

Supporting Creativity with Search Tools

Bill Kules, University of Maryland

Introduction

Is searching creative? Searching and information seeking are part of the creative process. An architect looking for “seed” ideas for a new project may search an architecture database. Novelists, journalists and artists may similarly search the web for new ideas. A historian will explore archival material for a research project. Even graduate students may employ search as they refine and narrow their research topic. Information seeking models of the writing process acknowledge the creative elements, identifying specific stages for topic exploration and formation (Kuhlthau, 1992). Advertising art directors search for images as part of their creative process (Garber & Grunes, 1992). Engineers and software developers search for creative solutions to technical problems, too.

Developers of search tools have traditionally focused on searches in which the objective is clearly defined, such as known-item or fact searches. Typical web search engines and databases are now very effective at satisfying these searches with a simple ranked list of results. In the context of a creative task, however, the information need may be only partially specified or ambiguous, and the searcher may not be familiar with terminology in the domain or collection being searched.

Four kinds of information have been proposed as aiding the creative process (Bawden, 1986): Interdisciplinary information, peripheral information, speculative information, and exceptions and inconsistencies. Creative searches can embody (at least) the following four characteristics. Within the characteristics we propose techniques that may support creative search by helping searchers encounter these types of information:

- Generative goals – Rather than finding a specific document or fact, the goal of the search can be to learn about a topic area, develop a question, generate ideas, identify unusual items (“outliers”) or even look for conflict or inconsistency. Queries may be ambiguous or partially defined.
 - Meaningful and stable categories
 - Variable categories, including clustering/clustered overviews
- Cross-context – Searches may extend across domains or collections. Indeed the searcher may deliberately search in unfamiliar domains to gain alternative perspectives on a topic or discover previously unnoticed connections.
 - Literature visualization
 - Literature linking
- Exploratory and iterative – Searchers interactively explore the search results by browsing, filtering or other techniques. They issue multiple, successive queries as their information need evolves, possibly across multiple sessions. They exhibit “berry-picking” or “information foraging” behavior as they collect useful bits of information, follow promising threads and identify new information sources.
 - History mechanisms
 - Workspaces
 - Enhanced bookmarks

- Serendipity, non-linearity – Serendipitous findings can provide valuable insight for the creative searcher

Because of the variety of creative tasks, this list is certainly incomplete, and no single task is likely to embody all characteristics.

Review of Search Interfaces

This section considers the first two bullets and illustrates features that we propose will support creative search. For a more comprehensive review of search visualization interfaces, see Hearst (1999).

1.1. Ranked Search Results

Typical web search engines are optimized to support known item and fact search by ranking documents according to query relevance, link analysis, popularity or combinations of several metrics. Google does provide a link to “similar pages,” which allows users to quickly find documents that satisfy a similarity metric.

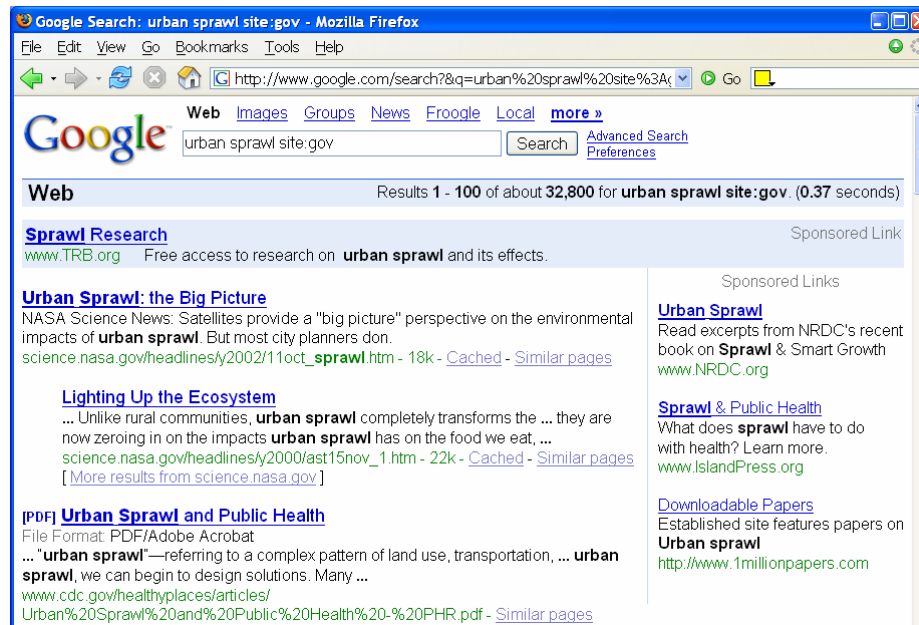


Figure 1. The Google interface shows the top search results for the query “urban sprawl site:gov” as a ranked list (www.google.com).

