Effects of Structure and Interaction Style on Distinct Search Tasks

Robert Capra, Gary Marchionini, Jung Sun Oh, Fred Stutzman, Yan Zhang
School of Information and Library Science
University of North Carolina at Chapel Hill
100 Manning Hall
rcapra3@unc.edu, march@ils.unc.edu, ohjs@email.unc.edu,
fred@metalab.unc.edu, yanz@email.unc.edu

ABSTRACT

In this paper we present the results of a study that investigates the relationships between search tasks, information architecture, and interaction style. Three kinds of search tasks (simple lookup, complex lookup and exploratory) were performed using three different user interfaces (standard web site, hierarchical text-based faceted interface, and dynamic query faceted interface) for a large-scale public corpus containing semi-structured statistical data and reports. Twenty-eight people conducted the three kinds of searches in a between-subjects study and twelve others conducted the three kinds of searches on all three systems in a within-subjects study. Quantitative results demonstrate that the alternative general-purpose user interfaces that accept automated structuring of data offer comparable effectiveness, efficiency, and aesthetics to manually constructed architectures. Qualitative results demonstrate the manual architectures are favored.

Categories & Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors, Design, Measurement.

Keywords

Browsing, facets, usability, information architecture.

INTRODUCTION

One of the defining characteristics of digital libraries (DLs) is direct access to digital primary materials. The integration of these primary materials (e.g., full or segmented files, series or aggregated collections of related materials) with library finding aids (e.g., online catalog, indexes, intermediate results and recommendations) is a double-edged sword for DL users. On one hand, the convenience of physical access overcomes spatial and temporal barriers, while on the other, the user interface becomes more complex as people move among different levels of representation without the benefits of those spatial and temporal boundaries that aid physical search. Successful access to sought information in DLs depends on three factors: searcher characteristics such as experience and knowledge, the information seeking goals as operationalized in search tasks, and the search system in the DL. In the work reported here, we investigate the interplay of the latter two factors.

There are two basic ways to find information on the web or in most DLs. First, information seekers can use queries (type words in a search box is by far the most common query tactic). Second, information seekers can browse or navigate using the site architecture and personal or social tags. There are many variants of each, and good search systems support agile use of both. It is apparent that search on the web has become dominated by query-based search such as Google and other search engines. In the work reported here, we focus on variations of interactive browsing and selection strategies.

Interactive search is strongly determined by how information is structured and by the interaction style presented in the user interface. We report results for a user study of how search is influenced by two different architectures (content-driven semi-hierarchical; and faceted structure) and three different interaction styles (hypertext selection, faceted navigation, and dynamic query) for three distinct kinds of search tasks (simple lookup, complex lookup, and exploratory). The paper first discusses issues of structure, interaction style, and search tasks; it then presents the methods used in the study, including details about the different user interfaces; next, it presents results for a between-subjects and a within-subjects portions of the overall study; and finally discusses results and implications for DL design.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
JCDL'07, June 18–23, 2007, Vancouver, British Columbia, Canada.
Copyright 2007 ACM 978-1-59593-644-8/07/0006...$5.00.

STRUCTURE, INTERACTION STYLE, SEARCH TASKS

DLs tend to layer highly structured metadata over the semi-structured primary materials in the collection to facilitate access. Thus, two grand challenges in DLs are metadata creation and representation to users. Thus, DLs are like large-scale websites in that they tend to have good structure at entry points (e.g., indexes) with less structure and metadata at primary content levels of representation (e.g., full text or data tables). Faceted structure has been shown to be effective when consistent, well-defined metadata is available. We expected that faceted structures should be especially helpful to casual users of DLs who are engaged in exploratory searches. We also expected that such users would prefer highly interactive user interface styles that support exploration.

Facets

Indian library scientist, S.R Ranganathan, first defined faceted classification, defining five mutually exclusive taxonomies of organization (personality, matter, energy, space, and time). A number of formal faceted classification systems have been developed for large-scale library collections (e.g., Colon classification) and the scope and variety of content on the WWW has naturally sparked interest in faceted organizational schemes for large websites [12]. A key notion of faceted organization is multiple representation for each item; thus rather than an object strictly belonging to only one storage bin, it can be put into many. Unlike physical books that can only be put on one shelf, digital items can be placed in many 'places,' making a faceted approach more realistic than rigid organizational schemes such as lists or simple hierarchies.

Facets are like categories or entity sets in database schemes, and each facet may have a number of attributes or sub-categories or fields (Ranganathan called them foci). A facet may have a simple list of attributes (e.g., cost ranges) or contain multi-leveled attributes (e.g., topics and subtopics). Regardless of the complexity of facets, search is facilitated because facets are mutually exclusive ensuring that result sets are in effect 'ANDed' across facets.

Hearst and her colleagues [6][17] have developed and evaluated faceted search systems for the WWW and demonstrated their value when good metadata is available. Other work has investigated the use of facets to aid navigation and search filtering on small screen displays [7]. Faceted search is especially applicable to e-commerce websites because different product features can be treated as facets. For example, in a clothing website, style, color, size, and price can each serve as facets that allow searchers (shoppers) to partition the offerings in multiple ways. DLs and large, specialized websites will require specialized facets and one goal of the study reported here is to investigate how faceted organization applies in these systems.

