Design Recommendations for Hierarchical Faceted Search Interfaces

Marti A. Hearst
School of Information
UC Berkeley
hearst@sims.berkeley.edu

ABSTRACT

This paper presents interface design recommendations for faceted navigation systems, based on 13 years of experience in experimenting with and evaluating such designs.

1. INTRODUCTION

Hierarchical faceted metadata has been found to be a highly understandable data model for search interfaces, intermediate in complexity between hierarchy and full knowledge representation [5, 10]. Although web sites, especially e-commerce sites, have made use of category information for navigation for quite some time, in many cases their application has tended to be inconsistent, incomplete, or otherwise problematic.

Experience shows that the success of the design of a user interface for search is highly dependent on and sensitive to the details of the design. For example, Platt et al. [8] found that algorithms presented in a first-try design for photo browsing failed from a usability perspective, but that a second careful design done in conjunction with usability testing, and using the same underlying technology, succeeded.

In the past, I have developed and/or collaborated on a number of interfaces that utilize categories for providing context for search results and navigation, including Cougar [3], DynaCat [9], Cat-a-Cone [6], Cha-Cha [1], and most recently, Flamenco [4, 5, 10]. The goals of the Flamenco project were to determine how exactly to aid navigation and browsing of information collections via the use of hierarchical faceted metadata. We addressed questions such as how to allow the user to navigate in several hierarchies simultaneously, how to display the query as it is built up, how to present the query previews, and so on.

In this paper I describe a set of design guidelines for faceted search interfaces, based on this experience. I illustrate the design lessons using both the Flamenco interface and a new commercial interface called eBay Express.[1] The focus is on design of the interface itself; space limitations prevent discussion of design issues for the metadata.

I start with the assumption that the overarching design goals are to support flexible navigation, seamless integration with directed (keyword) search, fluid alternation between refining and expanding, avoidance of empty results sets, and at all times retaining a feeling of control and understanding.

[1]express.ebay.com. This website was informed by ideas from the Flamenco project but was built independently of it.

2. TERMINOLOGY

For the purposes of the discussion below, this paper uses the following terminology.

Facets refer to categories used to characterize information items in a collection. Each facet has a name, such as *Ingredients* or *Cuisine* for a recipe collection. A facet can be *flat* or *hierarchical*. In both cases, associated with a facet is a set of *labels*. Examples for the hierarchical *Ingredients* facet include *Vegetables*, *Breads and Cereals*, and so on. There may be labels that are subsumed by *Vegetables*, such as *carrots*, *lettuce*, etc. I sometimes refer to portions of the hierarchies within a facet as the facet's *subcategory* or *subhierarchy*.

In a faceted search interface, labels are assigned to items from the collection. So a recipe for pasta primavera would receive labels of the form *Cuisine* > *Italian*, *Ingredient* > *Pasta*, *Ingredient* > *Vegetable*, and so on.

When a label within a facet is selected within the interface, all items that have been assigned to that label are retrieved, so selecting a label within a hierarchy is equivalent to performing a disjunction over all the labels beneath it. When labels from different parts of the interface are selected, the system in effect builds a query that is a conjunct of disjuncts over the selected labels and their subcategories.

3. FACET HIERARCHY NAVIGATION

In the early days of designing Flamenco, we were uncertain if users would be comfortable navigating in multiple facet hierarchies simultaneously; that is, selecting from one facet, refining within its hierarchy, switching to another facet, and then drilling down into its hierarchy, adding in keyword search, and so on.

Thus we conducted usability experiments that deliberately contrasted an interface that allowed drilling down in only one facet versus one that allowed moving around in multiple hierarchies simultaneously. A strong majority of participants preferred being allowed to navigate in multiple hierarchies; they felt they were in control and did not feel lost [2].

