Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Information Processing and Management xxx (2007) xxx–xxx
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

Users can change their web search tactics: Design guidelines for categorized overviews

Bill Kules *, Ben Shneiderman

Department of Computer Science, Human-Computer Interaction Laboratory, and Institute for Advanced Computer Studies, University of Maryland at College Park, College Park, MD 20742, United States

Received 29 December 2006; received in revised form 5 June 2007; accepted 24 July 2007

Abstract

Categorized overviews of web search results are a promising way to support user exploration, understanding, and discovery. These search interfaces combine a metadata-based overview with the list of search results to enable a rich form of interaction. A study of 24 sophisticated users carrying out complex tasks suggests how searchers may adapt their search tactics when using categorized overviews. This mixed methods study evaluated categorized overviews of web search results organized into thematic, geographic, and government categories. Participants conducted four exploratory searches during a 2-hour session to generate ideas for newspaper articles about specified topics such as "human smuggling." Results showed that subjects explored deeper while feeling more organized, and that the categorized overview helped subjects better assess their results, although no significant differences were detected in the quality of the article ideas. A qualitative analysis of searcher comments identified seven tactics that participants reported adopting when using categorized overviews. This paper concludes by proposing a set of guidelines for the design of exploratory search interfaces. An understanding of the impact of categorized overviews on search tactics will be useful to web search researchers, search interface designers, information architects and web developers.

Keywords: Exploratory search; Web search; Categorized overviews; Search user interfaces; Categorization; Categorized search results; Search result visualization; Information seeking; Human-computer interaction

1. Introduction

Categorized overviews of web search results are a promising way to support user exploration, understanding, and discovery. These search interfaces, also referred to as faceted or guided search interfaces, combine a metadata-based overview with the list of search results to enable a rich form of interaction. The strategies and tactics that searchers use are affected by the capabilities provided by the search interface (Bates, 1990;

* Corresponding author. Present address: School of Library and Information Science, The Catholic University of America, Washington, DC 20064, United States. Tel.: +1 202 319 6278; fax: +1 240 599 7671.
E-mail addresses: kules@cua.edu (B. Kules), ben@cs.umd.edu (B. Shneiderman).

0306-4573/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2007.07.014

Golovchinsky, 1997). Strategies are high level plans for the whole search, and tactics are individual actions or sequences of actions (often called moves) taken to further the search (Bates, 1979; Marchionini, 1995). Searchers can take numerous actions while examining search results (Bates, 1990; Fidel, 1985; Garcia & Sicilia, 2003; Marchionini, 1995; Shneiderman & Plaisant, 2004; Wildemuth, 2004).

Designers build interfaces to support specific strategies and tactics, based on intuition or analysis. But the effect of new capabilities on search tactics may not be what designers anticipate. Unexpected problems may negate expected benefits. Serendipitous possibilities may present to searchers. In response, searchers may adapt their tactics and strategies as they become familiar with the capabilities. Our research seeks to understand how exploratory search systems with rich user interfaces change the way that searchers think about and pursue their searches. What strategies and tactics do exploratory search interfaces enable? And, ultimately, do they enable searchers to achieve their higher-level objectives? One outcome of this research is a set of guidelines that search interface designers can use to support exploratory searchers through the design of categorized overview interfaces.

1.1. Research questions

Three research questions for this study were:

1. How do searchers think differently about their search tactics when categorized overviews are available to augment the result list?
2. What kinds of behaviors do searchers exhibit when categorized overviews are available?
3. In what ways could the presence of categorized overviews affect the quality of the search outcome?

Evaluation of exploratory search systems is an exciting research challenge (White, Muresan, & Marchionini, 2006; White, Kules, Drucker, & Schraefel, 2006). The situated nature of exploratory search tasks can lead to many different task outcomes for different searchers, making it difficult to specify quantitative performance measures like time to completion, error rates, precision, or recall. Completing an exploratory task often involves developing and refining an information need that is specific to the individual. Documents that have great utility or novelty to one person may have little value to another, because of variations in domain knowledge, interests, and previously encountered information, so establishing ground truth for a measure of relevance is problematic. To mitigate these problems, this study adopted a mixed methods approach. It shows how a combination of qualitative and quantitative methods can address research questions related to exploratory search.

