Top Ten Problems in Visual Interfaces to Digital Libraries

Chaomei Chen[1] and Katy Börner[2]

[1]College of Information Science and Technology, Drexel UniversityPhiladelphia, PA 19104-2875, USA
Chaomei.Chen@cis.drexel.edu
[2]School of Library and Information Science, Indiana University, Bloomington, IN, 47405, USA
katy@indiana.edu

Abstract. Many research communities periodically and collectively deliberate about their most significant and challenging problems as a way to agree upon pressing questions and promising research directions. Identifying top-ten problems is a useful way to reflect on what a community has achieved and define a new research agenda for the future. This chapter introduces a set of such problems for research in visual interfaces of digital libraries in order to simulate studies in this area.

1 Motivation

David Hilbert presented one of the most famous lists of challenging mathematical problems to the International Congress of Mathematicians in Paris in 1900. His twenty-three mathematical problems influenced and continue to influence mathematical research all over the world.

Research on visual interfaces of digital libraries grew out of several interrelated areas such as information visualization, digital libraries, and information retrieval. As a result, there is a limited consensus on the set of major research questions and their priority, making it hard for researchers in the field to define their research, to find collaborators and combine efforts, and to contribute to a general theory of information visualization. So far, there exists no central software code repositories and very few data sets are shared, which might be due to the fact that information visualization as a field itself is very young.

In this chapter, we first outline some top-ten problems identified in fields that are closely related to visual interfaces in particular and information visualization in general. Then we make our first attempt to crystallize the top-ten problems based on what we learned from the two workshops in 2001 and 2002[1]. Many stimulating ideas are presented in earlier chapters of this book; we conclude by listing which of these initiatives we think are most significant and what remains to be done in the future.

[1] http://vw.indiana.edu/visual01 and http://vw.indiana.edu/visual02/

K. Börner and C. Chen (Eds.): Visual Interfaces to Digital Libraries, LNCS 2539, pp. 226–231, 2002.

2 Related Top-Ten Problem Lists

Various "top-ten problems" lists have emerged in research communities related to visual interfaces of digital libraries and information visualization. For example, Foley identified the ten most challenging problems [1] in computer graphics. He emphasized that each of them shares common concerns and also represents a particular perspective of the discipline. To get a sense of what a top-ten list looks like, Foley's top-ten problems are shown below:

1. Fill the gap between image-based and geometric modeling techniques
2. Fill the gap between motion-capture animation and simulation/procedural animation
3. Creative information visualization
4. Automated creation of information and scientific visualizations
5. Abstracting away from reality
6. Display more pixels
7. Display fewer pixels
8. Unified graphics architectures
9. User interfaces for 3D creativity
10. Truly immersive virtual reality

Problem 3 is particularly relevant to research on visual interfaces of digital libraries. Foley defined information visualization as "creating representations of non-geometric information by adding geometry to the information." Foley had data warehousing in mind when he listed this problem, but this is equally valid for digital libraries. Information visualization will grow in importance as digital libraries become more and more common. The content of a digital library does not always come with an inherent geometry. The major challenges are how to extract new and more complex types of relationships and visualize them so that they can make the contents of digital libraries more accessible and manageable to users.

Information retrieval is another topic relevant to this book. Visual interfaces aim to support retrieval as well as browsing. Croft in 1995 listed the top ten research issues for information retrieval [2]:

1. Efficient, flexible indexing and retrieval
2. Integrated solutions
3. Distributed information retrieval
4. Vocabulary expansion
5. Interfaces and browsing
6. Routing and filtering
7. Effective retrieval
8. Multimedia retrieval
9. Information extraction
10. Relevance feedback

Problem 5 is particularly relevant to our mission. Croft states in [2] that, "The interface is a major part of how a system is evaluated, and as the retrieval and routing

algorithms become more complex to improve recall and precision, more emphasis is placed on the design of interfaces that make the system easy to use and understandable." Croft further points out that, "Interfaces must support a range of functions including query formation, presentation of retrieved information, feedback, and browsing. The challenge is to present this sophisticated functionality in a conceptually simple way." Until recently, this has been a relatively under-researched issue. However, Croft predicted this will change as more work in information visualization appears.

