponding jumps in the video (and vice versa). Unlike text-skimming with Scim and Paper Plain, video-skimming in Papeo combines multiple modalities to explain complex information. For example, instead of reading a long text description of a complex, dynamic system, readers can see the system’s behavior in a video recording or animation along with the author’s commentary. As observed in Papeo’s studies, readers can use these interactions to fluidly transition between watching video and reading text, using video to quickly get an overview and then selectively descend into the text when they desire a more detailed understanding of the paper. Finally, AI can

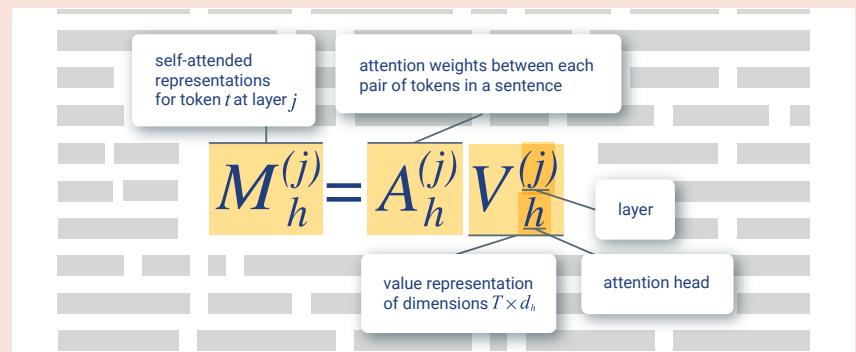
be an effective means to automate alignment between paper passages and video excerpts; in fact, Papeo developed an AI-supported authoring interface in which these links are derived using a pretrained language model and surfaced as suggestions for an author to interactively confirm or refine.

### Bootstrapping Literature Synthesis with Related Work Sections

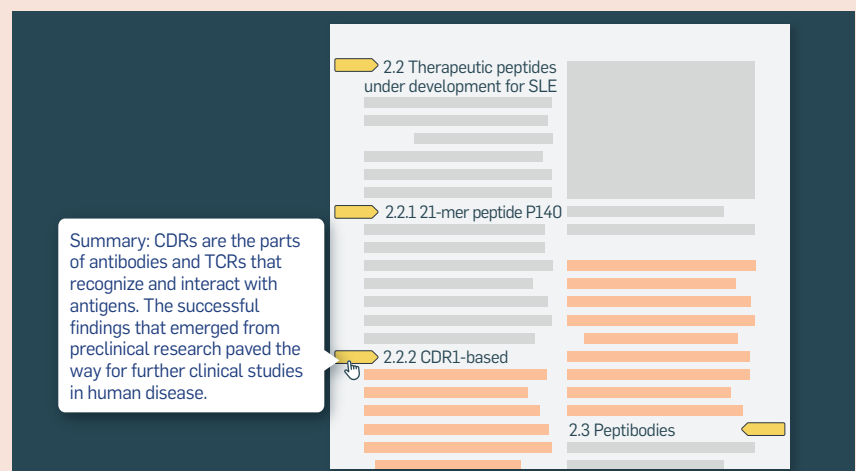
Scientific breakthroughs often rely upon scholars synthesizing multiple published works into broad overviews to identify gaps in the current literature.<sup>32</sup> For this, scholars periodically compile survey articles to help other scholars gain a comprehensive overview of important research topics. For example, some fields have dedicated

outlets for such articles (for instance, the *Psychological Bulletin*). However, survey articles require significant time and effort to synthesize, and they can quickly become outdated with the exponential growth of scientific publication.<sup>4</sup> Instead, scholars in fast-paced disciplines often rely on the related work section when they need to better understand the broader background when reading a paper. While related work sections also summarize multiple prior works, unlike comprehensive survey articles, they typically provide partial views of the larger research topic most relevant to a single paper. There is an opportunity to build better tooling for scholars to read and synthesize related work sections across many papers to gain richer and more comprehensive

**Figure 5. ScholarPhi<sup>12</sup> shows definitions of terms and symbols in pop-up tooltips. When a reader selects a formula, all known definitions of symbols are shown simultaneously. To let readers select nested symbols (for example, “ $h$ ” in “ $V_h^{(j)}$ ”), ScholarPhi supports “drill-down” sub-symbol selection.**



**Figure 6. Paper Plain<sup>1</sup> provides AI-generated plain-language summaries of passages called “gists” to help readers who are overwhelmed by complex textual passages. Readers access gists by clicking a flag next to a section header.**





overviews of unfamiliar research topics. For example, interactive reading interfaces might provide integrated tools for clipping and organizing research threads mentioned across papers, and even support readers directly exploring and reading related work sections extracted across many papers.

