Flexible Search and Navigation using Faceted Metadata

Jennifer English, Marti Hearst, Rashmi Sinha, Kirsten Swearingen, Ka-Ping Yee
School of Information Management & Systems
University of California, Berkeley
Berkeley, CA 94720-4600
{jenglish,hearst,sinha,kirstens,ping}@sims.berkeley.edu

ABSTRACT

We have developed an innovative search interface that allows non-expert users to move through large information spaces in a flexible manner without feeling lost. The design goal was to offer users a "browsing the shelves" experience seamlessly integrated with focused search. Key to achieving our goal is the explicit exposure of hierarchical faceted metadata in a manner that is intuitive and inviting to users. After several iterations of design and testing, the usability results are strikingly positive. We believe our approach marks a major step forward in search user interfaces and can serve as a model for web-based collections of up to 100,000 items.

Topics: Search User Interfaces, Faceted Metadata

INTRODUCTION

Although general Web search is steadily improving [30], studies show that search is still the primary usability problem in web site design. A recent report by Vividence Research analyzing 69 web sites found that the most common usability problem was poorly organized search results, affecting 53% of sites studied. The second most common problem was poor information architecture, affecting 32% of sites [27].

Studies of search behavior reveal that good search involves both broadening and narrowing of the query, appropriate selection of terminology, and the ability to modify the query [31]. Still others show that users often express a concern about online search systems since they do not allow a "browsing the shelves" experience afforded by physical libraries [6] and that users like well-structured hyperlinks but often feel lost when navigating through complex sites [23].

Our goals are to support search usability guidelines [28], while avoiding negative consequences like empty result sets or feelings of being lost. We are especially interested in large collections of similar-style items (such as product catalog sites, sites consisting of collections of images, or text documents on a topic such as medicine or law). Our approach is to follow iterative design practices from the field of human-computer interaction [29], meaning that we first assess the behavior of the target users, then prototype a system, then assess that system with target users, learn from and adjust to the problems found, and repeat until a successful interface is produced.

We have applied this method to the problem of creating an information architecture that seamlessly integrates navigation and free-text search into one interface. This system builds on earlier work that shows the importance of query previews [25] for indicating next choices (thus allowing the user to use recognition over recall) and avoiding empty result sets. The approach makes use of faceted hierarchical metadata (described below) as the basis for a navigation structure showing next choices, providing alternative views, and permitting refinement and expansion in new directions, while at the same time maintaining a consistent representation of the collection's structure [14]. This use of metadata is integrated with free-text search, allowing the user to follow links, then add search terms, then follow more links, without interrupting the interaction flow. Our most recent usability studies show strong, positive results along most measured variables.

An added advantage of this framework is that it can be built using off-the-shelf database technology, and it allows the contents of the collection to be changed without requiring the web site maintainer to change the system or the interface. For these reasons, we believe these results should influence the design of information architecture of information-centric web sites.

In the following sections we define the metadata-based terminology, describe the interface framework as applied to a collection of architectural images, report the results of usability studies, discuss related work, and discuss the implications of these results.

METADATA

Content-oriented category metadata has become more prevalent in the last few years, and many people are interested in standards for describing content in various fields (e.g., Dublin Core and the Semantic Web[1]). Web directories such as Yahoo and the Open Directory Project[2] are familiar examples of the use of metadata for navigation structures. Web search engines have begun to interleave search hits on category labels with other search results. Many individual collections already have rich metadata assigned to their contents; for example, biomedical journal articles have on average a dozen or more content attributes attached to them. Metadata for organizing content collections can be classified along several dimensions:

- The metadata may be *faceted*, that is, composed of orthogonal sets of categories. For example, in the domain of architectural images, some possible facets might be Materials (concrete, brick, wood, etc.), Styles (Baroque, Gothic, Ming, etc.), View Types, People (architects, artists, developers, etc.), Locations, Periods, and so on.
- The metadata (or an individual facet) may be *hierarchical* ("located in Berkeley, California, United States") or *flat* ("by Ansel Adams").
- The metadata (or an individual facet) may be *single-valued* or *multi-valued*. That is, the data may be constrained so that at most one value can be assigned to an item ("measures 36 cm tall") or it may allow multiple values to be assigned to an item ("uses oil paint, ink, and watercolor").

We note that there are a number of issues associated with creation of metadata itself which we are not addressing here. The most pressing problem is how to decide which descriptors are correct or at least most appropriate for a collection of information. Another problem relates to how to assign metadata descriptors to items that currently do not have metadata assigned. We will not be addressing these issues, in part because many other researchers already are, and because the fact remains that there are many existing, important collections whose contents have hierarchical metadata already assigned.

