CrossLit: Connecting Visual and Textual Sensemaking for Literature Review

Kiroong Choe
Seoul National University
Seoul, Republic of Korea
krchoe@hcil.snu.ac.kr

Eunhye Kim
KAIST
Daejeon, Republic of Korea
gracekim027@kaist.ac.kr

Min Hyeong Kim
Seoul National University
Seoul, Republic of Korea
mhkim@hcil.snu.ac.kr

Suyeon Hwang
Seoul National University
Seoul, Republic of Korea
stom.hwang@hcil.snu.ac.kr

Sangwon Park
Seoul National University
Seoul, Republic of Korea
sangwon.park@hcs.snu.ac.kr

Nam Wook Kim
Boston College
Boston, MA, U.S.A.
nam.wook.kim@bc.edu

Jinwook Seo*
Seoul National University
Seoul, Republic of Korea
jseo@snu.ac.kr

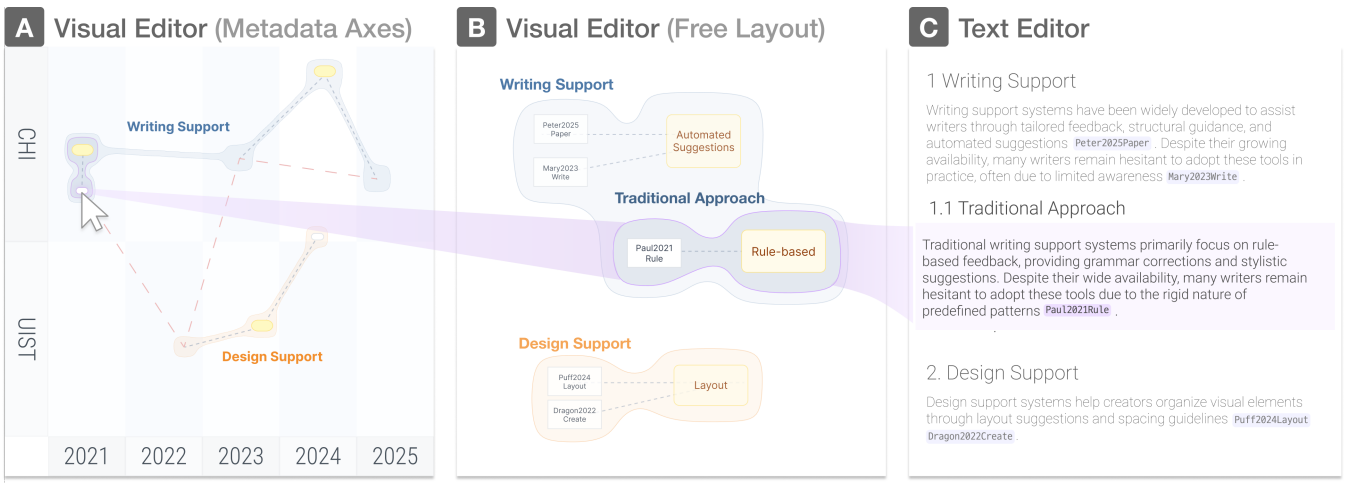

Figure 1: Conceptual overview of CrossLit, a literature review supporting system where visual and text editors are connected. In the visual editor, users can examine (A) metadata such as venue, year, and citation relationships or (B) arrange papers in a free layout according to their interpretation. In the text editor, users can (C) compose the manuscript directly.

Abstract

Conducting literature reviews is cognitively demanding, requiring researchers to navigate large volumes of work while constructing coherent narratives that position their contributions. The process unfolds through iterative stages of sensemaking, each demanding different support. Existing tools emphasize either visual interfaces that provide macroscopic overviews or textual interfaces that support thematic organization and narrative construction. However, keeping modalities separate forces researchers to switch between tools, disrupting workflow continuity. We present CrossLit, a system that integrates and synchronizes visual and textual interfaces to support the entire process from discovering papers to composing coherent narratives. CrossLit allows researchers to group and annotate papers visually while generating aligned textual structures, and to edit text that automatically updates visual representations. We find that CrossLit helps users develop and refine conceptual structures and build narratives iteratively through seamless cross-modal transitions. We conclude by discussing design implications for synchronizing visual and textual interfaces for sensemaking support.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; *Visualization systems and tools*; • **Information systems** → Users and interactive retrieval.

Keywords

Literature Review, Sensemaking, Crossmodal Interaction, Writing Support, Visual Interface, Large Language Models, Human-AI Collaboration

*Corresponding author

CHI '26, April 13–17, 2026, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791418

ACM Reference Format:
Kiroong Choe, Eunhye Kim, Min Hyeong Kim, Suyeon Hwang, Sangwon Park, Nam Wook Kim, and Jinwook Seo. 2026. CrossLit: Connecting Visual and Textual Sensemaking for Literature Review. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 30 pages. https://doi.org/10.1145/3772318.3791418

1 Introduction

Conducting a literature review is a central task in academic research, essential for situating one's work within the broader research landscape. It is not merely a summary of prior studies, but an evaluative process of searching, identifying, and synthesizing existing scholarship to uncover gaps and frame new contributions [13, 26, 56]. Researchers actively decide which studies to include and through which lens to interpret them, selecting one thread among many possible lines of inquiry [47]. This process unfolds iteratively across distinct stages of sensemaking, including exploration, understanding, organization, and discovery [102]. Each newly encountered paper reshapes the researcher's understanding of prior work and suggests new search directions [76]. Because it involves multiple stages, cognitive complexity, and vast numbers of papers, the literature review is both indispensable and cognitively demanding, motivating the design of tools that can better support it.

Previously, numerous systems have been developed to support this cognitively demanding process, typically falling into two broad categories: **visual-based interfaces** and **text-based interfaces**. Visual approaches, often employing bibliometric visualizations [40], primarily facilitate exploration of large literature collections by organizing academic metadata such as publication years, venues, and citations. These systems effectively provide macroscopic overviews of the research landscape, yet their large-scale, metadata-driven representations may not fully align with the contextual and interpretive processes inherent in literature reviews. On the other hand, text-based interfaces, such as paper-reading tools [88] and thematic structuring tools [66, 69], manage relatively smaller-scale collections, supporting detailed understanding, thematic organization, and narrative synthesis. However, text-based approaches struggle to reveal broader connections in the literature, such as how research clusters evolve across venues, years, or citation links.

As each modality addresses complementary research needs, integrating visual- and text-based interfaces can significantly enhance the literature review process. Smooth transitions between broad literature exploration and detailed comprehension are essential, yet prior systems often adopt a single focal modality to support a subset of stages. This separation can interrupt the iterative nature of literature review, which demands that both modalities remain available across all stages, from clustering based on academic metadata to critical synthesis and narrative coherence. These tensions remain unresolved, highlighting the need for an integrated interface that enables the two modalities to effectively support one another.

We present CrossLit, a literature review support system that integrates visual- and text-based interfaces to improve review workflows. By synchronizing the two views, CrossLit allows users to edit the same review outline—ranging from higher-level structures such as sections to finer details such as sentences—in both modalities. Building on prior studies of literature support tools, we distill the strengths and weaknesses of each modality into four design goals, and apply them to CrossLit. In CrossLit, the visual editor renders the review outline as paper nodes, note nodes, and hierarchical bubblesets, while the text editor presents it as headings, paragraphs, and cited papers. Edits in one interface are mirrored in the other; structural edits maintain a one-to-one correspondence, and an LLM aligns textual content with the revised structure. CrossLit also supports queries that exploit the unique strengths of each modality. In the visual editor, users can arrange papers along axes such as venue, author, and year, and inspect citation links; these actions reveal and express metadata-driven search intents. In the text editor, users can scrutinize narrative and logic to define textual search intents. Both intents can appear in a single query, and CrossLit fulfills them by unifying existing text-based and metadata-based paper retrieval approaches.

We conducted a user study with 16 researchers who used CrossLit for 45 minutes on their own literature review topics, following a think-aloud protocol. Participants predominantly relied on the visual modality in early stages to filter papers and form intuitive groupings, and turned to the textual modality in later stages to refine these intuitions into coherent narratives. The bidirectional synchronization feature allowed seamless transition between modalities, enabling more frequent and fine-grained iterations across exploratory and convergent processes than in their typical workflows. Notably, when constructing conceptual structures for literature review outlines, the visual and textual modalities worked in tight coordination, supporting effective sensemaking outcomes.

Our findings demonstrate the importance of integrating visual and textual modalities to support complex sensemaking tasks as a continuous process. Researchers could freely arrange and experiment with early-stage ideas in the visual space, without being forced into the immediate logical structure that linear text requires. This visual exploration, synchronized in real-time with text, enabled them to see how intuitive groupings might develop into coherent narratives and to commit to narrative construction at their chosen moment. We discuss how CrossLit's goal of synchronizing visual and text can be further advanced, and how this idea can extend to other cognitively demanding tasks where human agency remains essential.

Overall, our contributions are as follows:

- We present CrossLit, a system that integrates visual and text modalities to support literature review as a continuous process.
- We report findings from a study with 16 researchers, showing how CrossLit enables provisional exploration, fluid transitions between modalities, and improved discovery of relevant literature.
- We contribute the idea of synchronizing visual and text views, illustrating how cross-modal collaboration can support complex sensemaking tasks.

2 Related Work

In this section, we examine how existing literature review support systems employ visual and text modalities to assist researchers throughout the research process. We identify distinct strengths

and limitations of each modality and the inefficiencies of single-modality dominance patterns and modal isolation. We organize our review by modality and by stage of the literature review process: exploration, understanding, organization, and discovery—highlighting how each modality supports or limits researchers in these stages.

2.1 Visual Approaches to Literature Review

Visual modality demonstrates particular effectiveness in enabling researchers to comprehend and navigate large collections of literature. Its core strength lies in the simultaneous representation of large collections of papers, allowing immediate recognition of relationships, patterns, and structures. In this work, we use the term visual modality interfaces to refer to systems that primarily rely on visual representations of literature. Examples include citation network visualizations, temporal visualizations, and spatial layouts designed to convey relationships among papers.

2.1.1 Exploration: Landscape-Level Visualizations. The primary advantage of visual approaches emerges in the initial phases of literature review, where researchers must understand the overall landscape of a research field and identify relevant areas for deeper investigation.

Systems supporting macroscopic exploration employ various visualization approaches, including network-based representations [23, 25, 36, 89, 122], temporal visualizations [31, 32, 55, 82, 109], and custom layouts optimized for specific analytical purposes [33, 37, 110, 121, 127]. The core value these systems provide is discovering connections and identifying outliers within research fields through techniques such as co-citation analysis [24, 25], topic flow tracking [31, 55, 109], and collaboration network analysis [32, 80, 108]. By visualizing influence patterns and identifying key papers and researchers [91, 111, 121], these systems help researchers understand historical context and comprehend the structure and evolution of research communities.

However, the transition from macroscopic overview to specific paper selection presents a critical challenge. Researchers must narrow from thousands of papers to dozens of relevant ones for deeper examination. To address this, systems employ interactive filtering that progressively reveals finer structures within selected clusters or topics [28]. More sophisticated approaches integrate metadata-based filtering with multi-perspective visualization [7, 10, 120], combine temporal evolution with collaboration patterns [31, 32], or partition literature space by discipline or research topic [113, 132].

2.1.2 Organization: Spatial Structuring of Paper Collections. Visual modality provides a spatial canvas for organizing literature, enabling multidimensional structuring. Systems primarily adopt bottom-up approaches, supporting researchers in starting with small paper sets and progressively adding adjacent papers to form groups [23, 84, 89]. The researcher's intent is expressed through annotations or group nodes, allowing explicit representation of conceptual groups on visualizations [89, 118].

However, visual structuring shows clear limitations once researchers move from exploration to deeper knowledge construction. Adding layers of structure onto already dense visual channels quickly increases complexity, especially with large collections [98]. Information is also lost when spatial groupings must be converted into linear text, since no clear pathways support this transition [128]. While visualizations reveal relationships and distributions, they cannot assess logical consistency or argumentative quality [57]—both essential for literature review writing.

2.1.3 Discovery: Metadata-Based Expansion. Visual modality supports ongoing literature expansion through metadata-based structural approaches. The core strength of this approach is the explainability of recommendations and the resulting trustworthiness. Systems quantify relationships between papers through metadata such as citation relationships, co-author networks, and conference venues [67, 68], and transparently present these relationships through visualization [2, 9]. Visualization performs a key role in providing transparency and explainability of recommendations in this iterative discovery process. Through network visualization [6, 125] or weight-based score distributions [27], systems explicitly show how recommended papers relate structurally to existing collections.

However, this metadata-based approach primarily relies on quantitative relationships and may not sufficiently reflect actual paper content or qualitative aspects of research [123]. Particularly in interdisciplinary research or emerging fields, existing citation networks may not be formed, potentially missing important conceptual connections that could only be identified through content analysis [61].

2.2 Textual Approaches to Literature Review

Text modality demonstrates essential capabilities in areas where visual approaches have limitations, particularly in deep content understanding and the construction of coherent knowledge structures. The core content of an academic paper can only be conveyed through text, making text modality indispensable for thorough literature comprehension and synthesis. We consider text-based systems to include tools that support reading and annotating papers, structuring and writing to organize knowledge, and searching for new literature through textual queries.

2.2.1 Understanding: Deep Content Analysis. After discovering and narrowing relevant papers through exploration, researchers must understand the actual content of selected papers in detail. Text modality performs an essential role in this phase, as academic work requires careful textual analysis. Recent systems provide various features supporting a deep understanding of text beyond conventional PDF viewers [44]. Features such as restructuring paper abstracts into recursively expandable structures [42] or providing immediate re-references for terms and equations reduce the cognitive burden of complex academic text [54]. Non-linear exploration enables reading beyond individual paper boundaries. Systems can reorganize related work sections from multiple papers by topic [101] or show how specific papers are cited in subsequent research [103]. Particularly, by providing user-context-tailored explanations for inline citations[21], systems help researchers understand literature in connection with their interests.

2.2.2 Organization: Thematic Structuring in Text. Text modality transforms linear text streams into meaningful thematic structures. This capability addresses the critical limitation of visual approaches in converting spatial arrangements into coherent written arguments.

Recent systems allow researchers to physically manipulate and rearrange fragments extracted from multiple documents [53, 66], supporting building non-linear structures during linear reading [43, 126]. The introduction of LLMs has enabled hybrid approaches that effectively combine top-down automatic analysis with bottom-up manual composition [69, 99].

In transforming organized knowledge into academic text, text modality-based systems provide features for a multidimensional review of written content quality. Systems provide automated feedback on various aspects, including fluency, coherence, vocabulary choice, and structure [73, 85]. In academic writing contexts, domain-specific aspects such as citation appropriateness, logical connections between claims and evidence, and dialogue with existing literature are particularly important [18, 45]. These scrutinized features enable researchers to reflectively review the logical soundness and academic contribution of their constructed knowledge structures.

2.2.3 Discovery: Content-Based Expansion. Textual approaches support iterative literature expansion by capturing semantic similarities that may be invisible in metadata [51, 52]. Papers that appear unconnected by citation or venue can nonetheless share conceptual or methodological links, and text analysis can surface these connections [12]. This capability complements metadata-based methods by revealing relationships beyond structural ties.

The advent of large language models (LLMs) has advanced this approach significantly. LLMs enable natural language queries and semantic search beyond keyword matching [1, 114] uncovering unexpected conceptual connections [135] and providing contextualized explanations that ease the integration of new material into existing knowledge structures [83].

Yet content-based discovery has important limitations. Unlike metadata-based approaches, it does not provide structural context or explainability. Researchers may recognize that papers are conceptually related but remain uncertain about how they fit within broader citation networks or research communities.

2.3 The Need for Cross-Modal Integration

Visual and textual modalities each offer unique strengths for literature review, yet most existing tools support only one modality or one stage of the workflow [39]. As a result, researchers often switch between separate systems to explore, read, organize, and write, incurring context-switching costs and manually transferring insights across tools. Because literature review is inherently iterative—cycling across exploration, understanding, organization, and synthesis [115]—content-only approaches risk cherry-picking [39], while metadata-only approaches risk reinforcing disciplinary bubbles [96]. The lack of integrated, cross-modal support forces users to mentally combine incomplete views, increasing cognitive burden and reducing overall review quality.

Recent work in human–AI collaboration has similarly emphasized the promise of cross-modality [62]. Prior approaches have synchronized textual and visual representations to support argumentative writing [131] and creative writing [90], connected narrative text with structural mappings to scaffold metaphor creation [71], proposed compositional structures that link text, visuals, and timelines for creative workflows [16]. Multimodal reading interfaces similarly synchronize academic text with video to reduce switching costs [72]. Together, these works illustrate the broader potential of bridging modalities to deepen engagement and structure sensemaking.