**Collecting and curating research threads by clipping and synthesizing across papers.** Saving clips and organizing them is one common approach to supporting synthesis across multiple documents. This is especially important during literature review, in which scholars often save clips from related work sections that organize and summarize different relevant papers. Prior work has pointed to the importance of tightly integrating clipping and synthesis support in the reading process, and how incurring significant context-switching costs can be detrimental to sense-making. Therefore, recent work has developed tools aimed at reducing the cognitive and interaction costs of clipping and triaging information<sup>23</sup> to support everyday online researchers. However, designing clipping and synthesis support tools for research papers is relatively under-explored and introduces exciting new research opportunities.

Better clipping and synthesis support has the potential to lower interaction and cognitive costs, as well as improve awareness and discovery during literature review (for example,

to help readers form their view of the research landscape). Systems such as Threddy<sup>17</sup> and Synergi<sup>19</sup> allow readers to clip sentences from different papers and organize them into a hierarchy of “threads.” These reading interfaces maintain rich context for each clipped snippet, keeping track of its provenance and resolving any inline citations that appear in the snippet to their corresponding papers. Based on the text of the clips and their inline citations, Synergi<sup>19</sup> traverses thousands of neighboring papers in the citation graph and generates summaries of relevant research threads (and their associated papers) to expand the readers’ exploration and coverage. As AI capabilities improve, there is a corresponding opportunity for human-AI interaction research to develop novel systems and interactions that can better support complex synthesis tasks by allowing readers to express their research interests via natural language (for example, clipped sentences) and to generate richer summaries over larger collections of documents based on reader interests.

**Understanding research landscapes by reading and exploring related work sections across papers.** In contrast to providing synthesis support for *individual* papers, a complementary approach allows readers to directly search and extract related work passages across *many* papers. The intuition here is that while any

single related work section may be missing important citations relevant to a reader’s research topic, piecing together related work passages across many papers in a single reading interface can provide readers with a rich and comprehensive overview of a research topic. However, unlike reading full papers, reading a collection of extracted passages presents novel research challenges due to their lack of navigational structures (for example, sections). Further, related work passages relevant to the same research topic often cite overlapping prior work, making them hard to explore and read while keeping track of which papers are new versus already explored.

Intelligent reading interfaces can empower scholars to efficiently explore research landscapes by directly supporting this novel reading process over extracted related work passages. For example, systems such as Relatedly<sup>29</sup> give readers the ability to organize passages into meaningful topics and subtopics by augmenting each passage with language-model-generated descriptive titles and organizing them using a diversity-based ranking algorithm that highlights different research threads. For example, when exploring related work paragraphs about “misinformation”, the first few passages returned by Relatedly may be titled “Fact Checking Datasets”, “Social Media and Misinformation”, and “Fake News Detection Techniques” as opposed to highly similar passages all titled “Related Work” or “Background.” Additionally, when scholars explore inline citations across many passages, Relatedly provides cross-referencing support by keeping track of passages and references visited by the reader which, in turn, allows the reading interface to dynamically re-rank passages to promote unexplored threads.

### Dynamic Documents for Improved Accessibility

A range of disabilities cause people to read scientific documents using a wide variety of devices and reading tools. For example, blind and low-vision readers may use assistive reading technology, such as screen readers, screen magnification, or

**Figure 7. Papeo<sup>20</sup> enables authors to map segments of talk videos to relevant passages in the paper, allowing readers to fluidly switch between the two formats. Color-coded bars show the mapping between the two formats and allow readers to scrub through video segments for quick previews.**



text-to-speech to read documents.<sup>36</sup> Furthermore, people without disabilities face situational impairments, such as the inability to view a screen while driving or may prefer consuming content on a small, mobile device.

