Rich Interaction *in the Digital Library*

Ramana Rao, Jan O. Pedersen, Marti A. Hearst,

Jock D. Mackinlay, Stuart K. Card, Larry Masinter, Per-Kristian Halvorsen,

and George G. Robertson

Effective information access involves rich interactions between users and information residing in diverse locations. Users seek and retrieve information from the sources—for example, file servers, databases, and digital libraries—and use various tools to browse, manipulate, reuse, and generally process the information. We have developed a number of techniques that support various aspects of the process of user/information interaction. These techniques can be considered attempts to increase the bandwidth and quality of the interactions between users and information in an *information workspace*—an environment designed to support information work (see Figure 1).

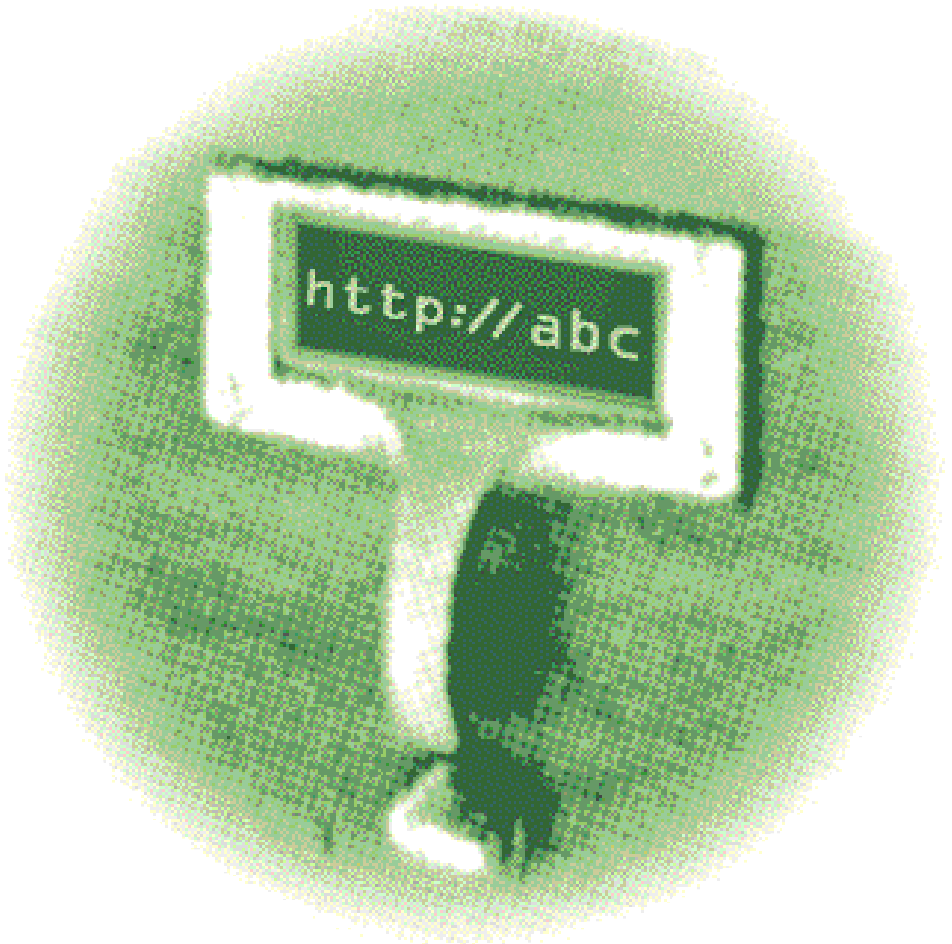


Table 1. Roadmap of Illustrated Techniques

Requirements	Techniques
Iterative Query Refinement	**Browsing Subsets of Source Iteratively** **Viewing Context of Query Match** **Visualizing Passages Within Documents**
Source Heterogeneity	Modeling Sources Rendering Sources and Results
Parallel, Interleaved Access	**Reflecting Time Costs of Interaction** **Managing Multiple Search Processes**
Larger Work Process	Integrating Multiple Search and Browsing Techniques Visualizing Large Information Sets

Workstation environments equipped with current access tools and applications could be viewed as information workspaces, but in a variety of ways they limit the effectiveness of informtion access and the larger process of information work. In particular, conventional retrieval interfaces are based on the view of information retrieval as an isolated task in which the user formulates a query against a homogenous collection to obtain matching documents. Observations of people using both digital and physical library services indicate a number of areas where this view misses the reality of users doing real work:

- Iterative Query Refinement. Users are often unable to formulate pointed questions or express them effectively using conventional retrieval systems. Users often learn during the course of a session what they are trying to ask and how to ask it.
- Source Heterogeneity. Users often access multiple sources with differing characteristics of content, form, and provenance, where each source has its own methods of access with differing functionality. Understanding these characteristics is an important part of a user's activity.
- Parallel, Interleaved Access. Users often switch among sources with slow or variable response. Though users may want to interleave access operations and track their progress, current systems are weak in their support for this process.
- Larger Work Process. Information access is usually intertwined with other parts of the overall work process, for example, analysis of the results. Users switch among different techniques for searching or browsing sources and visualizing or utilizing results.

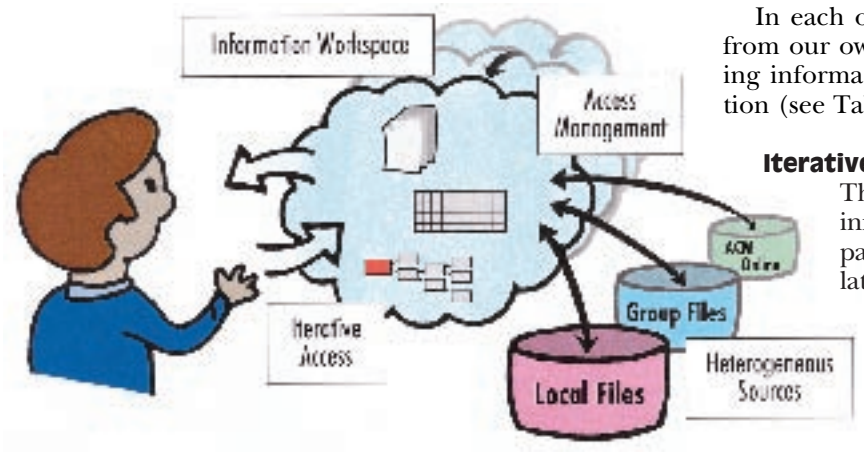


Figure 1. The user accesses information from heterogeneous sources and brings it into a workspace to utilize in some broader task

In each of these four areas, we present examples from our own work that point the way toward building information workspaces supporting rich interaction (see Table 1).