Another big question about how to support hierarchical faceted metadata is how to expose the hierarchical categories without crowding the display or confusing the user. One approach is to expose an entire facet subhierarchy when the user mouses over it, using a "fly-away" menu as is done in some operating system file browsers. One advantage of this approach is it allows the user to see all the options beneath a label, and drill down to precisely the desired label. This can turn into a disadvantage, however, if the number of choices is very large, (as in the case with names of

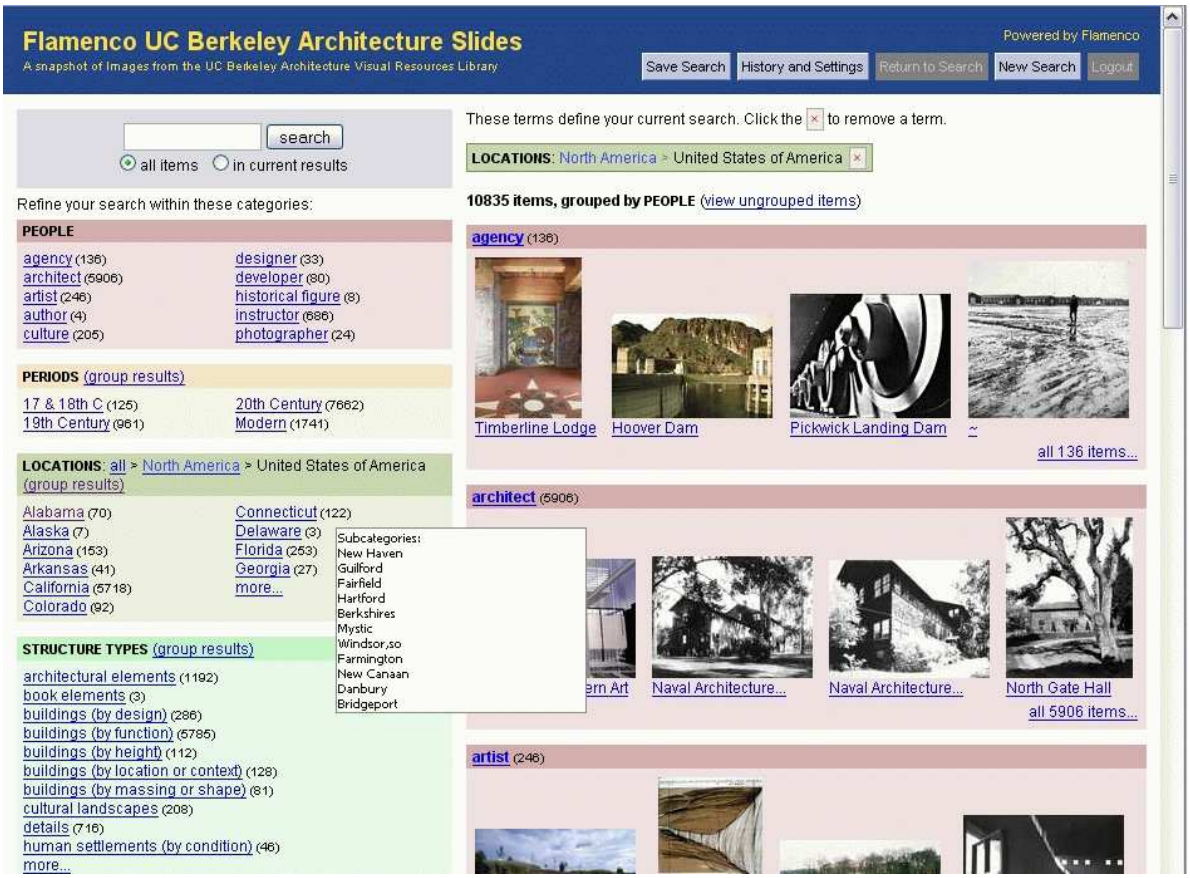

Figure 1: Hierarchical facet navigation in Flamenco.

artists, for example), the menu obscures everything else in the display, requires scrolling (which is bad practice for a menu), and makes multi-select difficult. Another downside to this approach is that query previews must be computed for every level of the hierarchy, lest the user unintentionally choose a branch for which the results are empty. Lastly, it precludes one of the benefits of progressive disclosure of hierarchy, which allows the user to see results grouped by higher-level concepts in order to obtain an understanding of the contents of the collection.

Another approach for revealing hierarchy is to build an interface similar to that of Windows Explorer with folders that expand and collapse, and which allow many different subhierarchies to be simultaneously viewable from root to the selected level. This has at least two major downsides. First, if many of the hierarchies are expanded, the navigation component can get very large and require extensive scrolling. Second, and more importantly, users would be unfamiliar with the idea of an item simultaneously residing in multiple folders, since Explorer does not support that functionality.