RECIPE USABILITY STUDY

We are particularly concerned with supporting non-professional searchers in rich information seeking tasks. Specifically we aim to answer the following questions: do users like and understand flexible organizations of metadata from different hierarchies? Are faceted hierarchies preferable to single hierarchies? Do people prefer to follow category-based hyperlinks or do they prefer to issue a keyword-based query and sort through results listings?

[1]http://dublincore.org, http://www.w3.org/2001/sw

[2]http://www.yahoo.com, http://dmoz.org

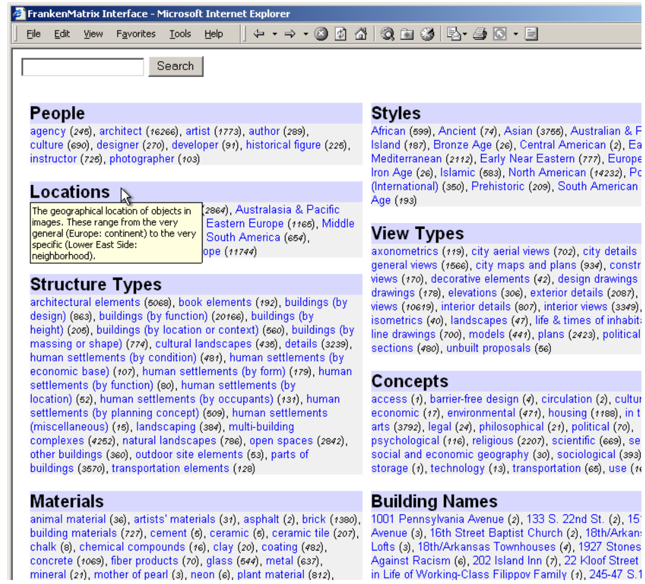

Figure 1: The opening page for both interfaces shows a text search box and the first level of metadata terms. Hovering over a facet name yields a tooltip (here shown below Locations) explaining the meaning of the facet.

Before developing our system, we tested the idea of using hierarchical faceted metadata on an existing interface that exemplified some of our design goals. This preliminary study was conducted using a commercial recipe web site called Epicurious[3] containing five flat facets, 93 metadata terms, and approximately 13,000 recipes. We compared the three available search interfaces:(1) Simple keyword search, with unsorted results list (2) Enhanced search form that exposes metadata using checkboxes and drop-down lists, with unsorted results list. (3) Browse interface that allows user to navigate through the collection, implicitly building up a query consisting of an AND across facets; Selecting a category within a facet (e.g., Pasta within Main Ingredient) narrows results set, and users are shown query previews at every step.

In the interests of space, we can only provide a brief summary of this small (9 participant) study: All the participants who liked the site (7 out of 9) said they were likely to use the browse interface again. Only 4 said this about enhanced search and 0 said this about simple search. Participants especially liked the browse interface for open-ended tasks such as "plan a dinner party." We took this as encouraging support for the faceted metadata approach. However, the recipe browse facility is lacking in several ways. Free-text search is not integrated with metadata browse, the collection and metadata are of only moderate size, and the metadata is organized into flat (non-hierarchical) facets. Finally users are only allowed to refine queries, they cannot broaden

[3]http://eat.epicurious.com/recipes/browse_home/

Figure 2: Matrix View middle game (items grouped).

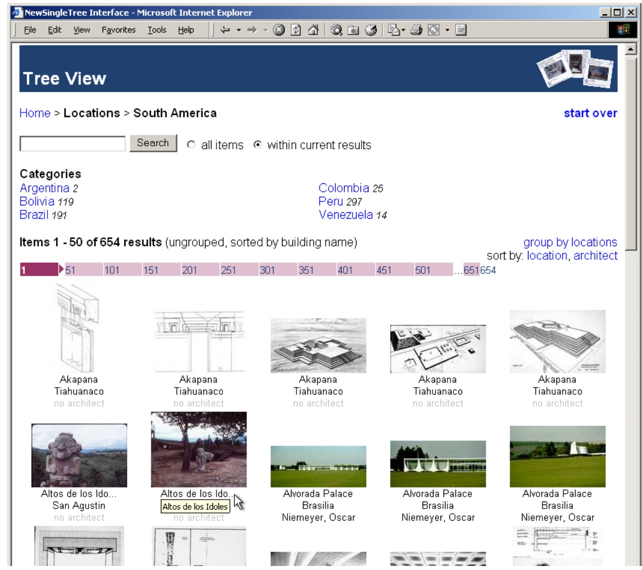

Figure 3: SingleTree View middle game (items ungrouped).

the query into new directions. Our goal is to make this kind of approach scale to larger collections (thus requiring hierarchical metadata facets), integrating keyword search fully, supporting search expansion as well as refinement. The result is the interface described in the next section.