Iterative Query Refinement

The Information Theatre paradigm [2] for information access emphasizes the participation of the user in a cycle of query formulation, presentation of search results, and query reformulation. Since the focus is on query repair, the information presented is typically not document descriptions, but rather intermediate information indicating relationships between the query and the retrieved documents. We have developed a number of techniques based on this paradigm. In this section, we illustrate some of these with a running example.

Suppose the user is interested in documents that discuss ways to counteract the effects of malicious programs. Initially the user has available a large, unfamiliar information source. In our example, this source is the 2.2GB TIPSTER text collection [6]. Because the collection is unfamiliar, the user will be unsure whether it contains relevant information, and if so, how to access it.

To address this situation, we have developed a browsing paradigm, called *Scatter/Gather* [3, 4], which allows a user to rapidly assess the general contents of a very large collection by scanning through a hierarchical representation that acts as a dynamic table of contents. Initially the system automatically *scatters*, or clusters, the collection into a small number of document groups, and presents short summaries of the groups to the user. These summaries consist of two types of information: topical titles (titles of documents close to the cluster centroid) and typical terms (terms of importance in the cluster). Based on these summaries, the user selects one or more of the groups for further study. The selected groups are *gathered*, or unioned, together to form a subcollection. The system then reapplies clustering to scatter the new subcollection into a small number of document groups, which are again presented to the user. With each successive iteration the clusters become smaller, and their contents more refined. The user may, at any time, switch to a more focused search method.

Figure 2 shows a three-part view of the example collection. Because the sixth cluster covers computer-related topics, the user chooses it for gathering and the scattering, yielding the refined set of clusters shown in an expanded view. The user next selects clusters 4 and 5, which cover networking issues (since networks are often associated with delivery of malicious programs), yielding the final set of clusters shown.

By browsing the collection in this manner, the user gains understanding of the computer-related contents of the corpus, and can choose whether or not to further explore here or to try another collection. From the titles and terms retrieved, it becomes apparent that this collection contains commercially oriented discussions of technology, rather than predominantly academic ones. From this overview information, the user can conclude that this is indeed a promising collection.

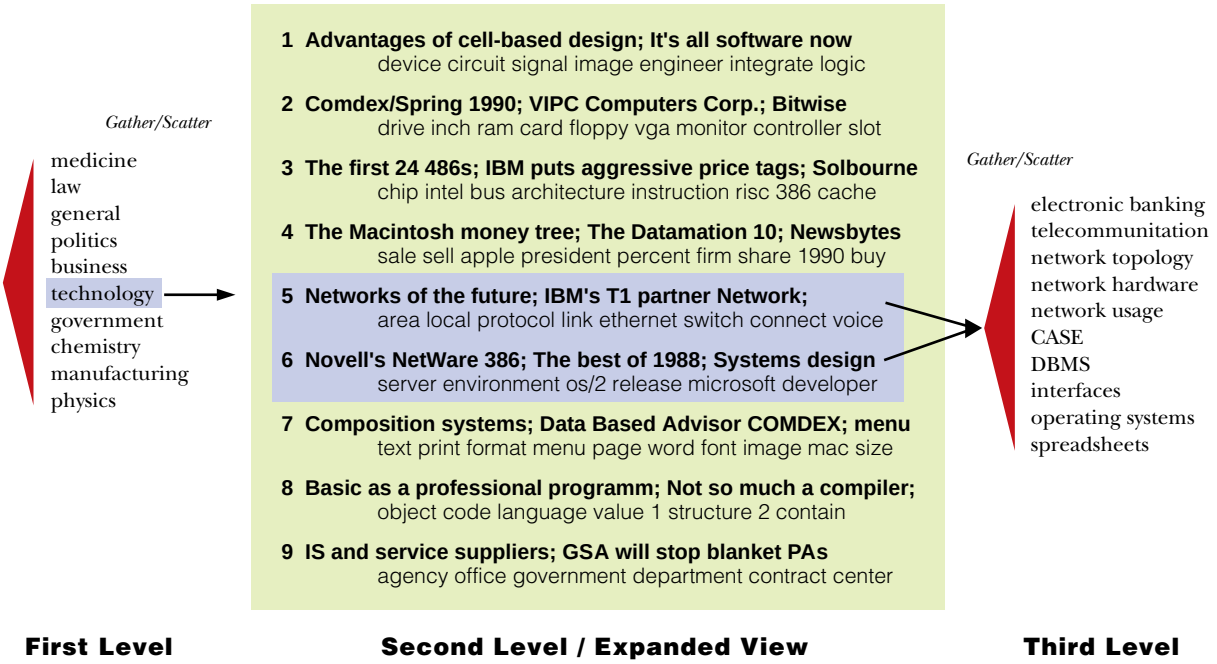


Figure 2. A schematization of the Scatter/Gather algorithm over the TIPSTER corpus

Suppose the user is not familiar with the jargon for malicious programs. A dictionary of available search terms can aid the search for alternative terminology, as can an online domain-specific thesaurus [9]. However, the help of a highly trained intermediary, such as a research librarian, is usually required to reach a desirable reformulation. Additionally, semantic relationships other than synonymy are often very useful, but conventional predefined thesauri do not account for the various ways terms are used with respect to one another.