1.2. Organizing by Meaningful and Stable Categories

Interactive overviews of categorized web search results using meaningful and stable classifications can support user exploration, understanding of large result sets, and discovery. The categories can be drawn from large thesauri, glossaries, or ontologies. Alternatively, they can be based on simple categorization schemes such as document type, country codes or ranges of document size. They support the “search and browse” process that is typical of exploratory searches by providing a consistent organizing structure for keyword searches. When used to filter ranked search results (Figure 2), users found pages of interest deeper within the results and noticed more unexpectedly missing results – that is categories with no associated results – compared to a control interface (Kules & Shneiderman, submitted).

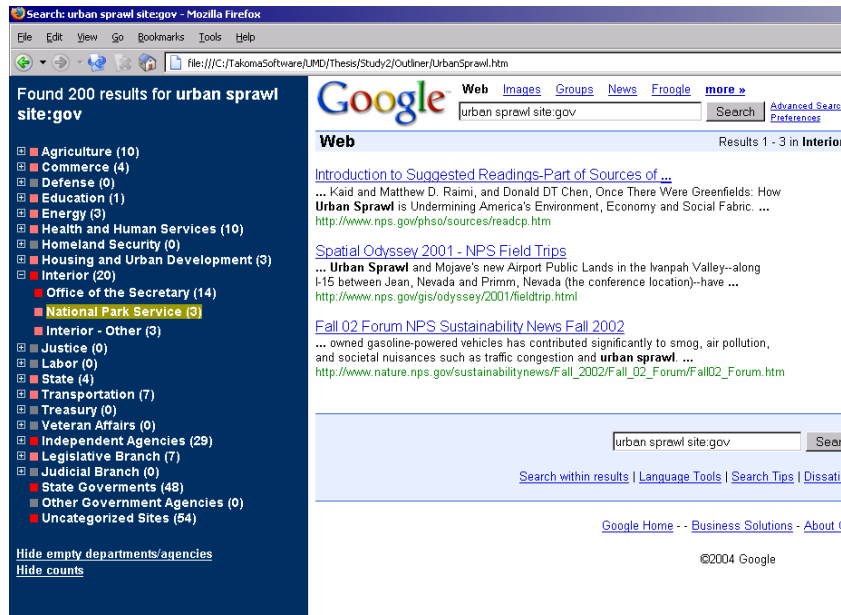


Figure 2. This overview+detail interface shows the top 200 results for the query “urban sprawl site:gov.” They have been categorized into a two-level government hierarchy, which is used to present a categorized overview on the left. The Interior Department, which has 20 results, has been expanded and the National Park Service has been selected. The effect on the right side is to show just the three results from the Park Service (Kules & Shneiderman, submitted).

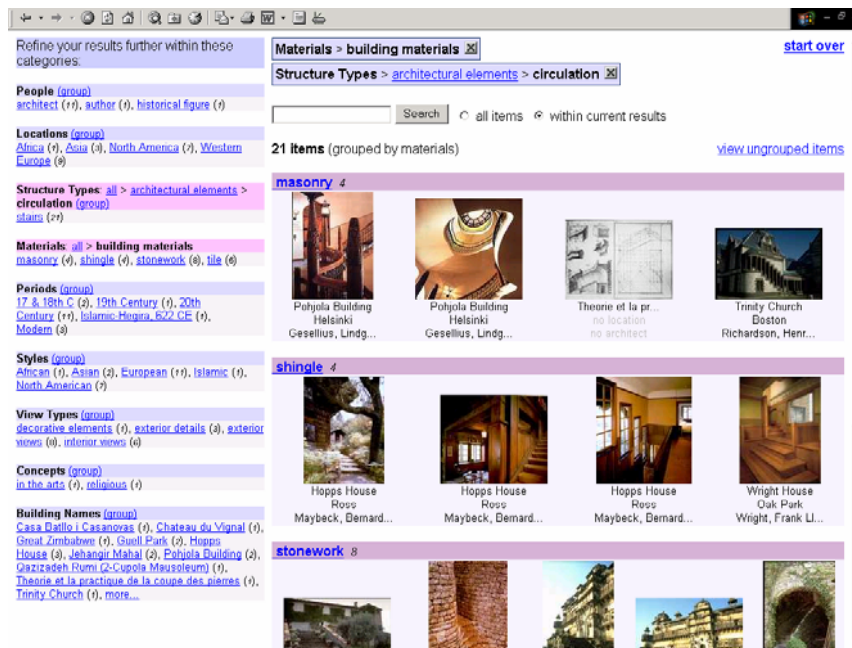


Figure 3. Flamenco uses multiple sets of hierarchical categories (“hierarchical faceted metadata”) to organize guide browsing and searching. In this figure, the user has filtered architectural images to show images that represent both *building materials* and *circulation elements* (Yee, Swearingen, Li, & Hearst, 2003).