Our system extends this vision to the literature domain. In this context, review practices require not only writing but also exploring external repositories, organizing references, and synthesizing evolving insights. To support these activities, we integrate original literature databases with synchronized visual and textual workspaces. We argue for a cross-modal bridge that bidirectionally connects the two modalities so that visual and textual representations continually reinforce each other throughout the process.

3 The CrossLit System

In this section, we outline the design goals derived from prior work and present CrossLit, a system built upon these rationales.

3.1 Design Goals

Based on the gaps and opportunities identified in prior works, we derived four design goals for our system. Our overarching objective is to help researchers seamlessly transition between visual and text modalities throughout the entire literature review process. Across activities such as discovering papers, comprehending content, organizing knowledge, and writing narratives, the system should ensure a continuous and integrated workflow, allowing researchers to employ the most suitable modality at each stage instead of fragmenting work into separate, modality-specific processes.

DG1: Supporting Literature Review through Visual Modality. Visual approaches excel at helping researchers overview large collections of papers and explore relationships that are not yet explicitly articulated. Our system should enable users to position and group papers spatially, gradually developing their understanding of how papers are related to each other. When needed, users can customize visual representations to arrange papers by metadata axes (such as year or venue), enabling rapid identification of patterns and trends in the literature.

DG2: Supporting Literature Review through Text Modality. Text approaches are essential for capturing the specific contributions and methods of individual papers and weaving them into coherent arguments. They allow researchers to articulate the nuanced reasons why certain papers belong in particular groups and to construct narrative structures that identify gaps in prior work and position new contributions, which is essential for writing a literature review. Our system should enable users to structure and refine their reviews directly within a text editor.

DG3: Bidirectional Synchronization between Visual and Text Modalities. As literature sensemaking progresses, researchers rely on both visual and textual modalities. Yet connecting them is difficult because the same organizational units take different forms

Figure 2: An example scenario of using CrossLit. (1) The user sorts papers into piles and adds short notes in the visual editor. (2) The user makes groups from the piles in the visual editor, and the system drafts a section for each group in the text editor. (3) The user writes a new section in the text editor, and the system creates a corresponding group in the visual editor. (4) The user organizes papers by venue and year in the visual editor. The user identifies the need for new papers from both editors. (5) Among the newly discovered papers, the user discovers a seminal work through citation patterns in the visual editor. (6) The user restructures sections into subsections in the text editor, and the visual editor updates accordingly.

in each. For example, grouping papers with a short note in the visual editor corresponds to creating sections, citing papers, and articulating themes in text. Visual representations favor spatial arrangement and concise labeling, while textual representations require detailed, contextualized descriptions. Our system should bridge these differences by synchronizing organizational structures across both views, ensuring that changes in one modality are consistently reflected in the other.

DG4: Integrating Discovery Cues across Visual and Text Modalities. During the literature review process, researchers frequently encounter the need to expand their collection with additional relevant papers. Prior visual systems primarily capture structural discovery cues, such as citation relationships, while text-based systems emphasize semantic cues, such as conceptual similarity. In practice, however, researchers rarely treat these intents as separate; instead, they seek papers that are simultaneously meaningful across multiple dimensions. Our system should integrate both structural and semantic cues as unified signals for discovery, enabling more comprehensive and contextually relevant paper exploration.

3.2 CrossLit User Scenario

Consider a researcher beginning a literature review on *AI collaboration tools for creative work*. At this early stage, they are unsure how to structure prior studies across domains such as creativity support, human–AI interaction, and collaborative systems, but already have a few papers in mind as starting points.

The researcher begins by importing about ten closely related papers already familiar to them into the system. To gain an exploratory sense of how the topic might be organized, they turn to the **visual editor**. Papers are arranged so that related works are positioned near one another, and the researcher adds short keyword annotations (two to three words) to capture emerging themes. Through this process, patterns begin to emerge: several papers cluster around writing support, while others center on design support (Figure 2-1). The researcher then defines two groups: `Writing Support` and `Design Support`, and assigns the corresponding papers to each. In the **text editor**, the system mirrors this action by dividing the outline into matching headers, each accompanied by a draft description that draws on the papers' descriptive information and the researcher's annotations (Figure 2-2).

To elaborate on these groups, the researcher shifts to the **text editor**. After reviewing the papers, they expand on each paper's specific contributions and methodologies, articulating why each paper belongs in its section. During this process, they realize that one paper aligns more with collaborative design than with individual design. This leads them to create a new, independent section, `Collaboration Tools`, alongside the existing sections. As the outline is reorganized in text, the **visual editor** updates simultaneously: the corresponding paper node moves to a new group and a note is added (Figure 2-3).

The researcher now looks for additional papers on creative collaboration beyond design. They shape this search intent by drawing on insights from both the visual and text editors. In the **visual editor**, they observe that most papers are published in CHI and UIST, whereas the collaborative brainstorming work appears in CSCW. In the **text editor**, they realize that the `Collaboration Tools` section does not sufficiently address pre-LLM paradigms (Figure 2-4). Integrating these insights, the researcher formulates a hybrid query for the system: search within the citation network of the current collection but restrict results to CSCW publications (a structural cue from the **visual editor**). Within that set, prioritize studies on collaboration methods that do not rely on LLMs (a semantic cue from the **text editor**). The system then executes this query and returns eight relevant papers that are surfaced through the combination of structural and semantic cues.

To integrate the eight newly discovered papers, the researcher alternates between the **visual editor** and the **text editor**. In the **visual editor**, they arrange papers by year on the x-axis and venue on the y-axis, revealing that all of the new papers are from CSCW and were published before 2021. Among them, one early paper stands out for being heavily cited by the CHI and UIST papers already in the collection, which leads the researcher to recognize it as a seminal work (Figure 2-5).

The researcher examines the seminal paper in a separate platform and encounters a framework that characterizes collaboration along multiple dimensions, such as temporal, spatial, and role structures. The researcher realizes that their earlier grouping into `Writing Support`, `Design Support`, and `Collaboration Tools` was overly simplistic. Returning to the **text editor**, they revisit `Writing Support` and observe that the papers assume different temporal dynamics—some systems support real-time collaboration, while others focus on asynchronous feedback. Recognizing that this distinction creates fundamental differences in user experience and system design, the researcher divides `Writing Support` into two subsections: `Synchronous Writing` and `Asynchronous Writing`. This incremental restructuring is immediately mirrored in the **visual editor**, where single clusters split into clearer subgroups (Figure 2-6).

3.3 System Overview

We implemented these four design goals in CrossLit (Figure 3). CrossLit provides two synchronized editors: a text editor, which represents the review as an outline of sections, paragraphs, and citations, and a visual editor, which represents the same structure as spatially organized nodes and groups. Edits made in one editor are immediately reflected in the other, maintaining consistency while accounting for the different ways organizational structures are expressed in each modality. Both editors also serve as entry points for discovering new papers: the visual editor highlights structural and metadata-based cues, while the text editor emphasizes semantic and narrative cues. Together, they help researchers shift easily between exploring the broader landscape and developing detailed arguments.

3.4 Visual Editor Design (DG1)

3.4.1 Representing Papers and Relationships Visually. The visual editor consists of three core elements: white rectangular paper nodes, yellow rectangular note nodes, and groups represented as bubble-sets. Paper nodes represent individual papers and display title and author information. Note nodes contain researcher interpretations and annotations, and can be connected to one or more paper nodes

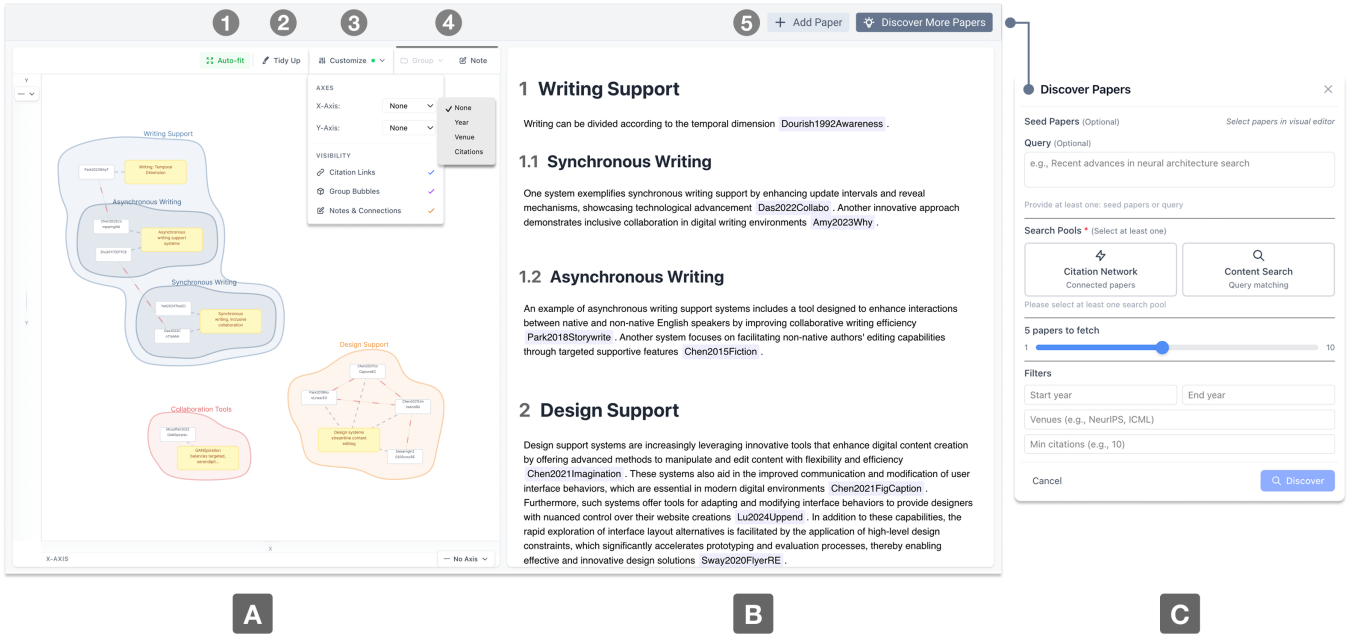

Figure 3: The full interface of CrossLit. (A) Visual editor allows users to freely position nodes and organize them into hierarchical groups represented by bubblesets. Citation relationships are displayed as red animated dashed lines. A toolbar allows users to (1) control zoom level, (2) automatically align elements according to the document-aligned layout, (3) customize visualization by organizing papers along metadata axes and toggling visual element display, and (4) add, edit, and remove elements. (B) Text editor provides a block-based interface for structuring the review into hierarchical sections, subsections, and paragraphs, with papers cited within the text. Users can (5) directly search for papers to add, or formulate hybrid search queries. (C) Hybrid query panel enables users to specify seed papers, compose natural language queries, select search pools (citation network pool for structural search or content search pool for semantic search), and apply metadata filters such as year range, venue, and minimum citation count.

by edges. These connections are used to provide brief interpretations of individual papers or to mark commonalities across multiple papers.

Groups are represented with bubblesets that semantically bind related papers and notes. The groups are explicitly created and organized by the researcher to reflect their own conceptual structure. Groups can also be nested hierarchically, supporting complex classification schemes such as organizing subtopics under larger themes.

The visual editor additionally supports zooming functions that let researchers adjust the level of detail of text according to their needs. When zoomed out, the view emphasizes overall structures by omitting textual information, helping users to see how groups and relationships are organized without clutter. When zoomed in, detailed information about individual papers and associated notes becomes accessible.

3.4.2 Supporting Metadata-Based Organization. By default, the visual editor allows free spatial arrangement, enabling researchers to drag papers and adjust their relative positions while exploring relationships. However, when working with large paper collections, systematic organization based on metadata becomes important. Users can assign attributes such as venue, year, or citation count to the x- and y-axes to structure papers systematically.

Once the axes are configured, the canvas is partitioned into regions defined by the metadata values, and each paper node is constrained to its corresponding region. For example, if the x-axis is set to year and the y-axis to venue, a 2023 CHI paper can only be positioned within the area where that year and venue intersect. Within these bounded regions, users can still reposition papers to highlight nuances in how they interpret their relationships.

Citation relationships are shown as connecting lines between nodes. Because bubblesets are rendered in color and note connections are also displayed, we employ animated edge textures [106] to clearly distinguish citation links. Both axis-based organization and citation line display can be toggled on or off according to user preference.

3.4.3 Aligning and Adjusting Visual Layouts by Semi-Automation. Our visualization involves two types of nodes (paper nodes and note nodes), two types of edges (citation edges and note-paper edges), and hierarchical bubblesets, which require effective strategies for handling visual complexity. Manually arranging all network elements would be tedious, while fully automated layouts may disregard the interpretive meaning that users embed in their spatial arrangements. To balance these needs, CrossLit employs a semi-automated approach in which users provide high-level guidance and the system handles detailed arrangement [133].

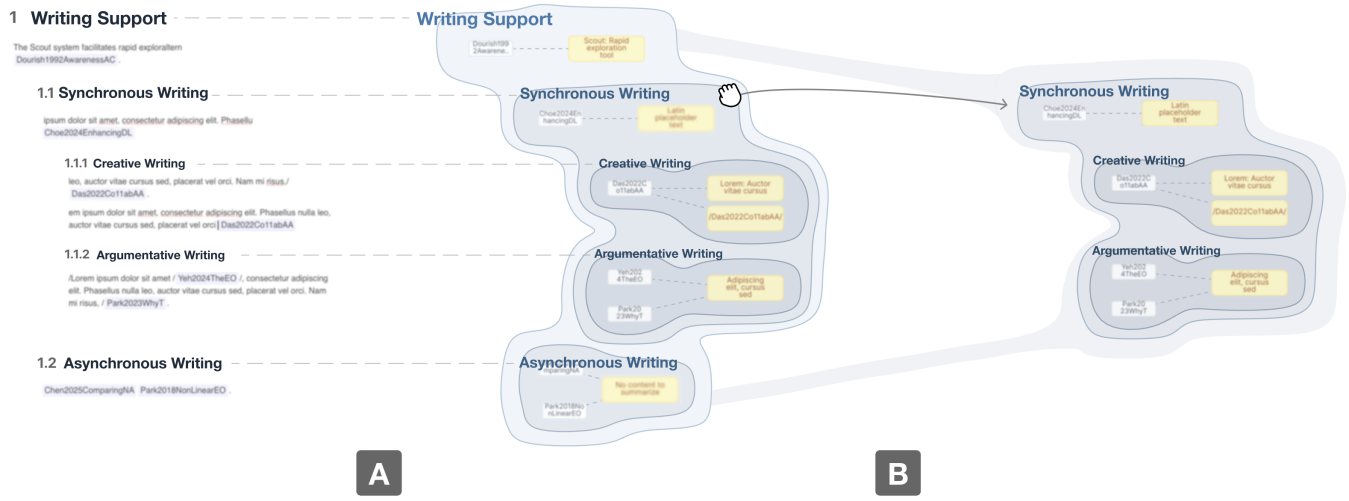

Figure 4: Visual editor features for layout organization. (A) Document-aligned layout mirrors the vertical hierarchy of sections, subsections, paragraphs, and citations from the text editor. (B) Users can reposition individual groups through drag-and-drop while preserving their internal structure.

CrossLit realizes this in two ways (Figure 4). First, a document-aligned layout aligns the visual arrangement with the structure of the text editor, providing a safe baseline that preserves correspondence between modalities while reflecting the organization users have developed. Second, users can reposition entire groups through drag-and-drop. When a group is repositioned, its papers and subgroups translate as a single block, preserving the internal spatial arrangement. This allows users to control placement where it matters most, while relying on consistent defaults for the rest.

3.5 Text Editor Design (DG2)

The text editor provides a block-based interface similar to commercial tools such as Notion[1]. Users can organize their writing into hierarchical sections with multi-level headers (e.g., using markdown-style markers such as ## and ###). Each section contains text blocks that can be independently moved, edited, or deleted. Papers can be referenced using a `@paper_id` notation, allowing them to be cited and manipulated consistently within the editor.

3.6 Bi-directional Synchronization between Editors (DG3)

The components of the visual editor and text editor maintain a strict one-to-one correspondence. Each note node in the visual editor corresponds to a text block in the text editor. When a note node is linked to paper nodes, the corresponding text block represents this by mentioning those papers. Similarly, the group structure in the visual editor maps directly to the header hierarchy in the text editor, ensuring that visual and textual organizations remain structurally identical. This correspondence enables users to quickly locate components across editors by highlighting them and ensures that editing in one modality automatically updates the other.