As an alternative to a predefined thesaurus, *Snippet Search* [15] allows the user to directly inspect the set of

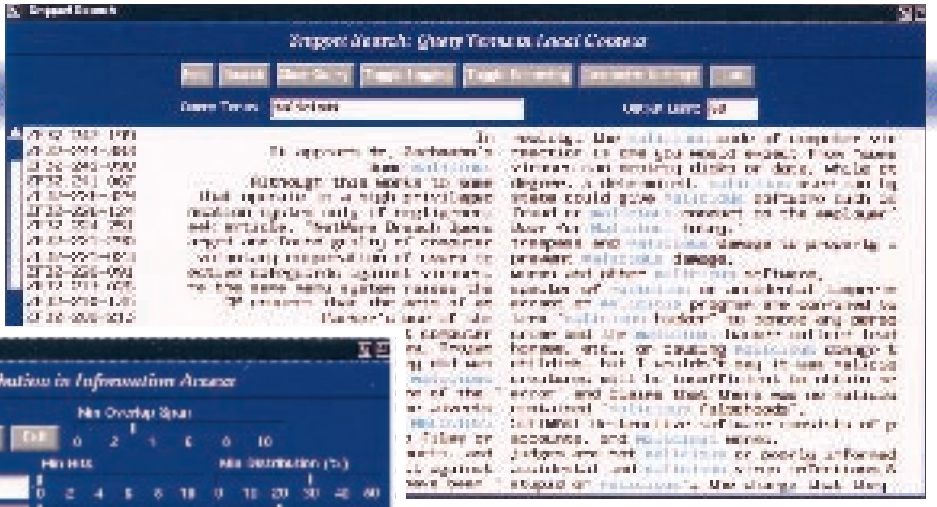


Figure 3. Snippet search on "malicious"

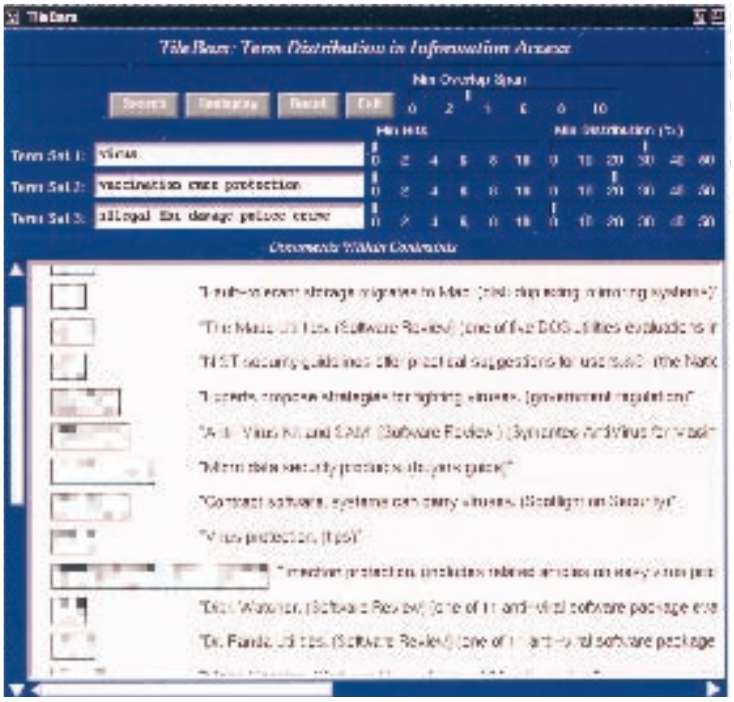


Figure 4. The TileBar display on a query consisting of three term sets

phrases containing focus terms. The intention is to aid query reformulation by showing *snippets* of context surrounding the terms in the variety of lexical environments present in the corpus. The current heuristic returns the text surrounding the search terms including some other "significant'' word, where significance is operationally defined by *not* being on some prespecified list of non-topic-influencing words (a "stop list''). These snippets are intended to contain sufficient context to distinguish usage, but not so much as to distract the reader or clutter the display.

For example, Figure 3 shows some of the local context surrounding snippet search on the term **malicious**. Terms such as **computer virus**, **Trojan horse**, and **worm** are seen to co-occur with this term, and therefore have potential as discriminating search terms. Note that these terms are not synonyms for malicious, although some closely affiliated terms are indeed present (e.g., **fraud**).

Based on this information, the user might decide to search on a conjunction of **virus** and terms indicating protection from these malicious programs (e.g., **vaccination**, **protection**, and **cure**). In a typical, system documents satisfying the query are returned and are rank-ordered according to some function of the number of hits for each term [24]. But this kind of ranking is opaque to the user; it is not clear how well each term is represented in the retrieved documents.

To address these issues, the **TileBars** interface [8] allows the user to make informed decisions about which documents and passages to view, based on the distributional behavior of the query terms in the documents. The goal is to simultaneously and compactly indicate (i) the relative length of the document, (ii) the frequency of the term sets in the document, and (iii) the distribution of the term sets with respect to the document and to one another. To facilitate display of distribution information, each document is partitioned in advance into a set of subtopical segments using an algorithm called *TextTiling* [7].

Figure 4 shows an example run on the query **(virus) AND (vaccination protection cure) AND (illegal fbi damage police crime)** with implicit **ORs** among the terms within each term set. Each large rectangle indicates a document, and each square within the document represents a coherent text segment (or tile). The darker the tile, the more frequent the term (white indicates 0, black indicates 8 or more hits, the frequencies of all the terms within a term set are added together). The top row of each rectangle corresponds to hits for Term Set 1, the second row to hits for Term Set 2, and the third for Term Set 3. Term overlap and term distribution are displayed in a manner in which both attributes together create recognizable patterns. When term sets are discussed simultaneously, their corresponding tiles blend together to cause a prominent dark block to appear. Scattered discussions have lightly colored tiles and large areas of white space. A mouse-click on a dark region brings up a view of the document starting at a passage with a large number of hits.

The figure shows a version of the interface that allows the user to impose several kinds of term distri-

bution constraints. In this example the user has specified that **virus** must occur in at least 1/3 of the tiles of each displayed document, and that the terms relating to **protection** must occur in at least 20% of the tiles, and that overlap among all three term sets must occur at least once within the span of three adjacent tiles. Judging from the titles displayed, this restriction is indeed useful in isolating documents that discuss strategies for protection from malicious programs. The figure also provides information about the extent to which the displayed documents discuss the criminal aspects of computer viruses.

TileBars are able to indicate that an article such as this one, if retrieved on the search terms **(virus)** AND **(protection, vaccination, cure)**, has only a passing reference or subtopical discussion of these concepts, and so may not be of interest. However, the interface does not prejudge the user's intention in this regard; for example, it could aid a user who wished to discover documents in which discussions of compromised programs occur in larger contexts.

As an example of iterative query refinement, the results of a TileBar search can be fed into a Scatter/Gather session, which can be useful to indicate which general topics were covered by the documents retrieved in the TileBar query. A subset of the clusters can be used as input to a refined TileBar query, and the cycle repeated.