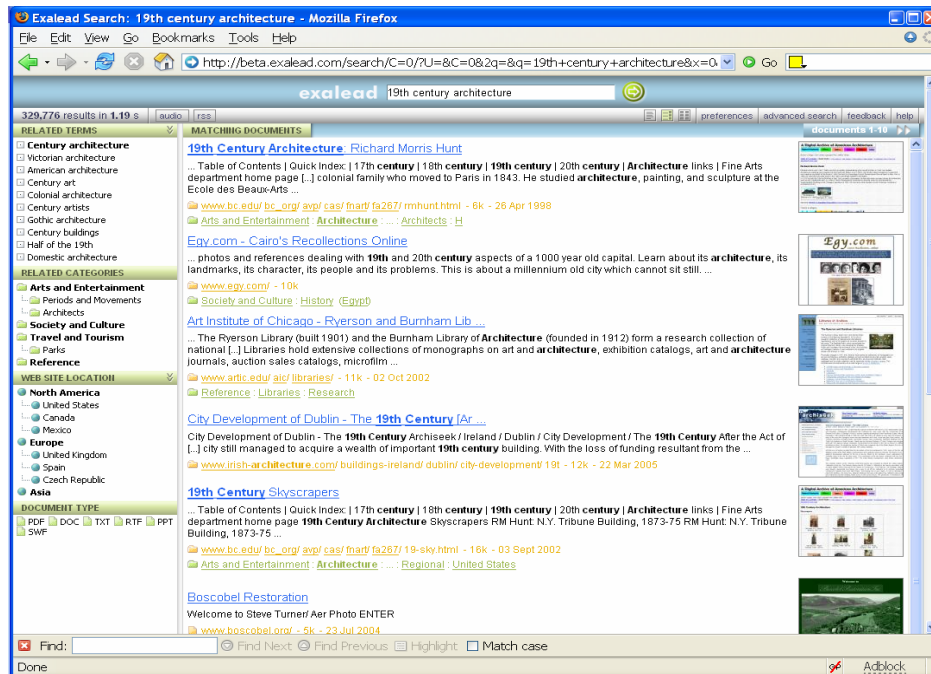


Figure 4. Exalead produces categorized overviews using topical categories as well as categories based on geography and document type. (www.exalead.com)

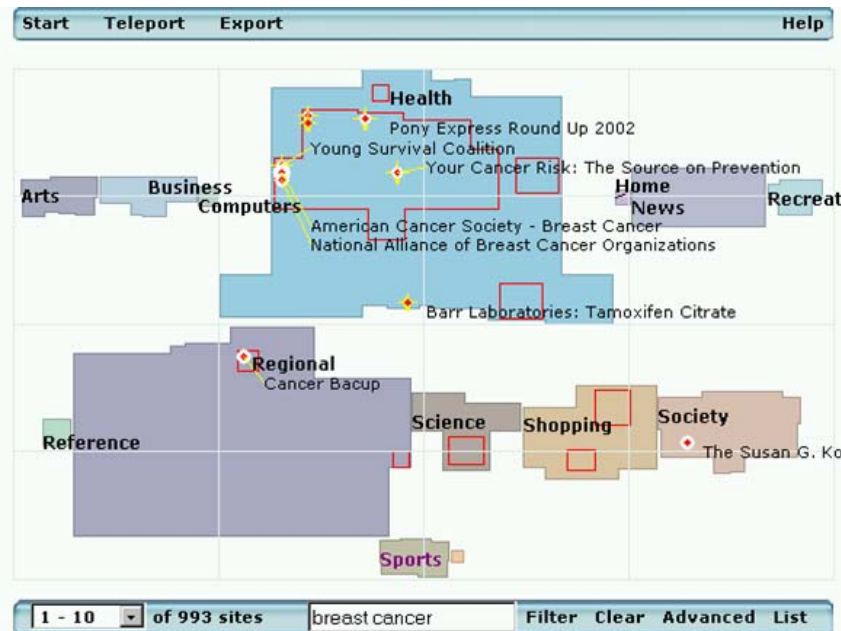


Figure 5. Antarctica uses an abstract two-dimensional map of hierarchical categories to display search results and support browsing. Here, the results of a query on “breast cancer” have been displayed. The top 10 results are plotted on the map using coded icons with titles. Users can click on a region to “zoom into” that category (www.antarcti.ca).