[1]https://www.notion.com/

Once structural alignment is established, CrossLit employs LLM-assisted adaptation to respect the expressive characteristics of each modality. Annotations in the visual editor are typically high-level and concise, whereas text requires more detailed and linguistically coherent descriptions. When a user assigns papers to a group and writes a short note in the visual editor, CrossLit expands the corresponding text block into sentences that incorporate the group name, paper titles, abstracts, and the note content. Conversely, when a text block is modified in the text editor, the system condenses it into keyword-like annotations for the visual side. If elements are added or moved such that positional information is invalidated, the system reinitializes their layout using the document-aligned arrangement described earlier. In CrossLit, LLMs are used primarily to maintain minimal semantic coherence after structural changes, rather than to generate full content. All generated text is presented for user review and revision, ensuring that final interpretations remain under the researcher's control.

3.7 Context-Aware Discovery (DG4)

When researchers want to discover new papers, they can create hybrid search queries by providing seed papers and/or natural language descriptions. The system incorporates contextual information from the seed papers' titles and abstracts, as well as from the entire text editor content, to transform the natural language input into a standalone query (Figure 11).

Users can draw from two search pools, either individually or in combination. The citation network pool is built using a co-citation–based approach similar to prior algorithms (e.g., [6]). Starting from the seed papers, it prioritizes works with high co-citation counts. In each iteration, the system extracts the top 50 cited papers, breaking ties with total citation count. This process is repeated three times, extending the search to papers within three degrees of separation.

The content search pool follows the approach in Singh et al. [114], adapted to our context: it retrieves the top 50 papers by issuing text-based queries over abstracts and snippets. To complement this, the system issues four additional relevance queries, each returning 10 papers, for a total of 40 results. These queries are generated alongside the initial query transformation.

After candidate papers are collected, the system reranks them using a cross-modal encoder to assess relevance. For each selected paper, an LLM generates concise explanatory content: a note in the visual editor and a corresponding text block in the text editor, describing why the paper is relevant to the query. This process also serves as a final validation step, excluding papers deemed irrelevant. The newly discovered papers then appear in the visual editor, each accompanied by corresponding notes and text blocks. Researchers can continue their work in whichever editor—visual or text—best suits their current focus, seamlessly integrating new material into their ongoing review.

Our retrieval pipeline operationalizes context-aware discovery by combining structural and semantic signals within a unified retrieval layer. This represents one feasible instantiation of the design goal; other retrieval approaches could be substituted while preserving the same cross-modal integration principle. The full retrieval method is described in Appendix B.1.

3.8 Design Considerations

We considered several design alternatives while designing CrossLit. Here we discuss key decisions and their rationales.

Separate Views vs. Mixed View. We chose to implement visual and textual modalities as separate, dedicated views rather than integrating them into a single mixed interface. While a unified view could potentially enhance cognitive immersion and enable creative interactions, we prioritized respecting the established ontologies that prior research has developed for each modality.

One-to-One Synchronization vs. Indirect Synchronization. When propagating changes from one modality to another, we opted for strict one-to-one correspondence between components rather than indirect mapping. While indirect synchronization offers greater flexibility, continuous visual-text synchronization is unprecedented in traditional literature review tools. We therefore prioritized making this novel concept consistent, predictable, and intuitive for users encountering this paradigm for the first time.

Visualization Design. The design space for representing bibliographic metadata and paper relationships through visual channels is vast. Prioritizing intuitive synchronization, we began by mapping the most common textual components (headers, paragraphs, and citations) to corresponding visual elements. As the relationships among these three elements already introduced substantial visual complexity, we minimized additional complexity for other features. For example, we enabled manual positioning of nodes as it adds no visual elements while supporting flexible annotation of document relationships, which is a core requirement for document organization [119].

Query Result Presentation. When incorporating newly discovered papers, we considered using LLMs to automatically modify relevant text sections. However, this would impose a dual cognitive burden on users who must track both synchronization-triggered changes and LLM modifications simultaneously. Instead, we chose to add new papers to fixed, predictable regions in both text and visualization views, allowing users to immediately grasp what has been added without ambiguity.

3.9 Implementation

CrossLit is implemented with a React frontend and a Python (FastAPI [2]) backend. The backend integrates three external services: OpenAI's `gpt-4o-2024-08-06` model for all LLM-based functionalities, the Semantic Scholar Academic Graph API[3] for paper retrieval, and Cohere's `rerank-v3.5` model[4] for cross-modal relevance ranking.

The visual editor employs a layered architecture with six components: (1) *Data layer* managing papers, notes, and groups; (2) Layout layer computing node positions through a hybrid pipeline; (3) *Bubbleset layer* rendering hierarchical group visualizations; (4) *Transform layer* handling viewport transformations; (5) *Presentation layer* rendering visual elements; and (6) Interaction layer managing user inputs and feedback.

Notably, the *Layout layer* coordinates multiple constraints to determine node positions: First, user-positioned nodes remain at their specified locations. Next, unpositioned nodes align to document structure coordinates. When metadata axes are active, nodes outside valid ranges snap to the nearest boundaries. Finally, d3's force simulation [5] resolves overlaps while minimizing displacement from intended positions.

4 User Study

We conducted an observational study using the think-aloud protocol with researchers who were actively performing literature reviews on their own topics, to understand how they engaged with CrossLit's visual and textual modalities.

4.1 User Study Design Rationale

The benefits of connecting visual and textual modalities have already been well established. Research shows that human cognition is dual-coded through visual and verbal systems [100], and that people learn more effectively from combined text and visuals than from text alone [92], particularly when corresponding elements are contiguously placed [93]. Recent cross-modal systems have demonstrated that connecting these modalities provides cognitive offloading [64, 117, 124], enables granular control [29, 64, 124, 131], and facilitates non-linear exploration [29, 64, 116, 117, 131] across domains including creative writing [29, 116], argumentative writing [131], information sensemaking [64, 117], and data analysis [124].

Building on these benefits, our user study investigates how visual and textual modalities interplay during the literature review sensemaking process. We address the following research questions:

RQ1 How well do features of CrossLit address users' literature review needs?

[2] https://fastapi.tiangolo.com/
[3] https://www.semanticscholar.org/product/api
[4] https://docs.cohere.com/docs/rerank
[5] https://d3js.org/

RQ2 What are common patterns of using visual and text modalities across the stages of literature review?

RQ3 How does synchronization between the two modalities support literature review?

RQ4 What improvements and unmet needs do users identify?

To answer these questions, we observed researchers conducting literature reviews on their own topics using CrossLit. By grounding the study in participants' genuine research needs rather than artificially assigned topics, we aimed to capture authentic sensemaking behaviors. We employed a think-aloud protocol [38] to surface participants' rationales for engaging with each modality, which would otherwise remain hidden if we examined only their final outputs.

We opted not to include a comparative baseline for several reasons. CrossLit is designed to integrate visual and textual modalities across all stages of literature review, whereas existing tools typically target a specific modality or stage. No single tool therefore serves as a fair point of comparison. We also considered combining multiple tools, such as Google Scholar[6], Connected Papers[7], and Notion[8]. However, such a combination would not represent a familiar workflow for most participants, potentially introducing confounds that could obscure the authentic sensemaking behaviors we sought to observe.

4.2 Participants

We recruited 16 researchers in HCI and adjacent fields through university mailing lists and research community channels. Participants were required to be currently conducting, or preparing to conduct, a literature review on their research topic. Fourteen participants were doctoral students and two were master's students, with an average of 4.5 years of research experience (range: 2–9 years). Their research areas covered a wide spectrum of HCI, including XR, human–AI interaction, data visualization, digital mental health, and privacy.

As part of the recruitment process, we collected information about the stage of participants' ongoing literature reviews. This was cross-checked with their actual activities during the study session, in case participants had made progress in their literature review since recruitment. They reported being at different points in the process: seven had formed topic ideas but had not yet engaged in substantive exploration; five were identifying key papers and drafting outlines; and four had produced an initial draft review that required revision.

4.3 Study Procedure and Data Collection

Each participant completed a 90-minute individual session consisting of a 25-minute pre-task phase, a 45-minute main task, a 5-minute questionnaire, and a 15-minute interview. Participants received a $20 compensation for their time.

4.3.1 Pre-task (25 min). Following an informed consent process (5 min), participants were introduced to the system through a 10-minute walkthrough. The walkthrough presented core functions aligned with the design goals: adding papers, visual editing, text editing, and their correspondence, hybrid queries, and axis-based visual analysis. To avoid biasing participants' usage strategies, the walkthrough focused only on demonstrating features rather than prescribing particular workflows. Next, participants engaged in a short hands-on practice (5 min) in which they accessed the system, added papers, and tried simple visual and textual editing tasks. This ensured they could operate the interface comfortably and provided an opportunity to ask questions. Finally, during goal setting (5 min), participants shared the current status of their literature review and identified objectives for the 45-minute session. They were instructed to pursue the activities most critical to their ongoing literature review within the allotted time.

4.3.2 Literature Review Session (45min). While participants used our system, we explained the think-aloud protocol, acknowledging that speaking while reading or writing can be difficult, and reassured participants that incomplete sentences or murmurs were acceptable. The facilitator intervened minimally, prompting participants to verbalize their thoughts if they remained silent for extended periods, and responding to questions only when asked about system functionality.

To obtain a comprehensive view of user experience, we collected multiple types of data during the session: (1) screen recordings, which captured not only system use but also external activities such as search engine queries, references to prepared documents, or full-paper PDFs; (2) audio recordings of think-aloud comments; and (3) researcher observation notes.

4.3.3 Questionnaire and Interview (20min). The session concluded with a 5-minute questionnaire and a 15-minute semi-structured interview. The questionnaire contained 20 items. To understand how usable the system was and how integratable with current workflows, we included the System Usability Scale [15] (10 items). Furthermore, to evaluate how effectively the system supported literature review, we included 10 items on sensemaking processes during literature review, such as exploration, narrowing, and structuring. The full survey is reported in Appendix C.1.

The interview explored participants' experiences in greater depth. We first asked them to elaborate on questionnaire items where they had given particularly high or low ratings, then proceeded with open-ended questions on (a) self-assessment of session outcomes, (b) differences from their typical workflow, (c) the role and appropriateness of the visual and text editors, (d) the effectiveness of synchronization, and (e) suggestions for improvement. When necessary, screenshots captured during the session were used to prompt reflection and clarify specific behaviors.

4.4 Analysis Methods

4.4.1 Video Segmentation and Annotation. To analyze participants' sessions, we drew on three complementary data sources: (1) screen recordings, which captured objective usage patterns; (2) think-aloud audio recordings, which revealed participants' reasoning and intentions; and (3) observation notes, which provided contextual cues. All data were synchronized on a common timeline and annotated using a custom-built interface. Audio recordings were transcribed with timestamps, and observation notes were recorded in a timestamped chat interface, allowing direct alignment with other data.

[6] https://scholar.google.com
[7] https://www.connectedpapers.com
[8] https://www.notion.com

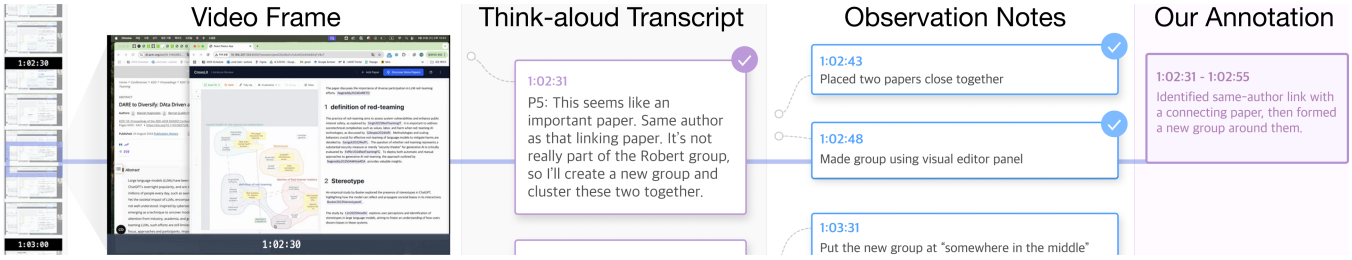

Figure 5: Custom annotation interface used for our analysis. The interface integrates video recordings, think-aloud transcripts, and observation notes to annotate time intervals of meaningful behaviors with interpretations.

The annotation interface displayed video frames as tiled thumbnails, with a video column for frame-by-frame review (hover to preview), alongside columns for think-aloud transcripts and observation notes. A dedicated annotation column allowed researchers to mark and describe specific time segments. We report the interface in detail in Appendix C.2.

4.4.2 Codebook Generation and Analysis. Using the annotation interface, the first author segmented the data from 6 of the 16 participants into meaningful time units and wrote integrated interpretation notes describing what occurred. This produced 330 interaction segments. These segments were open-coded, and the first and second authors discussed how best to classify and analyze different types of interactions.

In developing the codebook, we paid particular attention to how similar surface-level actions could reflect different sensemaking stages. For example, placing a paper into a group could simply indicate provisional categorization for later review, a paper-specific interpretation based on content, or a restructuring of higher-level schemas. One participant (P5) explained, *"As I assign papers to groups, I'm comparing alternative ways to structure the argument—deciding which group to present first and how to narrow to the other"*, suggesting a late-stage sensemaking activity oriented toward argument construction.

To capture such distinctions, we generated a codebook that maps observed actions to stages of sensemaking outcomes (Table 1). We report further detailed definitions of the terms used for our coding in Appendix C.3.

In addition to sensemaking outcomes, we also coded the primary modality (visual or textual) in which they occurred. For example, when a participant read an abstract and then grouped the paper, the action was coded as *textual* if the paper was inserted into a section via block dragging in the text editor, and *visual* if the corresponding node was moved into a group in the visual editor. External activities such as searching for new papers or reading full texts were not assigned a modality. A few observation segments were assigned as *"visual and textual"* when they closely referenced both visualization and text for the same sensemaking outcome.

Following this, the first author went through a segmentation process for all 16 participants, and obtained 731 observation segments. In this process, transitions were carefully segmented based on the definitions in the codebook. Closed coding was then conducted using the codebook. Since modality could be clearly derived from interaction logs, we omitted calculating inter-rater reliability [94]. For sensemaking outcomes, the first and third authors coded independently and achieved almost perfect agreement (Cohen's κ = 0.90) [79].

4.4.3 Sequential Pattern and Thematic Analysis. To understand which sensemaking stages participants navigated when using CrossLit for literature review and how they utilized visual and textual modalities in the process, we conducted both quantitative and qualitative analyses. Quantitatively, we first examined the overall frequency of participants' stage transitions (i.e., sensemaking stage & interaction modality), and also focused on transitions with positive contingency (i.e., after doing A, one is highly likely to do B). As our statistical methodology, we used lag sequential analysis (lag = 1) to identify statistically significant contingent transitions [4], along with Yule's Q as an effect size measure [5]. Yule's Q indicates how much more or less frequently a specific transition ($A \rightarrow B$) occurs compared to other transitions from the same origin ($A \rightarrow\sim B$) , while lag sequential analysis tests whether such bias exceeds chance levels.

Among the 731 observation segments, we analyzed a total of 573 transitions by considering only transitions between different types while ignoring consecutive segments of the same type. Transitions from or to *"visual and textual"* modality were calculated as 0.5 transitions for each modality. We did not apply Bonferroni correction for multiple comparisons given the exploratory nature of this analysis [8]. Yule's Q effect sizes were interpreted following Cohen's criteria for correlation coefficients [30], retaining only positive values of 0.1 or greater to focus on transitions with at least small effect sizes.

Qualitatively, we performed thematic analysis [14] by integrating participants' observation segments with post-interview content. Considering the context of each participant's entire session, we conducted in-depth analysis of what stage and modality transitions were important to them and why they made such transitions. Additionally, we identified recurring themes regarding perceived benefits and challenges of systems and workflows like CrossLit.

4.4.4 Discovery Intent Analysis. We applied a similar analysis methodology to examine whether users employed metadata-driven or content-driven intent when searching for new papers using CrossLit's hybrid query feature. Among the 731 observed segments, we analyzed segments where users utilized hybrid queries, examining the context and purpose of their queries to develop a codebook of four discovery intent types (Table 2).

Based on this codebook, the first and third authors performed closed coding and achieved almost perfect agreement (Cohen's κ

Table 1: Mapping of Pirolli & Card's Sensemaking Model [102] to our codebook adaptation for analyzing literature review activities observed in our user study.

	Pirolli & Card Model	Description	Codebook Adaptation	Description	Codebook Activity Examples
1	External Data Sources	Raw data from various sources	External Academic Databases	Papers in academic databases	—
2	**Shoebox**	Much smaller relevant subset filtered from external data	**Triaged Papers**	Pre-sorted paper collection awaiting detailed review	• Prioritizing papers via metadata • Pre-grouping by assumed similarity
3	**Evidence Files**	Extracted snippets from shoebox items	**Paper Interpretations**	Individual paper analyses with contextual meaning assignments	• Writing interpretation notes on a paper • Determining a paper belongs to a certain group
4	**Schema**	Implicit conceptual patterns for organizing domain knowledge	**Related Work Schema**	Implicit grouping of papers for organizing paper relationships	• Structuring grouping hierarchy for papers • Evaluating structural coherence
5	**Hypothesis**	Tentative conclusions with supporting arguments	**Related Work Narratives**	Argumentative synthesis with literature trends and gaps	• Identifying trends and limitations of prior work • Adding supporting citations

Table 2: Codebook for classifying hybrid query intent when users search for new papers in CrossLit.