Source Modeling

Information work often involves the use of multiple information sources with disparate content and access mechanisms. A common goal of current work on information access protocols is to provide uniform access to diverse collections. However, uniform access protocols should not hide the heterogeneity of information and access methods.

People working in an office make use of a rich set of visual and physical cues when arranging and seeking information. Many studies of office work [1, 14] show that people are surprisingly adept and resourceful at using such cues to organize, access, and use information. For this reason, we believe that representing information about collections and their contained items—so-called meta-information—is needed not just to support integration across disparate sources and services, but just as importantly, to support a number of other activities in an information workspace including selecting, understanding, utilizing, and remembering sources and their contents.

The practices of professional searchers who use multiple online sources is illustrative. Erickson and Salomon [5] observed that one group of expert online searchers spent a remarkable amount of time at their weekly status meeting sharing information about sources: topics included newly available sources, information quality, frequency of updates, timeliness of updates, costs, and exemplar situations for source use. It is this kind of information that should be provided through access protocols to support workspace interaction. Important categories of meta-information include the following:

- Content. What information is covered by the source. Various paradigms for representing this information include a textual description of the source contents, a statistical analysis of the words used in the source, or a more structured representation of coverage using a knowledge representation language.
- Provenance. The nature of the process that produced and/or maintains the source. For example, whether the source is a personal archive, a group collection, or a public source; the institutional source of the information; update frequency; and mutability.
- Form. The schemas for items contained in the source including available attributes and the types of values stored for those attributes.
- Functionality. The capabilities and properties of the access service including the kinds of searches supported, performance characteristics, and the nature of progress reporting supported.
- Usage Statistics. Statistics about source usage including previous use by the same user or other users. These statistics are often more useful if aggregated by categories of users or by organizational or other hierarchical groupings.

Each of these categories of meta-information is useful for a number of aspects of building workspaces. Source selection is an important problem for the user in dealing with multiple sources. Content coverage is one important criteria for selection, but often the provenance or form of the information is just as important in deciding whether a particular source meets the requirements imposed by the overall task. Usage statistics can also play a role; for example, a user may remember using a particular source for a particular purpose or decide to use a source endorsed by other workgroup members. Thus, tools for direct browsing and searching of this meta-information provide a powerful mechanism for source selection.

Another important use for meta-information is to drive the rendering and visualization of information in the electronic workspace, thus re-introducing the rich cues of physical workspaces. For example, visualizations, which map sources into spatial and graphical elements based on meta-information, can support interactions that allow users to select sources as well as build a spatial memory of sources. In addition, meta-information can be used in the rendering of items retrieved from the sources.

The use of meta-information requires support

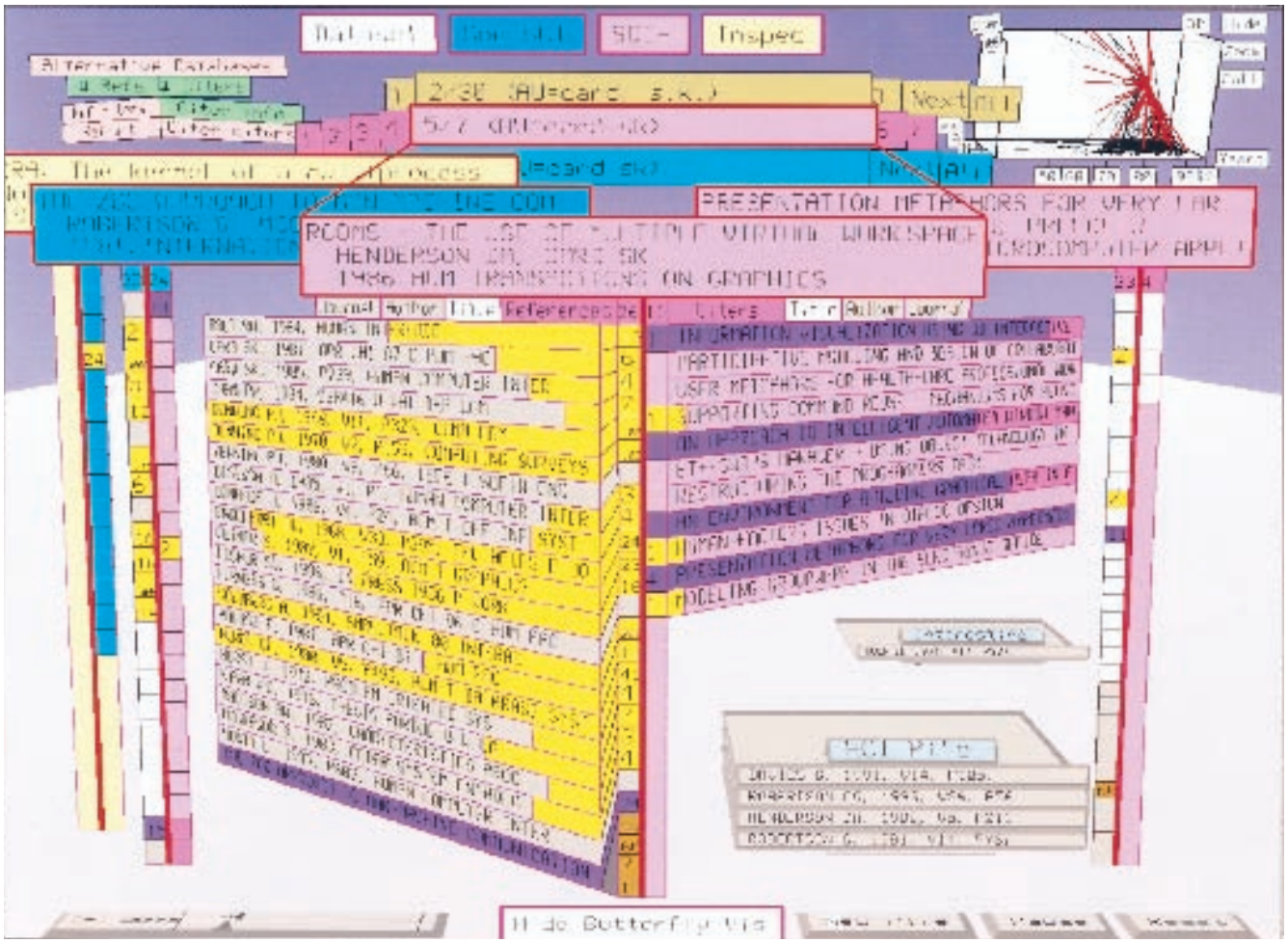


Figure 5. The Butterfly application for visualizing citations among bibliographic records

from the access infrastructure. The GAIA protocol [19, 20] defines an interface for uniform access to diverse sources and meta-information about those sources. We have built an intermediary server that supports the GAIA protocol for sources of various types including scanned document sources, WAIS sources, and sources served by the DIALOG online information service. Using GAIA, we have begun to explore the building of workspaces that exploit meta-information to provide support for source selection, rich visual cues about sources and their contents, performance, and cost of access.