THE USER INTERFACE

For our problem domain, we chose an architectural image database provided by our University. The database contains about 40,000 photographs and drawings of landscapes and buildings from a wide variety of historical periods, styles, and geographical regions. The images are classified under about 16,000 hierarchical metadata terms, which we manually reorganized into nine facets: people, locations, structure types, materials, periods, styles, view types, concepts, and building names.

We developed two different Web-based interfaces for finding images in this database, with the aim of testing the usefulness of faceted metadata as compared to a simpler and more conventional single-hierarchy approach. The two interfaces, which we will refer to as "Matrix View" and "SingleTree View", begin with the same opening, but provide very different intermediate pages (which we refer to as the "middle game"). The SingleTree View is simpler, but the Matrix View provides more flexibility. Figure 1 shows the opening screen for both, which provides a text search box and an overview of the first level of metadata terms under each facet. Query previews are shown next to each term.

Suppose the user now chooses "South America" under the Locations facet. This selects a result set of 654 images that are classified with this term. The Matrix interface yields the display in Figure 2 while the SingleTree interface yields Figure 3.

The Matrix View shows a column of metadata on the left, and the images in the current result set on the right. The column on the left shows query previews for all the terms that apply to the current result set. The user can select multiple terms from facets in any order, and can have the items grouped under any facet. The caption under each image gives the name of the building, the location, and the architect.

The SingleTree View shows a list of subcategories of the currently selected term at the top, with query previews. The items in the result set are listed below. The user can only drill down to subcategories of the current category, and cannot select terms from more than one facet. This view was modeled after Yahoo's directory, where items are arranged under a single tree hierarchy.

In Figures 2 and 3, we have only shown the Matrix interface in grouped mode and the SingleTree interface in ungrouped mode, though the user can click on a link to group or ungroup the items in both interfaces. In grouped mode, up to four sample items are shown in each subcategory. Selecting a group title or "all items" link drills into a subcategory. In ungrouped mode, 50 items are shown per page, and a "page bar," visible in Figure 3, allows moving from page to page. Both interfaces display the current query at the top of the page; the SingleTree view shows a single path, while the Matrix view may show multiple paths. The search can be expanded by selecting a breadcrumb in any path or by clicking an "X" button to remove a term entirely. Both interfaces provide a keyword search field at all times. When a keyword is entered, the query initially admits any item tagged with any matching term. A list of all matching metadata terms is shown above the result set,

Figure 4: Keyword search for "water".

as in Figure 4. Selecting a term in the list turns the keyword constraint into a category constraint.

In some situations there are too many subcategories or keyword matches to fit on the page. When this occurs, an alphabetized list is presented on a page of its own, so that the user can make a selection. The "more" links visible in Figures 2 and 4 also take the user to listing pages of this type.

Selecting a particular image yields an "endgame" page that is almost the same for both interfaces (Figure 5). The endgame page shows a detailed view of the selected image and its complete set of metadata. In the Matrix interface, hyperlinks let the user add any of these metadata terms to the current query to get a more specific query. Both interfaces let the user choose a more general query that clears all current constraints and shows all the items matching a single metadata term.

System Architecture

The two interfaces provide different navigation structures on top of exactly the same information structure, the same collection, and the same metadata. We used Python, MySQL, and the WebWare toolkit[4] to build the system, which we call Flamenco. The system is based on a highly generic database schema that we designed to accommodate any collection with hierarchical faceted metadata. The schema consists of:

- A table keyed by facet, listing all facets.
- A table keyed by item ID, containing all single-valued non-hierarchical metadata for each item.
- For each hierarchical or multi-valued facet:
  - A table keyed by facet value ID, listing the values of the facet, giving the hierarchy level and the path from the root for each value.
  - A table that associates item IDs with facet value IDs. If item x is assigned value y, a row appears associating x with every ancestor of y. The row associating x with y has a Boolean flag to indicate that it is the leaf value in the path.

[4] http://www.python.org, http://www.mysql.com, http://webware.sourceforge.net

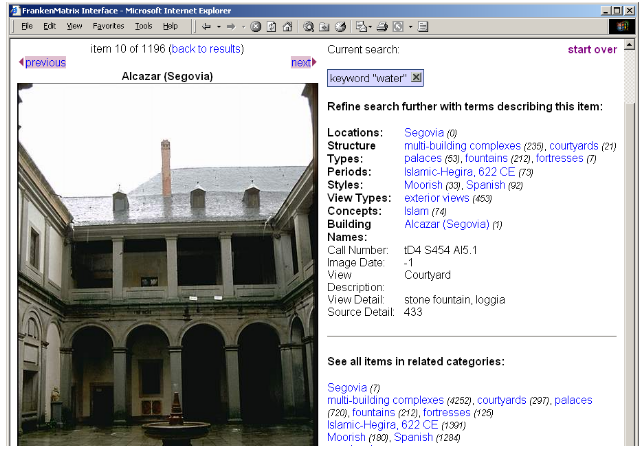

Figure 5: Endgame page for the Matrix interface.