Many of these reading tools, such as screen readers, do not function properly on document formats designed for print, such as PDF, unless the document has been manually post-processed to add information about reading order, content type, and so on, which is rarely performed on scientific documents.<sup>38</sup> Further, certain content elements, such as figures, require the addition of alternative text in order to be read aloud at all (figure captions typically assume the reader can see the figure and do not provide the same semantic content as alt text). High magnification reduces the viewport (the amount of visible content) and can dramatically increase the amount of scrolling and panning required, especially for multi-columnar formats commonly used by scientific documents. Visual scanning for information may be impacted or unavailable in these settings, making it more difficult to find and navigate between content in the document.<sup>31</sup>

One way to render legacy PDF content more accessibly is to parse and convert it into a more flexible format, such as XML or HTML, which can then be formatted for mobile devices and augmented for reading by screen readers. The SciA11y system<sup>d</sup> demonstrates this approach, automatically converting 12M academic PDFs to HTML.<sup>37</sup> In a user study with blind and low-vision participants, we observed strong user appreciation of the output, though some errors remain (for example, failing in certain cases to distinguish footnotes from body text, difficulty parsing math equations).<sup>38</sup> When available, alt text can be automatically categorized into semantic content types, enabling new reading experiences that allow skipping or prioritizing certain types.<sup>7</sup> Other approaches provide complementary benefits, such as interfaces tailored for low-vision read-



**The Semantic Reader project has made inroads into accelerating and improving the process of reading static scientific documents. However, much more research remains to be done.**



ers (for example, in Ocean), as well as the range of reading-support systems outlined above.

### Discussion and Future Work

The Semantic Reader project has made inroads into accelerating and improving the process of reading static scientific documents. However, much more research remains to be done.

**Advancing AI for scholarly documents.** The development of intelligent reading interfaces presents an opportunity for further AI research in scholarly document processing, especially when paired with human-centered research grounded in user-validated systems and scenarios. Until recently, interface design could require months of development of bespoke AI models, which creates a barrier for quickly iterating on different system designs. Recent advancements in scaling language models has altered this landscape by enabling researchers to experiment with a wide range of new natural-language processing (NLP) capabilities at relatively low cost. This has the potential of significantly lowering the cost of human-centered AI design by incorporating user feedback in earlier stages of interface and model development to create AI systems that work in symphony with the users beyond pure automation.<sup>35</sup> While recent work has shown that these models can occasionally make critical errors or generate factually incorrect text when processing scientific text,<sup>28</sup> we remain cautiously optimistic about developing ways to address their limitations.

**Collaborative reading experiences.** Research is often done in a collaborative manner, and to date we have mostly focused on improving the experience of individual readers. How might we develop support for the oftentimes social and collaborative nature of scholarly reading? Scholars frequently leverage their social networks and other social signals for paper discovery,<sup>18</sup> work in groups to conduct literature review triage and synthesis, or engage in reading group discussions to aid comprehension. Existing augmentations within the Semantic Reader product could imbue social information, such as

<sup>d</sup> A demo of a subsequent version is available at <https://papertohtml.org/>

## more online

To see the full list of authors, visit doi/10.1145/3659096.

providing signals from one's co-author network (for example, in CiteSee) or aggregate navigation traces (for example, in Scim). Our publicly released data and software for building reading interfaces should also scaffold the creation of novel crowd- or community-sourced content, such as author- or reader-provided explanations (for example, in Papeo), reading aids (for example, in Ocean), commentary (for example, in CiteRead), or verification of paper content. Finally, we wonder how the scholarly community can be empowered to step in where current AI systems fall short, such as by fixing improperly extracted content or incorrect generated text, which are especially problematic for interfaces such as SciA11y. In fact, one promising avenue is to mitigate or even eschew the problem of deriving augmentations post-publication by developing better tools for authors to tag or prepare their works ahead of time to be consumed by users of modern reading interfaces.

