

Drilling for Papers in INKE

Stan Ruecker

Illinois Institute of Technology

Geoffrey Rockwell, Lindsay Doll, Mark Bieber, & Shannon Lucky

University of Alberta

Stéfan Sinclair

McGill University

Milena Radzikowska

Mount Royal University

Christian Vandendorpe

University of Ottawa

Ray Siemens

University of Victoria

Teresa Dobson

University of British Columbia

Michael Eberle-Sinatra

Université de Montréal

Abstract

In this article, we discuss the first year research plan for the INKE interface design team, which focuses on a prototype for chaining. Interpretable as a subclass of Unsworth's scholarly primitive of "discovering", "chaining" is the process of beginning with an exemplary article, then finding the articles that it cites, the articles they cite, and so on until the reader begins to get a feel for the terrain. The chaining strategy is of particular utility for scholars working in new areas, either through doing background work for interdisciplinary interests or else by pursuing a subtopic in a domain that generates a paper storm of publications every year. In our prototype project, we plan to produce a system that accepts a seed article, tunnels through a number of levels of citation, and generates a summary report listing the most frequent authors and articles. One of the innovative features of this prototype is its use of the experimental "oil and water" interface effect, which uses text animation to provide the user with a sense of the underlying process.

Keywords

Interface design; INKE; Research plan; Prototyping; Citations; Content searching; Search tool

CCSP Press

Scholarly and Research Communication

Volume 3, Issue 1, Article ID 010110, 5 pages

Journal URL: www.src-online.ca

Received August 17, 2011, Accepted November 15, 2011, Published March 26, 2012

Ruecker, Stan, Rockwell, Geoffrey, Doll, Lindsay, Bieber, Mark, Lucky, Shannon, Sinclair, Stéfan, Radzikowska, Milena, Vandendorpe, Christian, Siemens, Ray, Dobson, Teresa, & Eberle-Sinatra, Michael. (2012). Drilling for Papers in INKE. *Scholarly and Research Communication*, 3(1): 010110, 5 pp.

© 2012 Stan Ruecker, Geoffrey Rockwell, Lindsay Doll, Mark Bieber, Shannon Lucky, Stéfan Sinclair, Milena Radzikowska, Christian Vandendorpe, Ray Siemens, Teresa Dobson, & Michael Eberle-Sinatra. This Open Access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc-nd/2.5/ca>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Scholarly and Research Communication

VOLUME 3 / ISSUE 1 / 2012

Stan Ruecker is Associate Professor at the Institute of Design, Illinois Institute of Technology, 350 North La Salle Street, 4th Floor, Chicago, IL, USA 60610
Email: sruecker@id.iit.edu

Lindsay Doll is a student in the Faculty of Education at the University of Alberta, 116 Street and 85 Avenue, Edmonton, AB, Canada T6G 2R3
Email: lin_ef_doll@hotmail.com

Mark Bieber is a student in the Department of Computer Science at the University of Alberta, 116 Street and 85 Avenue, Edmonton, AB, Canada T6G 2R3.
Email: bieber@ualberta.ca

Shannon Lucky is a graduate student in the Humanities Computing Program of Library and Information Studies at the University of Alberta, 116 Street and 85 Avenue, Edmonton, AB, Canada T6G 2R3
Email: lucky.shannon14@gmail.com

Geoffrey Rockwell is Professor of Philosophy and Humanities Computing in the Department of Philosophy at the University of Alberta, 67 Assiniboia Hall, Edmonton, AB, Canada T6G 2E7
Email: geoffrey.rockwell@ualberta.ca

Stéfan Sinclair is Associate Professor in Digital Humanities in the Department of Languages, Literatures & Cultures at McGill University, 688 Sherbrooke Street West, Room 425, Montreal, QC, Canada H3A 3R1
Email: sgsinclair@gmail.com

Milena Radzikowska is an Associate Professor in Information Design, Faculty of Communication Studies at Mount Royal University, 4825 Mount Royal Gate SW, Calgary, AB, Canada T3E 6K6 Email: mradzikowska@gmail.com

Christian Vandendorpe is Professor Emeritus in the Département de français at the University of Ottawa, 75 Laurier Avenue East, Ottawa, ON, Canada K1N 6N5 Email: christian.vandendorpe@gmail.com

Ray Siemens is Canada Research Chair in Humanities Computing and Distinguished Professor in the Faculty of Humanities in English with cross appointment in Computer Science at the University of Victoria, PO Box 3070 STN CSC, Victoria, BC, Canada V8W 3W1 Email: siemens@uvic.ca

Teresa Dobson is an Associate Professor in the Department of Language and Literacy Education and Director of the Digital Literacy Centre at the University of British Columbia, 2329 West Mall Vancouver, BC, Canada V6T 1Z4 Email: teresa.dobson@ubc.ca

Michael Eberle-Sinatra is an Associate Professor in the Département d'études anglaises at the Université de Montréal, CP 6128, Station Centreville, Montréal, QC, Canada H3C 3J7 Email: michael.eberle.sinatra@umontreal.ca

The INKE Research Group comprises over 35 researchers (and their research assistants and postdoctoral fellows) at more than 20 universities in Canada, England, the United States, and Ireland, and across 20 partners in the public and private sectors. INKE is a large-scale, long-term, interdisciplinary project to study the future of books and reading, supported by the Social Sciences and Humanities Research Council of Canada as well as contributions from participating universities and partners, and bringing together activities associated with book history and textual scholarship; user experience studies; interface design; and prototyping of digital reading environments.