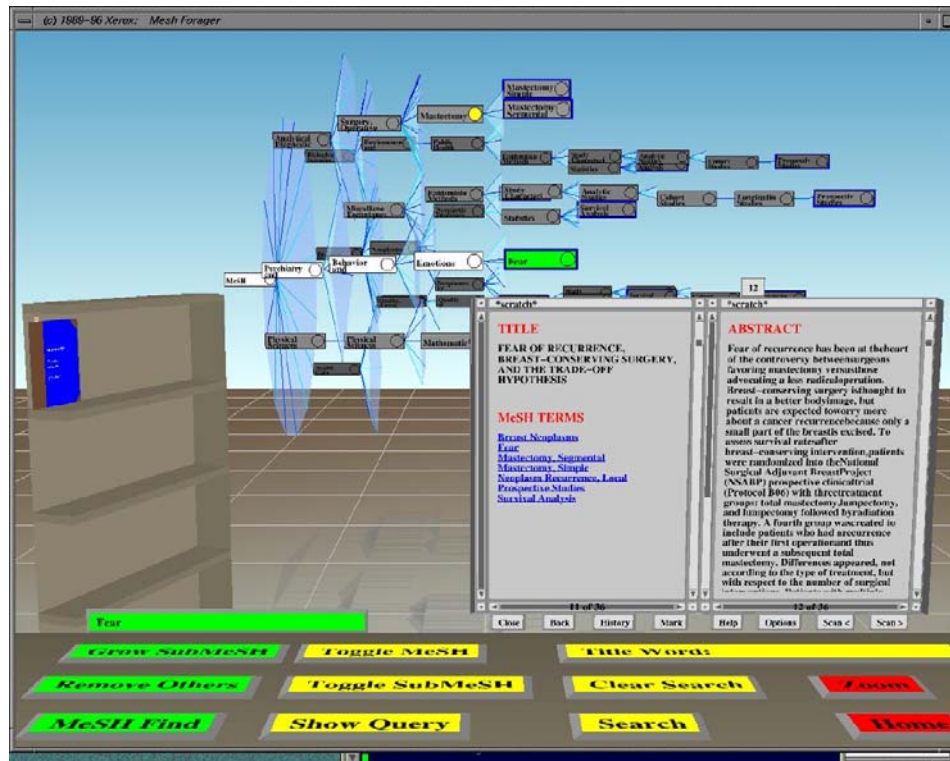


Figure 6. Cat-a-Cone supports search and browse in MeSH, a very large hierarchy of terms used to classify medical research reports. The hierarchy is displayed as a cone tree, which users can interactively navigate through. They can also issue a query, whereupon the tree is pruned to show only categories with matching documents (Hearst & Karadi, 1997).

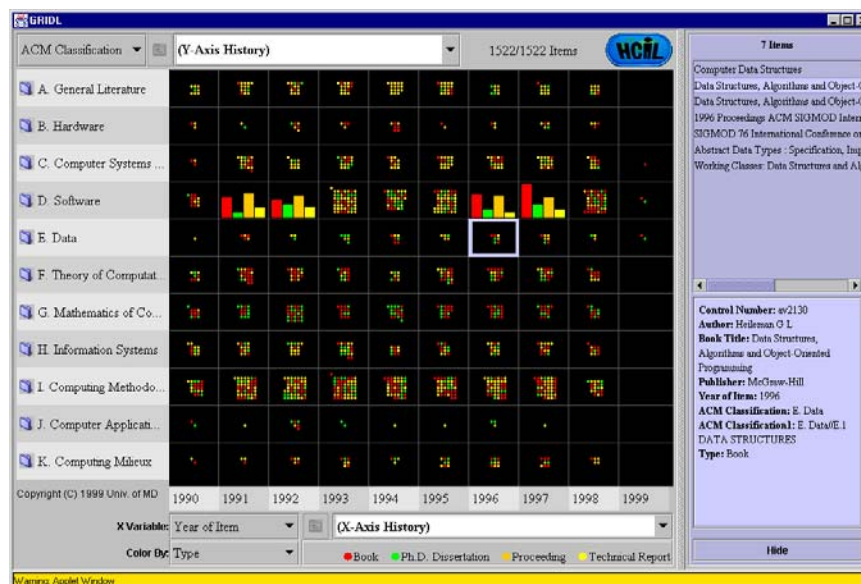


Figure 7. GRIDL uses categorical variables to organize search results on a two-dimensional grid. Here the user has organized computer science documents along according to the ACM classification (vertical) and year of publication (horizontal). At each grid point clusters of color-coded dots represent the documents and show a third attribute, the document type. If there are more than 49 documents in a grid point a bar chart summarizes the by document type. Users see the entire result set and can then click on labels to move down a level in the hierarchy.

Multiple sets of categories can be used to support conjunctive filters while retaining the benefit of stable organization, and helping users avoid feeling “lost” in the information space (English, 2002).

Categorical metadata can be represented using graphical displays. Search results can be displayed on an abstract map (a two-dimensional space) based on a hierarchy of categories (Figure 5). Three-dimensional displays have been used to visualize very large hierarchies

A matrix can be used to organize results along two categorical or numeric dimensions (Kunz, 2003; Kunz & Botsch, 2002; Shneiderman, Feldman, Rose, & Grau, 2000).

1.3. Organizing by Variable Categories

Variable categories, produced by clustering search results into dynamically generated categories, can be used in place of stable categories to produce similar displays of search results. Clustering has been found helpful for search tasks, although searchers sometimes fail to understand the clusters or their labels. Variable categories can be used in overview+detail interfaces (Figures 8 and 9) or visual maps (Figures 10 and 11).