Following a brief description of related work, the experiment design is described Section 3. Section 4 presents the results. Section 5 discusses the results, identifying seven tactics that searchers adopted. Section 6 proposes eight design guidelines suggested or refined by the study. Section 7 concludes with a summary of the contributions and suggestions for future research.

2. Related work

At least one commercial search engine (Exalead.com) has implemented categorized overviews of web search results. However we are not aware of any studies of this approach. Evaluations of categorized overviews in non-web domains have assessed and rated the quality of a task outcome to generate quantitative measures on a lesson plan creation task (Kabel, Hoog, Wielinga, & Anjewierden, 2004) or measured incidental learning that occurred during a search session (Pirolli, Schank, Hearst, & Diehl, 1996). Exploratory tasks have been decomposed or narrowed to constrain the task (Janecek & Pu, 2005). Log studies have been used to explore user needs and mistakes in web search (Jansen, Spink, & Saracevic, 2000).

Task-based evaluation of exploratory search systems using controlled experiments has been effective for showing subjective satisfaction differences between systems, but less effective at showing objective differences in task performance, particularly in task outcomes. (Kabel et al., 2004; Yee, Swearingen, Li, & Hearst, 2003). Evaluations have assessed and rated the quality of a task outcome to generate quantitative measures on a lesson plan creation task (Kabel et al., 2004) or measured incidental learning that occurred during a search session (Pirolli et al., 1996). Exploratory tasks have been decomposed or narrowed to constrain the task (Janecek & Pu, 2005). A combination of quantitative and qualitative evaluation methods have also been used (Toms, Freund, Kopak, & Bartlett, 2003; Yee et al., 2003).

Research prototypes and commercial search engines have incorporated categorized overviews, but there have been few, if any, user studies of categorized overviews for exploratory web search, and there is little research explaining whether they are effective, why, and under what circumstances. Research is needed to understand how categorized overviews change the way users conduct web searches, to guide the design of search engine interfaces, and to justify the entry and maintenance of category metadata.

3. Experimental design

Based on previous research (Kules and Shneiderman, submitted for publication), we expected to observe quantifiable and significant differences in specific behaviors and preferences. For example, we expected that searchers would explore deeper in their result lists using the categorized overview. Selected quantitative results are highlighted in Section 4.1; however this paper focuses on the qualitative results. We anticipated that the interface would prompt additional behavioral changes, but there was no a priori list. A qualitative approach used a combination of observation and semi-structured interview questions. The study was initially designed to also investigate the effect of broad and narrow topics, but that aspect was problematic and is not discussed here. Details of the complete study can be found in Kules (2006b).