Interaction Styles

Shneiderman & Plaisant [14] define three general classes of interaction style (direct manipulation, selection, and command), arguing that direct manipulation is highly engaging and especially effective for casual users. In the case of search systems, Shneiderman and his colleagues have demonstrated the usefulness of dynamic query user interfaces to support search [1][13]. Entering queries into a search box is a kind of command interaction style that provides highly discrete turn taking between the information seeker and system. Good query specification is much easier when the search task is well specified (e.g., simple lookup tasks). Hierarchical navigation of layered menus and hypertext navigation of visible links tend to help people with less specified queries because the menu labels and hyperlink anchors provide strong semantic cues. Dynamic query user interfaces provide the semantic cues but also add control mechanisms such as sliders and mouse activated pop ups that closely couple search specification and results examination with rapid and visible display and undo capabilities.

Our efforts have aimed to extend the dynamic query paradigm to a design framework that incorporates different easy to control views of collections, primary objects, and events with agile control mechanisms such as mouse brushing. The AgileViews framework [4] guided the development of a generic user interface known as the Relation Browser [9][18] that can be layered on top of most database structures.

Tasks

Different tasks require different kinds of search strategies, systems, and UIs [16]. Tasks also have significant effects on people's searching performance and behaviors [15]. In terms of the purpose and the cognitive activities involved, search tasks can be classified as lookup tasks and exploratory search tasks [8]. In lookup tasks, users often have a predefined goal and are able to use well-planed analytical search strategies to query the system. Finding the answer itself is the purpose of lookup tasks. The results of lookup tasks are often discrete facts such as numbers, names, or specific files. In the current literature, lookup tasks are also termed fact retrieval, known item search, close-ended question, or question answering.

The Web has become a source for learning and knowledge discovery. Finding information itself is not the major purpose; rather, it is a means to explore the unknown knowledge territory and a medium to improve people's understanding of the world. In this exploratory search scenario, searching, browsing, and trial-and-error tactics are integrated. Users tend to be more actively engaged in the search process. More complex cognitive processes such as analyzing, synthesizing, and evaluating information are required in the exploratory search process. The diversity and complexity of information activities involved in exploratory search processes call for more interactive tools to facilitate knowledge discovery.

We aimed to separate the effects of different structuring techniques from the user interaction effects by adopting the extant metadata in a large-scale public website that in many respects mirrors the properties of a DL. This strongly constrained the tasks and more importantly limits the advantages that faceted search interfaces offer. These results are therefore a strongly parsimonious examination of faceted search. Furthermore, we were curious about the roles that interaction style plays in faceted search. We therefore used two distinct faceted search interfaces using the same underlying website data (one dynamic graphical UI and one text-based).

METHOD

There are tradeoffs between pure experimental (between-subjects) and repeated measure (within-subject) user studies. In this work, we conducted two studies using the same experimental conditions, a between-subjects study with 28 participants and a within-subjects study with 12 participants. In both studies, three kinds of access to a large website were rendered (the standard website homepage, a simple faceted front end with no graphical embellishments, and a highly interactive graphical front end with the same facet structure) and participants were assigned three kinds of search tasks (simple lookup, complex lookup, and exploratory).

Stimuli and Instrumentation

In this section, we describe the systems (three UIs), the test instruments (search tasks and elicitation forms for user responses), and the experimental protocol used in the study.

User Interfaces and Facet Construction

The US Bureau of Labor Statistics (BLS) researches and publishes information about economic and labor conditions in the United States. Their web site, www.bls.gov, contains over 67,000 documents and serves as a specialized digital library for economic statistical data. The BLS data provides a rich set of semi-structured data that is used for both focused and exploratory searches.

Three interfaces were used in this study – the existing "standard" BLS web interface and two alternatives that are layered on top of the BLS; one is a text based faceted structure with hierarchical navigation interaction style—simple facet (SF) interface, and the other is a graphic based faceted structure with a dynamic query interaction style—Relation Browser (RB). The three interfaces shared an underlying topical organization of the BLS data, but differed in their presentation level and capabilities. Each interface is briefly summarized below.

The standard BLS web interface (Figure 1) has a broad polyhierarchical structure: there are two levels of topics displayed on the web page. The topics are hand crafted and fine grained. Some second-level topics overlap with each other. The advantage of multiple entrances provided by this broad structure costs the interface some simplicity: a large amount of information is displayed in a limited screen space. The BLS web interface encourages a selection-oriented interaction style. There is no search box on the front interface; users have to access search by clicking on "Advanced search" on the upper right menu ribbon of the page. Doing so provides full text search as well as metadata search. The web interface does not offer graphic illustrations of the current search or browse status and progress. The website UI evolved over several years and benefited from a series of user needs assessments and user studies [10].

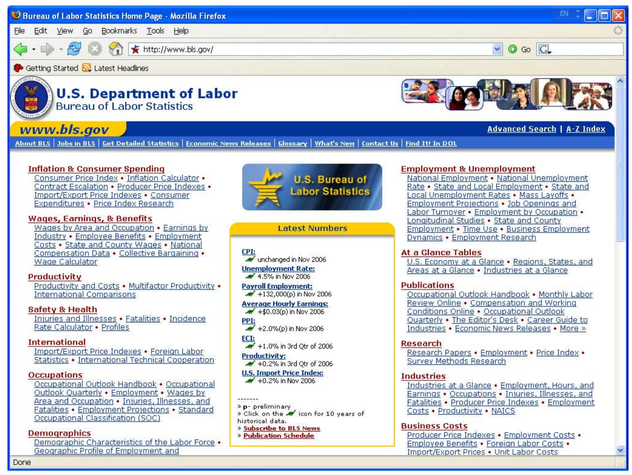

Figure 1. BLS Web interface

The SF and RB interfaces were constructed as "layers" on top of the BLS web data rather than containing copies of the data themselves. That is, the SF and RB interfaces organized the BLS documents into a facet structure and presented metadata about the documents, but linked back to the BLS web site to actually display the documents. In addition, the SF and RB interfaces only supported keyword search over metadata as described for each interface below. The structure for the SF and RB interfaces was created by scripts that crawled the BLS web site and generated indexes of documents. The scripts generated a master list of documents (including html, text, and pdf files). The master list along with a map of the URLs and link structure of the BLS site was used to create four facets: topic, genre, region, and format. The generation of each of these facets is described below.