All components of the interface are dynamically generated, based on the facets and facet values defined in the database. Query previews are generated using the SQL "group by" operator to count the number of items that fall into each subcategory.

To test the flexibility of the architecture, we loaded a portion of the MEDLINE database into our system. We wrote a few scripts to convert the MeSH classification tree into a set of facet hierarchies in our table schema, and were able to have both interfaces up and fully operational on the MEDLINE data within two days. The only modification we had to make to the user interface was to replace the component for displaying individual images with a component for displaying author names, titles, and abstracts.

ARCHITECTURE IMAGE USABILITY STUDY

To develop these interfaces, initially we performed a needs assessment (including an ethnographic analysis of how architects use and look for images as inspiration for their design work), a simple prototype and an informal usability test. After this we conducted two rounds of development and two formal usability studies. In Study 1, after participants assessed three experimental interfaces, they indicated that they liked having access to the metadata, but problems with the designs resulted in no interface design emerging as a clear success. The 11 participants divided evenly among the interfaces in their choice of a favorite, and many of the features received lukewarm or negative preference responses. However, a fourth design, shown only informally, received much interest; this was the precursor to the Matrix view.

We selected the most successful elements from each of the four interfaces to create two new versions of the interface for Study 2 (which is the focus of this section). Changes included: introducing the Matrix, modifying the topic descriptor hierarchy to make it more usable, displaying the hierarchical metadata on the left of the images to make the content available at the top of the screen, using intermediate pages to handle large category lists, providing multiple ways to perform every refine and expand function, changing the position and graphic display of key functions (such as the "X" to remove a search term), and offering an option to expand (as well as refine) from the image detail view. These changes were incorporated into two new interfaces described in the previous section.

Methodology

Others have argued for the importance of using *motivated* participants in search usability studies [29]; this is our experience as well, and so we recruited practicing architects, city planners, and graduate students in these fields as our 19 study participants. All participants indicated they regularly conducted searches on the Internet. About half stated that they were looking for images "all the time;" the other half said they searched for images on a monthly or yearly basis.

They received $12/hour for participating in a session that lasted about 1.5 hours. All participants were tested in a lab setting, using the Internet Explorer v6 browser on Windows 2000 workstations with 21 inch monitors set at 1280 x 1024 pixels in 24-bit color. Data was recorded with multiple methods: (a) server logs (b) behavioral logs (time-stamped observations), (c) online post-task questionnaires and (d) paper surveys at the end of the session. Two experienced usability analysts conducted each session. One took written notes while the other facilitated the session. Data from all the sources was collated to create a complete record of the test session.

The study used a within-subjects design. Each participant used both the SingleTree and the Matrix; each interface was the starting view for half the participants. They began with a brief "warm-up" period on the interface, performing a search of personal interest. After this, participants completed one highly structured task, with each action directed by the facilitator. This step-by-step task introduced participants to each interface's features and measured baseline understanding of those features. Following the warm-up, participants performed three types of tasks:

1 High Constraint Search: Find images with specific metadata assigned from three facets: View Types, Structure Types, and Locations (e.g., exterior views of temples in Lebanon). Participants were instructed to find as many images as they could that met the given criteria. This was done three different times:
   - 1.1 Start by using a Keyword Search to find such images
   - 1.2 Start by Browsing (clicking a hyperlink) to find such images
   - 1.3 Start by using method of their choice (Search or Browse) to find such images

2 Low Constraint Search: Find a low-constraint set of images (metadata in one facet)

3 Specific Image Search: Find a specific image. Participants were given a photograph, but no other information, and asked to find the same image in the 40,000 item collection based on the observable characteristics of the photograph.

For each of the five tasks, we created alternate versions of the task, and pre-tested these for equivalency. Order of tasks was carefully counterbalanced to take care of any order effects due to learning or fatigue. For both interfaces, participants filled out brief questionnaires after each task, and a longer questionnaire at the end of using the interface. Likert scales ranging from 1 (low) to 7 (high) were used throughout.

Results

Overall Impressions The majority of the participants said they preferred the power and flexibility of the Matrix to the simplicity of the SingleTree View. Participants found it easier to refine and expand their searches using the various features provided by the Matrix, and felt it was "easier to shift between searches" and to "troubleshoot research problems." Those who preferred the Matrix commented that they liked having the choices for refining the search displayed on the left side of the screen along with the images; participants referred to the metadata display as a "map," an "index," a "table of contents" and a "menu." They found this easier than struggling to guess (and spell) the correct keyword(s) for searching. Some participants were initially put off by the text-heavy appearance of the Matrix, but grew to like it after they had completed one or two searches. Four out of nineteen participants stated an overall preference for the "cleaner" and "simpler" SingleTree view. While SingleTree view was preferred for locating a specific image, the Matrix was perceived as more useful for browsing and exploring a collection; as one participant commented, "I could sit at the computer for hours!"

