

CiteSense: Supporting Sensemaking of Research Literature

Xiaolong Zhang¹, Yan Qu², C. Lee Giles¹, Piyou Song¹

¹ Pennsylvania State University
University Park, PA 16801, USA
{xuz14, clg20, pus137}@psu.edu

² University of Maryland
College Park, MD 20742, USA
yanqu@umd.edu

ABSTRACT

Making sense of research literature is a complicated process that involves various information seeking and comprehension tasks. The lack of support for sensemaking in existing systems presents important design challenges and opportunities. This research proposes the design of an integral environment to support literature search, selection, organization and comprehension. Our system prototype, CiteSense, offers lightweight interaction tools and a smooth transition among various information activities. This research deepens our understanding of the design of systems that support the sensemaking of research literature.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General Terms: Design

Keywords: Sensemaking, light weight interaction

INTRODUCTION

Literature Review is a critical part of research and involves tasks to "*describe, summarize, evaluate, clarify and/or integrate the content of primary reports*"[8]. As a vehicle for learning and as a research facilitator[6], literature review includes essential activities like search, sensemaking, and composition. Imagine a scholar writing the review section of a paper. She would take an iterative process of searching and identifying relevant papers, reading them, taking notes, organizing and synthesizing them, and finally, integrating and discussing them in her writing.

Providing effective support for reviewing literature can benefit the whole research community and ultimately improve the scientific productivity. Many systems provide support for literature search and preliminary organization of the search results. General search engines allow users to search papers. Services like CiteSeer[11] provide users with citation information and context information within citations. Systems such as Butterfly[16] enable various search strategies (e.g., keyword-based search, citation chasing, etc.) and organization of search results.

However, creating a literature review is much more than

finding and organizing a collection of research literature. Researchers need to answer questions such as: Does the reported work have significant contributions to the research field? Is the result accurate and valid? Are there any gaps in the knowledge of the subject? Is there any consensus or debate on the topic? Scholars need to understand and appraise individual articles, to understand relationships between different works, to categorize, synthesize, summarize a collection of articles, and to compose the review based on their understanding of research work in the area of interest. These are difficult tasks involving intensive sensemaking, information processing and representation construction, Researchers would need help.

The lack of support for comprehension and composition in existing systems presents important design challenges and opportunities. This research is an effort to address this issue. We propose a design, CiteSense, an information-rich environment providing integral support for literature search and sensemaking. This environment supports information search, selection, organization, comprehension, and composition without imposing strict task structures, and allows users to smoothly perform different types of tasks, such as searching, filtering, browsing, citation chasing, information organizing, note taking, and writing.

RELATED WORK

Sensemaking arises when people face new problems or unfamiliar situations and their current knowledge is insufficient. It involves finding important structures in a seemingly unstructured situation by developing successively more sophisticated representations and fitting information into the representations in service of a task[20].

Various systems have been developed to assist sensemaking tasks. Some systems[7,18] help formulate, structure, and manage ideas by providing representation construction tools. Tools like Sensemaker[3] and FireWork[13] allowed users to organize and manipulate search results to better understand collected information while mediating the iterative search sessions. Recently, increasing attention has been given to the systems that integrally support information seeking, sensemaking and learning [17,19]

Literature Sensemaking has some unique features. It involves information sources with special structures embedded in it, such as citation network, hierarchy of categories of the articles, etc. Also, it has many task-specific information behaviors, including tracing citations, judging the quality of a work based on other people's reviews, looking

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2008, April 5–10, 2008, Florence, Italy.

Copyright 2008 ACM 978-1-60558-011-1/08/04...\$5.00.

for inadequacy in existing work, etc. Some efforts have been made to help such tasks, including Citation and Co-Author Network[15], Cat-a-Cone[12], BiolioViz[21], Butterfly[16], and so on. However, design efforts to explore integral support for seeking and sensemaking of research literature have rarely been seen since the Butterfly system.

UI DESIGN FOR LITERATURE SENSEMAKING

Our research developed some core tasks related to information seeking and sensemaking in research literature review by drawing on existing research on seeking, organizing, and making sense of information [9,4,2,10,16,19]. When a researcher tries to find, understand, and review literature in an unfamiliar area, she often conducts query-based search, traces references and citers of papers, or browses structures provided by information repositories (e.g., the ACM category). She reads the abstract, content, or citers' comments, and selects papers of interests. She takes notes on papers, categorizes or groups papers, and then composes the review based on collection and personal notes. This scenario suggests key tasks to be supported:

Seek

Search Literature Based on Keywords Keyword-based search is a required component in a sense-making process. People rely on powerful search capabilities to locate a set of papers with desired keywords.