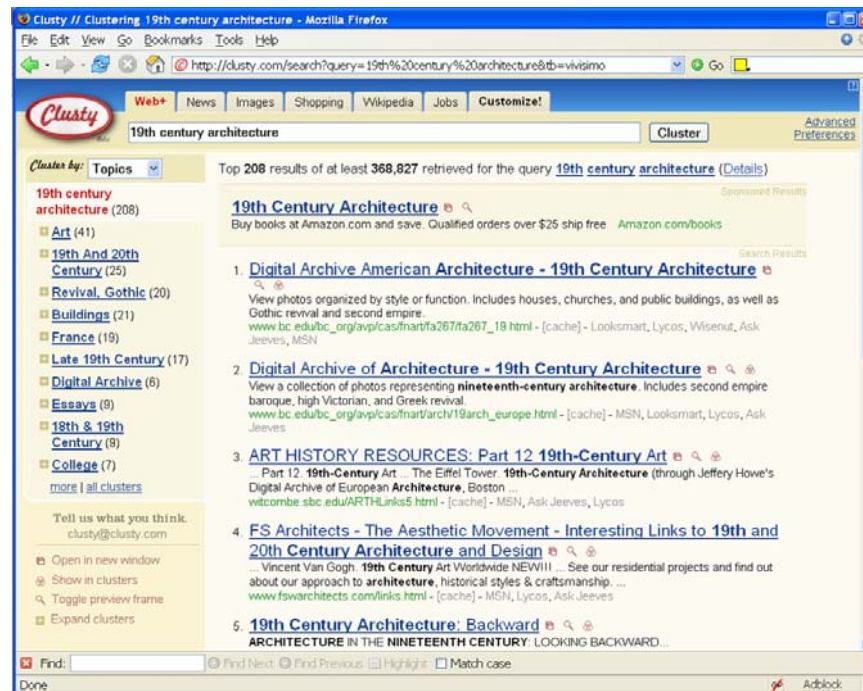


Figure 8. The metasearch engine Clusty (and its predecessor Vivisimo) uses a form of automated document clustering that generates hierarchies of concisely labeled clusters. In this example, the top 208 results have been clustered, and the cluster labels have been used to generate an overview with 10 categories initially visible. Users can show more categories or filter and navigate the results using the expandable outline (www.clusty.com).

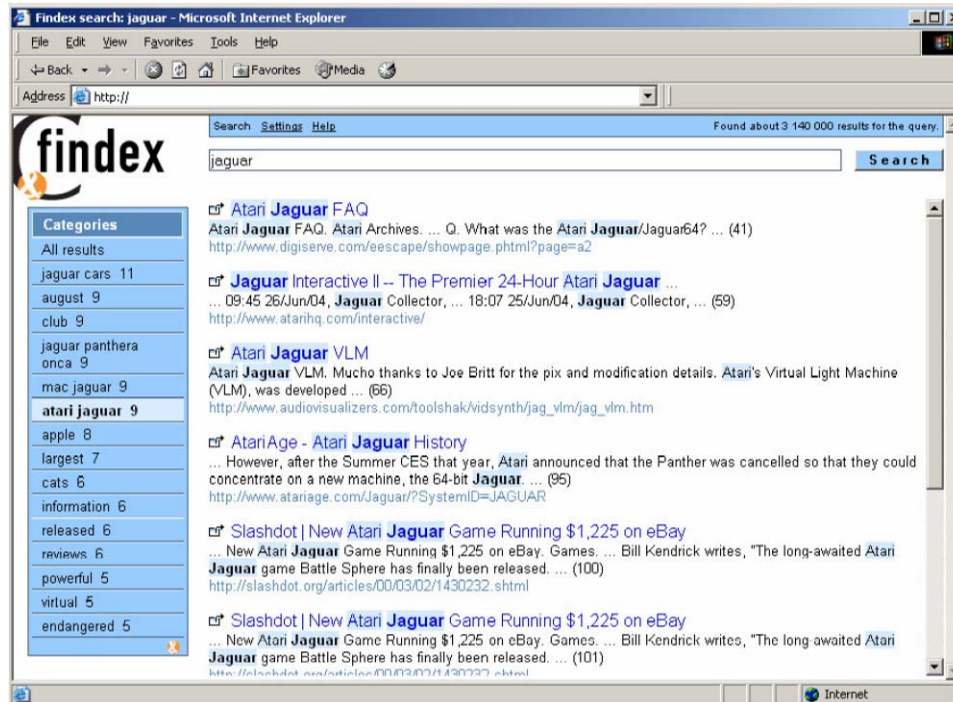


Figure 9. Findex clusters documents into a flat set of categories. Here the results from the query “jaguar” have been clustered into 15 categories. The “atari jaguar” category has been selected (Käki, 2005).

