Drilling for Papers in INKE

[1]Stan Ruecker, [2]Geoffrey Rockwell, [3]Milena Radzikowska, [4]Stéfan Sinclair, [5]Christian Vandendorpe, [6]Ray Siemens, [7]Teresa Dobson, [8]Lindsay Doll, [9]Mark Bieber, [10]Michael Eberle-Sinatra, Shannon Lucky[11], and the INKE Research Group

[1]Humanties Computing Program, University of Alberta, Edmonton, Alberta, Canada
sruecker@ualberta.ca
[2]Humanities Computing Centre, McMaster University, Hamilton, Ontario, Canada
georock@mcmaster.ca
[3]Information Design Department, Faculty of Communication Studies, Mount Royal University, Calgary, Alberta, Canada
mradzikowska@mtroyal.ca
[4]Department of Communication Studies, McMaster University, Hamilton, Ontario, Canada
sgs@mcmaster.ca
[5]christian.vandendorpe@gmail.com
[6]Faculty of Humanities, University of Victoria, Victoria, British Columbia, Canada
siemens@uvic.ca
[7]Department of Language and Literacy Education, University of British Columbia, Vancouver, British Columbia, Canada
teresa.dobson@ubc.ca
[8]Faculty of Education, University of Alberta, Edmonton, Alberta, Canada
lin_ef_doll@hotmail.com
[9]Department of Computer Science, University of Alberta, Edmonton, Alberta, Canada
bieber@ualberta.ca
[10]Department of English, Université de Montréal, Montréal, Quebec, Canada
michael.eberle.sinatra@umontreal.ca
[11]Humanities Computing Program and School of Library and Information Studies, University of Alberta, Alberta, Canada

Abstract. In this paper, we discuss the first year research plan for the INKE interface design team, which focuses on a prototype for "chaining" (Ellis and Oldman 2005). Interpretable as a subclass of Unsworth's (2000) 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 (Ruecker 2009), 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

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." 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." 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 et al. (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 et al. (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% internet use 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-centered 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 et al. (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, ACM, 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 any number of both 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 API or proxy layer that

will keep interface design separated from the database work. Our principal partner in this initiative is Synergies, 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.

Works Cited

Akin, L. "Methods for examining small literatures: explication, physical analysis, and citation patterns." *Library and Information Science Research* 20.3 (1998): 251-70. Print.

Buchanan, George et al. "Information Seeking by Humanities Scholars." *Research and Advanced Technology for Digital Libraries*, 2005. 218-229. Print.

Ellis, D. "A behavioural approach to information retrieval system design." *Journal of Documentation* 45.3 (1989): 171-212. Print.

Ellis, D. and Oldman, H. "The English literature researcher in the age of the Internet." *Journal of Information Science* 31.1 (2005): 29-36. Print.

George, C., Bright, A., Hurlbert, T., Linke, E.C., St. Clair, G., & Stein, J. "Scholarly use of information: graduate students' information seeking behaviour." *Information Research* 11.4 (2006): 272.; rpt. Web. <http://InformationR.net/ir/11-4/paper272.html>.

Kerne, A., & Smith, S.M. "The Information Discovery Framework." *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques.* Cambridge, MA, 2004. 357-360. Print.

Mackinlay, J. D., Rao, R., & Card, S.K. "An Organic User Interface for Searching Citation Links." *Proceedings of the ACM CHI 95 Human Factors in Computing Systems Conference.* Eds. Katz et al. Denver,1995. 67-73. Print; rpt. Web. <http://www.sigchi.org/chi95/proceedings/papers/jdm_bdy.htm>.

PaperScope. Web. 30 May 2009. <http://paperscope.sourceforge.net/index.htm>.

Ruecker, Stan. "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. 3 April 2009. Address.

Unsworth, John. "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. Web. 12 November 2006. <http://www.iath.virginia.edu/~jmu2m/Kings.5-00/primitives.html>.

Whitmire, E. "Epistemological beliefs and the information-seeking behavior of undergraduates." *Library & Information Science Research* 25.2 (2003): 127-42. Print.