Search, Browse, and Navigate Using Structures in the Data Literature resources often have *additional* structural information available. Structures could be classification system (e.g. the ACM classification), or citation relationships among articles.

Appraise, Comprehend and Compose

Appraise Collected Papers Initial examination on descriptive data, content analysis and comments from citers helps to evaluate the relevance and quality of a paper.

Organize Collected Information Organization of collected papers could help users understand the overview of and the relationships among different papers.

Take Notes and Compose Literature Review Note taking is often necessary to keep record of the important thoughts on a paper, or on relationships among different papers. Literature review is composed by consulting the notes and through the organization of information.

Task Management

Switch between Different Sub-tasks Seeking and comprehending literature involves various closely intertwined sub-tasks. Smoothly switching between sub-tasks is needed so that researchers will not be distracted by multiple operation steps or heavy cognitive loads.

Export Sensemaking Results A user can save the results of a sensemaking process – either a structured bibliographic list or a draft of literature review, and use them in other tasks or for other purposes.

THE CITESENSE SYSTEM

Based on the analysis above, we developed a prototype system, CiteSense, to support sensemaking of research literature. Figure 1 shows the view of its user interface. The

interface has three basic components: a query space (Panel A), a working space (Panel B), and a large area (Panel C) to provide in-depth context for search results.

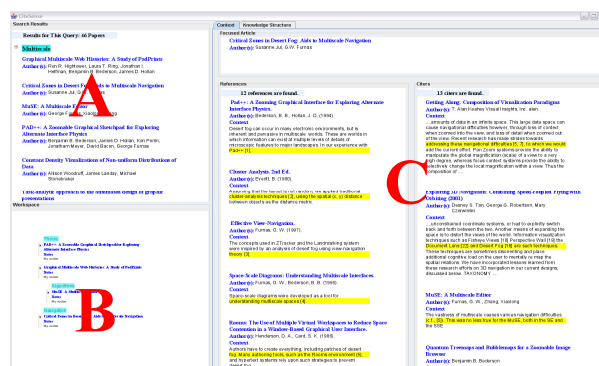


Figure 1: CiteSense User Interface

Query Space

Query space is where a sensemaking process starts. In this space, a user can directly type in a query and see search results. Descriptive information such as titles, authors, and abstracts of papers are shown in the results. New queries can be conducted on existing search results to get a smaller set of papers of interests.

Queries are organized as a tree structure so that the user can easily move between larger yet coarser search results and smaller but finer returns. Queries and results are visualized as an indented text structures to help the user navigate through the tree.

Context Space

The contextual space provides two kinds of contextual information of a paper of interest: citer/reference contexts and knowledge structures contexts.

Reference/Citer Context

Figure 2 shows a view on the reference/citation context information. On the top is a panel to display the paper of interest (Panel 1). Two panels below it are a reference panel (Panel 2) and a citer panel (Panel 3), showing the references and citations of the paper of interest respectively.

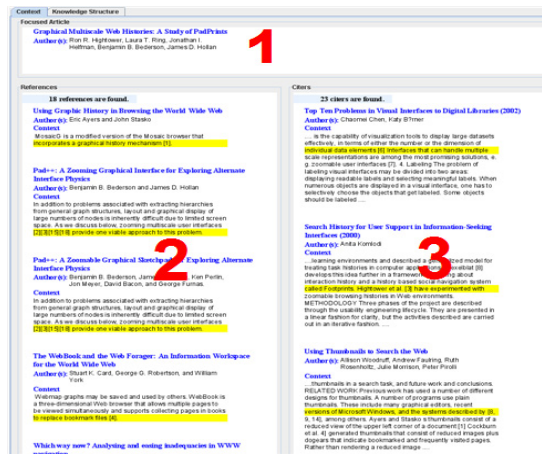


Figure 2: Contextual Space.

In addition to such information as title and authors of papers, the reference and citer panels also provide more in-depth information about how a reference was commented and analyzed by the paper of interest and how a citer commented and analyzed the paper of interest. Such information is highlighted for quick browse. The user can further trace a reference or a citer by clicking it. Expanding the scope of sensemaking is made easy.

Knowledge Structure Context

The contextual space can be switched to display knowledge structure. Figure 3 shows a view of a knowledge structure, a section of the ACM category tree, in the context area.

The ACM classification is chosen because it is authoritative in defining domain knowledge, created by experts, and accepted by users. Such structures are called categorical structures in information retrieval[14] and definitional structures in epistemology[1]. It is important to help people see how knowledge is organized in such formal structures and understand the difference between a formal structure and the structures they have in mind when interacting with the world and information[1,14].