Access Management

In a distributed system involving varying degrees of distance and delay between a user's workspace and an information service, a uniform interface cannot ensure uniform performance. A user model that assumes all operations take the same amount of time is unrealistic. A more workable user model includes upfront feedback on time costs and ongoing status of running operations. In addition, given varied performance and the desirable interleaving of multiple explorations, support for multiple search processes would be useful.

To these ends, long-running operations in the GAIA protocol including search, link traversal, and document access have three useful properties: (i) they are analyzed upfront to estimate time costs, (ii) they are performed asynchronously, and (iii) they generate regular status feedback to the client. The client before or upon operation initiation is notified of the estimated time to completion. During the asynchronous execution of the operation, the client is notified of significant events. In addition, the status of the operation—including estimations of percentage done, remaining time to completion, and state of the computation—is periodically reported to the client. Feedback on time behavior information allows users to manage their attention and form search strategies more effectively. Furthermore, asynchronous operations enables formulation and execution of multiple operations in parallel. GAIA defines operations for managing running processes. These protocol mechanisms collectively allow more sophisticated user interfaces than the typical paradigm of one operation at a time and percent-done indicators.

The Butterfly application, developed within the Information Visualizer environment [21], exploits the GAIA protocol to support the exploration of online bibliographic databases [12]. In particular, it

utilizes an intermediary server that supports the GAIA protocol for a number of bibliographic sources (e.g., Science Citation and Inspec databases as provided by DIALOG).

Butterfly, shown in Figure 5, combines search and browsing elements. The upper part of the space focuses on searches and the lower part focuses on browsing. Users typically start with queries to find articles in topic areas of interest and then browse reference and citation links to find related articles. The user can initiate multiple operations including queries, link traversals and record accesses, which proceed in parallel. Progress feedback allows the visualization to reflect the ongoing status of the operation.

Butterfly provides a pyramid of objects for visualizing the results of queries. Query results typically have an unpredictable size and require an unpredictable amount of time to enumerate. GAIA provides fine control of result enumeration, which Butterfly uses to disclose results incrementally. Each query result is visualized as a horizontal layer in the pyramid colored to indicate the source database. New items are retrieved by clicking on the "Next'' button. The "All'' button is used to initiate an access process that forces the complete enumeration of the result.

An item and its citation context is visualized using a *butterfly*. The head of the butterfly shows a summary of the bibliographic information for an article. The wings show the article's reference on the left and citers of the article on the right. Colors indicate previous interaction with those items and other properties (e.g., frequencies of citations). Clicking on items in these wings retrieve the corresponding bibliographic record for a reference or citer and a new butterfly is constructed and brought into the center for the user to focus on. The Butterfly application provides options for process management that deploy computational resources based on user attention; for example, search processes automatically fill in missing information for articles brought into focus by the user.

Information rarely exists in isolation, even though databases typically store information in discrete chunks and search applications typically view results as separate subcollections. Many attributes of records implicitly represent linking relationships (citations are just one example) that typically require the user to generate a special query to traverse. This kind of link-generating query is automatically provided by Butterfly through direct interaction with the visualization. As the user performs a number of queries and explores items, citations, and references, visualizations accumulate in the workspace. They collectively depict the search space and allow the user to control and refine the space through direct interaction. Butterfly integrates search, browsing, and access management using the information in the space.

Workspace Integration

The Butterfly application is a new kind of workspace based on interactive 3-D graphics and animation. Many of the properties of workspaces illustrated in Butterfly can also be applied in more conventional interfaces. In particular, Figure 5 shows other components which are important for supporting the broader workspace activity of using gathered information. The objects below the butterflies are piles of articles that the user has collected. The object in the upper-right corner is a 3-D scatterplot in which the points are articles plotted by time, first author, and number of citations and the links are citation links. Multiple visualizations can be useful during the process of access and subsequent use. A challenge in the design of the complete workspace is the integration of multiple general-purpose as well as task-spe-

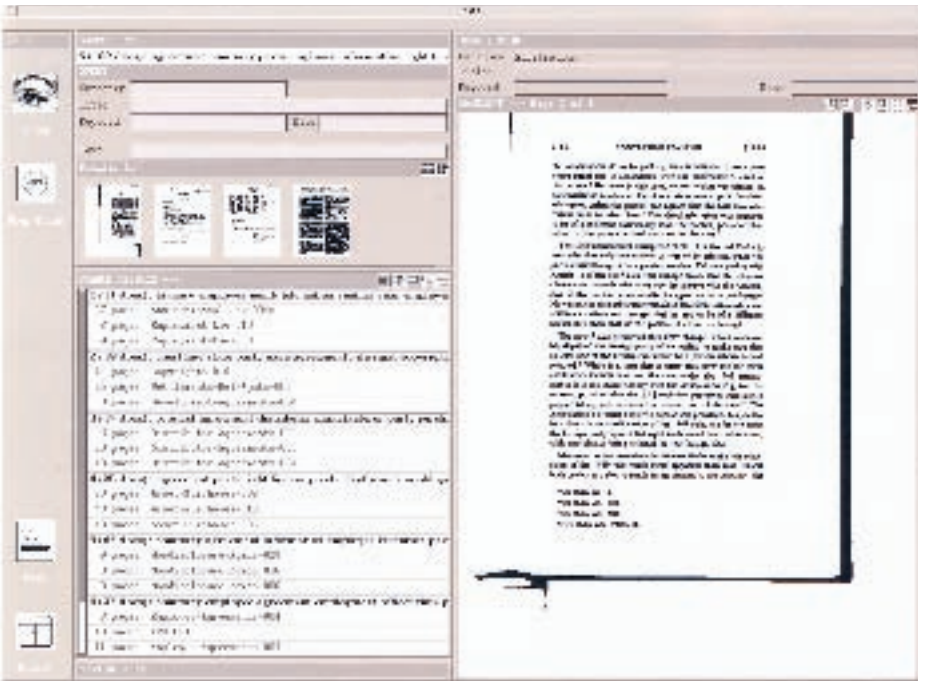


Figure 6. The Protofoil interface for accessing scanned documents using multiple search and browsing techniques

