BUILDING SOCIAL DIGITAL LIBRARIES

by Katy Borner,
Indiana University SLIS,
Bloomington, Indiana

INTRODUCTION

The accelerating rate of scientific and technical discovery, typified by the ever-shortening time period for the doubling of information – currently estimated at 18 months [1] – causes new topics to emerge at an increasing rate. Large amounts of human knowledge are available online – not only in the form of texts and images, but also as audio files, movies, software demos, etc.

In parallel, economic, organizational, and societal pressures, as well as the desire to reach shared goals more efficiently and effectively, are driving an increase in collaborative research. Research collaborations frequently occur among participants separated by temporal, geographical, organizational, disciplinary, and cultural boundaries. Increasingly complex collaborative projects focus attention on the question of how to enable researchers to more easily share expertise [2].

However, digital libraries (DLs) are very lonely places. Your best friend may query, e.g., the ACM library, with the identical search term and sift through the potentially large set of retrieved documents at the very same time. However, there is no way you will ever find out this happened. There is no means to annotate articles so that others can capitalize from your expertise; no indication how many people have read an article (besides page access counters or citation counts); nobody to ask for help.

In comparison, real world (public) libraries are really social places – crowded with people looking for a specific book, browsing audio-visual material under a certain topic, attending author readings, hanging out with their kids. There are multiple cues that aid social navigation. Among them are crowded parking around the library, a higher density of people during events in certain areas of the library, usage signs on books, tapes, toys (e.g., dog ears, annotations), and last but not least reference librarians providing assistance.

How can we create social digital libraries that provide easy access to massive amounts of data but also human expertise?

Today, search engines are our primary means to access data in digital form. However, search interfaces lack the ability to support information exploration, making it increasingly difficult for scientists and practitioners to gain a "big picture" view of DLs, to locate germane resources, to monitor the evolution of their own and other knowledge domains, to track the influence of theories within and across domains, etc.

Information Visualization (IV) techniques [3-7] have been applied to facilitate access to online data and to digital libraries. Visual interfaces to DLs exploit powerful human vision and spatial cognition to help humans mentally organize and electronically access and manage large, complex information spaces. The aim is to shift the user's mental load from slow reading to faster perceptual processes such as visual pattern recognition. The recently published book on "Visual Interfaces to Digital Libraries" [8] provides an overview of major work in this area.

Diverse data mining and visualization algorithms [9] have been applied to extract salient semantic structures and/or co-citation relationships among documents and to layout documents spatially; helping users to visualize, locate, and remember documents more quickly. Sample systems are Orendorf & Kacmar's 2-D map proposed as a method to structure DLs and their content and to ease document location and access [10]; Populated Information Terrains [11]; VR-VIBE [12]; and Bead [13], which use statistical techniques to analyze and group documents based on their semantic similarity and create visualizations of bibliographies. Crossley et al.'s Knowledge Garden [14] aims to provide an environment where people can meet colleagues and share relevant information.

Chen (1999a) used Latent Semantic Analysis and Pathfinder Network Scaling to create a semantically organized information space. The approach was implemented in StarWalker, a system that uses Blaxxun's community platform to display citation networks as a set union of all possible minimum spanning trees (Chen & Carr, 1999). StarWalker is the very first system that uses a tightly coupled spatial-semantic model as focal point in a multi-user environment. Chen, Thomas, Cole, and Chennawasin showed that the proliferation of IV models can play a significant role in extending and enriching the design of inhabited, multi-user virtual environments [15]. In StarWalker multiple users can examine a complex visualization together but they cannot rate, annotate, or contribute documents, etc. to adapt this space to their changing social and information needs. Social navigation is supported exclusively by the visibility of other participants and the ability to chat.

