California State University, San Bernardino

CSUSB ScholarWorks

1-2007

Toward a Design Space for Categorized Overviews of Search Results

Jonathan M. Smith
California State University - San Bernardino, jsmith@csusb.edu

Bill Kules
The Catholic University of America

Recommended Citation

Toward a Design Space for Categorized Overviews of Search Results

Jonathan M. Smith
School of Library and Information Science
The Catholic University of America
Washington, D.C.

smithjm@cua.edu

Bill Kules
School of Library and Information Science
The Catholic University of America
Washington, D.C.

kules@cua.edu

ABSTRACT

There is a burgeoning movement to create user interfaces that combine searching and browsing in order to facilitate knowledge discovery. Categorized overviews are promising techniques to improve the browsing capabilities and subject access of search results. Research is needed to examine the design space in order to promote the usability and understanding of such interfaces.

This paper proposes a set of 14 dimensions grouped into three areas of the design space for categorized overviews. These dimensions emerged from a review of recent literature about exploratory search interface design and categorized overviews, an analysis of five examples that are seeking to meet the needs of knowledge discovery, and personal experience. Finally, this paper discusses the application of the dimensions on a case study of a categorized overview created for a bibliographic database.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *graphical user interfaces (GUI), prototyping, screen design, theory and methods, user-centered design.*

General Terms

Design, Human Factors, Standardization.

Keywords

Categorized Overview, Design Space, Search Interface.

1. INTRODUCTION

As keyword search engines become more powerful the trend in user interface design has tended toward a simpler search interface. The single text entry box and a button that says "Search" or "Go" has become the standard presentation of a search feature for documents and web sites. In situations of known-item retrieval this method works quite well. If the user is searching for a company's web site, a specific document or a particular author, he can usually find what he is looking for near the top of the results list. However, this method fails when the user is conducting an exploratory search to investigate a topic.

In order to facilitate the exploratory searcher, rich interfaces are being developed that harvest metadata or keywords from the records. This metadata is then displayed along with the search results in a manner that encourages browsing and exploration in combination with keyword searching. The goal of this paper is to propose a set of dimensions of the design space for categorized overviews of search results. This will aid designers and researchers by identifying the areas of the design space they will have to address as well as providing them with options.

While there is a fair amount of recent literature discussing the mechanics of exploratory search a limited number of studies that investigate the interface design have been published. Identifying best practices in exploratory search interface design is important to further its development. By recognizing user expectations and establishing standards we will promote both the use and understanding of such interfaces.

A number of different approaches to the user interface design have been taken with varying results. Categorized overviews of search results can improve the browsing capabilities and subject access of databases [6]. The richer the descriptive metadata assigned to records are, the more access points can be exposed through the categorized overview. By using stable categories users can become familiar with the organization of the overview.

This paper will identify 14 dimensions of the design space grouped into three areas: Organization, Display, and Interaction. These dimensions emerged from a review of recent literature, an analysis of five examples, and personal experience. The first section of the paper will highlight recent literature on search interface design and categorized overviews of search results. The section following will identify the dimensions and note the design decisions some projects have made. The final section will illustrate the application of the dimensions on a case study designing a prototype categorized overview interface for a bibliographic database and the initial user feedback.

2. RELATED WORK

While studies conducted on exploratory search interfaces are beginning to emerge, there are few that identify a set of design dimensions. Marchionini states, "the search system designer aims to bring people more directly into the search process through highly interactive user interfaces that continuously engage human control over the information seeking process." [7] Identifying the best practices for the design will aid designers and advance usability of such interfaces.

Several approaches to exploratory search have been investigated and reported on. A digital library interface that supports browsing

and analytical search strategies was the goal of research reported by Marchionini, Plaisant, & Komlodi in "Interfaces and tools for the library of congress national digital library program." [8] While the paper focuses on digital libraries many of the ideas may be applied to exploratory search interfaces in general.