Intent Type	Subcategory	Definition
Content-driven	Exploratory	Broadly exploring papers on a topic using general keywords or topic names without specific targets.
	Confirmatory	Seeking papers that provide evidence or support for a specific, pre-formulated claim or proposition.
	Targeted	Combines both exploratory and confirmatory qualities: finding papers that meet specific content criteria while remaining open to variations within those constraints.
Metadata-driven	—	Searching based on non-content attributes such as publication year, venue, citation count, or citation relationships.

= 0.88). Using the same sequential pattern analysis methodology, we examined contingent transitions from previous sensemaking stages to current discovery intent.

4.5 Convenience Feature for Study Facilitation

We allocated a 45-minute session to observe as authentic a literature review experience as possible, but this was still a short time to include substantial writing work. In particular, since most of our participants were non-native English speakers, we anticipated additional friction related to writing. To ensure sufficient experience within the limited session time, we provided a convenience feature that allowed participants to request rewriting from an LLM in the text editor. When participants selected consecutive text blocks and wrote a request, the content was replaced with results rewritten by GPT-4o, given the entire text editor content, information about the selected blocks, and titles and abstracts of papers belonging to the selected blocks. As a convenience feature for writing, this function was limited to rewriting text blocks only. Restructuring across multiple section headers was not supported.

5 Findings

In this section, we present findings from our user study of CrossLit. Drawing on interaction logs, video annotations, and post-task interviews, we analyze how participants used the system and how it shaped their literature review workflows. The findings are organized into three themes: (1) overall usability and perceived usefulness, (2) how and why participants employed visual and textual modalities in sensemaking, and (3) opportunities and tensions arising from synchronization between the two modalities.

5.1 RQ 1. General Usability and Usefulness

5.1.1 Users found CrossLit's features generally usable and useful. The System Usability Scale (SUS) score for CrossLit averaged 73.2, indicating above-average usability and suggesting that the system

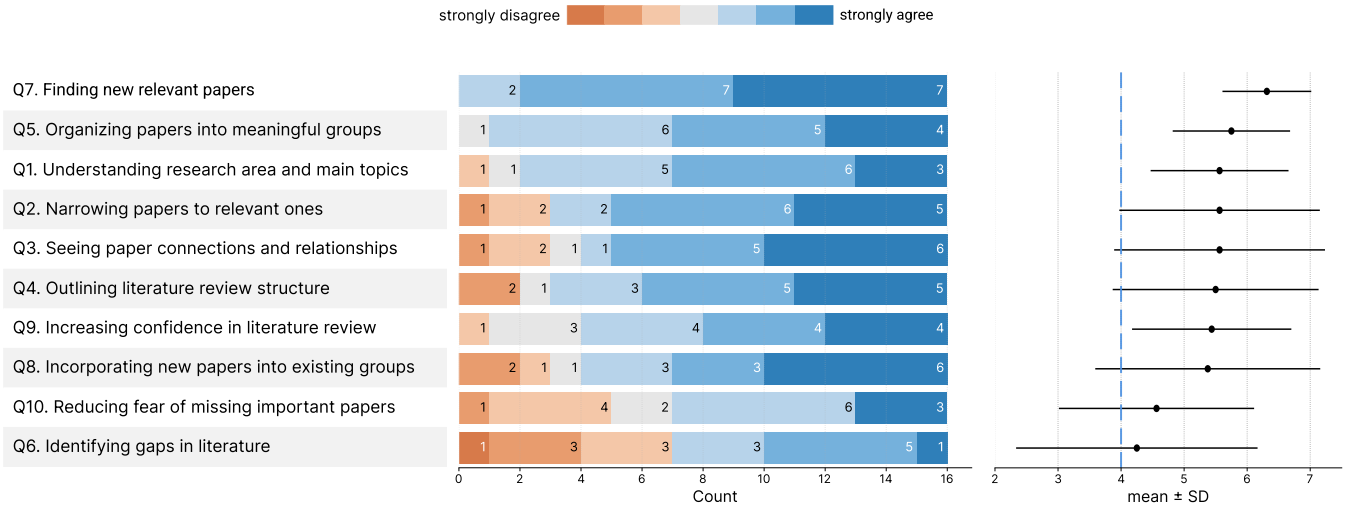

Figure 6: Participants' ratings of CrossLit's support for ten sensemaking activities in literature review, shown as stacked distributions and ordered by average score. Higher-rated activities include finding new relevant papers and organizing literature into meaningful groups, while identifying research gaps and reducing fear of missing important papers received comparatively lower ratings. Overall, participants evaluated the system positively across multiple stages of the literature review workflow.

generally provided a favorable user experience. In our custom questionnaire designed to assess system support for sensemaking activities in literature review, participants rated CrossLit as effective across a wide range of tasks (Figure 6).

Many participants particularly valued the integrated discovery features, emphasizing the flexibility to expand their collections using diverse strategies. For example, some preferred exploring papers closely connected to their seed articles (P7), while others deliberately sought out more distantly related works (P11). The integration of discovery functions directly into the workflow was also highlighted as a major strength. P4 and P16 noted that tasks they previously had to perform across multiple tools—such as finding papers, grouping them, and translating them int.o written text—could now be accomplished within a single system. A key factor enabling this integration was the complementarity between visual and textual modalities (Section 5.2). In addition, integrated LLM-based features were seen as helpful in supporting smooth transitions between task stages, allowing participants to balance bottom-up and top-down processes (P3).

However, participants gave lower ratings for Q10 (reducing fear of missing out) and Q6 (identifying research gaps). They explained that ensuring complete coverage in literature search is inherently difficult (P4, P6–8). Other critical feedback included concerns about visual clutter when citation edges became dense (Q3: seeing paper connections, raised by P4, P10, P13), as well as the lack of an automatic grouping suggestion feature for Q4 (outlining a literature review) and Q8 (incorporating new papers) (P3, P5–6, P12).

5.1.2 Integrated query feature encompassed distinct paper search needs. To examine whether participants' paper search needs encompassed both metadata and content-driven intents, and whether the two search pools (citation network pool and content search pool) of

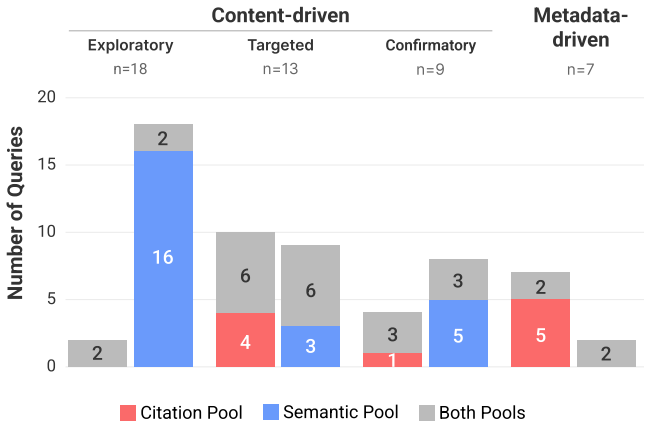

Figure 7: Distribution of search-pool usage across discovery intent categories. Exploratory searches predominantly used the semantic pool, while metadata-driven searches relied mainly on the citation pool. Targeted searches showed substantial use of both pools, though without a dominant pattern. Gray segments represent queries that drew on both pools. Overall, the figure illustrates how participants selected pools in alignment with their differing discovery needs.

CrossLit's hybrid query feature reflect such needs, we examined the distribution of discovery intents and search pool usage.

A chi-square test of independence revealed a statistically significant association between users' discovery needs and their use of search pools ($\chi^2 = 27.24$, $p < .001$), with Cramér's $V = .54$ indicating a large effect size. This suggests that users strategically

utilized the two pools differently according to their search intentions. Post-hoc analysis with adjusted residuals showed that exploratory searches were significantly more likely to use only the semantic pool than expected (adjusted residual = 2.25, $p < .05$), while metadata-driven searches were significantly more likely to use only the citation pool than expected (adjusted residual = 2.88, $p < .01$). Targeted searches most frequently employed a hybrid approach, using both semantic and citation pools simultaneously.

These search intents were also meaningfully incorporated into the final outcomes. The 16 participants' final literature review outcomes contained an average of $M = 18.31$ papers ($SD = 9.00$; range: 7–41), of which $M = 11.38$ papers ($SD = 8.60$; range: 5–40) were retrieved through hybrid discovery queries, accounting for approximately 62.2% of the total collection. All intent types contributed equally to participants' final paper collections, as a Kruskal-Wallis H-test comparing the mean number of incorporated papers across discovery intent types (exploratory: $M = 3.33$, $SD = 2.81$; targeted: $M = 3.54$, $SD = 2.63$; confirmatory: $M = 5.22$, $SD = 3.61$; metadata-driven: $M = 4.14$, $SD = 2.73$) revealed no significant differences ($H = 2.73$, $p = 0.43$).

5.2 RQ 2. How and Why Researchers Use Visual or Text Modality

Based on analyses of participants' individual patterns (Figure 8) as well as statistically significant patterns that occurred more frequently than chance (Figure 9), we found consistent and coherent patterns regarding sensemaking stages, visual/textual modalities, and paper discovery intents.

5.2.1 Visual Modality is Effective in Early Stage. Analysis of transition frequencies (Figure 8-A, C) and transition patterns (upper part of Figure 9) revealed that the use of the visual modality was concentrated in the earlier stages of sensemaking. Moreover, early-stage sensemaking in visual modality led to metadata-driven paper discovery intents. These results suggest that many participants employed the visual modality to filter papers at the outset, establish exploration priorities, and interpret individual works while intuitively grouping them (P1–5, P7–8, P10, P12, P16).

Inspecting metadata distributions in the visualization often provided cues for prioritizing which unexplored papers to examine or for discovering new papers. Citation relationships were interpreted as signals of "relevance." Participants sometimes grouped citing papers together in the same space for joint inspection (P2, P5), or prioritized exploring papers that cited or were cited by ones they had already understood (P3, P7). P8 sought to identify additional groups that were similar yet distinct from the clusters already formed. When encountering a new paper with citation links to one of these groups, P8 expanded the candidate pool from the citation network even before reading the paper, using queries such as *"find papers on [broader theme] but exclude [specific topic]."*

Temporal and venue distributions were also used to check whether a topic was properly grounded within participants' academic field. For instance, P1 reflected: *"I saw that papers on [specific topic] started appearing only from 2022, which made me realize it's a very trendy area. And regarding venues, they seemed clustered at CHI. [...] Since I think it's important to submit to a venue that will recognize my work, this convinced me that I should target CHI soon."* Similarly, P7 observed: *"Aside from the [theory papers] I included, the [application papers] are very recent. [...] Still, I don't think there would be absolutely none between 2014 and 2021. [...] That made me think I should also look for [application papers] from that period. And I noticed these are concentrated at VIS, which confirmed that searching within VIS would be a superior strategy."*

These uses of metadata visualization helped participants contextualize their topics within disciplinary trends and venues, supporting not only discovery but also the positioning of their own work in relation to the field. The ability to structure papers, notes, and groups and arrange them freely on the canvas enabled fluid and intuitive grouping. Participants often expressed implicit relationships by spatially positioning papers that were only loosely connected—those that were *"still just in my head"* (P5, P7) or that *"seemed related, even if I wasn't sure how to group them yet"* (P8). This provisional and flexible nature of visual grouping was particularly valuable in the early stages of developing a Related Work schema. Participants used this flexibility to pursue multiple schema directions simultaneously (P4), to subdivide groups with slightly different emphases (P7, P8), to discover intersections between groups, or to reconsider their schema entirely when group development diverged from their initial expectations (P5, P7).

5.2.2 Text Modality is Effective in Later Stages. In contrast, analyses of transition frequencies (Figure 8-B, D) and transition patterns (lower part of Figure 9) showed that the use of the text modality was concentrated in the later stages of sensemaking. Moreover, later-stage sensemaking in text modality led to content-driven paper discovery intents, specifically for finding confirmatory evidence papers. Many participants (P1–4, P6, P9, P11–16) employed text to refine interpretations, construct storylines, and strengthen argumentative structures. Participants explained that they preferred text for these tasks because it allowed them to develop ideas about "flow" without constraints. For example, when P1 was struggling with how to structure the related work schema, they began by *"just writing naturally"*, which then sparked an idea for three sections and led to drafting detailed content across them in one continuous burst. Similarly, P16 noted: *"Text gives you flow. [...] With visualization, you can indicate which papers are related, but it doesn't really capture the story."*

For participants whose related work schema was already fixed and partially outlined, the most pressing task was locating supporting papers for specific claims—a process they also found better supported in text (P9, P11, P13). Familiarity with text-based work and the constraints of final deliverables further reinforced participants' reliance on text. Those accustomed to selecting and grouping papers in text editors reported gravitating toward text even in the early phases of sensemaking (P14, P15). As P15 explained: *"Maybe the visual editor would have been more well-suited [...] I'm just a little bit more familiar with doing this kind of thing with a text editor."* Finally, because the end product of literature review must ultimately take textual form, several participants reported a preference for text as the modality most closely aligned with their final goals (P2, P16).

5.2.3 Sensemaking Stage Patterns Reflect Actual Literature Review Progress. To examine whether the patterns based on sensemaking model provide an appropriate proxy for actual literature review

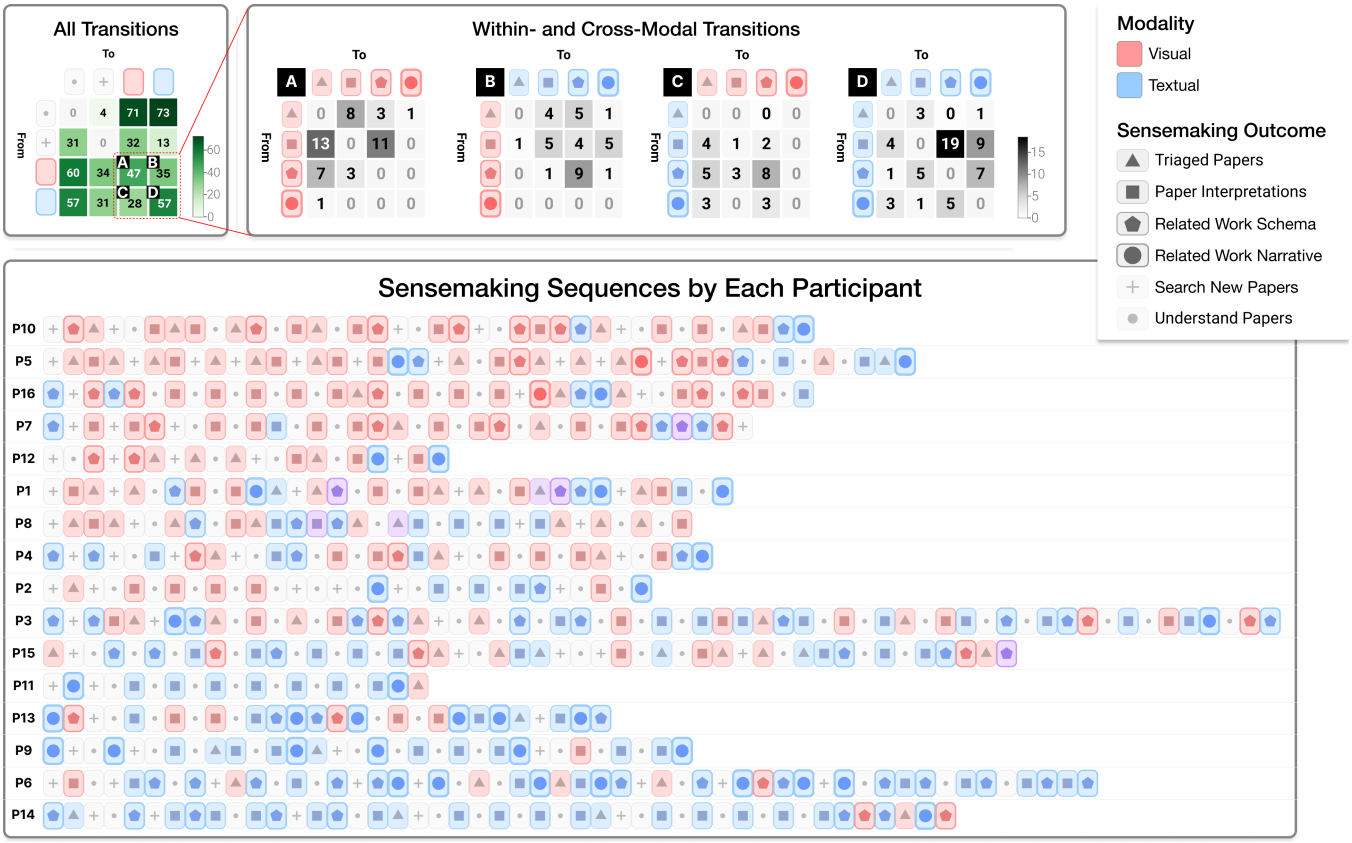

Figure 8: Overview of modality transitions and participant-level sensemaking sequences. The top matrices show transition frequencies between sensemaking states, with the left panel summarizing all transitions and the right panels detailing within- and cross-modal transitions for each outcome. Darker cells indicate more frequent transitions. The bottom panel displays each participant's sequence of sensemaking actions over time, where colors represent modality (visual, textual, or combined) and symbols represent outcome types. The visualization highlights variation in individual strategies—some participants remained primarily in one modality, while others switched rapidly between them. Visual-to-visual transitions were most common in early sensemaking stages, textual-to-textual transitions in later stages, and the schema-building stage showed the highest level of cross-modal transitions.

stages, we conducted a comparative analysis with participants' self-reported literature review stages collected prior to the session.