Comparing Matrix to SingleTree The Matrix view was rated slightly higher than SingleTree for all the measures (future use, ease of use, and for all types of design tasks). However, the differences are only significant for

Task Type	Matrix	SingleTree
High Constraint Tasks	14	4
Low Constraint Task	13	5
Specific Image Task	16	3
Overall Preference	16	3

Table 1: Overall interface preferences.

design work and seeing relationships (both t's > 1.7; both p's < .05). We asked participants which of the interfaces they preferred. As Table 1 shows, Matrix was generally preferred over SingleTree interface, both for specific tasks and overall.

Sense of Control Next we investigated how lost or on track participants felt during the search. Our concern was that with so many varied options (especially for Matrix) participants might find the interfaces too browsable, and feel "lost." Participants felt a high (average 5.65 on a 7 point scale) sense of control. Also, for all but the Specific Image tasks, participants felt a greater sense of control in the Matrix (differences are not significant except for the High Constraint Search Task).

We had also asked participants to describe the two interfaces with a series of adjectives using a semantic differential scale. Participants perceived the Matrix to be more flexible, powerful and browsable than the SingleTree. On average, the Matrix was perceived as 22% more flexible, 18% more powerful and 8% more browsable.

Task Success and Timing We gave participants a maximum of 6 minutes to complete each task. Success in a task implies that the participant indicated (within the allotted time) that he / she had reached an appropriate set of images / specific image in the collection. Success includes situations when the participant made the search too constrained and as a result missed some of the relevant images. Given this criteria for success, all participants were able to successfully complete the task within the given time for all but the image search task (discussed below).

Figure 6 shows the median task completion times for each task[5]. Task completion times were higher for the Matrix than for the SingleTree. This was due in part to the slower processing time – each page in the Matrix required the generation of numerous query previews while page generation in the simpler SingleTree was considerably faster.

In the image search task, participants were able to find the correct image within the allotted time only 22% of the time using the Matrix as compared to 66% of the time for the SingleTree. Note that the participants saw only a photograph; most problems arose when they made erroneous assumptions about the location of the photographed item. (When searching for *types* of images, they were highly successful.) After looking at the logs, we concluded that the participants were more successful in the SingleTree view because the Matrix view by default showed images in grouped view. This meant that participants would only see the first four images in each subhierarchy, and so were likely to drill down into the wrong subhierarchy. The default view for SingleTree view was ungrouped, so participants were more likely to simply see the correct image by scanning through all the items.

[5]We report median times in order to overcome the effect of outliers.

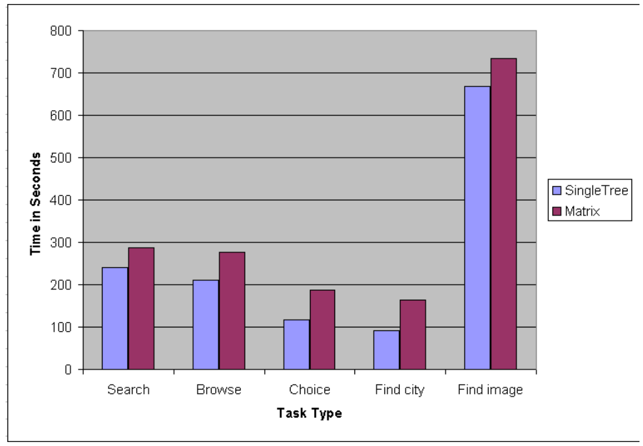

Figure 6: Median task completion time.

Precision	Search	Browse	Choice
Matrix	.85	.86	.96
Tree	.89	.82	.86
Recall	Search	Browse	Choice
Matrix	.79	.87	.96
Tree	.66	.61	1.0

Table 2: Recall and Precision.

Table 2 shows results for precision and recall, computed for tasks 1.1-1.3 only. The relevant item set was determined by pooling all the relevant results seen across participants. Precision was determined by inspecting the result set (all images together on one page) and counting the relevant ones. (Non-relevant documents would result from, for example, a task in which participants were asked to find interior views of churches in France, but the result set included other types of religious buildings as well.)