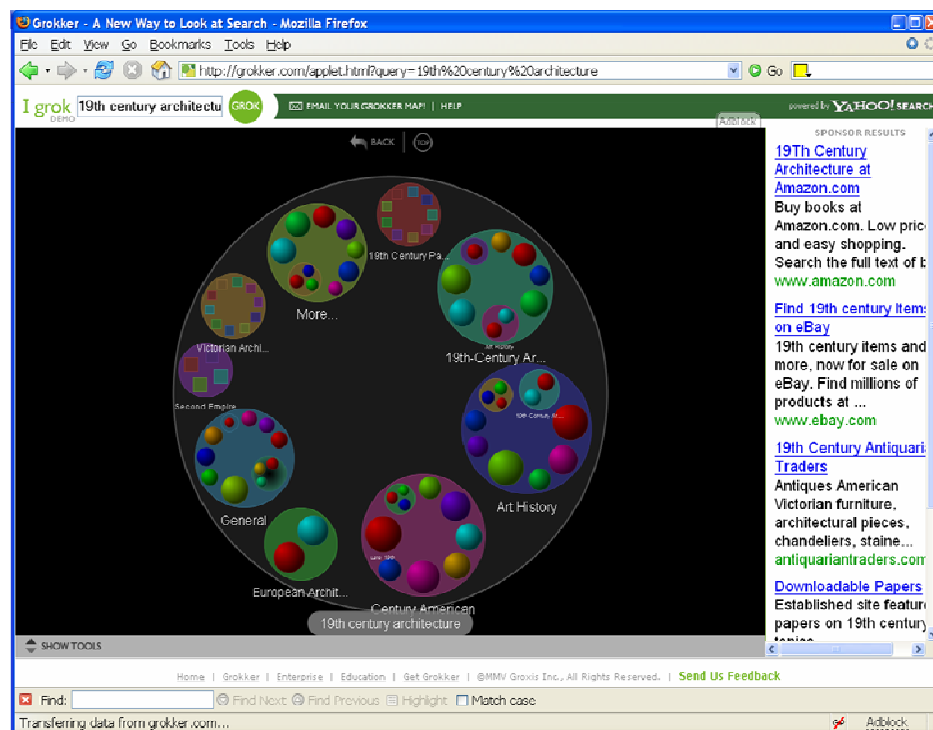
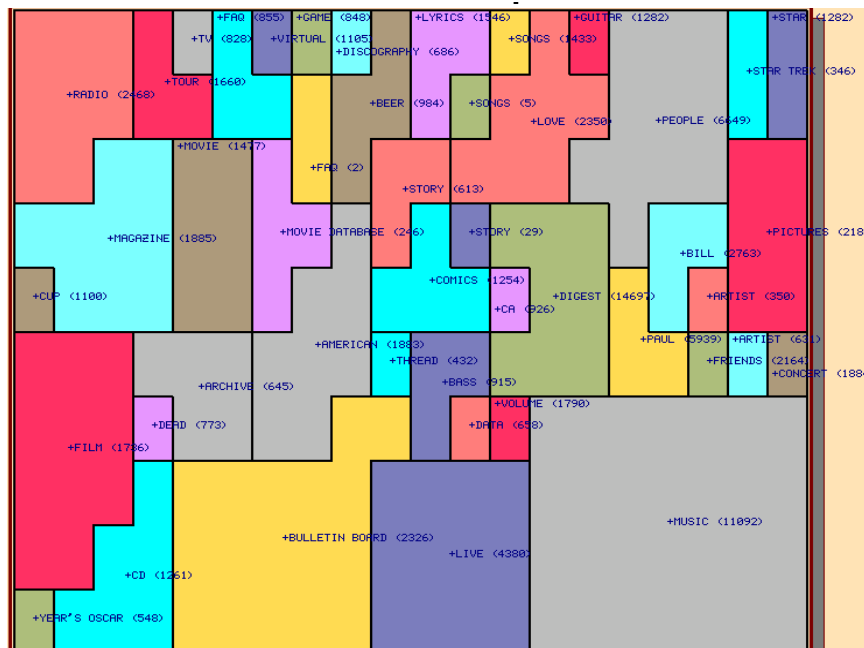


Figure 10. Grokker generates hierarchical clusters and displays those clusters using concentric circles. Users can drill down into clusters to explore the results.