Antelman, Lynema, & Pace report on North Carolina State University (NCSU) Libraries' decision to replace their catalog interface with a new interface powered by technology from Endeca, the same company behind the technology used on HomeDepot.com, Walmart.com and Barnes & Nobel's BookBrowser. They discuss the implementation and summarize the first two months of user statistics [1]. An interesting feature of their implementation is the ability to browse the collection without ever submitting a search query.

Hearst discusses user interface design and the interface developed at University of California, Berkeley, called Flamenco [2]. Hearst, Elliott, English, et al present results of a usability study conducted on Flamenco [4]. Yee, Swearingen, Li, et al present the findings of a usability study they conducted on an implementation of Flamenco used for image searching [9].

Findex is an implementation of categorization using a clustering algorithm. Käki reports on a longitudinal study on Findex to determine the usefulness of a categorization interface in web searching. He found that users would select categories in 25% of searches and an average of 2.3 categories in searches where categories were used [5]. Kules is also interested in categorization and asserts that stable categories "will allow searchers to reuse category knowledge on subsequent searches." [6] Both Käki and Kules have found that categorized overviews help searchers find relevant results deeper in the result list than they would have otherwise [5, 6].

Hearst makes detailed design recommendations for hierarchical faceted search interfaces. She states that, "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." [3] This paper explores some of the recommendations she provides, notably the dimensions we have identified as the display of subcategories, interface complexity and broadening the results.

Kules develops a set of eight design principles for exploratory search interfaces [6]. He also identifies a set of ten dimensions and their corresponding design options for categorized overviews:

- Display of facets
- Selection of facets
- Display and selection of categories within a facet
- Display and selection of categories
- Visible levels of hierarchy displayed
- Overall depth of hierarhcy
- Display of "uncategorized" pseudo-category
- Display of empty categories
- Sort order of categories
- Actions / operations on overview

In this paper we attempt to refine and build on the dimensions identified by Kules. It is in this spirit that we have set out to investigate the design space of a categorized overview for search results.

3. DESIGN SPACE
There are a number of recent projects that have implemented exploratory search interfaces similar to a categorized overview. This section will note some of their commonalities and differences in an attempt to identify important design features that might be applied to a categorized overview interface. A total of 14 dimensions emerged which can be grouped into three areas: Organization, Display and Interaction.

Five search interfaces were examined for their design dimensions. Clusty [http://clusty.com/] is a web search engine built on research done by the software company Vivisimo. EBSCOhost [http://search.ebscohost.com/] provides access to several databases of general and academic journals. Flamenco [http://flamenco.berkeley.edu/] is a search interface developed at the University of California, Berkeley featuring "hierarchical faceted categories". Librarians' Internet Index (LII.org) [http://www.lii.org/] is powered by Siderean's Seamark Navigator. NCSU Libraries' catalog interface http://www.lib.ncsu.edu/catalog/] is powered by the Endeca Profind™ platform. While not a comprehensive list of exploratory search interfaces, this list represents some of the implementations currently in use and available to the general public.

3.1 Organization of the Categorized Overview

Table 1. Organization of the categorized overview

Dimension	Options
Sort order of category links	- Alphanumeric - Number of hits - Classification scheme - No order
Structure of categories	- Flat - Hierarchical
Category classification method	- Clustering - Taxonomy
Facets	- Yes - No

3.1.1 Sort Order of Category Links
Ordering the category links in a manner that is meaningful and clear to the user will make the interface more useable. An obvious option for ordering any list is alphanumeric. Flamenco's interface displays many categories simultaneously; consequently the alphanumeric display brings order to an interface that could otherwise be overwhelming. Alphanumeric order could also make sense if the users are very familiar with the categories, e.g., geographic location or authors' names.

An option favored by NCSU Libraries and Clusty is ordering categories by the number of hits each has in the result set. Thus the highest populated category would reside at the top of the list. In this case we assume that the highest populated category is also the most relevant to the search. Both of these sites make their criteria for ordering clear by displaying the number of hits next to each category link. If this were not so, the user could very well be confused at the seemingly nonsensical order of the categories.