Feature Use and Understanding The hallmark of our approach to metadata-based search interfaces is offering the participants many movement options. The interface allows the participant to refine, expand, and arrange result sets in multiple ways. One critical question is: do participants notice, use and understand the various fea-

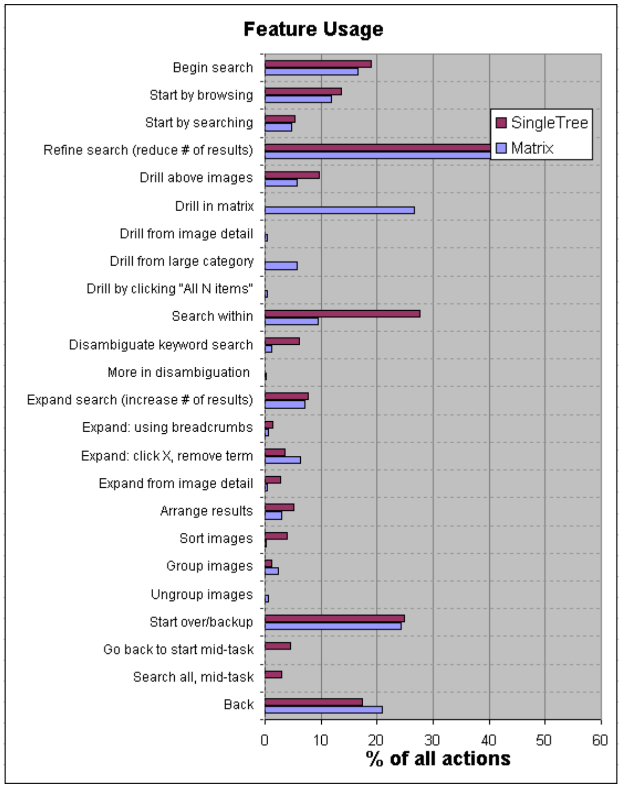

Figure 7: Percentage of time features were used. "Drill" means refine by descending a subhierarchy.

tures? Results from our previous studies had indicated that often participants did not notice, did not understand, or did not like some of the most powerful features. Our current design efforts are focused on making the features easy to understand. Results indicated that participants felt they understood the various features, and found them useful (most ratings ranged between 5.6 and 6.9). However, informal comments by participants indicate that in several cases some found certain features confusing or unintuitive; particularly, the free-text search term disambiguation and the method of expanding by removing terms from the current query (remove via "X"). However, in most cases these features became easy to use after the initial encounter.

A more direct measure of usefulness of various features is how often the features were actually used (see Figure 7). Across both interfaces, participants chose to begin more frequently by browsing (12.7%) than by searching (5%). For refining actions, participants refined by using "Drill in Matrix" 26.6% of the time. This facility was not available in the SingleTree, where participants chose to "Search Within" 27.6% of the time. The "Search Within" facility was used in the Matrix only 9% of the time. We think that this shows the power of the faceted hierarchies, which allow participants to flexibly modify their query, rather than forcing them to look for appropriate keywords for searching.

The option of expanding on a facet is not available in most search interfaces. Hence this is not something participants have experienced before. Nevertheless, about 7% of the actions are related to expanding a search. We think that the expand facility gives the interface added flexibility and our hypothesis is that once participants get more experienced with this interface, they will use it more often.

Participants chose to re-arrange results 2.9% of the time with the Matrix and 5.1% of the time with the SingleTree, mostly using the "sort" feature; this feature can only be used in ungrouped view, and the default sort order is designed to be the one most often desired.

Finally, participants chose to go back to start mid-task 4.5% of the time for SingleTree and only .02% of the time for the Matrix. We think that this also reflects well on our design, since participants started over less often for complex information needs.

RELATED WORK

Much work has been done on automating the mapping of query terms into metadata categories, for example, converting the query "heart attack" into the pre-defined term "myocardial infarction," and thus increasing the likelihood of making relevant matches [11, 19, 4, 17]. However, here we are interested in a different problem; namely, how to use the metadata directly in the web site user interface, both as a starting point for the search and as a structure upon which to organize the results of a keyword-based query.

Previous research has suggested a two-level architecture for linking documents and their "auxiliary data" [1, 5]. A related idea makes use of 3D visualization to integrate search and browsing of large category hierarchies with their associated text collections [15]. Other projects (e.g., [13, 18]) describe interfaces that allow the user to use thesaurus terms as part of the interface. Others have used table-of-contents views to provide context for retrieval results [20, 2, 9].

Many researchers have applied text clustering to retrieval results [16, 3, 22]. Clustering is attractive from the point of view of implementation, because the groupings can be determined automatically. Unfortunately, the usability studies that have been conducted show that ordinary users find the results of clustering to be difficult to interpret. Instead, they prefer the predictable organization of category hierarchies [26, 8, 7].

DISCUSSION

Our previous experiences with developing search user interfaces tell us that it is quite difficult to get positive reactions to novel approaches to search. In most cases, users end up preferring a simple keyword entry and title listing over more complex systems. Studies of existing systems show that users tend not to make use of advanced features of the search interface because they are too much work to use [31]. This makes the results of our studies all the more striking.