**Ethical consideration of augmenting papers.** Finally, this research opens new ethical considerations. For instance, any system that elevates certain papers or certain content over others introduces a form of bias. Systems that rely on citation graph signals, such as CiteSee or Relatedly, should carefully consider additional signals of relevance, such as semantic similarity to surface newer and overlooked papers. Another tension is the potential discrepancy between author desires and reader desires for how a work is presented and how much control to provide authors. Systems for more efficient reading or synthesis may encourage readers to take shortcuts that lead to incorrect understanding, sloppy research, or even outright plagiarism. Instead of simply seeking to increase reading throughput uniformly, our systems should enable *triage*, so that readers can dedicate time for thoughtful and careful reading when the content is important. For instance, our systems could design pathways that, while may be more efficient, do not obfuscate the full context (for exam-

**Recent trends in open science and datasets point to a promising future where we could continue to explore different ways to remix and reuse scholarly content across context so that future scientists can take fuller advantage of prior research.**

ple, Scim), and that encourage good practices, such as verification and provenance tracing. A final consideration is around what is considered ethical reuse of a paper's contents to support reader experiences outside of that paper and its licensing implications. For instance, CiteRead extracts paper citations and places them in the cited paper, and Relatedly extracts related work sections from different papers for users to explore. Recent trends in *open science and datasets*<sup>25</sup> point to a promising future where we could continue to explore different ways to *remix and reuse* scholarly content across context so that future scientists can take fuller advantage of prior research.

**Beyond the PDF.** The Semantic Reader project currently focuses on analyzing and augmenting research papers in the static PDF format because it has been the dominant format for scholarly publications. However, the publishing industry may be gradually moving to more flexible formats, such as HTML, in part to better support accessibility. While our focus has been on augmenting the millions of current and legacy PDF documents in support of current scholar reading practices, all of our reading interfaces are built with Web technologies—allowing these novel interactions to extend (possibly even more effectively) to future publication formats which likely will be rendered in Web browsers.

## Conclusion

With the Semantic Reader Project, we develop and evaluate AI-powered reading interfaces to support scholars around discovery,<sup>6,33</sup> efficiency,<sup>10,31</sup> comprehension,<sup>1,12,20</sup> synthesis,<sup>17,19,29</sup> and accessibility<sup>37</sup> when reading research papers. Validating our approach of augmenting existing PDFs of research papers, we have seen tremendous adoption of the freely available Semantic Reader<sup>e</sup> product, which has grown to more than 200,000 unique monthly users.<sup>f</sup> We plan to continue experimenting with novel AI-powered intelligent reading interfaces, as well as migrating successful interactive features into the product.

<sup>e</sup> Semantic Reader: <https://bit.ly/3Z6BW3L>

<sup>f</sup> As of early August 2024



Finally, we offer a collection of freely available resources to the larger research community, including datasets of open-access research papers,<sup>25</sup> APIs for accessing the academic citation graph,<sup>22</sup> machine-learning models for processing and understanding research papers,<sup>5,8,16,34</sup> and open source software for rendering and augmenting PDF documents for developing reading interfaces.<sup>8</sup> We hope by providing these resources we can enable and encourage the broader research community to work on exciting novel intelligent reading interfaces for research papers with us.

## Acknowledgments

This project is supported in part by NSF Grant OIA-2033558, NSF Grant CNS-2213656, NSF RAPID Award 2040196, ONR Grant N00014-21-1-2707, ONR Grant N00014-22-S-B001, and a grant from the Alfred P. Sloan Foundation.

g <https://openreader.semanticscholar.org>.