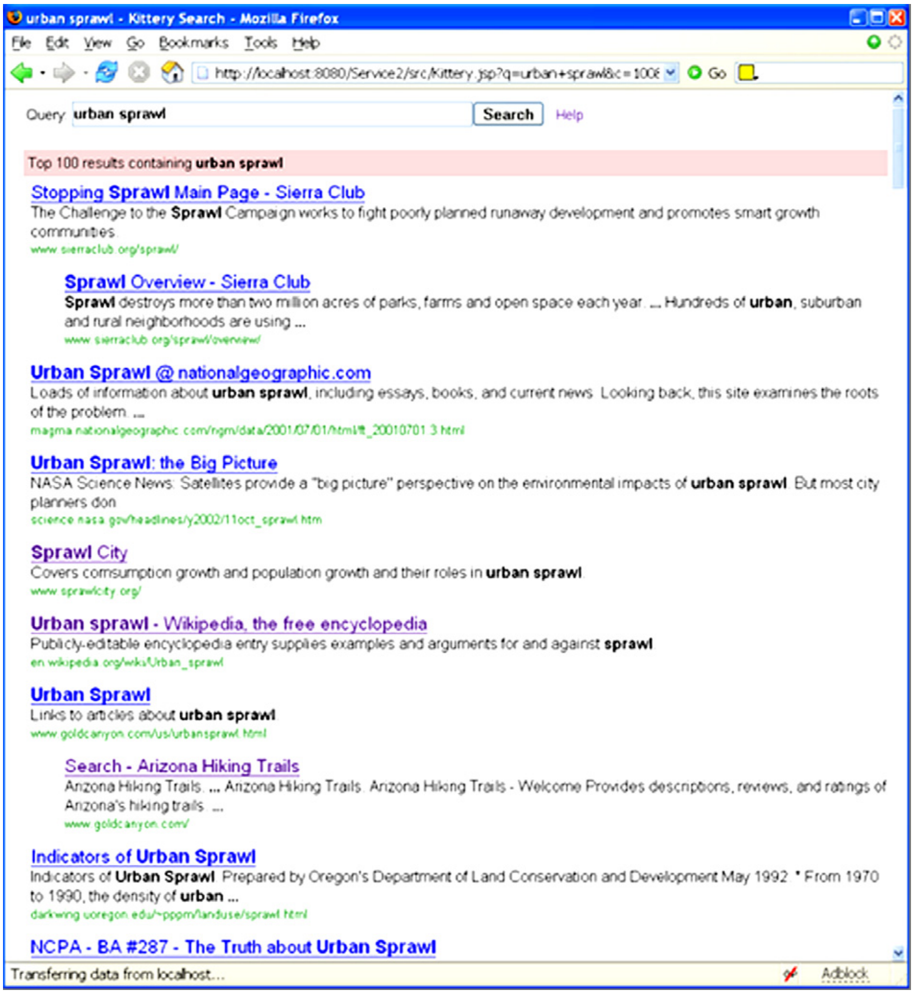
3.1. The SERVICE search system

This study used the SERVICE (SEarch Result Visualization and Interactive Categorized Exploration) search system (Kules, 2006b) (see http://www.cs.umd.edu/hcil/categorizedoverview). The SERVICE system was designed to be a flexible, extensible architecture and framework for research in categorizing search interfaces. For this study, it was configured to present two interfaces: a baseline interface that displays a typical list of search results, similar to Google (Fig. 1), and a categorized overview interface that displays the overview to the left of the result list (Fig. 2).

With the categorized overview enabled, clicking on a category filters (or narrows) the displayed results to just the pages within that category. Searchers can remove filters by clicking on a facet label (which removes filters based on that facet) or a special link (which removes all filters). Moving the pointer over a category highlights the visible search results in that category in yellow. It also opens a small pop-up window with a list of populated subcategories. Moving the pointer over a result highlights all the categories in the overview that contain the result.

Categories were drawn from the open directory project (ODP) and a database of US Government web sites (http://www.lib.lsu.edu/gov/tree), organized into three orthogonal, hierarchical sets, or *facets* (Hearst et al., 2002): Topic, Geography and US Government. The topical facet, extracted from the ODP web directory (www.dmoz.org), classified web sites according to 14 top-level categories (Table 1). The geographic facet was extracted from the ODP top-level category, Region. A database of federal government web sites was used to create the US Government facet. Web sites were categorized into the top two levels of each hierarchy. The categorized overviews were thus comprised of three 2-level facets.

When a user submits a query, the SERVICE system sends it to the Google search engine, retrieves the top 100 results, and categorizes them using fast-feature classifiers (Kules, Kustanowitz, & Shneiderman, 2006). The US government classifier uses an existing database that maps government web pages into a government hierarchy, for example mapping http://www.af.mil/ to the hierarchy node "/Executive/Executive_Agencies/Department_of_Defense/Department_of_the_Air_Force". On its own, this classifier has coverage that is limited to the list of URLs in the database. We extended coverage by using prefix matching, i.e., *any* URL beginning with www.af.mil would be mapped to this node, unless a more detailed match was found. The Topic and Geography classifier uses the ODP RDF data, applying a prefix matching technique similar to the US Government classifier. Since web sites can be cataloged in multiple categories, this yields a list of zero or more categories for each result. Although many web sites are not in either database, in initial tests an average of 66% of the top 100 search results were categorized within the ODP, and 62% of the top 100 search results were


urban sprawl - Kittery Search - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://localhost:8080/Service2/src/Kittery.jsp?q=urban+sprawl&c=100&
Go
Query
Search
Help
Adblock

Query: urban sprawl

Top 100 results containing **urban sprawl**

Stopping Sprawl Main Page - Sierra Club
The Challenge to the **Sprawl** Campaign works to fight poorly planned runaway development and promotes smart growth communities.
www.sierraclub.org/sprawl/