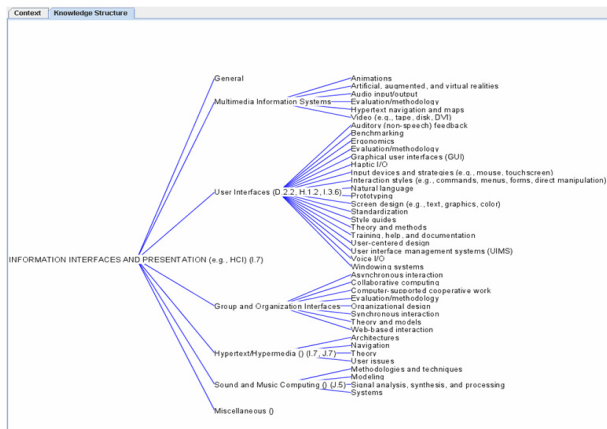


Figure 3: A Knowledge Structure in Context Area

The knowledge structure has three roles here. First, it helps the user to understand which categories a paper belongs to. Moving the cursor to a paper in the query space will highlight its ACM categories. Such structural information could help sensemaking processes by suggesting relevant areas to explore. The second role of the structure is to help the exploration of knowledge structures. Moving the cursor over a category in the structure, all relevant papers in the query space are highlighted. This allows the user to better understand the structural relationship among papers. The third role is to help people create filters. The user can drag and drop a structural node to the query space and get a filter automatically created in the query space. This approach could further simplify sensemaking processes by combining general search and domain-specific structure browsing.

Working Space

In the working space, a researcher can use lightweight operations to organize collected information, make sense of them, take notes, and draft a review.

Drag and Drop Papers into the Working Space

Any paper in the query space, the reference panel, or the citer panel can be transported into the working space by drag and drop. This direct manipulation simplifies the process to build personal paper collections.

Creating and Manipulating Collection Structures

Putting information artifacts of interest into a structure is important to sensemaking. In the early stage of sensemaking, when a user does not have a clear idea for structure, she can choose to lay the items in different areas in the working space as informal clusters. Later, she has a better understanding of the structure in a collection, and creates hierarchical sections to hold individual papers.

Drafting Personal Notes on Papers

A user can draft personal notes on collected papers (Figure 4). Each paper in the collection has a note field, which allows users to add annotations. All personal notes are highlighted so that the user can easily see which papers have been annotated and which ones have not. Moreover, a user can write a note on multiple papers and use this type of notes to organize papers. Such notes could be about relationships between different papers or be a summary of them. These different kinds of notes can potentially grow into a comprehensive literature review.

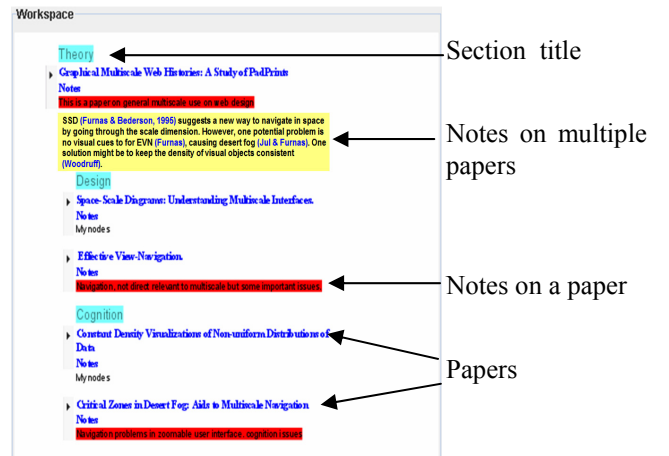


Figure 4: Personal Collection and Notes

Exporting Personal Collection

All information in the working area can be exported. Due to high cognitive costs of sensemaking processes, a user often wants to save the final results for future use. Our system allows saving personal collections and notes in various formats (e.g., simple text, EndNotes, and XML).

Our CiteSense prototype is a Java application built with the Piccolo toolkit[5]. Paper data is from the CiteSeer database.

PRELIMINARY EVALUATION

Three faculty members and four graduate students contributed their time to the evaluation of the CiteSense prototype.

All people gave positive comments on the system. In general, they like CiteSense and want it in their work. The integral environment is welcomed because it allows them to focus on finding, understanding, and collecting literature in one place, rather than switching between different appli-

cations. They indicated that the working space in the system can greatly improve the way they collect and manage literature. The flexibility of organizing papers with customized section titles and personal notes is appealing to them.

Some issues were also raised. One participant wanted to see other kinds of knowledge structures, rather than just the ACM category tree. More diverse export formats, such as BibTex, was also suggested.