NCSU Libraries uses a different method of ordering for an additional set of categories. This set of categories represents Library of Congress Classification, which feature a well established classification scheme. The user who is familiar with this classification scheme will have strong expectations as to how it should be ordered. Users who are not familiar with LC Classification have the opportunity to learn the classification scheme, as the order does not change.

3.1.2 Structure of Categories
The structural organization of the category terms can have an impact on the usefulness of the overview. A hierarchical structure of categories can aid exploration by providing an overview of broad categories first, then allowing the user to expand the hierarchy and refine the search using narrower subcategories.

Another consideration that should be studied further is how deep within a hierarchy users are willing to navigate. If it is determined, for example, that most users are unwilling to drill down further than 3 levels, hierarchies should be broadened in order to compensate.

3.1.3 Category Classification Method
The classification method has a large impact on the hierarchy structure. Clustering is a technique that groups similar records together, often based on common words found in records. Using a clustering algorithm to create a hierarchy on the fly can produce some unintuitive results, which can confuse the user. However, a predefined taxonomy can be designed in a logical manner that is learnable by the user. If a thesaurus were available it would also be interesting to study how to take advantage of features such as near terms, related terms and synonyms.

3.1.4 Facets
Grouping similar categories together into facets can be a way to bring order to a crowded interface while providing a large overview. Facets can bring additional meaning to the overview by exposing category relationships [6]. Facets can be any logical grouping for the collection. A library might use facets such as: location, format, language and author. The facets might also be topical. In the case of a music collection the facets might be: composer, artist, genre, era, ensemble type and instruments.

3.2 Display of the Categorized Overview
Visual design can have a large impact on the perceived complexity of the interface. Font type, size, weight and use of white space all play a large role. Aesthetic value and the interface's integration into the visual theme of the entire web site can also affect the user experience.

3.2.1 Terminology
Terminology used to describe the sections of the interface should be clear and direct. It may be obvious that shop speak should be avoided; yet many web interfaces use technical terms that are not known by the common user. Carefully consider the intended audience and their understanding of terminology. "Narrow" (Flamenco), "refine" (NCSU Libraries), "filter" (LII.org) and their synonyms are just the beginning.

Table 2. Display of the categorized overview

Dimension	Options
Terminology	▪ Common ▪ Specialized
Layout	▪ Standard ▪ Nonstandard
Complexity	▪ Simple ▪ Complex
Visual cues	▪ Color ▪ Symbols
Display of subcategories	▪ Expandable outline ▪ Display parent-child levels ▪ Display single level ▪ Pop-up bubble

3.2.2 Layout
The layout of web search interfaces are fairly standard. They usually have a header presenting general information and the majority of the page is taken up by the search results displayed below. It might include a narrow column to the side with controls to refine the search. Because this presentation is familiar, deviations can be used to highlight information.

To emphasize a particular facet, one could separate that facet from the rest of the categorized overview. NCSU Libraries locates the main category overview in a column on the left side of the web page as is standard in a two column design. The Library of Congress Classification facet is located in a box across the top of the page, placing it in a prominent position separate from the other facets. This is also significant since this facet behaves differently from the other facets. It uses a strict hierarchy in which only one category may be selected at a time.

3.2.3 Complexity
An interface that is complex may frustrate the user who may then decide that it is not worth the trouble. An interface that *appears* complex may persuade some users to give up before they have even tried. While a feature heavy-interface might provide the most flexibility, it is important to find the right balance of usability and power. EBSCOhost interface takes the simple approach. Ten subject categories are presented in a column alongside the search results -- no expandable hierarchy, no additional information, nor even discernable order. While the low complexity makes use of the interface seem low risk, it also appears limited in functionality.

There are a variety of ways to limit the number of categories displayed in order to avoid crowding the overview.

- When ordering categories by the number of hits, include only the most populous categories and provide a link to reveal the rest, such as the standard "*See more...*" link.