Topic facet – For this facet, the 18 top-level categories from the BLS web site homepage were collapsed into 10 "main" topic categories: inflation, wages, productivity, safety/health, occupations, demographics, employment, publications, industries, and business costs. Documents were mapped onto topics based on the document's own URL and on the URLs of all BLS pages that linked to it. A script matched known URL patterns with known topic areas. For example, all the URLs on the BLS site with "www.bls.gov/cpi" are in the consumer price index section, which was mapped to the Inflation topic. Using the URLs and parent link structure, documents could be mapped to multiple topics.

Genre facet – The genre facet includes categories reflective of the BLS's mission to make statistical data available: news releases, static tables, dynamic tables, index pages, and an occupational handbook. BLS documents were categorized into different genres using several heuristics and the BLS directory/link structure. For example, the proportion of the <td> cells containing numbers combined with the adjacency of table titles was used for detecting static table(s) within an html document. A script that examined the number of links on a page was used to identify index pages. Pages were considered to be an index page if the proportion of the lines containing a link was higher than 20% in the main content, after removing the header and the footer and site navigational links.

Region facet – This facet included state, national, international, and eight US geographical areas used at BLS (e.g. Mid-Atlantic, Southwest). Documents were mapped onto the region facet by matching known patterns in the URL. The state category included documents about individual states. The eight geographical area categories each contained documents about that particular region as a whole. International documents were identified based on URL matching. All remaining documents were included in the national category.

Format facet – This facet included categories: html, pdf, text, zip, and other. These were identified by the file extension of the document filename (e.g. doc1.html was placed in the html category).

We adopted rule-based approaches to faceting based on several years of collaborative work with federal statistical agencies and our previous efforts to apply clustering and machine learning that demonstrated the efficacy of leveraging human expertise [3]. Although we strongly believe that the resulting facets had semantic validity, they were forced upon a large corpus of 'documents' (pages) that had evolved over time without the very strong structuring that a database-driven corpus would have. The most difficult classifications came with tables, regions, and topics. Tables are very difficult to identify (e.g., see [11]), regions often overlap generally, and BLS data tends to be national with state or metropolitan area breakouts. Topics are inherently complex with data tables that are equally pertinent to employment, careers, or prices. Thus, although our rule-bases provided a general basis for classifying documents into facets, these results are far from perfect. Additionally, many documents have no title tag and make results displays by title awkward as well as minimizing the usefulness of title string searching in SF and RB results.

The RB interface is shown in Figure 2 with two levels of hierarchical facets. Both levels are at higher conceptual levels than the facets on the BLS web interface. All interaction is done through mouse brushing and clicking with the exception of an option to conduct string searches within three fields in the results panel (title, description, and URL). RB provides multiple dynamic visual cues to indicate the current status of searching/browsing. For example, the number beside each facet shows the number of records in the database that matches that facet and the blue bars on the sub-facets show the same information graphically. When multiple facets are combined, the number of records in the database that match the combination is displayed beside each sub-facet and the bars change correspondingly to indicate the size of the results.

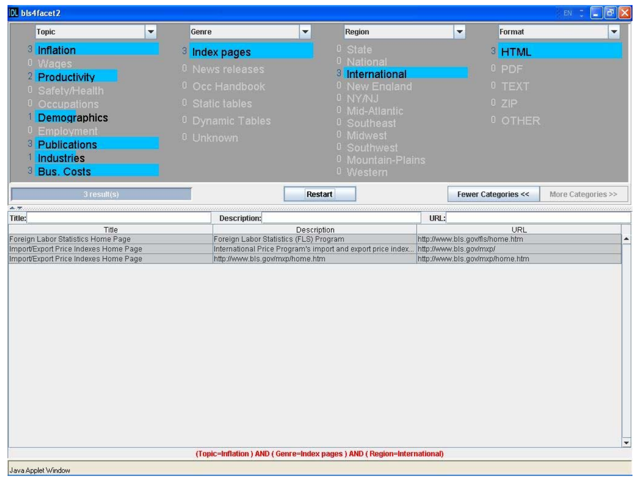

Figure 2. RB interface

The facets in the simple facet (SF) and RB are the same; but the SF interface shows only the top-level facets. Figure 3 shows a screen display for the SF. Users click on the facets to access categories The SF interface is also database-driven and text search is only active across the extracted fields rather than the full text. Visual cues under the "Current Query" panel indicate the current status of searching and browsing. Users also can change queries by removing selected facets from the current query.

Tasks

Three types of tasks with increasing levels of complexity were created for use in the study: simple lookup, complex lookup, and exploratory. In general, a lookup task involves finding a specific fact or an answer to a precisely defined question, and an exploratory task involves general or ill-defined questions and typically evolves as search progresses, possibly over multiple sessions [8]. In our study, lookup tasks were further divided into two groups based on complexity. The complexity of lookup tasks was determined by three contributing factors: 1) the number of facets to be combined to get the target page, 2) the extent to which a higher level thinking (rather than simple recognition) such as comparing pages in the result set is required, 3) the navigation path to a target BLS web page.