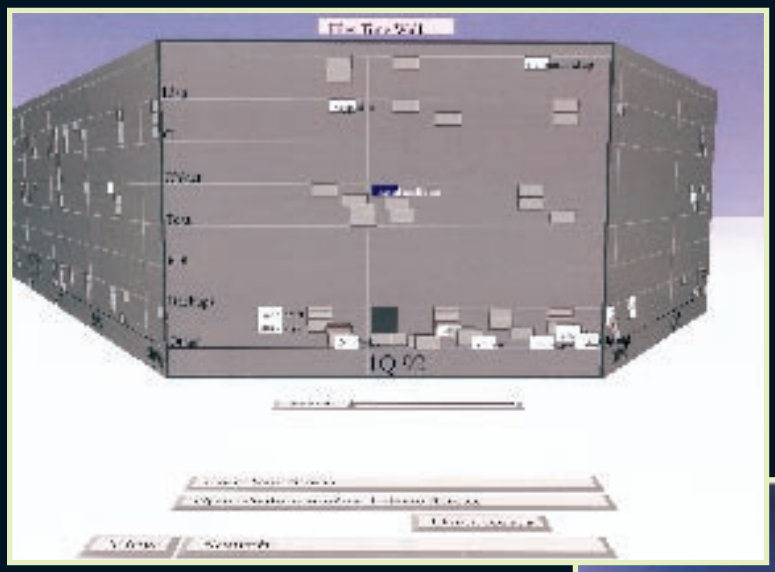


Figure 7.
Visualizations from the Information Visualizer: Perspective Wall

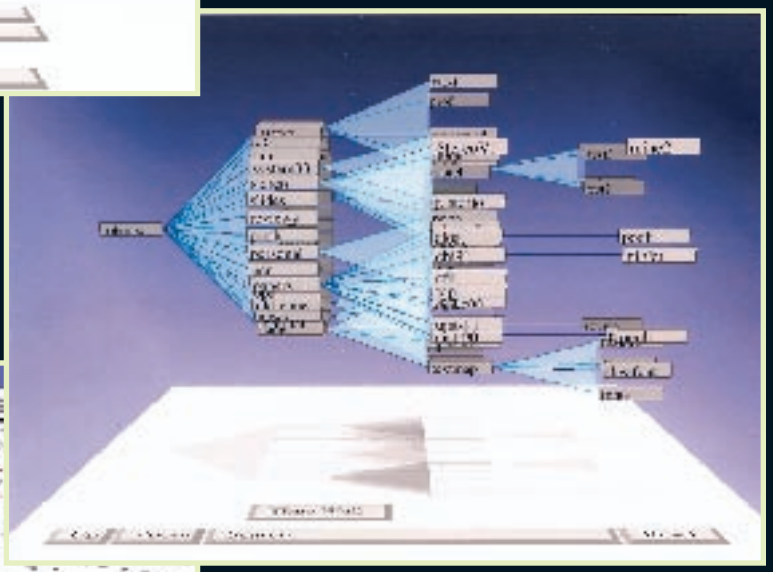


Figure 8.
Visualizations from the Information Visualizer: Cone Tree

Figure 9.
Visualizations from the Information Visualizer: Document Lens

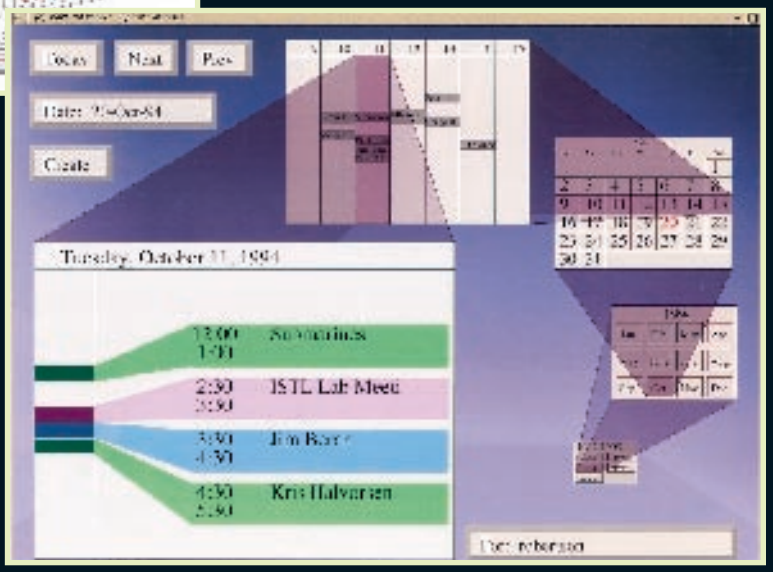


Figure 10.
Visualizations from the Information Visualizer: Spiral Calendar

cific visualizations with multiple search techniques.

In another line of work [17, 18, 20], we integrated a variety of retrieval and conventional result-viewing techniques. An incarnation of these ideas is Protofoil, a digital file cabinet application, which supports retrieval over the images and OCR-ed text of collections of scanned documents. The user interface, shown in Figure 6, has a tiled pane layout with three columns: the thin left column containing tools for invoking operations; the middle column for specifying retrieval requests and browsing through search results; and the right column for browsing documents. Document access is based on an iterative loop in which four user actions are variously interleaved: selecting a scope (initially the entire file cabinet); specifying a query; browsing results; and viewing and using documents.

Queries can be based on several search techniques, singly or jointly: category search, attribute search based on system- or user-supplied attributes of the items, textual similarity search based on new text or previous result items, and proximity search which matches items that contain a set or sequence of works within a given distance. Categories are the correlates of physical file folders or in a digital library context, perhaps a subject-based categorization system. Though category could be viewed as another attribute, it is often useful in a user interface to provide custom interactions based on metaphors of location or physicality.