Hibbard, in May 1999, listed top-ten problems for visualization [3]. He grouped the problems into visual quality, integration, information, interactions, and abstractions. The ten problems are as follows:

1. Realistic visual displays
2. Integrated virtual reality and physical reality
3. Integration of visualization with networking, voice, artificial vision, computation and data storage
4. Optimal visual interactions
5. Visualization of high-dimensional numerical information
6. Visualization of non-numerical information
7. Direct manipulation with visualizations
8. Visual idioms for collaborative interactions
9. Abstractions for visualization and user interaction processes
10. Reconciliation of expressiveness and easy of use

A number of problems listed here are related to the design of visual interfaces.

3 Top-Ten Problems in Visual Interfaces of Digital Libraries

The two workshops held in 2001 and 2002 on Visual Interfaces to Digital Libraries generated lively discussions. The participants offered many good suggestions and valuable feedback (see section 2.1 on Socio-Technical Challenges). Here we have assembled an initial list of top-ten problems and outline what should be addressed by future research in this area.

1. Theoretical Foundations

Research in visual interfaces of digital libraries as a whole lacks solid theoretical foundations. Although principles for perception and cognition, computer graphics, and human-computer interaction do exist, they do not readily lend themselves to formation of design principles. Many principles are tightly coupled with particular environments and are hard to generalize. More often, the same fundamental problem disguises itself in different forms, which also complicates the process of putting available theories into practice. Foundation works such as [4] are urgently needed. Theoretical contributions can significantly influence our practice.

The first problem is concerned with how we should build on the success of current? visual interfaces and explore promising application areas. Some exciting candidates include bibliometrics, scientometrics, knowledge tracking, and knowledge discovery. These areas have special requirements for their unique tasks and many have already developed their own approaches for handling complex information visually. Insights, experiences, and lessons learned from these fields are valuable sources of inspiration.

2. Empirical Foundations

This essence of this problem is, "to know where we are." It is crucial to make clear what has been empirically proven to be useful and beneficial. What are the common elements that have been found in every successful example of visual interfaces of digital libraries? Which features have worked well in some cases, but not others? Which features so far have shown no conclusive benefits? Examples in areas such as visual information retrieval, visual information exploration, and empirical studies of information visualization [5] are likely to provide valuable clues. New methodologies and taxonomies of exemplar systems should be carefully considered.

3. Scalability

Digital libraries often face problems of scalability because of their varying sizes and contents. Will the algorithms and solutions that have been tried and worked on small-scale digital libraries break down in large-scale digital libraries? Researchers in information visualization are striking for faster responses, incremental updates, and a scale-proof layout performance. Computing and data processing power is continually growing more powerful, as the volumes of the data we need to handle increase. Visual scalability is the capability of visualization tools to display large datasets effectively, in terms of either the number or the dimension of individual data elements [6]. Interfaces that can handle multiple scale representations are among the most promising solutions, e.g., zoomable user interfaces [7].

4. Labeling

The problem of labeling visual interfaces may be divided into two areas: displaying readable labels and selecting meaningful labels. When numerous objects are displayed in a visual interface, one has to selectively choose the objects that get labeled. Some objects should be labeled prominently, some should be labeled moderately. Still others should not be labeled at all. Making meaningful labels is equally challenging if not more challenging than making readable labels. However, natural language processing and automated indexing may provide useful techniques for this purpose. Users' interventions as part of the labeling process should also not be completely ruled out. A related question is, "meaningful to whom?" To create quality labels, one may need to take into account the nature of tasks and characteristics of the users of a particular digital library.

5. Individual Differences

The fifth problem concerns individual differences and how visual interfaces should accommodate such differences. In reality, one size can hardly fit all. One user's favorite visual interface may be another user's nightmare. Individuals have relatively stable cognitive preferences and abilities that can be measured by psychometric tests. For example, spatial ability indicates an individual's ability to recognize and handle spatial relationships of objects [8, 9]. Research in human-computer interaction has shown that individual differences can be the most significant factor in one's performance.