Because the BLS website is a highly specialized corpus devoted to economic data in the United States organized across very specific time periods (e.g., monthly releases of price or employment data), we decided to include the US as a geographic facet and a month or year as a temporal facet to provide context for all search tasks in our study. Thus, the simple lookup tasks were constructed around a single economic facet but also included the spatial and temporal facets to provide context for the searchers. The complex lookup tasks involve additional facets including genre (e.g. press release) and/or region.

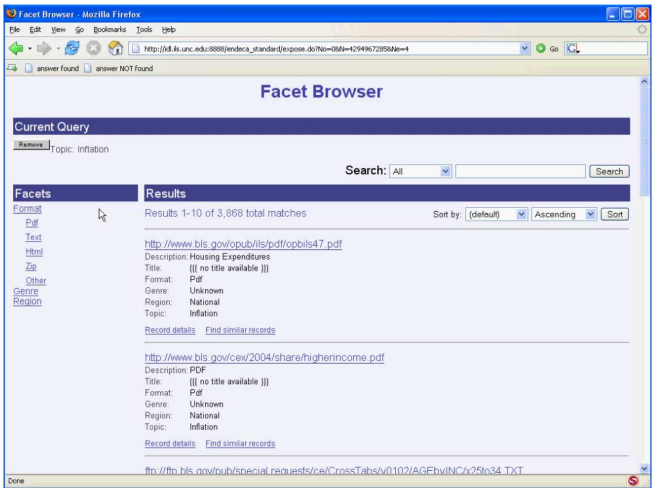

Figure 3. SF interface

As shown in Table 1, by not specifying a particular value for the temporal facet (e.g. 'the most recent') the complex lookup tasks also required subjects to examine and compare pages. In addition, the path to the target page in the BLS website was considered so that the simple lookup tasks could be resolved with a small number of steps (within 2 clicks) starting from the BLS homepage.

Task Type	Task Statement
Simple Lookup	How many job openings were there in the US in April, 2006?
Complex Lookup	Find the most recent press release that reports the average wage for employees in Dallas County, TX
Exploratory	Collecting data is not without controversy. Imagine you are writing a paper that assesses different criticisms that have been made on how BLS creates the consumer price index. Find several factors that would be addressed in your paper

Table 1. Task Statements.

Experimental Protocols

For the study operation and data collection, we developed a dedicated experimental protocol website that administered all stimuli and recorded participant activities. All the tasks and questionnaires, except for an informal debriefing interview, were administered within the system. For each task, a question and a link to one of three interfaces were presented to the participant according to a counter-balanced experimental settings stored in the database. As participants searched and examined BLS pages, they were instructed to highlight the text containing the answer and click on a bookmarklet icon that was developed to capture answers to the database. The URL, title of the page, and the highlighted text were recorded in a study administration page that participants reviewed, and optionally edited before submitting as their answer. Also included on this page were five-point Likert scales for confidence (how confident they were that they found the right answer to the task), satisfaction (how satisfied they were with the interface), and mental effort (how much mental effort they used to complete the task). Upon submission of the answer, time spent for the task was also recorded.

A series of questionnaires were administered as participants worked. As mentioned above, after each task trial, the 5-point Likert scales for confidence, mental load, and satisfaction were completed. Tasks were grouped into sets: for the between-subject study, there were 3 sets, one set for each task type; for the within-subjects study there were 3 sets: one for each interface. After each set of tasks, participants completed a questionnaire that included thirteen 5-point Likert scaled statements related to usefulness (e.g., Using this system enables me to find information quickly) and usability (e.g., I found this system easy to use). This instrument was adapted from the scales in Davis [2]. These responses were averaged into a usefulness score and a usability (ease-of-use) score for each participant-interface combination. Two 7-point semantic differential scales adapted from Ghani et al. [5] with five word pairs that focused on engagement and enjoyment (e.g. Using the retrieval system is interesting / not interesting; How you felt using the retrieval system: attention was focused / attention was not focused) were also completed and responses combined into engagement and enjoyment scores for analysis. At the end of each study, a questionnaire with specific questions related to structure and interaction style was completed. Additionally, if time and availability permitted, the experimenters conducted a brief post-session interview with participants about their experiences using the system(s).

Procedure

Two studies were conducted: a 3 X 3 between-subjects design with 28 participants where each participant used one of the three interfaces (BLS, SF, RB) for all three kinds of search task (simple, complex, exploratory); and a 3 X 3 within-subjects design with 12 participants where each participant used all three interfaces to complete the three kinds of search tasks. In the between-subject study, participants completed three trials of the simple and complex tasks and one exploratory task (seven tasks total). In the within-subjects study, one trial of each task type was presented for each interface for a total of 9 tasks.

This two-phase approach was used to give us the benefit of comparisons within a small number of individuals while gaining the statistical power a randomized trial with multiple trials offers in the between-subjects design.

Participants were recruited from the University of North Carolina at Chapel Hill campus and received a $20 incentive for participation in one of the studies. The sessions lasted approximately 1.5 to 2 hours and were conducted in a UNC computer lab where participants were positioned to minimize the chances of seeing another participant's screen. The between group included 9 male and 19 female participants ranging in ages from 18 to 35 from 17 different academic disciplines and the within group included 5 males and 7 females ranging in age from 18 to 66 from 10 different disciplines.