Sprawl Overview - Sierra Club
Sprawl destroys more than two million acres of parks, farms and open space each year. ... Hundreds of **urban**, suburban and rural neighborhoods are using ...
www.sierraclub.org/sprawl/overview/

Urban Sprawl @ nationalgeographic.com
Loads of information about **urban sprawl**, including essays, books, and current news. Looking back, this site examines the roots of the problem. ...
magma.nationalgeographic.com/ngm/data/2001/07/01/html/ft_20010701.3.html

Urban Sprawl: the Big Picture
NASA Science News: Satellites provide a "big picture" perspective on the environmental impacts of **urban sprawl**. But most city planners don
science.nasa.gov/headlines/y2002/11oct_sprawl.htm

Sprawl City
Covers consumption growth and population growth and their roles in **urban sprawl**.
www.sprawlcity.org/

Urban sprawl - Wikipedia, the free encyclopedia
Publicly-editable encyclopedia entry supplies examples and arguments for and against **sprawl**.
en.wikipedia.org/wiki/Urban_sprawl

Urban Sprawl
Links to articles about **urban sprawl**.
www.goldcanyon.com/us/urbansprawl.html

Search - Arizona Hiking Trails
Arizona Hiking Trails. ... Arizona Hiking Trails. Arizona Hiking Trails - Welcome Provides descriptions, reviews, and ratings of Arizona's hiking trails. ...
www.goldcanyon.com/

Indicators of Urban Sprawl
Indicators of **Urban Sprawl**. Prepared by Oregon's Department of Land Conservation and Development May 1992. * From 1970 to 1990, the density of **urban** ...
darkwing.uoregon.edu/~pppm/landuse/sprawl.html

NCPA - BA #287 - The Truth about Urban Sprawl

Transferring data from localhost...

Fig. 1. When SERVICE was configured as the baseline system (control condition), it presented search results as a typical ranked list, similar to Google. It was referred to as the Kittery system in the study.

categorized within the US Government hierarchy (Kules et al., 2006). This was acceptable for the study, although participant comments (Section 4.2) reveal some limitations of this approach.