In addition to multiple search methods, Protofoil supports multiple viewing methods for results and documents. Five views of search results are supported:

- Category Groups. A grouping of documents according to user-assigned category. This view is automatically selected on category search. The categories can be expanded to show the matching documents within each category. Since categories are assigned by the user, this view organizes search results in user-specific groupings.
- Clusters. A list of automatically-generated clusters of the result documents (as shown in the Results pane of Figure 6) displayed using the clustering technique of Scatter/Gather. Clusters can be expanded to show included documents. Clustering provides an alternative to manual categorization or groupings of documents.
- Thumbnails. An array of thumbnail images (as shown in the **Similar To** pane of Figure 6). Thumbnails (as opposed to icons) can recapture many of the visual cues of paper documents and can support rapid visual recognition and sometimes directly provide the information sought.
- Description. A list showing attributes of each document similar to the list view used commonly in file managers (e.g., the Macintosh Finder). This view is automatically selected on similarity searches and is augmented with scores and matching words.
- Snippets. A list of documents returned by a proximity search along with matching locations displayed using the Snippet Search design.

Other views can be incorporated into Protofoil, for example, the Tilebar display. An important aspect of this is that the views and search methods are integrated in the iterative access loop in several ways. First, the views are based on displaying items at varying degrees of granularity—document groups, documents, and pieces of documents—which corresponds to the granularity of the query operation. Second, appropriate views and customizations are automatically selected based on the query type. Finally, views also play a central role in scope and query refinement. For example, any result set or a subset (e.g., category groups or cluster groups) can be used as the scope of a further cycle of query. Thus, the Scatter/Gather interaction loop is a subcase of the general access loop in which clusters are added to the next scope and a **match all** search is performed into the clustering view.

Another important problem for workspaces is handling large information sets. Many query results or other information sets in digital library settings are too large to be effectively handled using conventional views. Cluster summaries are one solution to this problem, but another is the use of new kinds of visualizations that handle larger information sets. In the Information Visualization project, we have explored novel uses of 2-D and 3-D graphics and animation to increase both the amount of information that users

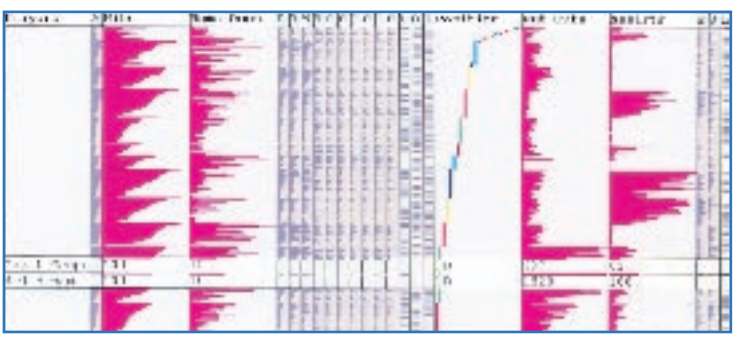
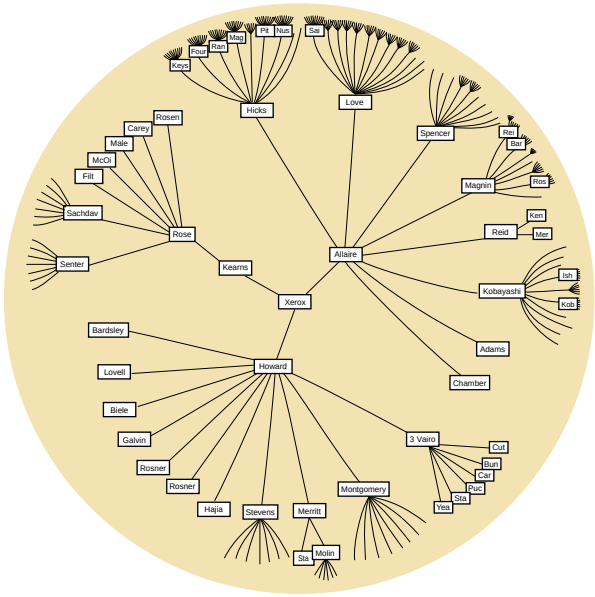


Figure 11. The Table Lens

can manipulate in a workspace at once and also the facility with which they can find patterns and make sense of the information. The essential principle underlying the effectiveness of many of these techniques is that many parts of a task can be off-loaded from the user's cognitive system onto their perceptual system, thus enabling more rapid processing.

Visualizations (shown in Figures 7 through 12) have been developed for displaying hierarchical, linear, document, temporal, and tabular structures [10, 11, 13, 16, 22, 23]. All these visualizations are focus+context techniques which display some portion of the information at a greater level of detail while still displaying all or much of the context. Dynamic control of the focus facilitates exploration of the information to find patterns in the big picture as well as to investigate interesting details. Since these visualizations are suited to canonical organizations of information, they can be applied broadly. In the digital library context, they can be used for visualizing sets of sources, documents, queries, and meta-information. For example, they could be integrated into the result viewing area of Protofoil or used to visualize the entire collection of articles pulled into the Butterfly workspace. By moving information sets across these various visualizations, different attributes can be examined and used to filter or understand the information.



D.T. Kearns, CEO, Chairman, Chief Executive Officer *667 — 135*

Figure 12. The Hyperbolic Tree Browser

In addition, the variety of visualizations allows flexible interactions with the information according to the needs of different tasks or parts of tasks. For example, page thumbnails have consistently proven their value during Protofoil usage. In a field study, Protofoil was placed in a lawyer's office with approximately 25% of his files scanned in. The lawyer and a colleague were in some cases able to use thumbnails to identify documents or directly address some information need. Besides identifying document types, the lawyer was able to recognize particular features of some documents (e.g., the shape of company letterheads). To provide further support for document image usage, intermediate page image sizes were introduced to allow the lawyer to browse through a document at a level appropriate to his purposes (e.g., locating a particular page vs. reading the text). Thus, it was clear that the direct physical rendering of pages at various reductions retained some of the visual advantages of paper documents. Similarly, we believe that three-dimensionality and animated transitions across interactions with a visualization provide additional opportunities for conveying the rich spatial, structural, and contextual properties of information.

Conclusion

The explosive growth in networked connectivity and rapid advances in computing power are replacing the older notion of standalone information utilities with newer notions of interconnected digital libraries. In the digital library, users will need to work simultaneously with multiple distributed information sources that differ in content, form, and source types: personal files, workgroup or enterprise resources, commercial databases, and public archives. They will need to understand displays of large sets of objects from these sources, organized in ways that allow them to make rapid decisions about what is relevant and what overall patterns exist among the objects. They will need to manipulate the access results and incorporate the information into their larger tasks. In short, the digital library will require effective means for the amplification of information-intensive work.