The distribution of sensemaking stages and modality involvement during sessions differed significantly based on participants' literature review stages ($\chi^2 = 51.37$, $p < .001$). Notably, participants who reported having "produced an initial draft review that required revision" engaged in writing related work narratives with text modality at a significantly higher proportion than other participants ($H = 6.69$, $p < .05$). Although not statistically significant, we observed that triaging papers in visual modality was most frequently used among participants who "had formed topic ideas but had not yet engaged in substantive exploration," while interpreting papers in visual modality was most commonly performed by those "identifying key papers and drafting outlines." These patterns correspond closely with the aforementioned patterns.

5.3 RQ 3. Benefits of Synchronizing Visual and Text Modalities

When the two strengths of modalities were treated as distinct roles, the bidirectional synchronization in CrossLit enabled much more frequent and fine-grained micro-iterations than in participants' usual workflows. More interestingly, however, during the process of developing a Related Work schema, the two modalities did not operate in separate roles but worked in close coordination.

5.3.1 Placing Visual and Text Together Enables Micro-Iterations. Some participants adopted a demarcation view, assigning distinct roles to each modality and using them only for those purposes (P3, P5–6, P8, P10, P12, P14–15). For example, P5, P10, and P12 explained that they needed to *"gather as much information as possible"* (P10) and *"get the big picture organized"* (P10, P12) in the visualization before moving into writing, after which *"the writing flows more easily"*(P5). P12 elaborated, *"The papers on the left (visual editor)*

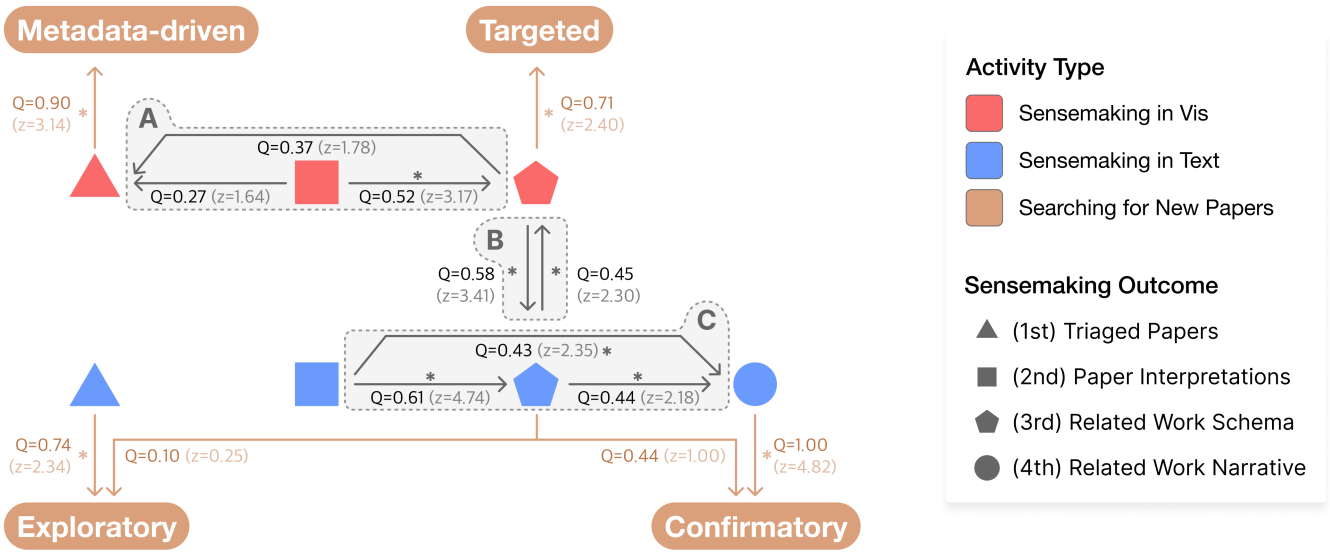

Figure 9: Transition patterns among sensemaking stages, modalities, and new paper discovery intent revealed by sequential pattern analyses. Each node represents a sensemaking stage (shape) performed in a specific modality (color), with brown nodes indicating search actions. Directed edges show observed transitions between states, annotated with effect size (Yule's Q) and statistical significance (*) $p < 0.05$; only edges with $Q \geq 0.1$, indicating meaningful contingency, are displayed. The analysis reveals coherent patterns: (A) visual modality is mainly used for early sensemaking stages, (B) both modalities for building the related work schema, and (C) text modality for later sensemaking stages. Patterns for paper search actions are consistent with these findings: in early stages, users conduct open-ended searches based on metadata in visualization and content in text; during schema building, they perform diverse types of searches; and in later stages, they conduct confirmatory searches in text. Edge colors differ to indicate that these results come from two similar but independent sequential pattern analyses.

are the ones I'm not very familiar with. [. . .] It felt a bit premature to insert them directly into the text editor." Their interaction logs (Figure 8-1, 2, 5th row) confirm that they spent most of their time organizing within the visual editor before eventually shifting to text.

Nevertheless, the synchronization between visualization and text allowed participants to iterate far more frequently—and at a finer granularity—than in their usual workflows. For participants primarily conducting selection and grouping in the visualization, having a synchronized draft in the text editor served as a preview of how their writing might take shape. This provided hints for structuring their arguments (P6, P8) or signals that they were ready to move into writing (P12). A particularly interesting case was P3: although they explicitly fixed the roles of visualization for selection/grouping and text for detailed structuring/writing, they still made the highest number of cross-modal switches—over 20 in a single 45-minute session. P3 reflected: *"Normally, I try to be very thorough. I might spend hours just finding papers and then hours reading before I even start structuring. But in 45 minutes, I already had this kind of outline, which I was satisfied with. [. . .] Since the system kept generating sentences, I needed to check them right away, then I thought about structure, and while finding papers, I was simultaneously doing later-stage work I usually save for much later."*

Participants who primarily worked in the text editor also benefited from the visual panel. Citation links shown in the visualization supported more efficient paper selection, especially when discovering new papers mid-writing. Across participants, papers with citation links to existing works consistently became priority candidates for review (P6, P14, P15). P15 even discovered one such paper and realized it was a highly important work they had already cited in earlier research.

5.3.2 Schema Building Relies on Both Visual and Text Modalities. Pattern analysis of transitions (bridge section in Figure 9) revealed that during the process of building a Related Work schema, switches between visual and text modalities occurred more often than chance. Moreover, this sensemaking stage led to various paper discovery intents, where visual modality led to targeted discovery needs while text modality led to content-driven needs. In contrast to the demarcation view, this finding suggests that schema construction requires tight cooperation between the two modalities.

Participants described how, while developing more detailed content in text, they simultaneously referred back to the visualization to gauge the overall outline and completeness of the schema. A key factor supporting this process was the abstracted representation unique to the visualization, which distilled detailed content into a form that made it easier to assess schema-level structures (P3, P4, P6–7, P10, P16). Reviewing these summarized notes during text work often sparked new insights into how to organize content (P3, P7, P14). P14 reflected that this back-and-forth felt natural: once sufficiently familiar with the text, seeing the material arranged in

another form provided new perspectives. Group counts and sizes became cues for whether enough information had been incorporated (P3, P7, P10, P15), while citation links were used to assess relationships between groups (P5, P7).

As P16 explained: *"People say there are visual thinkers and verbal thinkers, and I think I'm probably the former. When I only see text, it's actually hard for me to notice relationships. The sections are grouped, but I still have to dig through sentences to find the papers. In the visualization, each paper exists as an entity in the graph, so it's easier to see. [...] For example, I can clearly see that [Group 1] has many papers but [Group 2] is almost empty, which helps me explain the gap in prior work that my paper is addressing."*

P1 offered a similar perspective while drafting the second section after finishing the first. They noticed conceptual overlap between sections, which led them to alternate between hovering over papers in the visual editor and reviewing their text outline: *"Deciding how to set sections and place papers is the fundamental challenge. In the text editor, I can only see the scope of one section at a time. I can't really think about the bigger picture there. But here [in the visual editor], all the papers are visible on one screen, so I can reason about the larger structure. [...] Normally, when I read papers individually, I don't really think about their relationships. But when citation links are visualized, I realize that some papers I thought were unrelated are actually connected. That even made me consider merging sections I initially thought were separate."*

5.3.3 Synchronization Turns Visual and Text into Alternative Interaction Modalities. Through synchronization, participants often used one modality as an alternative interaction channel for tasks they would normally perform in the other. This reduced interaction costs and sometimes enabled actions that would otherwise have been impossible. For example, instead of explicitly linking or unlinking notes and papers one by one in the visual editor, participants chose to merge or split text blocks in the text editor simply by pressing backspace or enter (P3, P7, P8). Conversely, when text became lengthy and participants wanted to reposition entire blocks, they used the visual editor instead, since it required *"much less dragging"* (P4).

An especially interesting case came from P8: although the system did not provide automatic clustering of papers based on keywords or content in the visualization, P8 improvised a workaround. They used the browser's native search to locate keywords in the text editor, then gathered the corresponding visual elements in the visualization to approximate the effect of clustering.

5.4 RQ 4. Perceived Challenges and Improvement

5.4.1 Users Need Ways to Preserve Carefully Crafted Content from Synchronization. As shown in the previous section, synchronization between visual and text modalities had clear benefits for literature review, facilitating iterative refinement and enabling schema development from multiple perspectives. Yet interviews also revealed participants' concerns about preserving their carefully created content from influenced by synchronization.

Many participants reported being careful not to make major changes to the visualization in order to preserve the intentional structure and subtle nuances they had crafted in their text (P2, P3, P7, P9, P14). For instance, P2 emphasized that synchronization should not disturb the logical order they had explicitly arranged in the text, while P9 hoped it would preserve the subtle nuances of expression. Similarly, participants who carefully selected just a few keywords for visual notes to represent their related work schema (P4, P12, P16) expressed that they would have preferred them to remain unchanged by synchronization triggered by text changes.

Participants wanted their carefully crafted content to be protected from unintended modifications. Without such protection, synchronization occurring after every interaction made them worry about its integrity. P6 recalled an experience of synchronization happening before they finished what they treated as an atomic change, and said, *"Do I have to check every time it changes? [...] Can I trust that what I intended is really reflected?"*. Consequently, they hoped such content could be placed in separate, protected areas—for example, P2 suggested adding a *"final deliverable"* section.

5.4.2 Visualization Can Serve More Diverse Roles. Participants expressed a desire for the visual modality in CrossLit to evolve in more diverse and advanced ways. In the early stages of work, they valued being able to record preliminary expressions in both the visualization and the text. However, as their work progressed, most participants assigned a single role to text—continuously refined until finalized. Consequently, they hoped the visualization could instead serve as a space for *"all the ideas not yet reflected in the text."* The envisioned roles fell into three categories: (1) revealing deeper metadata relationships, (2) accumulating secondary labels, and (3) creating and comparing multiple versions of sensemaking.

Regarding metadata relationships, participants wanted more analytical capabilities for metadata distributions (P6, P7, P8). For example, P6 hoped for hierarchically arranged citation links to identify foundational papers.

Regarding secondary labels, participants wished to mark certain notes *"as memory aids"* (P3) rather than polished content intended for the final text. P8 and P9 wanted the visualization to function as a sandbox for storing tentative papers and notes, while P12 and P16 expressed a desire for figures from papers to be displayed directly in the visualization panel.

Regarding multiple versions, participants wanted the visualization panel to become a space for exploring and comparing alternatives. They hoped it could serve as a more sophisticated schematic representation, allowing flexible recording of meaning units at increasingly fine levels of granularity (P2, P5, P10, P16). From these perspectives emerged concrete feature suggestions: automatically proposing subgroups based on paper content or existing notes (P2, P16), and enabling switching between and reusing multiple visualization panels (P2).

6 Discussion

In this section, we reflect on how synchronization between visual and textual modalities reveals new opportunities for supporting complex sensemaking tasks. Our findings suggest that the key challenge is not choosing between modalities but enabling fluid transitions between them, with implications for how we design tools that preserve human agency while leveraging AI assistance.

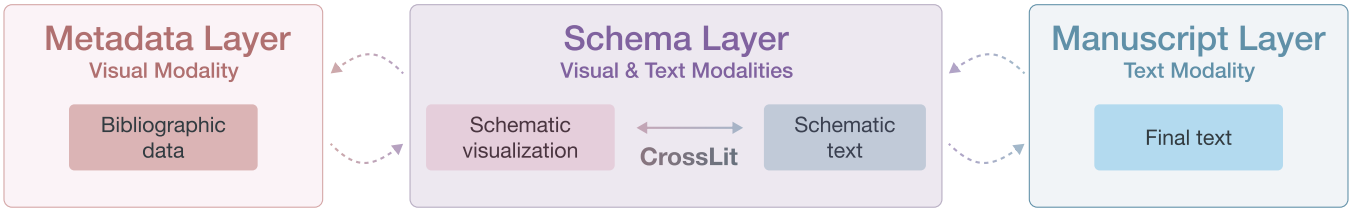

Figure 10: Design space for connecting visual and text modalities in literature review. While existing literature review support systems are distributed across three semantic layers, CrossLit enables automated synchronization within the schema layer. Future work should explore making such synchronization more versatile and intelligent, as well as bridging the boundaries between tools residing in different layers.

6.1 Design Space for Visual and Text Modalities in Literature Review

Our findings reveal that visual and text modalities serve distinct yet complementary roles across the literature review process. Participants progressed through three stages: lightly categorizing papers based on metadata (primarily visual), constructing schemas for their related work (visual and text in close cooperation), and refining arguments (primarily text). Because CrossLit maintained constant synchronization between modalities, participants could fluidly transition between stages, drawing from either or both as needed. This synchronization proved particularly effective during provisional grouping, experimental rearrangement, and recording of incomplete ideas. However, as work matured, participants increasingly produced sophisticated outputs that demanded more deliberate synchronization.

To structure these findings and propose directions for future systems, we conceptualize a design space where the roles of visual and text modalities are defined across three semantic layers (Figure 10). The **metadata layer** captures bibliographic information along with optional short notes that researchers may add based on initial impressions. The **schema layer** represents the intermediate structures that researchers construct to organize their review, including groups, categories, and themes. Finally, the **manuscript layer** encompasses the finalized prose, where logical flow and argumentative nuance take precedence. A key insight of this design space is the distinction between two types of synchronization: **synchronization within the schema layer** across visual and text modalities, and **cross-layer synchronization** that propagates changes between metadata, schema, and manuscript layers.

6.1.1 Synchronization within the Schema Layer. Our findings demonstrate that visual and text modalities collaborate closely during schema construction. The schema stage represents a pivotal point where new and existing information intersect—a critical juncture recognized across various sensemaking models [11, 49, 107], beyond the Pirolli & Card's framework we adopted. Because schemas evolve incrementally and multiple alternative schemas often coexist, managing flexible data representation across the spectrum from implicit to explicit is essential [11, 136]. Visual and text modalities served as effective representations for related work schemas because they align with the domain's primary axes—metadata and textual argumentation—while offering distinct representational affordances.

Therefore, synchronization between visual and text modalities within the schema layer should be automated to respect the intuitive, evolving, and multifaceted nature of schemas. This precisely defines CrossLit's position in our design space: providing automated synchronization between visual and text—a gap that previous literature review tools did not address. Building on this foundation, we identify two directions for future research.

First, future systems could support creation and comparison of multiple alternative schemas, which our one-to-one mapping currently constrains. Relevant foundations include research on managing slightly different versions of knowledge artifacts—such as computational notebooks [22], trigger-action programs [134], and generative AI outputs [46]—as well as ontology merging and alignment techniques from knowledge engineering, which address comparing and integrating similar knowledge structures [34, 97].