We suggest going beyond existing efforts by creating collaborative visual interfaces to DLs that users can shape by contributing new documents, annotate them or influence the layout and presentation of documents. The design of collaborative interfaces is facilitated by commercially available 3-D Online Browser Systems such as are Blaxxun's online community client-server architecture,[1] Microsoft's Virtual Worlds Platform,[2] Active Worlds technology by Activeworlds, Inc.,[3] and the new Adobe Atmosphere browser.[4] Each of these 3-D browser systems facilitates the creation of multi-modal, multi-user, navigable, and collaborative virtual worlds in 3-D that are interconnected with standard web pages and that are accessible from standard computer platforms via the Internet, 24 hours and 7 days a week.

We are in the process of developing a shared three-dimensional document space for a scholarly community – namely faculty, staff, and students at the School of Library and Information Science, Indiana University. Upon completion, about 530 people – including about 300 students in Bloomington and 200 students in Indianapolis – will have access to this space.

The space will provide access to an initial seed collection of about 8,000 links to online documents such as text, images, video, software demonstrations, etc. collected from personal favorites or bookmark lists from SLIS faculty and staff. The full text of all documents has been downloaded and semantically analyzed using data mining techniques such as Latent Semantic Analysis [16]. The resulting document-by-document similarity matrix was utilized to group semantically similar documents – see also our work on the LVis – Digital Library Visualizer project [17, 18]. A Semantic Treemap algorithm [19] was developed to layout documents in a 3-D space. Semantic Treemaps utilize the original treemap approach [20] to determine the size (dependent on the number of documents) and layout of document clusters. Subsequently, a force directed placement algorithm [21] is applied to the documents in each cluster to place documents spatially based on their semantic similarity.

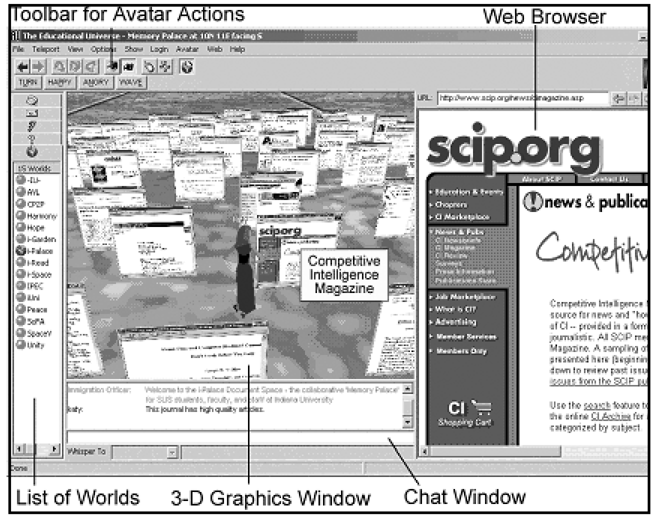

Figure 1. Desktop Interface to the SLIS Document Space

In the 3-D space (cf. Figure 1), each document is represented by a square panel textured by the corresponding web page's thumbnail image and augmented by a short description such as the web page title which appears when the user moves the mouse over the panel – similar to Robertson et al's, Data Mountain interface [22]. Upon clicking the panel the corresponding web page is displayed in the web browser interface. Users can collaboratively examine, discuss, and will be able to modify (add and annotate) documents, thereby converting this document space into an ever-evolving repository of the user community's collective knowledge that members can access, learn from, contribute to, and build upon. The space becomes a shared Memory Palace representing the knowledge of the community.

ACKNOWLEDGEMENTS

Over the years, a number of students have been involved in the design of the visualization tools discussed in this paper. Among them are Tamara McMahon, Ying Feng, Sy-Miaw Lin, and Min Xiao.

Support comes from a High Performance Network Application grant from Indiana University awarded in 2000/2001.

Part of this work was published in Katy Börner, Ying Feng & Tamara McMahon: Collaborative Visual Interfaces to Digital Libraries. Marchionini, G. & Hersh, W. (eds), Second ACM+IEEE Joint Conference on Digital Libraries, July 14-18, 2002, Portland, Oregon, USA, ACM Press, pp. 279-280.