Our design reflects bits and pieces of what can be found in existing web interfaces, especially on ecommerce sites, but none of these allow simultaneous navigation of several hierarchical facets, and those that come close either are confusing and cluttered, do not allow expansion, or do not successfully integrate search within the navigation metadata. However, the point of this paper is not to stress the novelty of the design, but rather to emphasize that after a number of careful design decisions, the interface tests very positively with users, and is flexible enough to be applied to many different domains and collections.

Existing studies show that non-expert searchers have difficulty with Boolean queries beyond simple conjunction [24, 12]. An advantage of our approach is that it allows users to fluidly compose queries consisting of ANDs of ORs, since selecting a category C at level L produces an OR of all of the terms beneath L. Selecting more than one facet produces an AND across facets. Research in the biomedical literature tells us that this style of ANDs of ORs of related terms is one of the more effective ways to search [21].

An interesting aspect of this interface is that it does not emphasize relevance ranking, and so far we have not received a request for a ranking feature. We believe this results in part from our emphasis on browsing rather than search, and in part from the fact that the faceted metadata allows the users to explicitly choose the dimensions along which the items are retrieved, which is in part what relevance ranking does. This combined with the ability to sort by important metadata facets (date, author, quality, and so on) may suffice to obviate the need for ranking.

In this design we have focused on how to improve the "middle game" of the search experience. We plan to assess some ideas about how to improve the "opening," that is, the first page that users see. One idea is to show a few universally familiar items for first-time users (e.g., photographs of famous buildings, survey articles) to help them get started. Another idea is to show only a small subset of the most familiar metadata categories on the start page.

CONCLUSIONS AND FUTURE WORK

We have successfully designed a search interface that uses faceted hierarchical metadata to allow users to flexibly explore a large collection without feeling lost or confused. After the third iteration of design, participants in usability studies exhibited strong positive feelings about the interface, sometimes to the point of pleading for immediate access to the system. After trying the interface, the head of our University's architecture department said, unprompted, "I feel like I'm browsing the shelves," and asked that our interface replace the current one used throughout the university.

This interface supports 6 out of the 8 design desiderata for usable search systems as outlined in [28]: strive for consistency, offer informative feedback, offer simple error handling, permit easy reversal of actions, support user control, and reduce short-term memory load. This is largely due to the use of query previews, which emphasize recognition over recall, and support the notion of providing "scent" of where to go next [10]. Error handling and reversal are provided by the various ways to go back and start over, and the use of metadata provides the view of consistency. (We do not support design for closure or provide shortcuts for experts.)

In the future we plan to address efficiency issues, try a non-specialized collection such as an ecommerce catalog, and add personalization, so return users would see examples of their own saved items as starting points. Finally, we plan to incorporate a standard relevance feedback mechanism; thus facilitating the "find more like this" task.

Even without these planned enhancements, we believe our approach marks a major step forward in search interfaces and can serve as a model for web-based collections of up to 100,000 items.

Acknowledgements. We thank Andrea Sahli for help with the usability tests, the UC Berkeley slide library for access to the dataset, and all our usability study participants. This research was funded by an NSF CAREER grant, NSF 9984741.

REFERENCES

1. M. Agosti, G. Gradenigo, and P.G. Marchetti. A hypertext environment for interacting with large textual databases. *Information Processing & Management*, 28(3):371–387, 1992.

2. Robert B. Allen. Two digital library interfaces that exploit hierarchical structure. In *Proceedings of DAGS95: Electronic Publishing and the Information Superhighway*, Boston, MA, May 1995.

3. Robert B. Allen, Pascal Obry, and Michael Littman. An interface for navigating clustered document sets returned by queries. In *Proceedings of ACM COOCS: Conference on Organizational Computing Systems*, Milpitis, CA, November 1993.

4. A. Aronson, T. Rindflesch, and A. Browne. Exploiting a large thesaurus for information retrieval. In *Proceedings of RIAO '94; Intelligent Multimedia Information Retrieval Systems and Management*, pages 197–216, October 1994.

5. N. Belkin, P. G. Marchetti, and C. Cool. Braque – design of an interface to support user interaction in information retrieval. *Information Processing and Management*, 29(3):325–344, 1993.