In Flamenco we adopted a step-by-step drill-down approach in which the level just below the currently selected level is visible, along with a trail indicating the higher level concepts positioned just above the labels. (See Figure 1.) In addition, when the mouse hovers over a label, its immediate children are displayed in a tooltip, so the user can in fact see three levels simultaneously. For instance, in Figure 1, if the user selects *Location* > *North America* > *United States*, the visible labels beneath it read *Alabama, Alaska*, etc. Hovering over one of these shows a list of cities within that state.[2] Thus three levels of information are simultaneously visible.

In the righthand pane, the results are grouped by the children of the level just selected. In this example the results were initially grouped according to US state. Note that the hierarchy trail is also reflected in the query breadcrumb. These results can in turn be grouped by the subcategories of a different facet, such as *People* as shown in Figure 1.

A different approach to hierarchy exposure is demonstrated in the eBay Express interface. In this system, if a label within a hierarchical facet is chosen, the next level is shown as a *separate* facet. For example, in Figure 3, the user has chosen two labels; the first is *Album Type* > *Box Set* and the second is *Genre* > *Jazz*. Below the query we see the available subcategories for the selected genre. The facet name is *Sub-Genre* and the relevant labels include *Big Band, Swing, Bop, Latin Jazz*, and so on. Users of the system understand the relation between the two [7]. One restriction currently is that the hierarchies do not go deeper than two levels.

4. LAYOUT OF LABELS WITHIN FACETS

In early Flamenco studies we asked participants about their preferences regarding the layout of the labels within

[2]In our current implementation, the tooltip contents do not compute the query previews, so it might misleadingly suggest there are results in subcategories when none are present, but this is an implementation detail that can be changed.

Figure 2: Before and after views of graphic design choices in Flamenco for (a) query breadcrumbs, and (b) facet navigation.

the facets. Although there were differences of opinion, the main tendency appears to be a preference for neatly laid out columns rather than a wrap-around list of labels.

One problem with the column-oriented layout is the variable length of the labels within the facets. If some facets have long labels and others are short, forcing a uniform choice for number of columns leads to wasted space or wrap-around views. In the Flamenco interface we resolve this problem by automatically computing the number of columns to show items within, based on the length of the longest visible label (see Figure 1).

The Flamenco code also provide a facility to adjust the order in which labels are shown, either alphabetical, numerical, or by number of hits for the label in the visible collection. Numerous studies of navigation interfaces suggest that in most cases users prefer the predictability of a well-known ordering (e.g., [9]).

However, in those cases in which there is room for only a few of many labels for a facet to be shown, it can be useful to show either the most salient or the most frequently occurring in the initial view, with a link to see more choices which are then shown in a predictable order. The eBay Express system uses such a strategy. In Figure 3 we see that the labels beneath the *Genre* facet lists the four most common genres of music in the *Boxed Set* albums. Within these four, the genres are listed in alphabetical order. By contrast, the *Price* facet shows price ranges in ascending numerical order.

The designers have made an interesting choice with the *Artist* facet. In this case the interface shows neither the most common artists nor the first in alphabetical order. Because this facet is very large, the system appears to select an alphabetical sequence of four labels, in order to give a varying sample of what is available within that facet. I have yet to learn of usability implications for this approach.

5. FACET EXPOSURE

Invariably, the question of how many facets to show arises, and in what order should they be shown. The current implementation of Flamenco exposes all facets, since it has primarily been applied to domains such as digital libraries and image collections where exploration of the dataset and learning about the category structure is a high priority. (The Flamenco code also includes a personalization feature that allows users to change their default ordering and hide those facets that are not of interest.)

However, exposing all facets can take up considerable screen real estate. If the facets are placed on the side, the user will be required to scroll to see the possibilities, but if placed across the top, the user will have to scroll before seeing any information items. It is a well-established search usability principle that users should be able to see results immediately after the initial query, even if using an interface that provides context. Requiring users to scroll before seeing any results can lead to site abandonment.