Between-subjects study. Upon arrival, participants were seated at a computer, given an overview of the study and asked to read and sign an informed consent form. Participants first completed a short demographic questionnaire that focused on search experience. Each participant was randomly assigned to one of the user interface conditions and asked to complete a series of three different tasks in a prescribed order (simple, complex, then exploratory). Before proceeding, participants watched a short video demonstration we created to describe the use of the data collection system. For the SF and RB conditions, participants watched a second video to describe the basic use of interface to which they were assigned. For each task, after eight minutes, a pop-up message was displayed informing the participant that they had worked on the task for eight minutes and should try to wrap up soon, however, no specific time limit was enforced. For the exploratory search tasks, the pop-up window did not appear until 20 minutes had passed. Participants were allowed to move on to the next question as soon as they completed each task and the short questionnaire to assess confidence, mental load, and satisfaction. Upon completion of the three trials for the single fact retrieval task, participants completed a questionnaire on their satisfaction for that system/task combination and then moved to the next type of task. After completing all the tasks, participants were presented with a final questionnaire about their overall experience with the search system.

Within-subjects study. The same procedure as above was used except that participants used each of the three different user interfaces but only completed one trial for each of the three task types for each of the user interfaces. A counterbalancing scheme (6 combinations of BLS, SF, and RB systems) was used to assign participants to the interfaces. The appropriate training videos were displayed before the participant began using each system. The final questionnaire included an added set of questions that focused on eliciting comparative data on the participant's experiences with the three systems.

Both studies – dependent measures. The main dependent measures for both studies are: task completion time (measured from the time the task was displayed until the answer was submitted using the bookmarklet form); confidence, mental load, and satisfaction (as measured on the 5-point Likert scales presented after each task); usefulness, usability, engagement, and enjoyment (as measured by the Likert and semantic differential scales presented after each set of tasks); self-reported level of understanding of the content and the structure of the overall BLS website (post session-questionnaire); and self-reported choice of system for specific and general search tasks (post-system questionnaire). Open-ended data collected on the questionnaires was analyzed qualitatively to investigate the relationships among system and task.

Accuracy was measured for the simple and complex lookup tasks. Two coders each rated participants' responses on an integer scale from 0 (low) to 2 (high). Inter-coder agreement was high: 83% for the between-subjects study. The average of the two coders' accuracy scores was used as the accuracy measure in the analysis. Participant responses to the exploratory tasks were always considered accurate since the task was an open-ended task to find information that would help in writing a report about how BLS computes statistics. Thus the exploratory tasks were not included in the statistical analyses of accuracy.

RESULTS

Results are first given for the between-subjects study, focusing on the quantitative findings. Then, the primarily qualitative results for the within-subjects study are given.

During the study, we were encouraged that, overall, participants took the tasks seriously and used the interfaces to find answers to the tasks. Users were instructed not to use web search engines to look for information outside of the BLS web site and in our observations, we did not observe anyone doing so. In addition, we observed that participants made good use of the SF and RB interfaces and did not simply start the task with the "layered" interface and then spend most of their time browsing the BLS site. Rather, participants would refer back to the SF and RB to look for additional matching pages and to issue and re-formulate their queries. For the exploratory tasks, participants overall made a good effort to locate several pieces of information – on average across both studies, participants found and submitted 2.7 items. The overall average accuracy for the simple and complex tasks was also high – 1.66 (stdev = 0.34) out of a scale from 0 to 2.

Between-Subjects Data

For each task, participants clicked one of two bookmarklets to indicate that they had "found" or "not found" an answer for the task. In the between-subjects study, participants reported finding an answer for an average of 5.96 (sd = 1.17) tasks out of the 7 tasks. One participant reported finding an answer for only 2 of the 7 tasks. This participant was deemed to be an outlier on the basis of a z-test (z=3.38) and not included in the analysis.

Task Completion Time: Different amounts of time were suggested to participants for the three task types (eight minutes for the simple and complex lookup tasks, 20 minutes for the exploratory task). Therefore, separate one-way ANOVAs were conducted across the three systems for each of the three different task types. No significant differences (α = 0.05) in task completion times were found among the three interfaces. Table 2 summarizes the mean task completion times for each interface and task type combination.

	n	**BLS** Mean (SD)	**SF** Mean (SD)	**RB** Mean (SD)
Simple	9	300 (99)	300 (94)	292 (112)
Complex	9	275 (91)	260 (84)	282 (130)
Exploratory	9	588 (295)	518 (354)	438 (251)

Table 2. Task Completion Time in Seconds

Accuracy, Confidence, Satisfaction, and Mental Effort: These measures were comparable across the task types, so analysis was conducted using two-way ANOVAs. There were three trials each for the simple and complex tasks, and one trial of the exploratory task type. Trials were averaged to provide an overall measure for each task type for use in the ANOVAs.

		n	**BLS** Mean (SD)	**SF** Mean (SD)	**RB** Mean (SD)
Accy	**Simp**	9	1.94 (0.13)	1.34 (0.74)	1.61 (0.70)
0-2	**Comp**	9	1.41 (0.60)	1.46 (0.50)	1.71 (0.43)
	Exp	9	*	*	*
Conf	**Simp**	9	4.28 (0.47)	3.87 (0.81)	3.87 (0.40)
1-5	**Comp**	9	3.78 (0.67)	3.80 (0.85)	3.31 (0.45)
	Exp	9	3.85 (0.93)	3.69 (0.74)	3.59 (0.49)
Satsf	**Simp**	9	2.89 (0.47)	3.15 (0.93)	3.22 (1.01)
1-5	**Comp**	9	2.96 (0.56)	3.26 (0.66)	3.22 (0.85)
	Exp	9	3.43 (0.64)	3.50 (0.73)	3.15 (1.11)
Meff	**Simp**	9	3.04 (0.70)	3.07 (0.70)	3.11 (0.37)
1-5	**Comp**	9	3.04 (0.73)	3.48 (0.60)	3.19 (0.69)
	Exp	9	3.30 (0.62)	3.44 (0.45)	3.00 (0.56)