Literature review

Ellis (1989) defined citation chaining as the practice of “following citation connections between materials.” He suggested that academics commonly perform “backward chaining—following up references or sources cited in material consulted,” while less commonly engaging in “forward chaining—identifying citations to material consulted or known” (p. 183). To analyze the literature on information overload, Akin (1998) applied backward chaining and citation patterning (among other methodologies). She found that the former method revealed inaccurate and incorrect citations, while supporting the discovery of “a cognitive trail of thought” (p. 254). On the other hand, citation patterning—systematically compiling and comparing bibliographies—facilitated the identification of integral information sources, academic collaboration, linking and missing citations, and citing behaviours such as peer and self-citing.

In a later study, Whitmire (2003) discovered that citation chaining is a dominant and effective information seeking strategy of undergraduate students, particularly among those rated at a medium-high or high level of epistemological development. Investigating the discrepancy between the skills of humanities scholars and available information retrieval technologies, Buchanan, Cunningham, Blandford, Rimmer, and Warwick (2005) observed that the use of references and citations in known sources to find unknown sources (i.e., citation chaining) was the most commonly reported research practice. They also noted that other behaviours previously defined by Ellis (1989) were reported by humanities academics, with monitoring (i.e., tracking particular authors, articles, or journals) being the second most frequently mentioned strategy, followed by browsing (i.e., semi-focused searching).

Furthermore, through in-depth interviews with 100 graduate students across disciplines, George, Bright, Hurlbert, Linke, St. Clair, and Stein (2006) found that almost half reported using the chaining process to establish a body of literature. Their study demonstrated that citation chaining was used most by computer science students (64%), followed closely by science and humanities students (62% and 60%, respectively). Those in art/architecture reported using it the least frequently (25%). They also noted that this information behaviour is supported by both human (e.g., professors/advisors often recommended the initial source(s) from which chaining was conducted) and computer resources (100% use of the internet for this research was reported).

However, despite evidence that citation chaining is commonly utilized, there is a consensus that better digital tools are needed to facilitate this practice (e.g., Ellis 1989; Buchanan et al., 2005). Among the efforts to enhance information systems, Kerne and Smith (2004), for instance, take a human-centred approach. They

proposed an information discovery (ID) framework—which combines cognitive and digital processes—to inform the design of more user-friendly and effective tools for information seeking, foraging, discovery, and usage. To specifically support citation chaining, Mackinlay, Rao, and Card (1995) developed the Butterfly, a visualization application for the exploration of multiple bibliographic repositories. This system enables rapid and comprehensive search and browsing activities through the integration of the following techniques: visualization, created “link-generating” queries, asynchronous query processes, and process controllers. In the Butterfly, bibliographic material is fastened to interface objects (called “butterflies”) that list an article’s references on one “wing” and its citers on the opposite “wing.” Users can perform backward and forward chaining simply by following combinations of related butterflies (c.f. the ISI Web of Science).

The Paper Drill prototype

The previous work in this area has resulted in a variety of online citation tools connected to specific kinds of data, typically but not exclusively in the sciences, such as CiteSeer, the ISI Web of Science, Association for Computing Machinery, and PaperScope. These systems provide a means of carrying out the process of chaining, allowing the user to select a “seed” article as a starting point, then seeing all the articles that cite it and all the articles that are cited. In some cases (such as the ISI Web of Science), a level can be assigned so that the visualization can include citations at more than a single remove.

What remains to be done is to create a system that helps to provide not only a simpler process to do chaining, but also a preliminary result of the chaining activity, in the form of a summary that shows the most commonly cited authors and articles. Our intention is therefore to build on these ideas in order to provide a tool that will allow the user to select a “seed” article, indicate how many levels deep to go, then have the system traverse the available metadata and articles to produce a summary report of the authors and articles most cited starting from that seed, as well as links (where possible) back to the articles.

A variety of controls should be useful. For instance, by allowing the user to set the threshold number of items necessary for an author or item to be included in the report, the scope of the results could be dynamically adjusted to accommodate frequency. Similarly, the user should be able to decide whether or not to include various different indications of authorship. For instance, should the system include articles in the total count for an author if that person is not the first author, or not the sole author? Similarly, there may be cases where the user would only wish to see cases where the author is the last author, since one of the conventions in the sciences is that the last author is often the senior scientist who runs the lab. This strategy would therefore allow the possibility of identifying papers emerging from specific research labs.