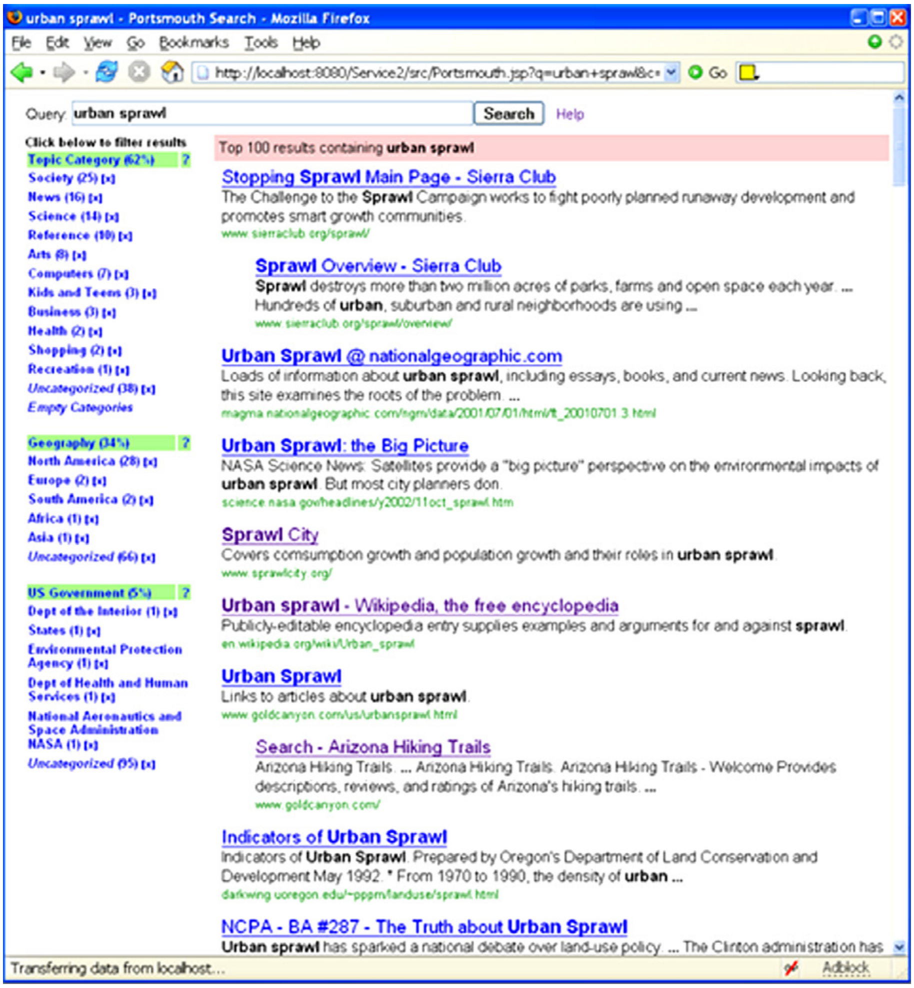
3.2. Experimental conditions

This study used a 2 × 2 within-subjects comparative design ($N = 24$), with System (baseline or categorized overview) and Topic Type (broad or narrow) as the independent variables. Each participant used both systems. System presentation order was counterbalanced.

3.3. Scenario and task design

A high-level scenario was constructed around an exploratory search task for journalists. A simulated work task (Borlund, 2003) provided a 'cover story' and an indicative task that instructed participants to conduct a short web search to generate 8–10 ideas for newspaper articles on a given topic. Journalists' information needs are often uncertain, and can change in response to external events (such as breaking news) or internal needs (e.g., increasing or decreasing the desired story length). Journalists work under tight deadlines, often with only hours between story assignment and filing. These characteristics guided design of the scenario and task, which

Please cite this article in press as: Kules, B., & Shneiderman, B. , Users can change their web search tactics: Design guidelines ..., *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.07.014


urban sprawl - Portsmouth Search - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://localhost:8080/Service2/src/Portsmouth.jsp?q=urban+sprawl&c=
Go
Search
Help
Adblock

Query: **urban sprawl**

Click below to filter results

Topic Category (62%) ?
- Society (25) [x]
- News (16) [x]
- Science (14) [x]
- Reference (10) [x]
- Arts (8) [x]
- Computers (7) [x]
- Kids and Teens (3) [x]
- Business (3) [x]
- Health (2) [x]
- Shopping (2) [x]
- Recreation (1) [x]
- *Uncategorized (38)* [x]
- *Empty Categories*

Geography (34%) ?
- North America (28) [x]
- Europe (2) [x]
- South America (2) [x]
- Africa (1) [x]
- Asia (1) [x]
- *Uncategorized (66)* [x]

US Government (5%) ?
- Dept of the Interior (1) [x]
- States (1) [x]
- Environmental Protection Agency (1) [x]
- Dept of Health and Human Services (1) [x]
- National Aeronautics and Space Administration NASA (1) [x]
- *Uncategorized (95)* [x]

Top 100 results containing **urban sprawl**

Stopping **Sprawl** Main Page - Sierra Club
The Challenge to the **Sprawl** Campaign works to fight poorly planned runaway development and promotes smart growth communities.
www.sierraclub.org/sprawl/

Sprawl Overview - Sierra Club
Sprawl destroys more than two million acres of parks, farms and open space each year. ... Hundreds of **urban**, suburban and rural neighborhoods are using ...
www.sierraclub.org/sprawl/overview/

Urban Sprawl @ nationalgeographic.com
Loads of information about **urban sprawl**, including essays, books, and current news. Looking back, this site examines the roots of the problem. ...
magma.nationalgeographic.com/ngm/data/2001/07/01/html/ft_20010701.3.html

Urban Sprawl: the Big Picture
NASA Science News: Satellites provide a "big picture" perspective on the environmental impacts of **urban sprawl**. But most city planners don.
science.nasa.gov/headlines/y2002/11oct_sprawl.htm

Sprawl City
Covers consumption growth and population growth and their roles in **urban sprawl**.
www.sprawlcity.org/

Urban sprawl - Wikipedia, the free encyclopedia
Publicly-editable encyclopedia entry supplies examples and arguments for and against **sprawl**.
en.wikipedia.org/wiki/Urban_sprawl

Urban Sprawl
Links to articles about **urban sprawl**.
www.goldcanyon.com/us/urbansprawl.html

Search - Arizona Hiking Trails
Arizona Hiking Trails. ... Arizona Hiking Trails. Arizona Hiking Trails - Welcome Provides descriptions, reviews, and ratings of Arizona's hiking trails. ...
www.goldcanyon.com/

Indicators of **Urban Sprawl**
Indicators of **Urban Sprawl**. Prepared by Oregon's Department of Land Conservation and Development May 1992. * From 1970 to 1990, the density of **urban** ...
darkwing.uoregon.edu/~pppm/landuse/sprawl.html

NCPA - BA #287 - The Truth about **Urban Sprawl**
Urban sprawl has sparked a national debate over land-use policy. ... The Clinton administration has

Transferring data from localhost...