6. Christine L. Borgman. Why are online catalogs still hard to use? *Journal of the American Society for Information Science*, 47(7):493–503, 1996.
7. Hao Chen and Susan Dumais. Bringing order to the web: automatically categorizing search results. In *Proceedings of CHI 2000*, The Hague, Netherlands, 2000.
8. Hsinchen Chen, Andrea L. Houston, Robin R. Sewell, and Bruce R. Schatz. Internet browsing and searching: User evaluations of category map and concept space techniques. *Journal of the American Society for Information Sciences (JASIS)*, 49(7), 1998.
9. Michael Chen, Marti A. Hearst, Jason Hong, and James Lin. Cha-cha: A system for organizing intranet search results. In *Proceedings of the 2nd USENIX Symposium on Internet Technologies and Systems*, Boulder, CO, October 11-14 1999.
10. George W. Furnas. Effective view navigation. In *Proceedings of ACM CHI 97 Conference on Human Factors in Computing Systems*, volume 1 of *PAPERS: Information Structures*, pages 367–374, 1997.
11. Fredric Gey, Hui-Min Chen, Barbara Norgard, Michael Buckland, Youngin Kim, Aitao Chen, Byron Lam, Jacek Purat, and Ray Larson. Advanced search technologies for unfamiliar metadata. In *Meta-Data '99 Third IEEE Meta-Data Conference*, Bethesda, MD, April 1999.
12. Sharon L. Greene, Susan J. Devlin, Philip E. Cannata, and Louis M. Gomez. No ifs, ands, or ors: A study of database querying. *International Journal of Man-Machine Studies*, 32(3):303–326, 1990.
13. Micheline Hancock-Beaulieu, Margaret Fieldhouse, and Thien Do. An evaluation of interactive query expansion in an online library catalogue with a graphical user interface. *Journal of Documentation*, 51(3):225–243, 1995.
14. Marti A. Hearst. Next generation web search: Setting our sites. *IEEE Data Engineering Bulletin*, 23(3), September 2000.
15. Marti A. Hearst and Chandu Karadi. Cat-a-cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. In *Proceedings of the 20th Annual International ACM/SIGIR Conference*, Philadelphia, PA, 1997.
16. Marti A. Hearst and Jan O. Pedersen. Reexamining the cluster hypothesis: Scatter/gather on retrieval results. In *Proceedings of the 19th Annual International ACM/SIGIR Conference*, pages 76–84, Zurich, Switzerland, 1996.
17. William R. Hersh, David H. Hickman, Brian Haynes, and K. Ann McKibbon. A performance and failure analysis of SAPHIRE with a MEDLINE test collection. *Journal of the American Medical Informatics Association*, 1(1):51–60, 1994.
18. Peter Ingwersen and Irene Wormell. Improved subject access, browsing, and scanning mechanisms in modern online IR. In *Proceedings of the 9th Annual International ACM/SIGIR Conference*, pages 68–76, Pisa, Italy, 1986.
19. Yufeng Jing and W. Bruce Croft. An association thesaurus for information retrieval. In *Proceedings of RIAO*, pages 146–160, Rockefeller University, New York, 1994.
20. Thomas K. Landauer, Dennis E. Egan, Joel R. Remde, Michael Lesk, Carol C. Lochbaum, and Daniel Ketchum. Enhancing the usability of text through computer delivery and formative evaluation: the superbook project. In C. McKnight, A. Dillon, and J. Richardson, editors, *Hypertext: A Psychological Perspective*, pages 71–136. Ellis Horwood, 1993.
21. Henry J. Lowe and G. Octo Barnett. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *Journal of the American Medical Assocation (JAMA)*, 271(4):1103–1108, 1994.
22. Y. S. Maarek and A.J. Wecker. The librarian's assistant: Automatically assembling books into dynamic bookshelves. In *Proceedings of RIAO '94; Intelligent Multimedia Information Retrieval Systems and Management*, October 1994.
23. Gary Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Press, 1995.
24. A. Michard. Graphical presentation of Boolean expressions in a database query language: design notes and an ergonomic evaluation. *Behaviour and Information Technology*, 1(3), 1982.
25. Catherine Plaisant, Ben Shneiderman, Khoa Doan, and Tom Bruns. Interface and data architecture for query preview in networked information systems. *ACM Transactions on Information Systems*, 17(3):320–341, 1999.
26. Wanda Pratt, Marti Hearst, and Larry Fagan. A knowledge-based approach to organizing retrieved documents. In *Proceedings of 16th Annual Conference on Artificial Intelligence (AAAI 99)*, Orlando, FL, 1999.
27. Vividence Research. Tangled web 2001. http://www.vividence.com, June 2001.
28. Ben Shneiderman, Donald Byrd, and W. Bruce Croft. Sorting out searching: A user-interface framework for text searches. *Communications of the ACM*, 41(4):95–98, 1998.
29. Jared Spool. *Web Site Usability: A Designer's Guide*. Morgan Kaufmann, 1998.
30. Danny Sullivan. Npd search and portal site study. http://searchenginewatch.internet.com/reports/npd.html, July 6 2000. NPD's URL is http://www.npd.com.
31. A. G. Sutcliffe, M. Ennis, and S.J. Watkinson. Empirical studies of end-user information searching. *Journal of the American Society for Information Science*, 51(13), 2000.