## References

- August, T. et al. Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing. *ACM Transactions on Computer-Human Interaction* (2023).
- Bazerman, C. Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written Communication* 2, 1 (Jan. 1985), 3–23.
- Beltagy, I., Lo, K., and Cohan, A. SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Intern. Joint Conf. on Natural Language Processing*.
- Bornmann, L., Mutz, R., and Haunschild, R. Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanities and Social Sciences Communications* 8 (2020), 1–15.
- Cachola, I., Lo, K., Cohan, A., and Weld, D.S. TLDR: Extreme summarization of scientific documents. In *Findings of the Association for Computational Linguistics: EMNLP 2020*.
- Chee Chang, J. et al. CiteSee: Augmenting citations in scientific papers with persistent and personalized historical context. In *Proceedings of the 2023 CHI Conf. on Human Factors in Computing Systems*. Association for Computing Machinery, Article 737, 15; 10.1145/3544548.3580847
- Chintalapati, S.S., Bragg, J., and Lu Wang, L. A dataset of all texts from HCI publications: Analyses and uses towards producing more descriptive alt texts of data visualizations in scientific papers. In *Proceedings of the 24th Intern. ACM SIGACCESS Conf. on Computers and Accessibility*. Association for Computing Machinery, (2022), Article 30, 12; 10.1145/3517428.3544796
- Cohan, A. et al. SPECTER: Document-level representation learning using citation-informed transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (July 2020), 2270–2282; 10.18653/v1/2020.aclmain.207
- Denney, A.S. and Allan Tewksbury, R. How to write a literature review. *J. of Criminal Justice Education* 24 (2013), 218–234.
- Fok, R. et al. Scim: Intelligent skimming support for scientific papers. In *Proceedings of the 28th Annual Conf. on Intelligent User Interfaces* (2023).
- Halliday, M.A. Some grammatical problems in scientific english. *Applying English Grammar*. Routledge, (2014), 77–94.
- Head, A. et al. Augmenting scientific papers with just-in-time, position-sensitive definitions of terms and symbols. In *Proceedings of the 2021 CHI Conf. on Human Factors in Computing Systems* (2021).
- Head, A., Xie, A., and Hearst, M.A. Math augmentation: How authors enhance the readability of formulas using novel visual design practices. In *Proceedings of the CHI Conf. on Human Factors in Computing Systems* (2022), 1–18.
- Huang, Y. et al. Layoutlmv3: Pre-training for document AI with unified text and image masking. In *Proceedings of the 30th ACM Intern. Conf. on Multimedia* (2022).
- Huang, Z., Chung, W., Ong, T., and Chen, H. A graph-based recommender system for digital library. In *Proceedings of JCDL '02*, (2002).
- Kang, D. et al. Document-level definition detection in scholarly documents: Existing models, error analyses, and future directions. In *Proceedings of the First Workshop on Scholarly Document Processing*. Association for Computational Linguistics, Online, (2020), 196–206; 10.18653/v1/2020.sdp.1.22
- Kang, H., Chee Chang, J., Kim, Y., and Kittur, A. Threddy: An interactive system for personalized thread-based exploration and organization of scientific literature. In *Proceedings of the 35th Annual ACM Symp. on User Interface Software and Technology*. Association for Computing Machinery (2022), Article 94, 15; 10.1145/3526113.3545660
- Kang, H.B. et al. From who you know to what you read: Augmenting scientific recommendations with implicit social networks. In *Proceedings of the 2022 CHI Conf. on Human Factors in Computing Systems*. Association for Computing Machinery, (2022), Article 302, 23; 10.1145/3491102.3517470
- Kang, H.B., Wu, T., Chee Chang, J., and Kittur, A. Synergi: A mixed-initiative system for scholarly synthesis and sensemaking. In *Proceedings of the 36th Annual ACM Symp. on User Interface Software and Technology*. Association for Computing Machinery (2023), Article 43, 19; 10.1145/3586183.3606759
- Kim, T.S. et al. Papeos: Augmenting research papers with talk videos. In *Proceedings of the 36th Annual ACM Symp. on User Interface Software and Technology*. Association for Computing Machinery (2023), Article 15, 19; 10.1145/3586183.3606770
- King, D.W., Tenopir, C., Choemprayong, S., and Wu, L. Scholarly journal information-seeking and reading patterns of faculty at five U.S. universities. *Learned Publishing* 22 (2009).
- Kinney, R. et al. The semantic scholar open data platform. *arXiv preprint arXiv:2301.10140* (2023).
- Liu, M.X. et al. Wiggly: Low-Cost Information Collection and Triage. In *Proceedings of the 35th Annual ACM Symp. on User Interface Software and Technology*. Association for Computing Machinery (2022), Article 32, 16; 10.1145/3526113.3545661