Fig. 2. When SERVICE was configured as the experimental condition it coupled the ranked result list with a categorized overview based on topical, geographical and US government classifications. This was referred to as the Portsmouth system in the study.

Table 1
Fourteen top-level categories were extracted from the ODP for the Topic facet

Arts	Business	Computers
Games	Health	Home
Kids and Teens	News	Recreation
Reference	Science	Shopping
Society	Sports	

were reviewed by a journalism professor to ensure that they were appropriate for the journalism students we would recruit as study participants. They were also verified as part of the exit interview. The complete scenario and task are included in the Appendix.

The four topics used for the study were:

- Workplace allergies (WA)
- The aging workforce (AW)
- Human smuggling (HS)
- International art crime (IAC)

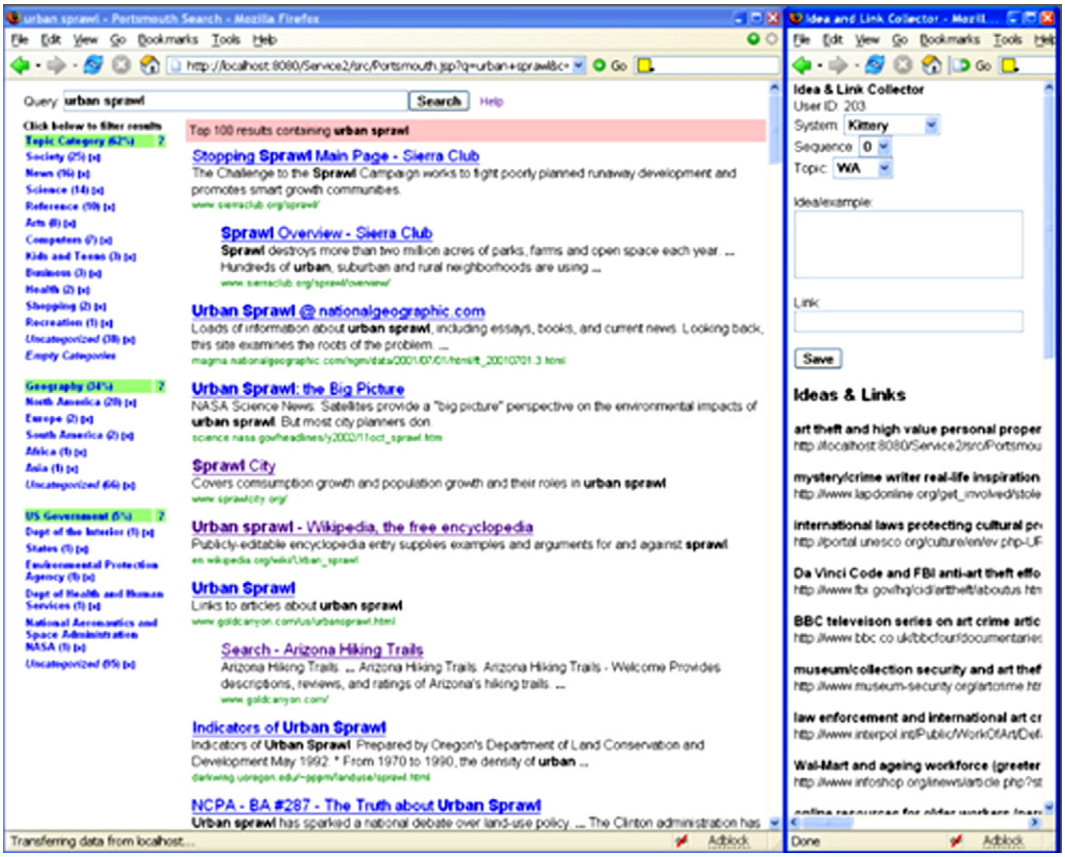
3.4. Participants

Twenty-four experienced web searchers (5 male, 19 female, primarily journalism students) were recruited and paid $30 for their participation. They ranged in age from 18 to 27 years, with a median age of 20. Twenty-one were undergraduate students, one was a graduate student and two had graduate degrees. All reported at least three years of search experience, and all but two reported searching at least once per day. All used the Google search engine.