- Instead of displaying empty categories provide an advanced user option to reveal them.

- Allowing the user to hide categories will have the added benefit of more user control over the interface.

- Consider the uniqueness of the categories. Synonyms shouldn't be grouped as the top category results.

3.2.4 Visual Cues
Visual cues can be used to imply meaning much like underlined blue text is understood to signify a hyperlink. Flamenco uses color to extend the value of the facets to other areas of the interface besides the category overview. When grouping the search results by category, the appropriate facet color appears in the background. Additionally, categories selected during the exploratory process are displayed along with their facet color in an area showing the current search history.

3.2.5 Display of Subcategories
The display of subcategories within a hierarchy is a design question that needs more study. A common model throughout computing is the expandable and collapsible outline. When a category or an associated icon is selected, the subcategories are displayed beneath. Clusty leaves the expanded category open showing both the parent and child levels. This may present a problem if multiple categories are expanded. The overview might increase in size until the user is required to scroll the page to view the entire list.

A solution used by Flamenco and studied by Kules [6] is to provide a pop-up bubble that lists the subcategories when hovering the mouse over a category. There are two drawbacks to be considered. A portion of the interface will temporarily be hidden by the pop-up, possibly causing inconvenience if it is large. Computing all of the populated subcategories in anticipation will also require additional processing time [3].

Another option is to replace the entire set of categories with the subcategories so that only the current level is displayed. NCSU Libraries' Library of Congress Classification facet and Flamenco both use this method.

3.3 Interaction with the Interface

3.3.1 Tightly Couple Categorized Overview to Results List
One of the design principles for exploratory search interfaces proposed by Kules is to "Tightly couple category labels to result list." [6] When a category is selected and the results are refined the categorized overview should update to reflect the new results set. The category link is expected to not only refine the search by itself, but to also include the original query and any previously selected categories. Clusty's overview not only serves to refine the search results but also to broaden if a broader category is selected than the active one. Kules also notes "it is important to provide clear feedback indicating which categories are currently applied." [6] Otherwise the user may not discern what filters are refining the results or even realize that a refined result set is being displayed.

Table 3. Interaction with the Interface

Dimension	Options
Tightly couple categorized overview to results list	- Tight - Loose
Selecting categories from facets	- Select from multiple facets - Mutually exclusive facets
Selection of multiple categories	- Multiple categories - Single category
Broaden the results	- Linear breadcrumbs - Removable filters
Sort / Group results list	- Sort alphanumeric, etc. - Group by category within a facet

3.3.2 Selecting Categories From Facets
If the categories are grouped into facets it must be determined whether to allow the selection of categories from multiple facets. The alternative would be mutually exclusive facets. Of the web sites reviewed only Clusty restricts the user to selecting a category from a single facet.

3.3.3 Selection of Multiple Categories
Selecting multiple categories from within a single facet must also be considered. In the case of a classification scheme like Library of Congress this might not be desirable as each record is only assigned one LC classification. However the ability to refine results using multiple categories can work well if the records have multiple categories assigned to them, as in subject headings.

3.3.4 Broaden the Results
Much of the focus of this paper has been on using categorized overviews to refine a search. Given that the user has begun exploring his search results the ability to broaden the result list can be just as useful. Conventional web site design uses "breadcrumbs", a navigation history tracing the sequence of options the user has selected up to the current point. The user can visit any point already visited along their journey, then branch into a new direction. This method restricts the user to moving in a linear path, whether forward or backwards.

Flamenco, NCSU Libraries and LII.org use a method that is not restricted to the sequence taken. Their navigation history is represented by links as in the breadcrumb example, however the user can remove any category previously applied to the results regardless of the order it was selected in thus producing a new result set. This means that the user is not restricted to moving backwards along the path taken but may leave the path in a new direction.