- Lo, K. et al. Papermage: A unified toolkit for processing, representing, and manipulating visually rich scientific documents. In *Proceedings of the 2023 Conf. on Empirical Methods in Natural Language Processing: System Demonstrations* (2023) 495–507.
- Lo, K. et al. S2ORC: The Semantic Scholar Open Research Corpus. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics (2020), 4969–4983; 10.18653/v1/2020.acl-main.447
- Murthy, S. et al. ACCORD: A multi-document approach to generating diverse descriptions of scientific concepts. In *Proceedings of the 2022 Conf. on Empirical Methods in Natural Language Processing: System Demonstrations*, W. Che and E. Shutova (Eds.). Association for Computational Linguistics, (2022), 200–213; 10.18653/v1/2022.emnlp-demos.20
- Okamura, K. Interdisciplinarity revisited: Evidence for research impact and dynamism. *Palgrave Communications* 5, 1 (2019), 1–9.
- Otmakhova, Y., Verspoor, K., Baldwin, T., and Han Lau, J. The patient is more dead than alive: Exploring the current state of the multidocument summarisation of the biomedical literature. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. (2022), 5098–5111; 10.18653/v1/2022.acl-long.350
- Palani, S. et al. Relatedly: Scaffolding literature reviews with existing related work sections. In *Proceedings of the 2023 CHI Conf. on Human Factors in Computing Systems*, Article 742, ACM, 1–20.
- Palmer, C.L., Teffau, L.C., and Pirmann, C.M. Scholarly information practices in the online environment. *Report commissioned by OCLC Research* (2022).
- Park, S. et al. Exploring team-sourced hyperlinks to address navigation challenges for low-vision readers of scientific papers. In *Proceedings of the 25th ACM Conf. on Computer-Supported Cooperative Work And Social Computing* (2022).
- Portenoy, J. et al. Bursting scientific filter bubbles: Boosting innovation via novel author discovery. In *Proceedings of the 2022 CHI Conf. on Human Factors in Computing Systems* (2022), 1–13.
- Rachatasumrit, N., Bragg, J., Zhang, A.X., and Weld, D.S. CiteRead: Integrating localized citation contexts into scientific paper reading. *27th Intern. Conf. on Intelligent User Interfaces* (2022).
- Shen, Z. et al. VILA: Improving structured content extraction from scientific pdfs using visual layout groups. *Transactions of the Association for Computational Linguistics* 10 (2022), 376–392; 10.1162/tacl\_a.00466
- Shneiderman, B. *Human-centered AI*. Oxford University Press, (2022).
- Szipro, S.F.A., Hashash, S., Zhao, Y., and Azenkot, S. How people with low vision access computing devices: Understanding challenges and opportunities. In *Proceedings of the 18th Intern. ACM SIGACCESS Conf. on Computers and Accessibility* (2016).
- Wang, L.L. et al. SciA11y: Converting scientific papers to accessible HTML. In *Proceedings of the 23rd Intern. ACM SIGACCESS Conf. on Computers and Accessibility*. Association for Computing Machinery (2021), Article 85, 4; 10.1145/3441852.3476545
- Wang, L.L. et al. Improving the accessibility of scientific documents: Current state, user needs, and a system solution to enhance scientific PDF accessibility for blind and low vision users. *arXiv: 2105.00076 [cs.DL]* (2021).
- Zellweger, P.T., Chang, B., and Mackinlay, J.D. Fluid links for informed and incremental link transitions. In *Proceedings of the Conf. on Hypertext and Hypermedia*. ACM, (1998), 50–57.
- Zyto, S., Karger, D., Ackerman, M., and Mahajan, S. Successful classroom deployment of a social document annotation system. In *Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems*, (2012), 1883–1892.

**Kyle Lo** is lead applied research scientist, Allen Institute for AI, Seattle, WA, USA.

**Joseph Chee Chang** is a senior research scientist, Allen Institute for AI, Seattle, WA, USA.

**Andrew Head** is assistant professor, Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA.

**Jonathan Bragg** is a senior research scientist, Allen Institute for AI, Seattle, WA, USA.

**Amy X. Zhang** is an assistant professor at the Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA.

**Cassidy Trier** is a senior product designer, Allen Institute for AI, Seattle, WA, USA.

**Marti A. Hearst** is a professor in the School of Information at the University of California, Berkeley, USA.

**Daniel S. Weld** is Professor Emeritus at the Allen School of Computer Science & Engineering, University of Washington, and general manager and chief scientist of Semantic Scholar at the Allen Institute for AI, Seattle, WA, USA.

This work is licensed under a Creative Commons Attribution International 4.0 License.



Watch the authors discuss this work in the exclusive *Communications* video. <https://cacm.acm.org/videos/semantic-reader-project>