For some tasks such as online shopping, full exploration of the collection is not desirable; the task can be more directed. Thus in the eBay Express interface, the designers have determined in advance which subset of facets are of most interest to most users, and initially expose only four of these fully, listing additional choices on one compact line below (see Figure 3). They also position the expanded facets across the top of the screen, in the "sweet spot" (the region of a web page that users tend to look at first since it usually contains the most important content on the page). After the user selects a facet, one of the compressed facets from the list below is expanded and moved up to the end of the line (righthand side) of the expanded facets. For instance, in Figure 3, the initial expanded facets were *Artist, Genre, Album Type*, and *Price*, but the list has been adjusted to account for the fact that two of these have been selected. The ordering of the facets in the query breadcrumbs reflects their order of selection by the user. This approach is being well-received by the user base [7].

Once the result set gets small, the population of labels within the facets gets small. Even when a facet no longer has any active labels, it is important to show its name in the interface; this is equivalent to graying out menu selections that are not currently active; numerous usability results indicate the importance of retaining consistency in availability of selections.

In the eBay Express interface, when the results set gets very small (say, under 10 items), the facets are no longer useful for refining, and so they are retained but temporarily hidden behind a link. It continues to keep the query breadcrumb visible in order to allow users to expand the query easily.

6. GRAPHIC DESIGN

The importance of small details in the graphic design of a search interface should not be underestimated. In interface design generally, the layout and graphic design suggests to the user what to do and how to do it.

Search interfaces, especially those searching text collections, must show detailed information, and so it is critical try to avoid a cluttered or confusing appearance. A search interface designer must balance the choices of layout, placement and amount of white space, font style, weight, and size, and the use of color to distinguish components without overwhelming.

Figure 2 shows before and after views of the design of two widgets in the Flamenco system. We experimented extensively with how to represent the numbers in the query previews, whether the categories should be hyperlinked alone or with the numbers, whether the numbers should be bolded, italicized or not, enclosed in parentheses or not, and so on. A good solution must balance many different choices simultaneously to maximize legibility while simultaneously presented as much information as possible, without appearing overwhelming. The final combination has become somewhat standard in website design.

Figures 3 and 4 show after and before screenshots for eBay Express graphic design of query breadcrumbs and of facet presentation and navigation. In the query breadcrumbs, note the change in emphasis; in Figure 4 the facet name is emphasized in bold and the label is in light grey; in the later launched version in Figure 3, the facet name is deemphasized in favor of the selected category label. Note also the small changes in the graphic design, bringing the query breadcrumbs closer to the rest of the facets to be selected to more clearly indicate their relationship, removal of the bullets from the facet labels, relying instead on the whitespace and column layout and font contrast between the facet name and the items labels to indicate the important relationships.

Over the last few years, many visual conventions have arisen[3], thus leading to more predictable and understandable interfaces. One such example is a link labeled *More ...*, used to indicate that a list or paragraph is incomplete and more material follows, and usually shown italicized or in a slightly smaller font than the main content. The latest design of eBay Express makes use of this convention, although the earlier design in Figure 4 strayed too far away, and so was less understandable.

7. INCORPORATING KEYWORD SEARCH

A key component to successful faceted search interfaces (which unfortunately is rarely implemented properly) is the seamless integration of keyword search. In the Flamenco system, a keyword search can be issued at any time, and is incorporated into the query breadcrumb in the same way as link selections are (see Figure 2(a), righthand side). The query can match categories or free text describing the item; the user can select the category that matches or simply retain the keyword in the query, matching all items containing the term. Our results suggest that in most cases users prefer the broader match, most likely in part because they can organize the results of a broad keyword query using the facet system.

A perennial search problem is that of "search within results". The Flamenco design does not solve this problem satisfactorily; currently it is a modal selection, which means the setting is easy forgotten or overlooked.

eBay Express has a particularly interesting approach to handling keyword queries. The system attempts to map the user-entered keywords into the corresponding facet label, and simply adds that label to the query breadcrumb. For example, a search on "Ella Fitzgerald" creates a query consisting of the *Artist* facet seleted with the *Ella Fitzgerald* label. Search within results is accomplished by nesting an entry form within the query region.

8. BREADCRUMB DESIGN

Common parlance refers to the users' query history as a "breadcrumb." Typically breadcrumbs simply record the sequence of actions that the user has done within the query session, and thus mix and match fields of various types. Faceted systems should instead keep the path within each facet in a separate visual component. This both reinforces the notion of the query consisting of a conjunction of different categories at different levels of hierarchy, and allows for flexible expansion of the query, since the user can eliminate an entire facet by clicking on the iconic *X* or "delete" link, or they can expand up within a category by clicking on a parent term (in Figure 1 this would mean generalizing from "US" to "North America" by clicking on the latter link in the breadcrumb). The eBay designers have taken this separation still further by dividing up a hierarchical facet into two pieces within the query, as discussed above.