However Flamenco also provides the more traditional use of breadcrumbs within each facet. As each facet's hierarchy are independent of each other this "reinforces the notion of the query consisting of a conjunction of different categories at different levels of hierarchy." [3]
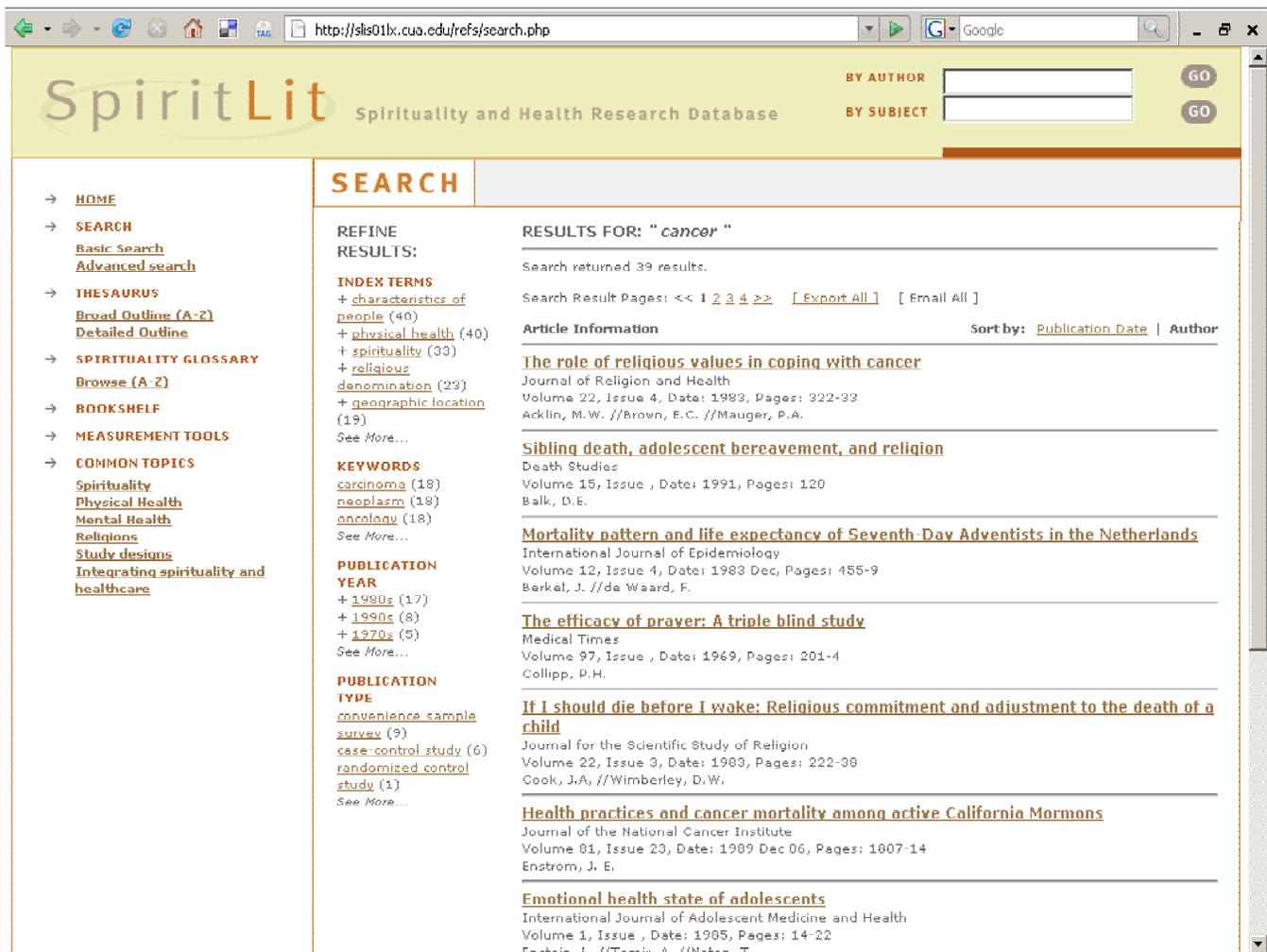
Figure 1. SpiritLit prototype search results interface featuring a categorized overview with four facets.