Second, future systems could preserve carefully crafted user refinements while maintaining automated synchronization, a challenging balance as these goals often conflict. One approach is to explicitly distinguish automatic update zones from protected areas and visualize which portions have been transformed, promoting user ownership and control over AI-mediated changes [58, 77, 104]. Future work could also explore synchronization that occurs not constantly but at appropriate moments and in appropriate amounts, for example by detecting and classifying user intent [78, 137].

6.1.2 Cross-Layer Synchronization. Our findings also revealed participants frequently switching between visual and text modalities to leverage their unique strengths. This was particularly pronounced for tasks other than schema construction. Unlike the schema layer where multiple data elements are simultaneously intertwined, data across different layers maintain indirect complementary relationships that influence each other but maintain independent value. This suggests that various literature review tools need not—and perhaps should not—be unified into a single system.

Numerous literature review supporting systems specialize in specific layers or modality aspects, as discussed in Section 2. Since dedicated tools address practical concerns about user characteristics, roles, and tasks (e.g., processing huge citation networks, compiling final manuscript PDFs), literature review workflows typically span multiple tools. Creating seamless integration between tools is a common challenge across adjacent domains including data analysis [22, 50, 129] and design [41].

Future work can draw from existing technical attempts to integrate fragmented knowledge across tools, which operate at multiple

levels: collection and curation through web extensions [19, 75, 86, 87], browser-level page bundling [20, 65], or even desktop-level application bundling [59, 60]. With advancing generative AI capabilities, research on malleable system experiences that transcend fixed interface boundaries is also gaining traction [17, 95].

Looking toward a future where tool integration is realized and interface boundaries dissolve, what becomes critical is conceptually defining the unit components of work and their operations [16, 70]. We suggest that the sensemaking stage outcomes discovered in our user study serve as key conceptual elements mediating connections between literature review tools. For example, related papers discovered in Connected Papers may become triaged papers, which are read through Mendeley with interpretation notes. These interconnect in CrossLit to transform into schemas, which import into Overleaf[9] for manuscript reflection. Creating standardized specifications for each outcome represents a technical challenge for future work.

6.2 Preserving Human Agency through Visual Modality

Incorporating AI into literature review allows AI to handle simple filtering and summarization, enabling humans to concentrate more resources on high-level cognitive work such as making connections, identifying patterns, and constructing arguments [43, 69]. However, extensive research in human–AI interaction emphasizes that incorporating AI into knowledge work requires careful consideration to avoid compromising human agency [130], control [112], ownership [105], and integrity [35, 63].

Our study shows that rather than AI intervening in human knowledge processes, simply enabling seamless transitions to visual modality can help people focus on high-value cognitive work. While careful argument construction in literature review was primarily supported through textual formats, the visual space became a site for provisional sensemaking, where organizational decisions could be tested without the linear commitment that text demands.

This insight extends to other domains such as qualitative analysis. Researchers similarly move between close examination of individual data and identification of broader patterns, each requiring different cognitive modes. The ability to maintain provisional groupings, such as themes or codes that might shift or overlap, while seeing how they would translate into findings could fundamentally change how researchers approach early-stage analysis.

6.3 Limitations and Future Work

6.3.1 Lack of Comparable Baseline Systems. Our study did not include a baseline because no existing system provides an integrated, synchronized environment comparable to CrossLit. Prior work on novel multimodal and sensemaking tools has similarly adopted single-condition, exploratory evaluations when comparable baselines are unavailable or when introducing a new interaction paradigm [3, 16, 48, 81]. In line with this tradition, our goal was to observe authentic sensemaking behaviors and understand how researchers coordinate visual and textual modalities during literature review, rather than to measure superiority over any specific alternative.

[9]https://www.overleaf.com

Nevertheless, comparative evaluation remains an important direction. As multimodal literature-review tools mature, CrossLit can serve as a reference point, similar to how formative systems in prior work later became baselines for follow-up designs. Future work should explore controlled comparisons against purpose-built variants, such as alternative synchronization designs or modality-only configurations, to more precisely characterize the benefits and tradeoffs of cross-modal integration.

6.3.2 Session Duration and Advanced Review Phases. Because our sessions were limited to 45 minutes, questions about longer and more advanced review phases remain open. From the perspective of Bloom's Taxonomy [74], later stages of literature review involve evaluation—examining methodological rigor, result validity, and argumentative persuasiveness to determine what to accept or reject [47]. These judgments extend beyond organization to require scholarly discernment and domain expertise.

Future work should investigate these later phases through longitudinal studies that capture complete review cycles, from initial exploration to final evaluation. In these extended contexts, collaborative settings could become important considerations for observation. Additionally, the three semantic layers we identified suggest opportunities for more adaptive synchronization strategies—for example, tighter coupling during exploration but greater independence once distinct organizational needs arise. Collectively, these directions outline a research agenda for designing tools that support the literature review process end to end, from short-term exploration to long-term evaluation and collaboration.

6.3.3 Visualization Scalability. Our system was designed to accommodate approximately 100 papers, reflecting the typical reference count in HCI publications, and study participants imported up to 41 papers during their sessions. However, real-world literature reviews often require managing hundreds of papers, including candidates not incorporated into the final manuscript. In such cases, the visualization would need additional levels of abstraction—for instance, semantic zooming with finer hierarchical granularity [117], allowing researchers to navigate between overview and detail while maintaining schematic coherence. Furthermore, displaying metadata visualizations alongside schematic representations can quickly produce visual clutter, potentially requiring techniques to select which citation links to show [113]. According to the future direction envisioned in our design space, however, when collections reach such scales, these challenges belong to the metadata layer and would be better addressed through seamless transitions to dedicated tools for large-scale bibliometric analysis.

7 Conclusion

We introduce CrossLit, a system that integrates visual and text modalities for literature review sensemaking. Our observational study with 16 researchers shows that visual modality enabled provisional exploration and intuitive grouping in early stages, while text modality supported argumentation and narrative refinement in later stages. Connecting the two modalities enabled fluid transitions between literature review stages, supporting more fine-grained iterations than traditional workflows. A notable finding is the importance of the schema building stage, where visual and text modalities

work in tight coordination. Future research should explore versatile synchronization within the schema layer and seamless cross-layer integration across specialized tools, toward unified support for the full spectrum of complex sensemaking tasks.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2023R1A2C200520911), the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)], Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) [RS-2024-00436124, Human Resource Development Program for Industrial Innovation(Global)], and by the SNU-Global Excellence Research Center establishment project. The ICT at Seoul National University provided research facilities for this study.

References

[1] Ai2. 2025. Introducing Ai2 Paper Finder. https://allenai.org/blog/paper-finder. Accessed: 2025-09-12.

[2] Hanadi Alfraidi, Won-Sook Lee, and David Sankoff. 2015. Literature visualization and similarity measurement based on citation relations. In *2015 19th International Conference on Information Visualisation*. IEEE, 217–222.

[3] Ian Arawjo, Chelse Swoopes, Priyan Vaithilingam, Martin Wattenberg, and Elena L Glassman. 2024. Chainforge: A visual toolkit for prompt engineering and llm hypothesis testing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

[4] Roger Bakeman and John M Gottman. 1997. *Observing interaction: An introduction to sequential analysis*. Cambridge university press.

[5] Roger Bakeman, Duncan McArthur, and Vicenç Quera. 1996. Detecting group differences in sequential association using sampled permutations: Log odds, kappa, and phi compared. *Behavior Research Methods, Instruments, & Computers* 28, 3 (1996), 446–457.

[6] Fabian Beck. 2024. PUREsuggest: citation-based literature search and visual exploration with keyword-controlled rankings. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[7] Fabian Beck, Sebastian Koch, and Daniel Weiskopf. 2015. Visual analysis and dissemination of scientific literature collections with SurVis. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 180–189.

[8] Ralf Bender and Stefan Lange. 2001. Adjusting for multiple testing—when and how? *Journal of clinical epidemiology* 54, 4 (2001), 343–349.

[9] Matthew Berger, Katherine McDonough, and Lee M Seversky. 2016. cite2vec: Citation-driven document exploration via word embeddings. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 691–700.

[10] Peter Bergström and Darren C Atkinson. 2009. Augmenting the exploration of digital libraries with web-based visualizations. In *2009 fourth international conference on digital information management*. IEEE, 1–7.

[11] Charles Berret and Tamara Munzner. 2024. Iceberg Sensemaking: A Process Model for Critical Data Analysis. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[12] Chandra Bhagavatula, Sergey Feldman, Russell Power, and Waleed Ammar. 2018. Content-Based Citation Recommendation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Marilyn Walker, Heng Ji, and Amanda Stent (Eds.). Association for Computational Linguistics, New Orleans, Louisiana, 238–251. https://doi.org/10.18653/v1/N18-1022

[13] David N Boote and Penny Beile. 2005. Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational researcher* 34, 6 (2005), 3–15.

[14] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[15] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[16] Yining Cao, Yiyi Huang, Anh Truong, Hijung Valentina Shin, and Haijun Xia. 2025. Compositional Structures as Substrates for Human-AI Co-creation Environment: A Design Approach and A Case Study. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–25.

[17] Yining Cao, Peiling Jiang, and Haijun Xia. 2025. Generative and Malleable User Interfaces with Generative and Evolving Task-Driven Data Model. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[18] Joel Chan, Joseph Chee Chang, Tom Hope, Dafna Shahaf, and Aniket Kittur. 2018. SOLVENT: A Mixed Initiative System for Finding Analogies between Research Papers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 31 (Nov. 2018), 21 pages. https://doi.org/10.1145/3274300

[19] Joseph Chee Chang, Nathan Hahn, and Aniket Kittur. 2020. Mesh: Scaffolding comparison tables for online decision making. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 391–405.

[20] Joseph Chee Chang, Yongsung Kim, Victor Miller, Michael Xieyang Liu, Brad A Myers, and Aniket Kittur. 2021. Tabs. do: Task-centric browser tab management. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 663–676.

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

[22] Souti Chattopadhyay, Ishita Prasad, Austin Z Henley, Anita Sarma, and Titus Barik. 2020. What's wrong with computational notebooks? Pain points, needs, and design opportunities. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.

[23] Duen Horng Chau, Aniket Kittur, Jason I Hong, and Christos Faloutsos. 2011. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 167–176.

[24] Chaomei Chen. 1999. Visualising semantic spaces and author co-citation networks in digital libraries. *Information processing & management* 35, 3 (1999), 401–420.

[25] Chaomei Chen, Fidelia Ibekwe-SanJuan, and Jianhua Hou. 2010. The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for information Science and Technology* 61, 7 (2010), 1386–1409.

[26] Uchendu Eugene Chigbu, Sulaiman Olusegun Atiku, and Cherley C. Du Plessis. 2023. The Science of Literature Reviews: Searching, Identifying, Selecting, and Synthesising. *Publ.* 11 (2023), 2. https://api.semanticscholar.org/CorpusID:255630635

[27] Kiroong Choe, Seokweon Jung, Seokhyeon Park, Hwajung Hong, and Jinwook Seo. 2021. Papers101: Supporting the discovery process in the literature review workflow for novice researchers. In *2021 IEEE 14th Pacific Visualization Symposium (PacificVis)*. IEEE, 176–180.

[28] Jaegul Choo, Hannah Kim, Edward Clarkson, Zhicheng Liu, Changhyun Lee, Fuxin Li, Hanseung Lee, Ramakrishnan Kannan, Charles D Stolper, John Stasko, et al. 2018. VisIRR: A visual analytics system for information retrieval and recommendation for large-scale document data. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 12, 1 (2018), 1–20.

[29] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

[30] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. routledge.

[31] Antonina Dattolo and Marco Corbatto. 2019. VisualBib: a novel Web app for supporting researchers in the creation, visualization and sharing of bibliographies. *Knowledge-Based Systems* 182 (2019), 104860.

[32] Antonina Dattolo and Marco Corbatto. 2022. Assisting researchers in bibliographic tasks: A new usable, real-time tool for analyzing bibliographies. *Journal of the Association for Information Science and Technology* 73, 6 (2022), 757–776.

[33] Ao Dong, Wei Zeng, Xi Chen, and Zhanglin Cheng. 2019. VIStory: Interactive storyboard for exploring visual information in scientific publications. In *Proceedings of the 12th International Symposium on Visual Information Communication and Interaction*. 1–8.

[34] Dimitrios Doumanas, Georgios Bouchouras, Andreas Soularidis, Konstantinos Kotis, and George Vouros. 2024. From human-to LLM-centered collaborative ontology engineering. *Applied Ontology* 19, 4 (2024), 334–367.

[35] Fiona Draxler, Anna Werner, Florian Lehmann, Matthias Hoppe, Albrecht Schmidt, Daniel Buschek, and Robin Welsch. 2024. The AI ghostwriter effect: When users do not perceive ownership of AI-generated text but self-declare as authors. *ACM Transactions on Computer-Human Interaction* 31, 2 (2024), 1–40.

[36] Cody Dunne, Ben Shneiderman, Robert Gove, Judith Klavans, and Bonnie Dorr. 2012. Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. *Journal of the American Society for Information Science and Technology* 63, 12 (2012), 2351–2369.

[37] Niklas Elmqvist and Philippas Tsigas. 2007. CiteWiz: a tool for the visualization of scientific citation networks. *Information Visualization* 6, 3 (2007), 215–232.

[38] K Anders Ericsson and Herbert A Simon. 1993. Protocol Analysis: Verbal Reports as Data. (1993).

[39] Xinrui Fang, Anran Xu, Sylvain Malacria, and Koji Yatani. 2025. Exploring Practices, Challenges, and Design Implications for Citation Foraging, Management, and Synthesis. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[40] Paolo Federico, Florian Heimerl, Steffen Koch, and Silvia Miksch. 2016. A survey on visual approaches for analyzing scientific literature and patents. *IEEE transactions on visualization and computer graphics* 23, 9 (2016), 2179–2198.

[41] KJ Kevin Feng, Tony W Li, and Amy X Zhang. 2023. Understanding collaborative practices and tools of professional UX practitioners in software organizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[42] Raymond Fok, Joseph Chee Chang, Tal August, Amy X Zhang, and Daniel S Weld. 2023. Qlarify: Bridging scholarly abstracts and papers with recursively expandable summaries. *arXiv preprint arXiv:2310.07581* 6, 3 (2023).

[43] Raymond Fok, Joseph Chee Chang, Marissa Radensky, Pao Siangliulue, Jonathan Bragg, Amy X. Zhang, and Daniel S. Weld. 2025. Facets, Taxonomies, and Syntheses: Navigating Structured Representations in LLM-Assisted Literature Review. arXiv:2504.18496 [cs.HC] https://arxiv.org/abs/2504.18496

[44] Raymond Fok, Hita Kambhamettu, Luca Soldaini, Jonathan Bragg, Kyle Lo, Marti Hearst, Andrew Head, and Daniel S Weld. 2023. Scim: Intelligent Skimming Support for Scientific Papers. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 476–490. https://doi.org/10.1145/3581641.3584034

[45] Yubin Ge, Ly Dinh, Xiaofeng Liu, Jinsong Su, Ziyao Lu, Ante Wang, and Jana Diesner. 2021. BACO: A Background Knowledge- and Content-Based Framework for Citing Sentence Generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 1466–1478. https://doi.org/10.18653/v1/2021.acl-long.116

[46] Katy Ilonka Gero, Chelse Swoopes, Ziwei Gu, Jonathan K Kummerfeld, and Elena L Glassman. 2024. Supporting sensemaking of large language model outputs at scale. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.

[47] Darcy Haag Granello. 2001. Promoting cognitive complexity in graduate written work: Using Bloom's taxonomy as a pedagogical tool to improve literature reviews. *Counselor Education and Supervision* 40, 4 (2001), 292–307.

[48] Saul Greenberg and Bill Buxton. 2008. Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 111–120.

[49] Garrett Grolemund and Hadley Wickham. 2014. A cognitive interpretation of data analysis. *International Statistical Review* 82, 2 (2014), 184–204.

[50] Ken Gu, Ruoxi Shang, Tim Althoff, Chenglong Wang, and Steven M Drucker. 2024. How do analysts understand and verify ai-assisted data analyses?. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–22.

[51] Nianlong Gu, Yingqiang Gao, and Richard H. R. Hahnloser. 2022. Local Citation Recommendation with Hierarchical-Attention Text Encoder and SciBERT-Based Reranking. In *Advances in Information Retrieval*, Matthias Hagen, Suzan Verberne, Craig Macdonald, Christin Seifert, Krisztian Balog, Kjetil Nørvåg, and Vinay Setty (Eds.). Springer International Publishing, Cham, 274–288.