In this context, conventional notions of information retrieval limit the kind of interaction that is typical of work as observed in actual work. The wide variety of ways documents are arranged, sought, and used in physical workspaces strongly suggests the importance of providing an equally rich and flexible environment for interacting with information in electronic workspaces. We have illustrated some of our results to date in conceptualizing and developing components for rich interaction.

Achieving the goal of rich interaction is not limited to the design of a user interface narrowly construed, but rather has consequences that permeate the design of all components of the digital library. The examples in this article cross-cut a number of the usual layers of digital library architectures including retrieval engines, access protocols, intermediate services, and user inter-

faces. Retrieval engines need to support a much broader range of operations that allow the user to understand properties of entire collections, groups of documents, or documents themselves, thus more effectively supporting query refinement as well as other stages of the task. Access protocols and intermediate servers should provide various forms of bridge or meta-services that explicitly reveal the characteristics of the information and the mediating access mechanisms, thus allowing effective and memorable rendering of sources and their contents. Finally, the user interface should be thought of as a complete workspace in which a variety of tools are available for flexibly interacting with richly-conveyed information.

Acknowledgments.
A number of current and former colleagues at PARC have contributed to the work which we describe here including Doug Cutting, David Karger, John Lamping, Peter Pirolli, Steve Putz, Anand Rajaraman, Dan Russell, and John Tukey. The Protofoil field study was conducted with Jeanette Blomberg, Lucy Suchman, and Randy Trigg. ◻

References

1. Case, D.O. Conceptual organization and retrieval of text by historians: The role of memory and metaphor. *J. Amer. Soc. for Info. Sci. 42*, 9 (1991), 657–668.
2. Cutting, D.R., Halvorsen, P.K., Pedersen, J.O., and Withgott, M. Information theater versus information refinery. In *AAAI Spring Symposium on Text-based Intelligent Systems*. Stanford University, Stanford, CA, March 1990. Also available as Xerox PARC Tech. Rep. SSL-89-101.
3. Cutting, D.R., Karger, D.R., and Pedersen, J.O. Constant interaction-time scatter/gather browsing of very large document collections. In *Proceedings of SIGIR'93*, June 1993.
4. Cutting, D.R., Karger, D.R., Pedersen, J.O., and Tukey, J.W. Scatter/gather: A cluster-based approach to browsing large document collections. In *Proceedings of SIGIR'92*. June 1992. Also available as Xerox PARC Tech. Rep. SSL-92-02.
5. Erickson, T., and Salomon, G. Designing a desktop information system: Observations and issues. In *Proceedings of ACM Conference on Human Factors in Computing Systems*. (1991), pp. 49–54.
6. Harman, D. Overview of the first text retrieval conference. In *Proceedings of the 16th Annual International ACM/SIGIR Conference*. (Pittsburgh, PA, 1993), pp. 36-48.
7. Hearst, M.A. Multi-paragraph segmentation of expository text. In *Proceedings of the 32nd Meeting of the Association for Computational Linguistics*, June 1994.
8. Hearst, M.A. Tilebars: Visualization of term distribution information in full text information access. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Denver, CO, May 1995.
9. Frei, H.P., Bärtschi, M., and Jauslin, J.F. Caliban: Its user-interface and retrieval algorithm. Tech. Rep. 62, Institut für Informatik, ETH Zürich, April 1985.
10. Lamping, J., Rao, R., and Pirolli, P. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. (ACM, May 1995).
11. Mackinlay, J.D., Robertson, G.G., and Card, S.K. The perspective wall: Detail and context smoothly integrated. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, (ACM, April 1991), pp. 173-179.
12. Mackinlay, J.D., Rao, R., and Card, S.K. An organic user interface for searching citation links. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. (ACM, May 1995).
13. Mackinlay, J., Robertson, G., and Deline, R. Developing calendar visualizers for the information visualizer. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, Nov. 1994.
14. Malone, T.W. How do people organize their desks? Implications for the design of office information systems. *ACM Trans. Off. Info. Syst. 1*, 1 (1983), 99–112.
15. Pedersen, J.O., Cutting, D.R., and Tukey, J.W. Snippet search: A single phrase approach to text access. In *Proceedings of the 1991 Joint Statistical Meetings*. American Statistical Association, 1991. Also available as Xerox PARC Tech. Rep. SSL-91-08.
16. Rao, R., and Card, S.K. The table lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, April 1994.
17. Rao, R., Card, S.K., Jellinek, H.D., Mackinlay, J.D., and Robertson, G.G. The information grid: A framework for building information retrieval and retrieval-centered applications. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, Nov. 1992.
18. Rao, R., Card, S.K., Johnson, W., Klotz, L., and Trigg, R. Protofoil: Storing and finding the information worker's paper documents in an electronic file cabinet. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, April 1994.
19. Rao, R., Janssen, B., and Rajaraman, A. Gaia technical overview. Tech. Rep., Xerox PARC, ISTL, 1994.
20. Rao, R., Russell, D.M., and Mackinlay, J.D. System components for embedded information retrieval from multiple disparate information sources. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, Nov. 1993.
21. Robertson, G.G., Card, S.K., and Mackinlay, J.D. Information visualization using 3-D interactive animation. *Commun. ACM 36*, 4 (Apr. 1993), 57–71.
22. Robertson, G.G., Mackinlay, J.D., and Card, S.K. Cone trees: Animated 3-D visualizations of hierarchical information. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, April 1991.
23. Robertson, G., and Mackinlay, J.D. The document lens. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM Press, Nov. 1993.
24. Salton, G. Automatic text processing: the transformation, analysis, and retrieval of information by computer. Addison-Wesley, Reading, Mass., 1988.

About the Authors:

The authors are researchers at the Xerox Palo Alto Research Center (PARC), where they participate in the Intelligent Information Access and Information Visualization projects. There is currently a broad set of activities at PARC related to digital libraries, particularly in the areas of document capture and analysis, multimedia, information access, and user interfaces. PARC researchers, including some of the authors, are involved in three of the NSF/DARPA/NASA Digital Libraries projects. **Authors' Present Address:** Xerox PARC, 3333 Coyote Hill Road, Palo Alto, CA 94304; email: {lastname}@parc.xerox.com