Table 3. Accuracy, Confidence, Satisfaction, Mental Effort

For all four measures, no significant differences (α = 0.05) were found for either of the two main effects (task type, interface) or for the task type by interface interaction. Table 3 summarizes the means and standard deviations for accuracy (0=low, 2=high), confidence, satisfaction, and mental effort (1=low, 5=high).

* Note that for the exploratory condition, responses were always considered accurate, so exploratory tasks were not included in the accuracy analysis.

Post-trial Questionnaire: The post-trial questionnaire was presented after each set of task type trials. The questionnaire measured: task usefulness, ease of use, enjoyment and engagement (1=low, 5=high). Analysis using two-way ANOVAs (task type x interface) reported no significant main effects or interactions. Table 4 summarizes the results of the post-trial questionnaire.

		n	**BLS** Mean (SD)	**SF** Mean (SD)	**RB** Mean (SD)
Usef	**Simp**	9	3.30 (1.02)	2.98 (0.88)	2.85 (1.07)
1-5	**Comp**	9	3.54 (0.61)	3.09 (1.06)	2.87 (0.82)
	Exp	9	3.43 (0.84)	3.09 (0.98)	2.93 (0.91)
Easeu	**Simp**	9	3.50 (0.92)	2.46 (1.32)	3.00 (1.23)
1-5	**Comp**	9	3.46 (0.61)	2.81 (1.36)	2.61 (0.91)
	Exp	9	3.76 (0.83)	2.87 (0.98)	2.76 (1.21)
Enjoy	**Simp**	9	4.75 (1.51)	4.39 (1.71)	4.81 (1.16)
1-7	**Comp**	9	5.50 (1.11)	4.61 (1.50)	4.39 (1.15)
	Exp	9	5.11 (1.41)	4.69 (1.49)	4.61 (1.54)
Engm	**Simp**	9	3.95 (2.06)	3.81 (1.02)	3.81 (1.47)
1-7	**Comp**	9	3.64 (1.40)	3.58 (0.78)	3.25 (0.84)
	Exp	9	3.36 (1.56)	3.64 (0.71)	3.56 (1.56)

Table 4. Usefulness, Ease of use, Enjoyment, Engagement

Within-Subjects Data

Quantitative statistical analysis of the within-subjects data revealed results similar to those found in the between-subject analysis with one exception: a main effect of task type was observed for satisfaction with both simple and complex conditions having greater satisfaction ratings than the exploratory condition. Qualitative data from the within-subjects study is reported in the remainder of this section.

Post-Session Questionnaire: After completing all the tasks for all the interface conditions, the within-subject participants completed a post-session questionnaire that asked free-response questions about their experiences with the interfaces. Table 5 summarizes the responses. Due to a limitation in the length of the responses recorded by the data collection system, preferences could not be determined for all questions and thus for some questions the counts sum to less than 12. Responses to the post-session questions summarized in Table 5 show a general preference for the standard BLS web site over the SF or RB interfaces.

To gain more insight into the features and qualities that participants found important, a simple content analysis was performed on the responses to Q51 through Q54. A total of 62 specific mentions of features were identified in these four questions from the 12 participants. The most frequently mentioned features are described in the remainder of this section.

Responses to these questions indicated that users recognized good organization of topics and structure. The good organization and visible topical layout of the BLS homepage were the most frequently mentioned factors in questions Q51 to Q54. Eight out of the 12 participants indicated the visible structure of the topical categories as a factor and seven of the 12 noted that good organization was a factor in their system preferences. As one participant noted, "All BLS categories were clearly spelled out on their front page and thus that page was easiest to understand in terms of organization."

Post-Session Questions	**BLS**	**SF**	**RB**
(Q51) Which of the three systems was easiest to use? Why?	7	3	1
(Q52) Which of the three systems helped you complete the tasks more effectively? Why?	5	3	2
(Q53) Which of the three systems helped you to develop and understanding of what kinds of information is available at the BLS? Why?	10	0	1
(Q54) Which of the three systems helped you to develop an understanding of how the BLS web site is organized? Why?	11	0	1
(Q56) Which of the three systems helped you to understand the organization of the BLS web site better?	11	0	1
(Q57) Which of the three systems helped you explore the BLS website better?	8	2	2
(Q58) Which of the three systems helped you to assess the BLS content better?	6	4	2
(Q60) If you could use one of these three different user interfaces for all websites, which one would you select? Why?	6	2	3

Table 5. Interface Preferences

However, some users found the large number of links on the BLS page overwhelming, even though the display was effective at showing the organization of the site. One user commented, "I hated the BLS website because I felt overwhelmed with possibilities and left without any real sense of how to navigate them effectively. The vast quantities of information presented there, however, gave me a better sense of how much is available."