[52] Nianlong Gu and Richard H.R. Hahnloser. 2023. SciLit: A Platform for Joint Scientific Literature Discovery, Summarization and Citation Generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, Danushka Bollegala, Ruihong Huang, and Alan Ritter (Eds.). Association for Computational Linguistics, Toronto, Canada, 235–246. https://doi.org/10.18653/v1/2023.acl-demo.22

[53] Han L Han, Junhang Yu, Raphael Bournet, Alexandre Ciorascu, Wendy E Mackay, and Michel Beaudouin-Lafon. 2022. Passages: Interacting with text across documents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[54] 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*. 1–18.

[55] Florian Heimerl, Qi Han, Steffen Koch, and Thomas Ertl. 2015. CiteRivers: Visual analytics of citation patterns. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 190–199.

[56] Allyson Holbrook. 2007. 'Levels' of success in the use of the literature in a doctorate. *South African Journal of Higher Education* 21, 8 (2007), 1020–1041.

[57] Thomas F. Homer-Dixon and Roger S. Karapin. 1989. Graphical Argument Analysis: A New Approach to Understanding Arguments, Applied to a Debate about the Window of Vulnerability. *International Studies Quarterly* 33, 4 (1989), 389–410. http://www.jstor.org/stable/2600519

[58] Md Naimul Hoque, Tasfia Mashiat, Bhavya Ghai, Cecilia D Shelton, Fanny Chevalier, Kari Kraus, and Niklas Elmqvist. 2024. The HaLLMark effect: Supporting provenance and transparent use of large language models in writing with interactive visualization. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–15.

[59] Donghan Hu and Sang Won Lee. 2020. ScreenTrack: Using a visual history of a computer screen to retrieve documents and Web pages. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[60] Donghan Hu and Sang Won Lee. 2022. Scrapbook: Screenshot-based bookmarks for effective digital resource curation across applications. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[61] Lu Huang, Yijie Cai, Erdong Zhao, Shengting Zhang, Yue Shu, and Jiao Fan. 2022. Measuring the interdisciplinarity of Information and Library Science interactions using citation analysis and semantic analysis. *Scientometrics* 127, 11 (01 Nov 2022), 6733–6761. https://doi.org/10.1007/s11192-022-04401-x

[62] Run Huang, Anna Katherine Zhao, Zeinabsadat Saghi, Sadra Sabouri, and Souti Chattopadhyay. 2025. Beyond the Page: Enriching Academic Paper Reading with Social Media Discussions. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–25.

[63] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. 2023. Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–15.

[64] Peiling Jiang, Jude Rayan, Steven P Dow, and Haijun Xia. 2023. Graphologue: Exploring large language model responses with interactive diagrams. In *Proceedings of the 36th annual ACM symposium on user interface software and technology*. 1–20.

[65] Peiling Jiang and Haijun Xia. 2025. Orca: Browsing at Scale Through User-Driven and AI-Facilitated Orchestration Across Malleable Webpages. *arXiv preprint arXiv:2505.22831* (2025).

[66] Hyeonsu Kang, Joseph Chee Chang, Yongsung Kim, and Aniket Kittur. 2022. Threddy: An interactive system for personalized thread-based exploration and organization of scientific literature. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[67] Hyeonsu B Kang, Rafal Kocielnik, Andrew Head, Jiangjiang Yang, Matt Latzke, Aniket Kittur, Daniel S Weld, Doug Downey, and Jonathan Bragg. 2022. From Who You Know to What You Read: Augmenting Scientific Recommendations with Implicit Social Networks. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 302, 23 pages. https://doi.org/10.1145/3491102.3517470

[68] Hyeonsu B Kang, Nouran Soliman, Matt Latzke, Joseph Chee Chang, and Jonathan Bragg. 2023. Comlittee: Literature discovery with personal elected author committees. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[69] Hyeonsu B Kang, Tongshuang Wu, Joseph Chee Chang, and Aniket Kittur. 2023. Synergi: A mixed-initiative system for scholarly synthesis and sensemaking. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–19.

[70] Sora Kanosue, Xiaorui Liu, Parker Ziegler, Eric Rawn, and Sarah E Chasins. 2025. HiLT: A Library for Generating Human-in-the-Loop Data Transformation GUIs. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[71] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. 2023. Metaphorian: Leveraging large language models to support extended metaphor creation for science writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 115–135.

[72] Tae Soo Kim, Matt Latzke, Jonathan Bragg, Amy X Zhang, and Joseph Chee Chang. 2023. Papeos: Augmenting research papers with talk videos. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–19.

[73] Simon Knight, Antonette Shibani, Sophie Abel, Andrew Gibson, and Philippa Ryan. 2020. AcaWriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research* (2020).

[74] David R Krathwohl. 2002. A revision of Bloom's taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218.

[75] Andrew Kuznetsov, Joseph Chee Chang, Nathan Hahn, Napol Rachatasumrit, Bradley Breneisen, Julina Coupland, and Aniket Kittur. 2022. Fuse: In-situ sensemaking support in the browser. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[76] Becky SC Kwan. 2008. The nexus of reading, writing and researching in the doctoral undertaking of humanities and social sciences: Implications for literature reviewing. *English for Specific Purposes* 27, 1 (2008), 42–56.

[77] Vivian Lai, Samuel Carton, Rajat Bhatnagar, Q Vera Liao, Yunfeng Zhang, and Chenhao Tan. 2022. Human-ai collaboration via conditional delegation: A case study of content moderation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[78] Michelle S Lam, Omar Shaikh, Hallie Xu, Alice Guo, Diyi Yang, Jeffrey Heer, James A Landay, and Michael S Bernstein. 2025. Just-In-Time Objectives: A General Approach for Specialized AI Interactions. *arXiv preprint arXiv:2510.14591* (2025).

[79] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[80] Shahid Latif and Fabian Beck. 2018. VIS Author Profiles: Interactive descriptions of publication records combining text and visualization. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 152–161.

[81] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation strategies for HCI toolkit research. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–17.

[82] Bongshin Lee, Mary Czerwinski, George Robertson, and Benjamin B Bederson. 2005. Understanding research trends in conferences using PaperLens. In *CHI'05 extended abstracts on Human factors in computing systems*. 1969–1972.

[83] Yoonjoo Lee, Hyeonsu B Kang, Matt Latzke, Juho Kim, Jonathan Bragg, Joseph Chee Chang, and Pao Siangliulue. 2024. Paperweaver: Enriching topical paper alerts by contextualizing recommended papers with user-collected papers. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.

[84] Kevin Li, Haoyang Yang, Evan Montoya, Anish Upadhayay, Zhiyan Zhou, Jon Saad-Falcon, and Duen Horng Chau. 2022. Visual exploration of literature with Argo Scholar. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4912–4916.

[85] Ming Liu, Rafael A Calvo, and Vasile Rus. 2012. G-Asks: An intelligent automatic question generation system for academic writing support. *Dialogue & Discourse* 3, 2 (2012), 101–124.

[86] Michael Xieyang Liu, Jane Hsieh, Nathan Hahn, Angelina Zhou, Emily Deng, Shaun Burley, Cynthia Taylor, Aniket Kittur, and Brad A Myers. 2019. Unakite: Scaffolding developers' decision-making using the web. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 67–80.

[87] Michael Xieyang Liu, Andrew Kuznetsov, Yongsung Kim, Joseph Chee Chang, Aniket Kittur, and Brad A Myers. 2022. Wigglite: Low-cost information collection and triage. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[88] Kyle Lo, Joseph Chee Chang, Andrew Head, Jonathan Bragg, Amy X Zhang, Cassidy Trier, Chloe Anastasiades, Tal August, Russell Authur, Danielle Bragg, et al. 2023. The semantic reader project: Augmenting scholarly documents through ai-powered interactive reading interfaces. *arXiv preprint arXiv:2303.14334* (2023).

[89] Donghyeok Ma, Hanbee Jang, Joon Hyub Lee, and Seok-Hyung Bae. 2025. Garden of Papers: Finding, Reading, and Organizing Research Papers in a Visual, Integrated, and Flexible Workspace. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[90] Damien Masson, Zixin Zhao, and Fanny Chevalier. 2025. Visual Story-Writing: Writing by Manipulating Visual Representations of Stories. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[91] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2012. Citeology: visualizing paper genealogy. In *CHI'12 extended abstracts on human factors in computing systems*. 181–190.

[92] Richard E Mayer. 2002. Multimedia learning. In *Psychology of learning and motivation*. Vol. 41. Elsevier, 85–139.

[93] Richard E Mayer and Roxana Moreno. 2003. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist* 38, 1 (2003), 43–52.

[94] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–23.

[95] Bryan Min and Haijun Xia. 2025. Meridian: A Design Framework for Malleable Overview-Detail Interfaces. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[96] Sheshera Mysore, Mahmood Jasim, Haoru Song, Sarah Akbar, Andre Kenneth Chase Randall, and Narges Mahyar. 2023. How data scientists review the scholarly literature. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 137–152.

[97] Natalya Fridman Noy, Mark A Musen, et al. 2000. Algorithm and tool for automated ontology merging and alignment. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-00). Available as SMI technical report SMI-2000-0831*, Vol. 115. sn.

[98] Lace M. Padilla, Sarah H. Creem-Regehr, Mary Hegarty, and Jeanine K. Stefanucci. 2018. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications* 3, 1 (11 Jul 2018), 29. https://doi.org/10.1186/s41235-018-0120-9

[99] Vishakh Padmakumar, Joseph Chee Chang, Kyle Lo, Doug Downey, and Aakanksha Naik. 2025. Intent-Aware Schema Generation And Refinement For Literature Review Tables. *Findings of the Association for Computational Linguistics: EMNLP 2025* (2025), 23450–23472.

[100] Allan Paivio. 1990. *Mental representations: A dual coding approach*. Oxford university press.

[101] Srishti Palani, Aakanksha Naik, Doug Downey, Amy X Zhang, Jonathan Bragg, and Joseph Chee Chang. 2023. Relatedly: Scaffolding literature reviews with existing related work sections. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[102] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.

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

[104] Mohi Reza, Nathan M Laundry, Ilya Musabirov, Peter Dushniku, Zhi Yuan "Michael" Yu, Kashish Mittal, Tovi Grossman, Michael Liut, Anastasia Kuzminykh, and Joseph Jay Williams. 2024. Abscribe: Rapid exploration & organization of multiple writing variations in human-ai co-writing tasks using large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

[105] Mohi Reza, Jeb Thomas-Mitchell, Peter Dushniku, Nathan Laundry, Joseph Jay Williams, and Anastasia Kuzminykh. 2025. Co-writing with ai, on human terms: Aligning research with user demands across the writing process. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–37.

[106] Hugo Romat, Caroline Appert, Benjamin Bach, Nathalie Henry-Riche, and Emmanuel Pietriga. 2018. Animated edge textures in node-link diagrams: A design space and initial evaluation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[107] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey Ellis, and Daniel A Keim. 2014. Knowledge generation model for visual analytics. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 1604–1613.

[108] Mario Salinas, Daniela Giorgi, Federico Ponchio, and Paolo Cignoni. 2020. ReviewerNet: A visualization platform for the selection of academic reviewers. *Computers & Graphics* 89 (2020), 77–87.

[109] Dafna Shahaf, Carlos Guestrin, and Eric Horvitz. 2012. Metro maps of science. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1122–1130.

[110] Zeqian Shen, Michael Ogawa, Soon Tee Teoh, and Kwan-Liu Ma. 2006. BiblioViz: a system for visualizing bibliography information. In *Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation-Volume 60*. Citeseer, 93–102.

[111] Chen Shi, Haoxuan Wang, Binjie Chen, Yuhua Liu, and Zhiguang Zhou. 2019. Visual analysis of citation context-based article influence ranking. *IEEE Access* 7 (2019), 113853–113866.

[112] Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504.

[113] Ben Shneiderman and Aleks Aris. 2006. Network visualization by semantic substrates. *IEEE transactions on visualization and computer graphics* 12, 5 (2006), 733–740.

[114] Amanpreet Singh, Joseph Chee Chang, Chloe Anastasiades, Dany Haddad, Aakanksha Naik, Amber Tanaka, Angele Zamarron, Cecile Nguyen, Jena D. Hwang, Jason Dunkleberger, Matt Latzke, Smita Rao, Jaron Lochner, Rob Evans, Rodney Kinney, Daniel S. Weld, Doug Downey, and Sergey Feldman. 2025. Ai2 Scholar QA: Organized Literature Synthesis with Attribution. https://api.semanticscholar.org/CorpusID:277786810

[115] Ayah Soufan, Ian Ruthven, and Leif Azzopardi. 2024. Uncharted territory: understanding exploratory search behaviours in literature reviews. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. 23–33.

[116] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured generation and exploration of design space with large language models for human-ai co-creation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–26.

[117] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling multilevel exploration and sensemaking with large language models. In *Proceedings of the 36th annual ACM symposium on user interface software and technology*. 1–18.

[118] Nicole Sultanum, Christine Murad, and Daniel Wigdor. 2020. Understanding and supporting academic literature review workflows with litsense. In *Proceedings of the 2020 International Conference on Advanced Visual Interfaces*. 1–5.

[119] 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*. 2927–2936.

[120] Min Tian, Guozheng Li, and Xiaoru Yuan. 2023. LitVis: a visual analytics approach for managing and exploring literature. *Journal of Visualization* 26, 6 (2023), 1445–1458.

[121] Yong Wang, Conglei Shi, Liangyue Li, Hanghang Tong, and Huamin Qu. 2018. Visualizing research impact through citation data. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 1 (2018), 1–24.

[122] Jason Wilkins, Jaakko Järvi, Ajit Jain, Gaurav Kejriwal, Andruid Kerne, and Vijay Gumudavelly. 2015. EvolutionWorks: Towards improved visualization of citation networks. In *Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015,*

Proceedings, Part IV 15. Springer, 213–230.

[123] John L. Worrall and Ellen G. Cohn. 2023. Citation Data and Analysis: Limitations and Shortcomings. *Journal of Contemporary Criminal Justice* 39, 3 (2023), 327–340. https://doi.org/10.1177/10439862231170972 arXiv:https://doi.org/10.1177/10439862231170972

[124] Liwenhan Xie, Chengbo Zheng, Haijun Xia, Huamin Qu, and Chen Zhu-Tian. 2024. Waitgpt: Monitoring and steering conversational llm agent in data analysis with on-the-fly code visualization. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[125] Haoyang Yang, Elliott H. Faa, Weijian Liu, Shunan Guo, Duen Horng Chau, and Yalong Yang. 2025. LitForager: Exploring Multimodal Literature Foraging Strategies in Immersive Sensemaking. arXiv:2508.15043 [cs.HC] https://arxiv.org/abs/2508.15043

[126] Runlong Ye, Patrick Lee, Matthew Varona, Oliver Huang, and Carolina Nobre. 2025. ScholarMate: A Mixed-Initiative Tool for Qualitative Knowledge Work and Information Sensemaking. In *Adjunct Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '25 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 7, 7 pages. https://doi.org/10.1145/3707640.3731913

[127] Taerin Yoon, Hyunwoo Han, Hyoji Ha, Juwon Hong, and Kyungwon Lee. 2020. A conference paper exploring system based on citing motivation and topic. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 231–235.

[128] Shuo Yu, Yingbo Wang, Ruolin Li, Guchun Liu, Yanming Shen, Shaoxiong Ji, Bowen Li, Fengling Han, Xiuzhen Zhang, and Feng Xia. 2025. Graph2text or Graph2token: A Perspective of Large Language Models for Graph Learning. arXiv:2501.01124 [cs.LG] https://arxiv.org/abs/2501.01124

[129] Amy X Zhang, Michael Muller, and Dakuo Wang. 2020. How do data science workers collaborate? roles, workflows, and tools. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–23.

[130] Shuning Zhang, Hui Wang, and Xin Yi. 2025. Exploring collaboration patterns and strategies in human-ai co-creation through the lens of agency: A scoping review of the top-tier hci literature. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–43.

[131] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. 2023. Visar: A human-ai argumentative writing assistant with visual programming and rapid draft prototyping. In *Proceedings of the 36th annual ACM symposium on user interface software and technology*. 1–30.

[132] Jian Zhao, Christopher Collins, Fanny Chevalier, and Ravin Balakrishnan. 2013. Interactive exploration of implicit and explicit relations in faceted datasets. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2080–2089.

[133] Jian Zhao, Michael Glueck, Simon Breslav, Fanny Chevalier, and Azam Khan. 2016. Annotation graphs: A graph-based visualization for meta-analysis of data based on user-authored annotations. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 261–270.