Acknowledgements: Portions of this research were funded by NSF IIS-9984741. Portions of this work appeared in Powerpoint form in [7]. I gratefully acknowledges the contributions of the Flamenco team: Ping Yee, Kirsten Swearingen, Kevin Li, Rashmi Sinha, and Jennifer English. I also thank Preston Smalley and Corey Chandler of eBay for the information and insights about the eBay Express design.

9. REFERENCES

[1] M. Chen, M. A. Hearst, J. Hong, and J. 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.

[2] J. English, M. Hearst, R. Sinha, K. Swearingen, and P. Yee. Flexible search and browsing using faceted metadata, 2002. Unpublished Manuscript.

[3] M. A. Hearst. Using categories to provide context for full-text retrieval results. In *Proceedings of RIAO '94; Intelligent Multimedia Information Retrieval Systems and Management*, pages 115–130, October 1994.

[4] M. A. Hearst. Next generation web search: Setting our sites. *IEEE Data Engineering Bulletin*, 23(3), September 2000.

[3]See the Yahoo Design Pattern Library, http://developer.yahoo.com/ypatterns/

Figure 3: Hierarchical Facet navigation in eBay Express, showing variation in label ordering within a facet as well as separation of subcategories into separate selection boxes.

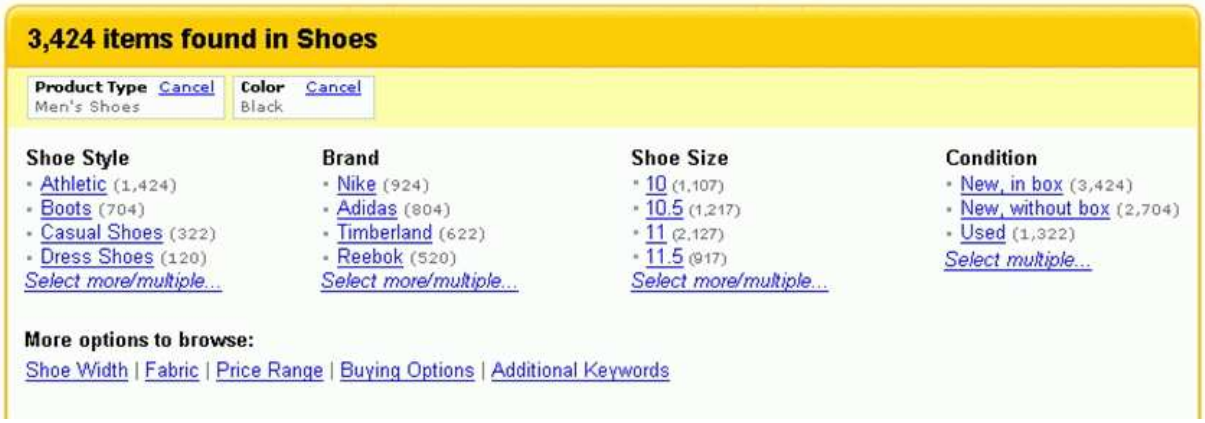

Figure 4: Earlier eBay graphic design.

[5] M. A. Hearst, J. English, R. Sinha, K. Swearingen, and K.-P. Yee. Finding the flow in web site search. *Communications of the ACM*, 45(9), September 2002.

[6] M. A. Hearst and C. 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.

[7] M. A. Hearst, P. Smalley, and C. Chandler. Chi course on faceted metadata for information architecture and search, 2006.

[8] J. C. Platt, M. Czerwinski, and B. a. Field. Phototoc: automatic clustering for browsing personal photographs. In *Proceedings of the Fourth IEEE Pacific Rim Conference on Multimedia*. IEEE, 2003.

[9] W. Pratt, M. Hearst, and L. Fagan. A knowledge-based approach to organizing retrieved documents. In *Proceedings of 16th Annual Conference on Artificial Intelligence (AAAI 99)*, Orlando, FL, 1999.

[10] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proceedings of the CHI 2003*, Fort Lauderdale, FL, April 2003.