Participants also indicated that it is important to have familiar, powerful interfaces for keyword search and presentation of relevant results (5 participants mentioned one of these aspects of search). One response in particular typified this sentiment: "The BLS website itself was the easiest to use because I am used to search engines where you can just type in a phrase or a bunch of keywords and find a bunch of results automatically."

Participants also appreciated options present in the RB to easily narrow results, see the number of results, and visually see relationships in the data (4 participants mentioned one of these features). For example, one user noted that they liked, "The relation system because I was able to narrow down my search much more quickly and able to find exactly what I was looking for." Other comments reflected an understanding of the unique features of the RB interface: "The relational [interface] allowed me to see how many results were present for a certain topic, and then their relationship to other topics."

Several additional questions were contained in the post-session questionnaire and were analyzed with frequency analysis. One question asked, "Was any interface good for a particular type of task?" Seven responses noted that SF and RB would be good for finding when you know a specific type of information (e.g. press release) or when you start with a narrow search. One participant summed up the differences between the BLS and RB: "The Relational Browser was clearly best for looking up very specific information; the website itself for browsing/exploring."

Another question asked, "Did the graphics in the Relation Browser add anything to your search experience? What?" Responses to this question were split: seven participants not finding the graphics helpful and five reporting that they were. Responses ranged from, "Yes, in terms of conveying to me in a quick manner the amounts of documents I'd be looking at once I sifted through with a keyword," to, "No! They were visually distracting, and I tried to ignore them."

Two questions asked participants to describe what they found to be the best and worst aspects of the interface. For the BLS web site, the best features included: good information, search functionality, clear topic categories, and page organization. The worst aspects were too much information, poor search results, and difficulty manipulating data. For the SF interface, useful facets and the ability to narrow down the search were the most mentioned best features. The worst features for SF included limited search capabilities and lack of page descriptions/context in search results. For the RB interface, participants listed the use of facets to narrow down a search, and support for gaining an overview of the data as the best features. The worst aspects of the RB included limited search capabilities and problems with the facets (e.g. not intuitive, unbalanced amount of information under the categories). It is interesting to note that there was no overlap in the most liked features of the BLS interface and the best features of either facet system. Although this is not surprising considering the differences in interaction styles, it illustrates that participants recognized and appreciated the respective strengths of each interaction style.

Overall, the comments suggest that the BLS was often preferred for its good organization and visual layout of the site contents, and because it offers a familiar look and feel. It is interesting to note that no difference was found between the BLS, SF and RB in task completion time, accuracy, satisfaction and confidence measures for the between-subjects participants even though the SF and RB did not have as much manual design and were not as familiar as the BLS site.

DISCUSSION

Structure. These results suggest that handcrafted structure is effective in supporting the full range of search tasks even when the overall design is complex and information intensive. However, the results also show that semi-automated methods can be used to construct usable and effective facet-based interfaces.

A significant challenge in both handcrafted and semi-automated methods is defining useful facets (a classification task). In previous work, we used clustering and machine learning techniques to identify topical facets with limited success [3]. In collections with consistent properties across all items, this is relatively straightforward and faceted structures can be highly effective (e.g., e-commerce sites where products can be classified with common and precise features like price, color, and size).

The BLS website lays out the overall topical structure of the site, unfolding details within hierarchical contexts. On the other hand, facet systems provide 'slicing and dicing' of the corpus, giving users control over what is being searched. In designing this study, we expected that facet systems should facilitate exploration whereas the BLS site should work well for lookup tasks. However, many subjects in this study found the BLS website itself more useful for browsing purposes than the other two facet systems. This result seems in part to be due to the dominance of the topic facet over other facets in the BLS website. It suggests that the characteristics of the corpus are important factors to consider when offering alternative interaction styles or interfaces. Even though the flexible and adaptable organization provided by a facet system works well in many applications, if there is a predominant model of the information space (as in our case of topical structure in the BLS corpus) a well designed hierarchical organization might be preferred.

Interaction Style. Participants expressed the importance of familiar interaction styles and interfaces in the questionnaires and post-session interviews. Participants were clearly familiar with and influenced by web search engines in which they could type in search terms and quickly get a list of relevant documents. Even in the SF and RB interfaces, participants often used keyword search in the results. However, participants did note many positive benefits of the topical organization of the BLS web site, and the ability to quickly narrow searches using the facets in the SF and RB interfaces.

It was somewhat surprising that the interaction style and graphical representations in the RB conditions were not better liked. This may be due to the layering effects of the RB over the BLS baseline where the primary content is found (the SF condition also had this disadvantage). As a participant noted, "BLS … was most directly related to the inquiry topics. The other two browsers frequently refered [sic] the searcher to the BLS site." In the post-session interviews, many participants expressed a desire to be "in control" of the information seeking process. Some felt that the SF and RB interfaces imposed the use of facets, limiting their control or usual search tactics.

There may also have been a familiarity effect at play in that people are used to website UIs and faceted selection styles. As more web sites embrace interactive user interfaces such as those enabled by AJAX (e.g. Google maps), the expectations and tactics of users will likely evolve.

Overall, these results suggest that interfaces should support familiar interaction styles such as keyword search, but that users gain benefits from support for facets and topic organization implemented in a flexible fashion.

Tasks. The assigned tasks in this study were designed to be realistic but also to reflect multiple facets. Because the BLS system was the base for answers, all tasks were carefully designed so that at least one path from the BLS homepage would lead to a page with pertinent information. Mapping these to the classifications in the other two conditions was difficult without strongly biasing answers to one of the facet categories.