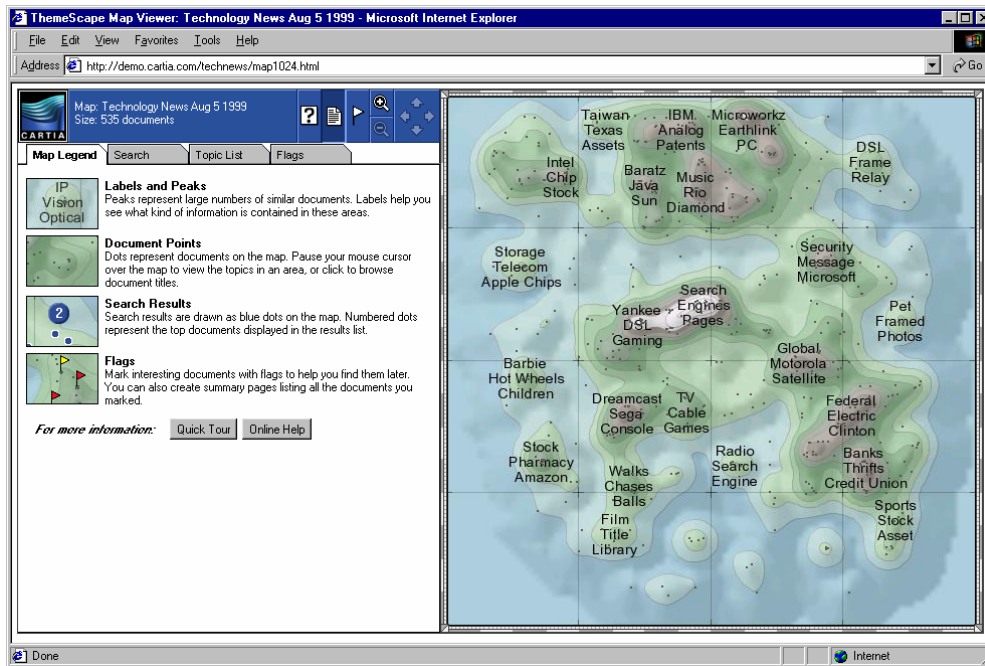


Figure 13. Themescape uses a topographic map metaphor to plot keywords extracted from a corpus.

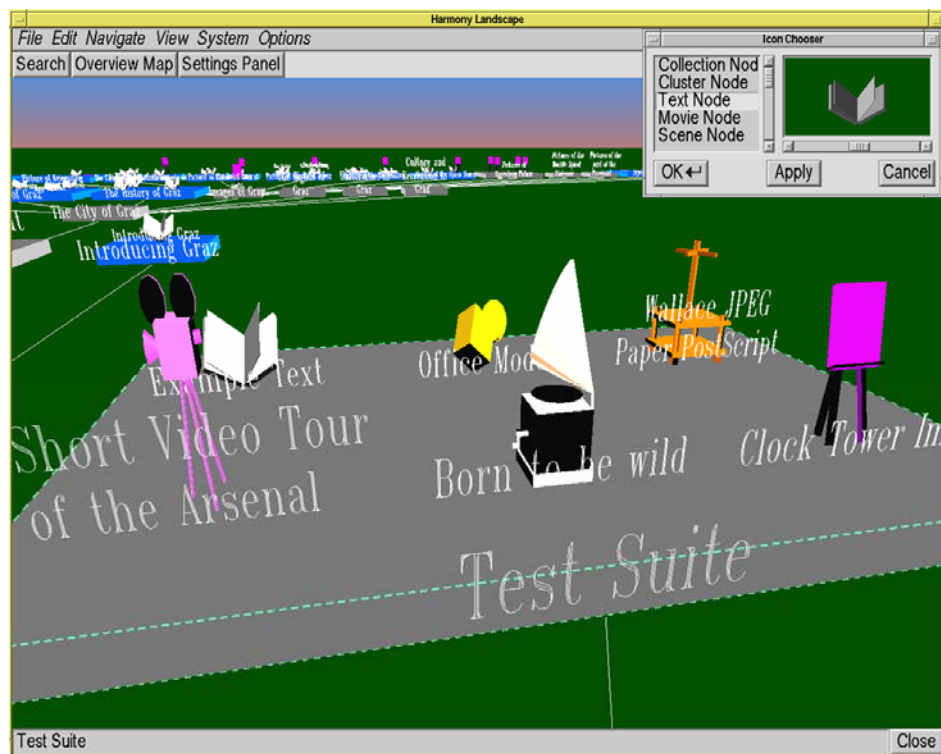


Figure 14. The Harmony Landscape visualizes an information space on a receding plane.

1.4. Literature visualization

Literature visualization tools provide overviews of a knowledge domain or field of research by visualizing bibliographic attributes such as citations between articles or common themes. Co-citation networks visualize citations between papers by significant authors in a field. They can graphically illustrate major topics and sub-fields (Chen, 1999). These maps may help searchers bridge multiple fields or identify trends.