3.3.5 Sort / Group Results List

Options in organizing the result list allow a user to examine the results in multiple ways and hopefully gain a more complete understanding. The ability to sort results might be a common function however the importance should not be overlooked. Consider the collection in question and what information might be exposed by taking advantage of unique sort options. NCSU Libraries have a sort option labeled "most popular". It sorts the records according to circulation data and is the second most frequently used sort option [1].

Flamenco and LII.org have a feature that groups items within the results list. LII.org is a web directory and as such allows the option to group items by site domain, publisher or date added. In Flamenco a user can group search results by the categories within a particular facet.

4. CASE STUDY

A categorized overview for a prototype of SpiritLit, a bibliographic database was designed utilizing the three design areas. SpiritLit is a collection of bibliographic records regarding the intersection of health and spirituality. The web site and the database were already in place so the focus was on integrating a categorized overview with the existing design. The database records have extensive metadata which provide a rich set of information to expose through the interface.

4.1 Interface Design

When designing the layout certain restrictions emerged as the existing design was not flexible and would require a major overhaul to allow much freedom. For an initial prototype it was decided to work with the current layout and try to integrate the categorized overview as best as possible. The layout features that could not change consisted of a header across the top and a navigational column on the left. The remaining area was to be

used for the categorized overview and search results. The header includes a quick search by author and by subject providing instant search access at all times. The navigational column applies to the entire site and would have to be modified extensively to create space for the categorized overview. The design for the remaining space is similar to Findex and other software interfaces where the overview resides in a narrow column on the left and the contents in a wide column to the right [5]. The narrow column is for the categorized overview and the wide column for the actual search results. A row also goes across the top of the results for the purpose of the search history feature.

The categorized overview features four facets in this order: index terms, keywords, publication year and publication type. Initially only the top populated categories are displayed, along with the number of hits listed next to them. Beneath the categories for each facet is a "*see more…*" link. Clicking on it will add the rest of the populated categories for that facet to the overview. Each facet orders their categories by number of hits.

The index terms are organized in a hierarchy according to a thesaurus that was developed for the collection. Each record is assigned multiple index terms, therefore multiple index terms may be selected to refine the search results. Rather than use a strict hierarchical display, plus signs next to each index term allow the user to expand the category and display subcategories beneath the term. A drawback of this design is that the index terms have the potential of taking up a large amount of space when they are expanded. A benefit is that multiple hierarchical levels may be displayed simultaneously.

Keywords are flat categories assigned to the records in addition to the index terms. Displaying index terms and keywords as separate facets rather than mixed is the best way to avoid confusion, as one is hierarchical and the other flat. As the keywords are not from a controlled vocabulary many synonyms are used which devalues the uniqueness of each keyword. As a result the top populated categories are often synonyms. A possible solution to this would be to check a thesaurus for synonyms when generating the categorized overview. Though this would lengthen the overview generation time.

There is a two level hierarchy for publication year where decades are the first level and individual years form the second level. It is worth considering changing the order of this facet so that the years are listed chronological instead of by number of hits.

Publication type is a facet that applies specifically to this collection. Flat categories are used to expose terms such as: case control study, convenience sample survey, and randomized controlled trial. While it remains to be seen how useful this particular facet is, being able to take advantage of collection specific metadata can certainly aid exploration beyond clusters, which use automated term extraction.

The search history interface is displayed in a bar across the top of the search results. The space is labeled "Results for:" and followed by the query terms entered into the search box. As each category is selected to refine the results, the categories appear following the query terms in the manner one would expect breadcrumbs to function. Following each term is a link displayed as a red 'X' that when clicked will remove the category from the search refinement and recalculate the results.

Implementation of the design was accomplished using PHP to formulate queries to access the MySQL database and display the results.