All the tasks in our study asked participants to find particular pieces of information, either with a closed set of answers (for the simple and complex tasks), or in an open style for the exploratory task. These types of tasks focus primarily on finding information in local contexts – the documents themselves. The BLS site provided much greater local context for documents than the SF and RB, which listed only metadata fields. However, one of the strengths of the RB is that it illustrates more global context for how documents are situated within the facet structure. Although we did ask questions in the post-session questionnaires about what people had learned about the BLS data structure and organization, it is difficult to assess this type of knowledge and our tasks did not specifically focus on learning site structure.

Design Recommendations. Overall, these results demonstrate that automated metadata and facet extraction presented in generic screen display forms is a feasible alternative to handcrafted structures because they did not penalize effectiveness, efficiency, or aesthetics for the participants in this study. DL designers should consider offering alternative UIs that use rule-based automated systems that are tuned to corpus characteristics.

CONCLUSION AND FUTURE WORK

In our study, we considered a hand-crafted web site along with two facet systems using semi-automatically generated facet data. Future work in this area could further explore how the quality and construction of facets affects the search process, comparing multiple facet structures in a single facet-based interface.

Our study showed that participants liked features of the facet interfaces, but preferred the familiarity afforded by the BLS interface (e.g. hyperlinks and keyword search). Our results also provide evidence from a controlled study that while faceted search may not be the "best" in all situations, it can support a search interface that is comparable to handcrafted structures in terms of effectiveness and efficiency; even when the dataset is semi-structured and the metadata were created using semi-automated methods.

Future work should explore systems that support multiple interaction styles to see how the benefits of facet structure can be integrated in a flexible fashion with familiar interfaces. In such systems, how and when are facets / search / hyperlinks used? How are the interaction styles used together?

Imagine what would be possible if the extracted data were mapped onto a full-featured facet-based content management system like Endeca or Flamenco with full-text search and facile facet manipulation features enabled. It is feasible to use rule bases and possibly machine learning techniques to automate large portions of the classification and indexing for large amounts of semi-structured data in a DL to map it onto the highly structured architectures that these general-purpose front-end systems can deliver.

ACKNOWLEDGMENTS

The authors wish to thank the participants in the user study. This work was partially supported by NSF grant EIA 0131824.

REFERENCES

1. Ahlberg, C., Williamson, C., and Shneiderman, B. (1992). Dynamic queries for information exploration: An implementation and evaluation. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 619-626.
2. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.
3. Efron, M., Elsas, J., Marchionini, G., and Zhang J. (2004). Machine learning for information architecture in a large governmental website. *ACM/IEEE Joint Conference on Digital Libraries 2004* (Tuscon, AZ, June 7-11, 2004), 151-159.
4. Geisler, G. (2003). AgileViews: A Framework for Creating More Effective Information Seeking Interfaces. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.
5. Ghani, J. A., Supnick, R., & Rooney, P. (1991). The experience of flow in computer-mediated and in face-to-face groups. *Proceedings of the Twelfth International Conference on Information Systems (December 16-18, 1991, New York),* 229-237.
6. Hearst, M. (2006). Clustering versus faceted categories for information exploration. *CACM*, 49(4), 59-61.
7. Karlson, A., Robertson, G., Robbins, D., Czerwinski, M. & Smith, G. FaThumb: A facet-based interface for mobile search. *Proc. CHI 2006*, ACM Press (2006), 711-720. NY: ACM Press.
8. Marchionini, G. (2006). Exploratory search: From finding to understanding*.* *CACM*, 49(4), 41-46.
9. Marchionini, G. and Brunk, B. (2003). Towards a general relation browser: A GUI for information architects. *Journal of Digital Information*, 4(1).
10. Marchionini, G. & Levi, M. (2003). Digital government services: The Bureau of Labor Statistics Case. *Interactions*: *New visions of human-computer interaction*, 10(4), July-August p. 18-27.
11. Pinto, D., McCallum, A., Wei, X., and Croft, W. B. 2003. Table extraction using conditional random fields. In *Proceedings of the 26th Annual international ACM SIGIR Conference on Research and Development in informaion Retrieval* (Toronto, Canada, July 28 - August 01, 2003). SIGIR '03. ACM Press, New York, NY, 235-242.
12. Rosenfeld, L. & Morville, P. (2002*). Information architecture for the World Wide Web*. Sepastopol, CA: O'Reilly.
13. Shneiderman, B. (1994), Dynamic Queries for Visual Information Seeking, *IEEE Software*, 11(6): 70-77.
14. Shneiderman, B. & Plaisant. C. (2005). Designing the user interface: Strategies for effective human-computer interaction (4^{th} Ed.). Boston: Addison-Wesley.
15. Vakkari, P. (1999). Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. Information Processing and Management 35(6):819-837.
16. Woodruff, A., Rosenholtz, R., Morrison, J., Faulring, A., and Pirolli, P. (2002). A Comparison of the Use of Text Summaries, Plain Thumbnails, and Enhanced Thumbnails for Web Search Tasks. Journal of the American Society for Information Science and Technology, 53(2):172-185.
17. Yee, K., Swearingen, K., Li, K., and Hearst, M. (2003). Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA, April 05 - 10, 2003). CHI '03. ACM Press, New York, NY, 401-408.
18. Zhang, J. and Marchionini, G. (2005). Evaluation and evolution of a browse and search interface: relation browser. In *Proceedings of the 2005 National Conference on Digital Government Research* (Atlanta, Georgia, May 15 - 18, 2005). ACM International Conference Proceeding Series, vol. 89. Digital Government Research Center, 179-188.