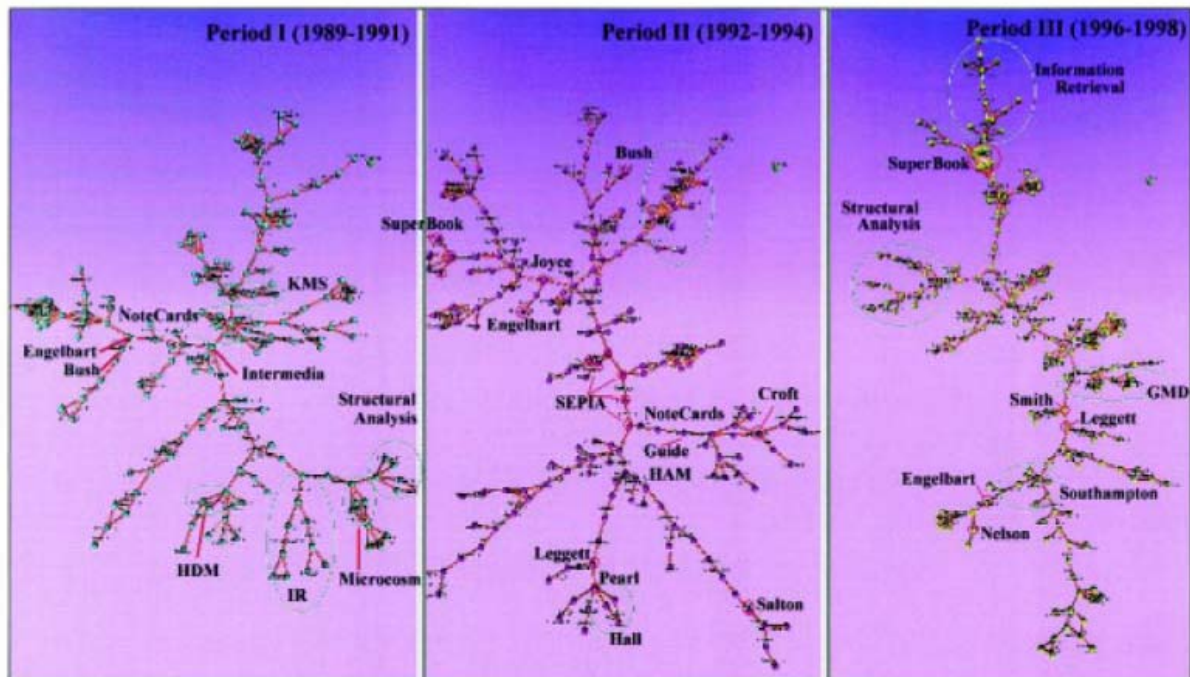


Figure 15. This co-citation map is derived from a collection of papers on hypertext. It shows three snapshots the hypertext field. Relationships between major topics and authors can be seen. (Chen, 1999)

1.5. Literature linking

Literature linking is a specialized form of creative information seeking that attempts to discover new connections between two literatures. It has been used to identify hidden connections in the medical literature between migraines and magnesium by citation analysis and manual review of terms common in both literatures (Swanson, 1988). The LitLinker system (Pratt & Yetisgen-Yildiz, 2003) is a recent example of this technique. It provides a user interface that allows searchers to select starting and target terms within a set of literatures, and interactively explore potential links.



Figure 16. LitLinker (<http://litlinker.ischool.washington.edu>)

Conclusion

The interfaces described above may help information seekers by doing more than simply displaying ranked lists of search results – by exposing them to information that will help the creative process. The visual presentation of information has a powerful impact on what is perceived, particularly with the information visualization techniques illustrated here, and could prove to be useful tools for the creative information seeker.

References

- Bawden, D. (1986). Information systems and the stimulation of creativity. *Journal of Information Science*, 12, 203-216.
- Chen, C. (1999). Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing and Management*, 35(3), 401-420.
- English, J., Hearst, M., Sinha, R., Swearington, K., and Yee, P. (2002). *Flexible Search and Navigation using Faceted Metadata*. Unpublished manuscript.
- Garber, S. R., & Grunes, M. B. (1992). The art of search: a study of art directors. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 157-163.
- Hearst, M. (1999). User interfaces and visualization. In R. Baeza-Yates & B. Ribeiro-Neto (Eds.), *Modern Information Retrieval* (pp. 257-323). Reading, MA: Addison-Wesley.
- Hearst, M. A., & Karadi, C. (1997). Cat-a-Cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. *Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval*, 246-255.
- Käki, M. (2005). Findex: search result categories help users when document ranking fails, *Proceeding of the SIGCHI conference on Human factors in computing systems*. Portland, Oregon, USA: ACM Press.

- Kuhlthau, C. (1992). *Seeking meaning: a process approach to library and information services*. Norwood, New Jersey: Ablex Publishing.
- Kules, B., & Shneiderman, B. (submitted). Using meaningful and stable categories to support exploratory web search: Two formative studies.
- Kunz, C. (2003). SERGIO - An Interface for context driven Knowledge Retrieval. *Proceedings of eChallenges, Bologna, Italy, 2003*.
- Kunz, C., & Botsch, V. (2002). Visual Representation and Contextualization of Search Results – List and Matrix Browser. *Proceedings of Dublin Core '02*.
- Pratt, W., & Yetisgen-Yildiz, M. (2003). LitLinker: capturing connections across the biomedical literature. *Proceedings of the international conference on Knowledge capture*, 105-112.
- Shneiderman, B., Feldman, D., Rose, A., & Grau, X. F. (2000). *Visualizing Digital Library Search Results with Categorical and Hierarchical Axes*. Paper presented at the Proc. 5th ACM International Conference on Digital Libraries (San Antonio, TX, June 2-7, 2000).
- Swanson, D. R. (1988). Migraine and magnesium: eleven neglected connections. *Perspect. Biol. Med.*, 31, 526-557.
- Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 401-408.