REFERENCES

1. Scholtz, J., DARPA/ITO Information Management Program Background. 2000.

2. Schleyer, T.K.L., Collaboratories: Leveraging information technology for cooperative research. Journal of Dental Research, 2001. 80(6): p. 1508-1512.

3. Card, S., J. Mackinlay, and B. Shneiderman, eds. Readings in Information Visualization: Using Vision to Think. 1999, Morgan Kaufmann.

4. Chen, C., Information Visualisation and Virtual Environments. 1999, London: Springer-Verlag London.

5. Ware, C., Information Visualization: Perception for Design. 2000: Morgan Kaufmann.

6. Spence, B., Information Visualization. 2000: Addison-Wesley.

7. Dodge, M. and R. Kitchin, Mapping Cyberspace. 2000: Routledge.

8. Börner, K. and C. Chen, eds. Visual Interfaces to Digital Libraries. LNCS 2539. 2002, Springer Verlag.

9. Börner, K., C. Chen, and K. Boyack, Visualizing Knowledge Domains, in Annual Review of Information Science & Technology, B. Cronin, Editor. 2003, Information Today, Inc./American Society for Information Science and Technology: Medford, NJ. p. 179-255.

10. Orendorf, J. and C. Kacmar. A spatial approach to organizing and locating digital libraries and their content. in 1st ACM international Conference on Digital Libraries. 1996: ACM Press.

11. Benford, S., D. Snowdon, and J. Mariani, Populated Information Terrains: First Steps, in Virtual Reality Application, R.A. Earnshaw, J.A. Vince, and H. Jones, Editors. 1995, Academic Press. p. 27-39.

12. Benford, S., et al., VR-VIBE: A virtual environment for co-operative information retrieval. Computer Graphics Forum, 1995. 14(3): p. 349-360.

13. Chalmers, M. BEAD: Explorations in information visualisation. in SIGIR '92. 1992. Copenhagen, Denmark: ACM Press.

14. Crossley, M., et al., The knowledge garden. BT Technology Journal, 1999. 17(1).

15. Chen, C., et al., Representing the semantics of virtual spaces. IEEE Multimedia, 1999. 6(2).

16. Landauer, T.K., P.W. Foltz, and D. Laham, Introduction to Latent Semantic Analysis. Discourse Processes, 1998. 25: p. 259-284.

17. Börner, K., A. Dillon, and M. Dolinsky. LVis - Digital Library Visualizer. in Fourth International Conference on Information Visualisation, Symposium on Digital Libraries. 2000. London, England.

18. Börner, K. Extracting and visualizing semantic structures in retrieval results for browsing. in ACM Digital Libraries. 2000. San Antonio, Texas.

19. Feng, Y. and K. Börner. Using Semantic Treemaps to Categorize and Visualize Bookmark Files. in SPIE Conference on Visualization and Data Analysis. 2002. San Jose, CA.

20. Shneiderman, B., Tree visualization with tree-maps: A 2-d space filling approach. ACM Transactions on Graphics, 1992.

21. Fruchterman, T.M.J. and E.M. Reingold, Graph Drawing by Force-Directed Placement. Software-Practice & Experience, 1991. 21(11): p. 1129-1164.

22. Robertson, G., et al. Data mountain: Using spatial memory for document management. in 11th annual ACM Symposium on User Interface Software and Technology (UIST '98),. 1998.

FOOTNOTES

1 http://www.blaxxun.com/community

2 http://www.vworlds.org/

3 http://www.activeworlds.com/

4 http://www.adobe.com/products/atmosphere/

Part of this article was previously published in Katy Börner, Ying Feng & Tamara McMahon: Collaborative Visual Interfaces to Digital Libraries. Marchionini, G. & Hersh, W. (eds), Second ACM+IEEE Joint Conference on Digital Libraries, July 14-18, 2002, Portland, Oregon, USA, ACM Press, pp. 279-280.