[134] Valerie Zhao, Lefan Zhang, Bo Wang, Michael L Littman, Shan Lu, and Blase Ur. 2021. Understanding trigger-action programs through novel visualizations of program differences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[135] Chengbo Zheng, Yuanhao Zhang, Zeyu Huang, Chuhan Shi, Minrui Xu, and Xiaojuan Ma. 2024. DiscipLink: Unfolding Interdisciplinary Information Seeking Process via Human-AI Co-Exploration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 91, 20 pages. https://doi.org/10.1145/3654777.3676366

[136] Siyi Zhu, Robert Haisfield, Brendan Langen, and Joel Chan. 2024. Patterns of Hypertext-Augmented Sensemaking. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–17.

[137] Chen Zhu-Tian and Haijun Xia. 2022. CrossData: Leveraging text-data connections for authoring data documents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

A Prompts

A.1 Visual-to-Text Rephrasing Module

```
**System Prompt**:
"You are an expert academic writer. Create a coherent text
paragraph for literature reviews.

RULES:
1. MUST preserve ALL paper citations in format /bibtexId/
2. Write exactly one sentence per paper citation
3. Each sentence MUST include its corresponding /bibtexId/
   reference
4. Be concise and focused on the given context
5. Maintain academic tone"

**User Prompt**:
"Rewrite this text block coherently:
Current text: {current_text}
Hierarchical location: {hierarchy_context}
Visual note: {current_note}
Linked papers: {paper_list}
Write exactly {num_papers} sentence(s). Each sentence MUST
include the /bibtexId/ reference for its paper."

**Output**: Returns `rephrased_text`, `sentence_count`, and
`citations_preserved`:
rephrased_text: The rephrased text incorporating visual note
  as well as previous contexts"
sentence_count: Number of sentences in the output"
citations_preserved: List of paper bibtexIds preserved in
  the text"
**Retry**: Up to 3 attempts, adds missing citations manually
on final attempt
```

A.2 Text-to-Visual Condensing Module

```
**System Prompt**:
"You are an expert at writing concise notes for academic
research.

RULES:
1. Create extremely concise notes (3-10 words max)
2. Capture the core idea or concept
3. Use keywords rather than full sentences
4. Omit articles (a, an, the) when possible
5. Focus on the main topic or finding"

**User Prompt**:
"Convert this text to a concise visual note:
Current text: {current_text}
Current note: {current_note}
Hierarchical location: {hierarchy_context}
Linked papers: {paper_list}
Create a 3-10 word note capturing the essence."

**Output**: Returns `concise_note`
concise_note: Concise note generated from current_text as
  well as previous contexts (3-10 words)
**Fallback**: Uses first 5 words if API fails
```

A.3 Discovery Query Module

A.3.1 Natural Language Query Synthesis.

```
"You are an expert research assistant specializing in
academic literature discovery. Your task is to synthesize
search queries from the user's natural language input,
considering the context of seed papers and text editor
sections.

The user's query is often written assuming context from
their current work. For example:
- 'What can I add more?' should be interpreted in the
  context of the existing papers and sections
- 'What's the latest discussion about this paper?' should
  focus on recent developments related to the seed papers
- 'Related work in this area' should consider both the seed
  papers and the editor sections

You need to provide TWO outputs:

1. A comprehensive natural language query (for text snippet
   search):
   - Expand vague or contextual queries into specific,
     searchable terms
   - Include key concepts from seed papers when relevant
   - Consider the section titles from the text editor as
     topic areas of interest
   - Maintain the user's original intent while making it
     more explicit
   - Use academic terminology when appropriate
   - Keep the query focused and under 200 words
   - Write as a single paragraph of natural English

2. Four traditional search queries (for paper relevance
   search):
   - Each should be 2-5 keywords or phrases
   - Should capture different aspects of the user's intent
   - Use simpler terms suitable for keyword-based search
   - Focus on core concepts and variations"

**Context Provided**: Up to 5 seed papers (title + 300 char
abstract) and 3 text editor sections
**Output**: Returns `natural_language_query` and
`traditional_queries` (list of 4)
natural_language_query: Comprehensive natural language query
  for semantic search
traditional_queries: List of 4 traditional keyword search
  queries
```

A.3.2 Paper Relevance Interpretation.

```
**System Prompt**:
"You are a research assistant creating concise paper
interpretations.

RULES:
1. Write ONE sentence (maximum 15 words) explaining the
   paper's relevance
2. Focus on the key contribution or finding
3. Be direct and concise
4. If irrelevant to query, set is_relevant to false
5. DO NOT include any citations or references in your text"

**User Prompt**:
"Query: {query}
Paper Title: {title}
Abstract: {abstract}
Write ONE sentence (max 15 words) explaining this paper's
relevance to the query.
If irrelevant, set is_relevant to false."

**Output**: Returns `interpretation` and `is_relevant` flag
interpretation: One concise sentence explaining the paper's
  relevance (max 15 words)
is_relevant: Whether the paper is relevant to the query"
**Fallback**: If the number of relevant papers is less than
required, return paper titles with "(?)" prefix for the rest.
```

A.4 Convenience Feature for Experiment

A.4.1 Rewriting in Text Editor.

```
**System Prompt**:
"You are an expert academic writer tasked with rewriting
multiple text blocks into a single coherent paragraph.

CRITICAL RULES:
1. MUST preserve ALL paper citations in format /bibtexId/
2. Combine all blocks into ONE flowing paragraph
3. Follow the user's rewriting instruction carefully
4. Maintain academic tone and clarity
5. Ensure natural flow between merged content
6. All paper citations MUST be preserved exactly as they
   appear"

**User Prompt**:
```

```
"Rewrite these text blocks according to the instruction
below:
Selected blocks to merge: {combined_text}
User instruction: {rewrite_instruction}
Paper citations to preserve (MUST include all):
  {citation_list}
{paper_details_for_context}
Rewrite into ONE coherent paragraph that follows the
instruction while preserving ALL paper citations."

**Processing**: Extracts citations from HTML, provides paper
abstracts (max 500 chars) as context
**Output**: Returns `rewritten_text` and
`citations_preserved`
rewritten_text: The rewritten single paragraph combining all
  selected blocks
citations_preserved: List of paper bibtexIds preserved in
  the text
```

B Hybrid Query Pipeline and Evaluation

B.1 Context-aware Retrieval Pipeline

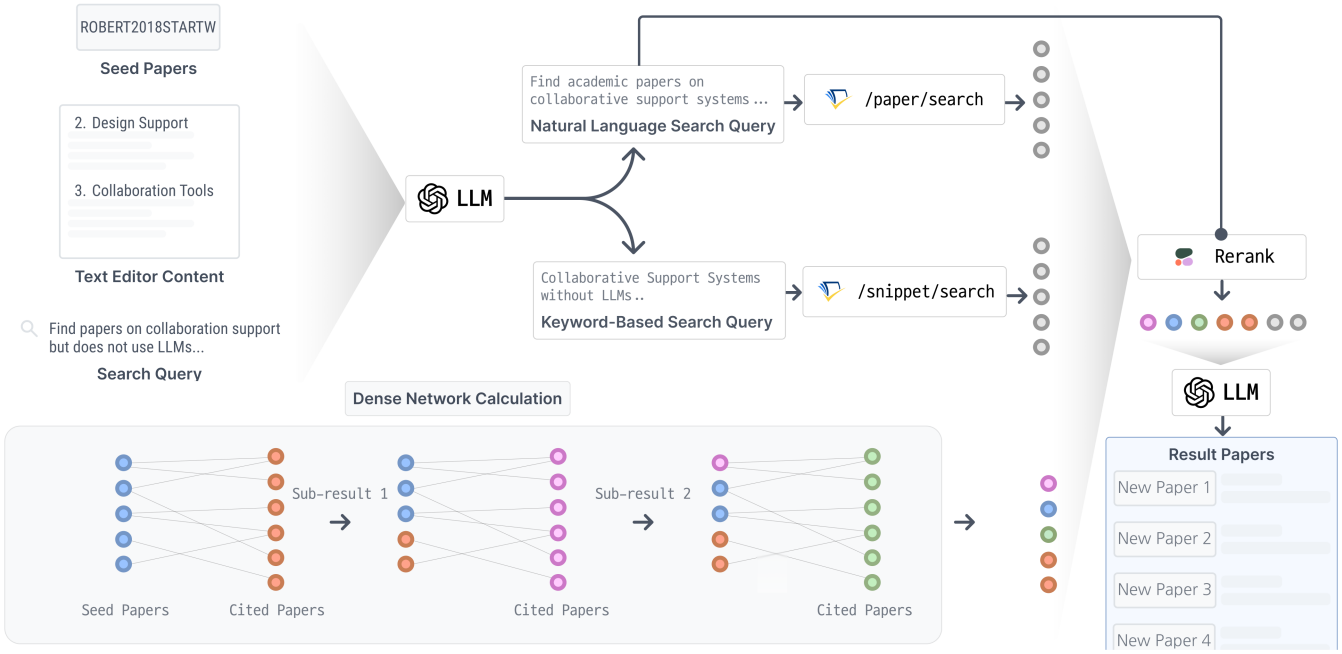

Figure 11: Technical pipeline for context-aware paper discovery in CrossLit. The pipeline combines three retrieval strategies. First, seed papers, text editor content, and search query are provided to an LLM, which generates both a natural language search query and a keyword-based search query. The natural language query is submitted to the Semantic Scholar `/paper/search` API to retrieve candidate papers, while the keyword-based query is used with `/snippet/search`. Second, the dense network calculation expands the seed set by iteratively collecting the most frequently co-cited papers (up to 50 per iteration, prioritizing higher citation counts in case of ties), repeated three times to form a network-based result set. Finally, results from all three sources are pooled and reranked using a semantic model and cross-checked by an LLM, which filters out irrelevant items. The output is a curated set of relevant papers, each presented with contextualized interpretations for the user.

B.2 Context-aware Retrieval Evaluation

To examine whether CrossLit's hybrid retrieval method integrates complementary structural (citation-based) and semantic (content-based) cues into a unified retrieval process (DG4), we conducted a quantitative evaluation comparing: (1) a citation-only baseline, (2) a semantic-only baseline, and (3) CrossLit's hybrid retrieval method.

B.2.1 Data Preparation. We constructed a ground-truth dataset for evaluating retrieval behavior by treating each subsection of the related work sections in recent HCI papers as an independent query unit. We first identified the top 10 most-cited papers from each of three major HCI venues (CHI, UIST, and CSCW) from the years 2024–2025, resulting in 30 papers. For each paper, we extracted all subsections in its related work section and retained only those that cited at least ten papers. For each retained subsection, we examined its citation map and selected seed papers by identifying the three most-cited referenced papers published within ±3 years of the source paper. Sections where a clear top three could not be determined (e.g., similar citation counts with no distinct top) were excluded. This resulted in 79 valid related-work subsections across 30 HCI papers.

Each subsection was represented by two textual expressions of author intent: the subsection heading (primary query) and, when necessary, the subsection's first sentence (used only when the heading returned no semantic matches). These queries were issued to Semantic Scholar's semantic search endpoints (`/graph/v1/snippet/search` and `/graph/v1/paper/search`). Candidate-paper metadata—including titles, abstracts, and citation relationships—was then collected via the Semantic Scholar Academic Graph API.

B.2.2 Retrieval Pipelines. We compared CrossLit's hybrid retrieval pipeline against two single-modality baselines—citation-only (structural) and semantic-only (content-based)—to assess whether it effectively integrates cues from both modalities.

Baseline 1: Citation-Only. Starting from the seed papers, we expanded the citation graph using a co-citation–based approach similar to prior work. At each hop, we selected the top 50 most co-cited papers (ties broken by global citation count), repeated for three iterations.

Baseline 2: Semantic-Only. We retrieved papers using Semantic Scholar's snippet search and paper relevance search endpoints (50 papers each), removed duplicated results, and returned the top 50 unique candidates.

CrossLit. We combined the union of the citation and semantic candidate pools and reranked all candidates using Cohere `rerank-v3.5`. For each paper, we computed both structural and semantic scores and produced a final ranking that integrates both modalities. Only the top-ranked papers were kept for evaluation (fixed at $k = 10$ per section for comparability).

B.2.3 Evaluation Metrics. We evaluated each retrieval method using two complementary metrics. **Co-Citation Density** measures structural coherence, computed as Density $= \#\{(i, j) : i \to j \vee j \to i\} \,/\, \lfloor n(n-1)/2 \rfloor$, where higher values indicate tighter citation-connected clusters. **Semantic Similarity** measures topical relevance, computed as Similarity $= \frac{1}{n} \sum_{i=1}^{n} \cos(q, p_i)$ between the query embedding q and retrieved paper embeddings p_i. The embeddings are calculated using OpenAI's `/v1/embeddings` API.

Method	Co-Citation Density	Semantic Similarity
Citation-only	0.0562 ± 0.0406	0.3565 ± 0.0724
Semantic-only	0.0085 ± 0.0180	0.4477 ± 0.1417
Hybrid (CrossLit)	0.0358 ± 0.0264	0.3818 ± 0.0653

Table 3: Average evaluation metrics across 79 sections.

B.2.4 Results. Paired t-tests showed clear differences among the three retrieval methods: citation-only produced the highest structural coherence (Citation-only > Hybrid > Semantic-only), while semantic-only produced the highest semantic relevance (Semantic-only > Hybrid > Citation-only), with all pairwise comparisons significant ($p < .05$). The hybrid method consistently fell between the two baselines on both metrics, indicating that it draws from both structural and semantic signals. These results suggest that the hybrid retrieval behaves as intended by combining cues from both modalities and appropriately supports context-aware discovery.

C User Study Details

C.1 Questionnaire for User Study

Table 4: Post-questionnaire items used in our user study.

Category	Question Item
System Usability Scale (5-point scale; 1: Strongly Disagree, 5: Strongly Agree)	
1. Frequency	I think that I would like to use this system frequently.
2. Complexity	I found the system unnecessarily complex.
3. Ease of Use	I thought the system was easy to use.
4. Technical Support	I think that I would need the support of a technical person to be able to use this system.
5. Integration	I found the various functions in this system were well integrated.
6. Inconsistency	I thought there was too much inconsistency in this system.
7. Learnability	I would imagine that most people would learn to use this system very quickly.
8. Awkwardness	I found the system very awkward to use.
9. Confidence	I felt very confident using the system.
10. Prior Learning	I needed to learn a lot of things before I could get going with this system.
Literature Review Support (7-point scale; 1: Strongly Disagree, 7: Strongly Agree)	
11. Overview	The system helped me understand the overall research area and its main topics.
12. Narrowing	The system helped me narrow a broad set of papers to those relevant to my work.
13. Connections	The system helped me see how the papers were connected or related to each other.
14. Outlining	The system helped me outline a literature review (e.g., sections or a clear storyline).
15. Grouping	The system helped me organize the papers I collected into meaningful groups.
16. Identifying Gaps	The system helped me identify what is missing or under-studied in the literature.
17. Discovery	The system helped me keep finding new, relevant papers as I worked.
18. Incorporation	The system helped me incorporate new papers into the groups I had already made.
19. Confidence	The system increased my confidence in reviewing the literature.
20. Fear of Missing Out	The system reduced my fear of missing out on important papers.

C.2 Annotation Interface

Figure 12: A full screenshot of our custom annotation interface used for analysis of the user study.

C.3 Sensemaking Outcome Codebook

Table 5: Codebook definitions used for analyzing participants' sensemaking outcome during our study.

Code	Definition
Shoebox	An evaluative activity analogous to mail sorting, where researchers assess the quality and relevance of a paper set without fully engaging with their content. Judgments are based on metadata or surface-level cues such as: • whether a paper is newly discovered, • citation links with previously interpreted papers, • venue, year, interpreted/uninterpreted status, • title cues, or assumptions from the query that retrieved it. Also includes classifying papers into groups for later examination and marking exploration priorities (e.g., immediate reading, deferred review, exclusion). *Coded as Shoebox* when reviewing outputs (self- or LLM-generated) that align with this evaluative process, or when constructing discovery queries guided by metadata gaps.
Interpretation	The act of analyzing individual papers by considering their content and explicitly assigning meaning within the researcher's literature review product (e.g., schema or narrative). Examples include: • creating notes, • assigning papers to groups, • excluding papers from analysis. *Coded as Interpretation* when using an LLM to elaborate paper content, or when reviewing outputs (self- or LLM-generated) that reflect interpretive work.
Schema	The process of creating, refining, and revising hierarchical grouping structures for organizing papers. This is not a rigid taxonomy but a dynamic process of schema formation and evaluation, often intuitive before explicit articulation. Typical cases include: • renaming groups or sections, • creating new subgroups, • restructuring to explore alternative organizational logics. *Coded as Schema* when using an LLM to summarize papers grouped together, when reviewing schema-related outputs, or when constructing queries for divergent exploration.
Argumentation	The activity of constructing, refining, or revising the narrative of a literature review, such as: • developing opening statements, • identifying trends and gaps, • articulating a position statement. *Coded as Argumentation* when using an LLM to generate claims and supporting evidence, when reviewing argumentative outputs, or when querying for papers to substantiate specific claims.