Some suggestions for system improvement were made. One suggestion is to make the structure of a personal collection available for other people to use. Another suggestion is to support the manipulation of multiple personal collections that share some common documents.

The result from the preliminary evaluation is encouraging. It indicates that in making sense of research literature, researchers may benefit from an integral environment that supports a wide range of tasks in sensemaking. Our system offers lightweight interaction tools and does not impose strict task structures on users. For example, queries and filters can be initiated by directly clicking on the space in the search panel; paper collections can be created by directly dragging papers from other panels; and different tasks, such as context browsing and reference/citer chasing, as well as knowledge structure browsing and filter creating, can be easily done without the need for switching to different tools. Lightweight interaction design can help to reduce cognitive costs in sensemaking and facilitate the smooth transition among different tasks.

CONCLUSION

In this paper, we presented the design of an integral environment to support the sensemaking of research literature. Our design provides synergies among various information seeking strategies such as searching, filtering, browsing, citer or reference chasing, and knowledge structure browsing. It also facilitates the evaluation and comprehension of retrieved information, and supports information organization, note taking, and writing. Our preliminary evaluation results show such an integrated approach is welcomed by researchers.

Some unaddressed issues exist in this research. Our future efforts will focus on two aspects. First, we will conduct formal user studies on the benefits and costs of this system. Second, we are also interested in making personal collections and their structures reusable and sharable.

REFERENCES

- Ahl, V., & Allen, T. (1996). *Hierarchy Theory: A Vision, Vocabulary and Epistemology*. Columbia University Press, New York.
- Attfield, S. (2005) Information seeking, gathering and review: Journalism as a case study for the design of search and authoring systems. Unpublished doctoral dissertation, University College London,
- Baldonado, M., Winograd, T. (1997). SenseMaker: An Information-Exploration Interface Supporting the Contextual Evolution of a User's Interests. *ACM CHI'97*.
- Bates, M. (1989). The design of browsing and berry-picking techniques for the online search interface. *Online Review*, 13, 5, pp. 407-424.
- Bederson, B., Grosjean, J., & Meyer, J. (2004). Toolkit Design for Interactive Structured Graphics. *IEEE Trans. on Software Engineering*, 30 (8), pp. 535-546.
- Bruce, C. S. (1994). Research student's early experiences of the dissertation literature review' *Studies in Higher Education*, vol. 19, no. 2, pp. 217-229.
- Conklin, J., Begeman, M. (1988). gIBIS: A Hypertext Tool for Exploratory Policy Discussion. *ACM Trans. on Office Information System*, 4(6), pp. 303-331.
- Cooper, H. (1998). *Synthesizing Research: A Guide for Literature Reviews*, p 107.
- Dervin, B. (1992). From the Mind's Eye of the User: the Sense-Making Qualitative-Quantitative Methodology. In Glazier, J.D. & Powell, R. (Eds.) *Qualitative Research in Information Management*, pp. 61-84.
- Faisal, S., Cairns, P., Blandford, A. (2006). Developing User Requirements for Visualizations of Literature Knowledge Domains. *IV'06*, pp. 264-269.
- Giles, C., Bollacker, K., Lawrence, S. (1998). CiteSeer: an Automatic Citation Indexing System. In *Proceedings of 3rd ACM Conference on Digital Libraries*, pp. 89-98.
- Hearst, M., & Karadi, C. (1997). Cat-a-Cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. *ACM SIGIR'97*, pp. 246-255.
- Hendry, D., & Harper, D. (1997). An informal information-seeking environment. *JASIS*. 48, pp.1036-1048.
- Ingwersen, P. (1992). *Information Retrieval Interaction*. London: Taylor Graham Publishing.
- Ke, W., Borner, K., & Viswanath, L. (2004). Major Information Visualization Authors, Papers and Topics in the ACM Library. *IEEE INFOVIS'04*.
- Mackinlay, J. D., Rao, R., & Card, S. K. (1995). An organic user interface for searching citation links. *ACM CHI'95*, pp. 67-73.
- Marchionini, G.(2006). Exploratory Search: From finding to understanding. *Comm. of the ACM*, 49(4) 41-46.
- Mindjet Inc., <http://www.mindjet.com>
- Qu, Y., & Furnas, G.W. (2005). Source of Structure in Sensemaking. *ACM CHI'05*, pp.1989-1992.
- Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993). The Cost Structure of Sensemaking. *ACM INTERCHI'93*, pp. 269-276.
- Shen, Z., Ogawa, M., Teoh, S. T., and Ma, K. (2006). BiblioViz: a system for visualizing bibliography information. In *Proc. of 2006 Asia-Pacific Symposium on information Visualisation*, pp. 93-102