6. Supporting Collaborative Work

Given the individual differences we need to accommodate and the diversity of social norms in cyberspace, supporting collaborative work is a challenging task in its own right. Translation of collaborative work into a visual process entails overcoming a variety of obstacles. For instance, how should we represent participating parties through a visual interface? How should we integrate social structures with the organization of the underlying digital library? How should we evaluate whether a particular visual interface is useful in a collaborative setting?

7. Benchmarking and Standardization

The provision of commonly accessible and comparable test collections has been proven useful in several fields, especially test collections in information retrieval and associated text retrieval conferences (e.g., TREC at http://trec.nist.gov). A test collection must be similar to a real digital library in terms of size and content. At the same time, it must be simple enough to provide a sound base for performance analysis. Standardization is a closely related issue, which involves more efficient and effective integrations of services, protocols, and interfaces at various levels. The lack of benchmarks and standards in part contributes to the current suboptimal situation.

8. Evaluation

Evaluation is an integral part of the design and development of visual interfaces as well as digital libraries as a whole. Evaluative studies are needed to find out what has worked/is working? for both users and designers?. The development of research methods may benefit from the huge literature in human-computer interaction. This problem is also closely related to problem 7. Common test collections are necessary to make evaluative studies more comparable in terms of the strengths of evidence and conclusions.

9. Personalization

Pro-active, customized, and personalized information delivery is an increasing trend in digital libraries. Visual interfaces are in a good position to organize and reorganize the way an underlying digital library is presented to a client, for instance, tailored according to the client's background and access history. This is also related to collaborative support, or collaborative recommendation. The online bookstore Amazon.com provides a collection of recommended books for each user based on the

titles purchased by the user and what others bought along with those same titles. While personalization and individual differences are a pair of closely interrelated problems, they are markedly different. Individual differences emphasize the types of abilities that are relatively stable, whereas personalization is driven by what information is needed. Although individual differences could be a factor, personalization focuses more on the type of information, the content, rather than the style.

10. Standardization

Modularization and standardization of digital library and information visualization services will save valuable resources and prevent designers from constantly "reinventing the wheel"[10]. Interfaces which link data storage and representation formats need to be standardized (see previous chapter in this book). As well, interoperation ability and cross-platform performance can be considerably improved.

We conclude this book with these challenging problems and hope that they will stimulate the creation and examination of more "top-ten" problems for this young, interdisciplinary field.

Acknowledgements

Special thanks to Stephen Eick and Sougata Mukherjea for their valuable feedback on an earlier version of the chapter.

References

1. Foley, J., *Getting there: The ten top problems left.* IEEE Computer Graphics and Applications, 1999.
2. Croft, W.B., *What do people want from information retrieval?* D-Lib Magazine, November 1995 (http://www.dlib.org/dlib/november95/11croft.html).
3. Hibbard, B., *Top ten visualization problems.* SIGGRAPH Newsletter, 1999. **33**(2).
4. Furnas, G.W. *Effective view navigation*. in *CHI '97*. 1997. Atlanta, Georgia: ACM Press.
5. Chen, C. and M. Czerwinski, *Empirical evaluation of information visualizations.* International Journal of Human-Computer Studies, 2000. **35**(5): p. 631-635.
6. Eick, S.G. and A.F. Karr, *Visual scalability.* Journal of Computational Graphics and Statistics, 2002. **11**(1): p. 22-43.
7. Hightower, R.R., et al. *Graphical multiscale Web histories: A study of PadPrints*. in *9th ACM Conference on Hypertext and Hypermedia (Hypertext '98)*. 1998. New York, NY: ACM Press.
8. Chen, C. and M. Czerwinski, *Spatial ability and visual navigation: An empirical study.* New Review of Hypermedia and Multimedia, 1997. **3**: p. 67-89.
9. Chen, C., M. Czerwinski, and R. Macredie, *Individual differences in virtual environments: Introduction and overview.* Journal of the American Society for Information Science, 2000. **51**(6): p. 499-507.
10. Börner, K. and Y. Zhou. *A Software Repository for Education and Research in Information Visualization*. in *Fifth International Conference on Information Visualisation*. 2001. London, England: IEEE Press.