There are numerous technical and theoretical issues to be worked through with this approach, including the need for consistent data, the benefits of separating concerns by using a proxy layer to isolate interface design from collections, and the implications for researchers in the humanities of having such a process automated. Consistent data is essential in helping to distinguish between similar or even identical author names, as well as author names in various locations on co-authored papers. It is also important to be able to identify identical articles cited under slightly different titles (e.g. using an

ampersand instead of the word “and” or the Oxford comma rather than no comma). Development of a proxy layer is essential in that it allows the interface design to proceed in the absence of “real” data being available from the databases. It is a matter of negotiation, however, in terms of how much processing, filtering, sorting, and so on is carried out at the server and delivered through the proxy, and how much is handled at the proxy layer or even up at the interface.

We intend that the implications for humanities researchers of having the process automated will be the subject of further research. Ideally, having fewer steps to carry out in chaining will allow researchers to do more of it and to spend more time in looking at the results rather than in producing this initial overview that the software will now provide.

Finally, although it is possible to envision a variety of conventional interface designs that would provide the affordances we outline for the Paper Drill, we are also experimenting with providing the Paper Drill functionality within the context of the oil and water browser, where the seed article is literally dragged, along with its settings, onto a visual representation of a collection, so that the process of selection can be animated.

Future work

Our goal is to create a working prototype of the system and collaborate with the INKE User Experience team on setting up user studies to help us better understand some of these issues. We are also working with the INKE Information Management team on developing a proper application programming interface (API) or proxy layer that will keep interface design separated from the database work. Our principal partner in this initiative is Synergies, a platform for the publication of research results which provides an extensive database of journals in the humanities and social sciences. Finally, the INKE Reader Studies team provides a context for the Paper Drill within the history of various citation systems and formats.

References

- Akin, Lynn. (1998). Methods for Examining Small Literatures: Explication, physical analysis, and citation patterns. *Library and Information Science Research*, 20(3), 251-270.
- Buchanan, George, Cunningham, Sally Jo, Blandford, Ann, Rimmer, Jon, & Warwick, Claire. (2005). Information Seeking by Humanities Scholars. In Andreas Rauber, Stavros Christodoulakis, A Min Tjoa (Eds.), *Research and Advanced Technology for Digital Libraries: 9th European Conference, ECDL 2005, Vienna, Austria, September 18-23, 2005*. Heidelberg: Springer Berlin.
- CiteSeerx. (n.d.). Scientific Literature Digital Library and Search Engine. URL: <http://cite-seerx.ist.psu.edu> [August 15, 2009].
- Ellis, David. (1989). A Behavioural Approach to Information Retrieval System Design. *Journal of Documentation*, 45(3), 171-212.
- Ellis, David, & Oldman, Hanna. (2005). The English Literature Researcher in the Age of the Internet. *Journal of Information Science*, 31(1), 29-36.
- George, Carole, Bright, Alice, Hurlbert, Terry, Linke, Erika C., St. Clair, Gloriana, & Stein, Joan. (2006). Scholarly Use of Information: Graduate students' information seeking behaviour. *Information Research*, 11(4), 272.
- Kerne, Andruid, & Smith, Steven M. (2004). The Information Discovery Framework. *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. Cambridge, MA.

- ISI Web of Science. URL: <http://apps.webofknowledge.com> [August 15, 2009].
- Mackinlay, Jock D., Rao, Ramana, & Card, Stuart K. An Organic User Interface for Searching Citation Links. In Irvin R. Katz, Robert L. Mack, Linn Marks, Mary Beth Rosson, & Jakob Nielsen (Eds.), *Proceedings of the ACM CHI 95 Human Factors in Computing Systems Conference*. URL: http://www.sigchi.org/chi95/proceedings/papers/jdm_bdy.htm [August 15, 2009].
- PaperScope*. URL: <http://paperscope.sourceforge.net/index.htm> [May 30, 2009].
- Ruecker, Stan. (2009). From SQL to Mandalas, From Spreadsheets to Oil & Water: The Practice of Humanities Interface Design. *The second symposium of TRUTH: Teaching and Research Using Technology in the Humanities*. University of Victoria.
- Unsworth, John. (2000). Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? *Humanities Computing: formal methods, experimental practice*. King's College, London. 13 May 2000. URL: <http://www.iath.virginia.edu/~jmu2m/Kings.5-00/primitives.html> [November 12, 2006].
- Whitmire, Ethelene. (2003). Epistemological beliefs and the Information-Seeking Behavior of Undergraduates. *Library & Information Science Research*, 25(2), 127-142.