4.2 Initial Feedback

Three graduate students were invited to give their feedback on the prototype categorized overview for SpiritLit. As the features were not all fully functioning at the time, they were not allowed the opportunity to use the interface – rather it was displayed for them and reactions were solicited. To begin they were given only the basic information that the prototype was of a search results interface for a bibliographic database. The subject of the database was also explained. The proctor then did a search for the subject "cancer" and the search results were displayed. At this point the proctor asked the student to describe the search interface and what outcome they expected from interacting with the interface. No prompting was given at this point.

The students identified many of the interface features correctly. They understood the ordering of the categories, the ability to refine results and the facet divisions. What was surprising was that they did not confuse the web site navigation column on the far left as having an effect on the search results. When prompted they explained that it obviously applied to the overall website.

While all of the students had to be prompted to explain their expectations for clicking on the plus signs located next to the index terms they correctly identified their ability to display the corresponding subcategories. Perhaps a graphic that is more prominent than the text plus sign "+" currently used would bring more attention to the feature. One student said that he only noticed the quick search boxes in the header after studying the page for a while. Being able to create a new search at any moment is certainly an important ability. Whether this feature needs to be more prominent is another matter to investigate.

Overall the student reactions to the categorized overview were very positive and they were excited for the capabilities it offered. It was also encouraging that the interface was indeed designed to behave as they expected it would. While much more study is needed this affirms that the design is on the right track.

5. CONCLUSION

The categorized overview has been shown to improve information retrieval and knowledge discovery [5, 6]. Recent literature and usability studies have also begun to identify some useful trends in interface design for categorized overviews. This work needs to continue in order to establish the best practices for designing these interfaces.

While reviewing recent literature and examining currently deployed search interfaces 14 dimensions of the design space emerged. These dimensions can be grouped into three areas: organization of the categorized overview, display of the categorized overview and interaction with the interface. These 14 dimensions and three areas can help to organize the design space of the interface for researchers and designers.

Designing a prototype implementation of a categorized overview proved to be a practical way to test some of the design dimensions previously identified. Working with a bibliographic database verified how valuable the descriptive metadata is in providing multiple access points to the collection. Initial user feedback was

positive and encourages further exploration of the categorized overview design.

With the increasing popularity of the simple search interface it is imperative that work be done to support exploratory search tasks. As known item retrieval is only one type of search activity it does not merit a monopoly on interfaces. Knowledge discovery will greatly benefit from the development of standards for the design space of categorized overviews.

6. REFERENCES

[1] Antelman, K., Lynema, E., & Pace, A.K. (2006). Toward a Twenty-First Century Library Catalog. *Information Technology & Libraries*, *25*(3), 128-139.

[2] Hearst, M. (1999). *User interfaces and visualization*. In *Modern Information Retrieval*. Addison-Wesley (Pp. 257-323).

[3] Hearst, M.A. (2006). Design Recommendations for Hierarchical Faceted Search Interfaces. *ACM SIGIR Workshop on Faceted Search*. Retrieved from http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf.

[4] Hearst, M., Elliott, A., English, J., et al. (2002). Finding the flow in web site search. *Commun. ACM*, *45*(9), 42-49.

[5] Käki, M. (2005). Findex: search result categories help users when document ranking fails. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 131 – 140.

[6] Kules, B. (2006). *Supporting exploratory web search with meaningful and stable categorized overviews*. (Doctoral Dissertation, University of Maryland at College Park, 2006). Retrieved from http://hcil.cs.umd.edu/trs/2006-14/2006-14.pdf.

[7] Marchionini, G. (2006). Exploratory search: from finding to understanding. *Commun. ACM*, *49*(4), 41-46.

[8] Marchionini, G., Plaisant, C., & Komlodi, A. (1998). Interfaces and tools for the library of congress national digital library program. *Information Processing and Management*, *34*(5), 535-555.

[9] Yee, K., Swearingen, K., Li, K., et al. (2003). Faceted metadata for image search and browsing. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 401-408.