3.5. Materials

The search interfaces were assigned neutral names (Kittery for the baseline and Portsmouth for the experimental) and displayed alongside a small web application, the Collector form (Fig. 3). The Collector form provided fields to capture the ideas to be generated as part of the indicative task for later assessment by the researcher, as well as the relevant URLs. It listed them in reverse chronological order so participants could refer to them during the session. The screen resolution was 1280 × 1024 pixels. Before each search, the search window was set to 1024 pixels wide and the collector window to 256 pixels wide.

A written script provided participants with background information on the study, to describe the scenario and task and to introduce the training task. Three short (1–3 min) training videos introduced participants to the two interfaces and the Collector form. An entry questionnaire collected participants' demographic and search experience data. A pre-search questionnaire captured knowledge of each topic prior to the search. A post-search questionnaire repeated the pre-search questions and collected reactions to the topic, interface and search process. The exit interview questions were read to the participants from a paper form.


urban sprawl - Portsmouth Search - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://localhost:8080/Service2/src/Portsmouth.jsp?q=urban+sprawl&c=
Go
Query
Search
Help
Transferring data from localhost...
Adblock
Idea and Link Collector - Mozil...
File Edit View Go Bookmarks Tools Help
Go
Save
Done
Adblock

Query: urban sprawl

Click below to filter results

Topic Category (62%) ?
Society (25) [+]
News (16) [+]
Science (14) [+]
Reference (10) [+]
Arts (8) [+]
Computers (7) [+]
Kids and Teens (3) [+]
Business (3) [+]
Health (2) [+]
Shopping (2) [+]
Recreation (1) [+]
Uncategorized (38) [+]
Empty Categories

Geography (34%) ?
North America (28) [+]
Europe (2) [+]
South America (2) [+]
Africa (1) [+]
Asia (1) [+]
Uncategorized (66) [+]

US Government (5%) ?
Dept of the Interior (1) [+]
States (1) [+]
Environmental Protection Agency (1) [+]
Dept of Health and Human Services (1) [+]
National Aeronautics and Space Administration NASA (1) [+]
Uncategorized (95) [+]

Top 100 results containing **urban sprawl**

Stopping Sprawl Main Page - Sierra Club
The Challenge to the **Sprawl** Campaign works to fight poorly planned runaway development and promotes smart growth communities.
www.sierraclub.org/sprawl/

Sprawl Overview - Sierra Club
Sprawl destroys more than two million acres of parks, farms and open space each year. ... Hundreds of **urban**, suburban and rural neighborhoods are using ...
www.sierraclub.org/sprawl/overview/

Urban Sprawl @ nationalgeographic.com
Loads of information about **urban sprawl**, including essays, books, and current news. Looking back, this site examines the roots of the problem. ...
magma.nationalgeographic.com/ngm/data/2001/07/01/html/ft_20010701.3.html

Urban Sprawl: the Big Picture
NASA Science News: Satellites provide a "big picture" perspective on the environmental impacts of **urban sprawl**. But most city planners don.
science.nasa.gov/headlines/y2002/11oct_sprawl.htm

Sprawl City
Covers consumption growth and population growth and their roles in **urban sprawl**.
www.sprawlcity.org/

Urban sprawl - Wikipedia, the free encyclopedia
Publicly-editable encyclopedia entry supplies examples and arguments for and against **sprawl**.
en.wikipedia.org/wiki/Urban_sprawl

Urban Sprawl
Links to articles about **urban sprawl**.
www.goldcanyon.com/us/urbansprawl.html

Search - Arizona Hiking Trails
Arizona Hiking Trails. ... Arizona Hiking Trails. Arizona Hiking Trails - Welcome Provides descriptions, reviews, and ratings of Arizona's hiking trails. ...
www.goldcanyon.com/

Indicators of Urban Sprawl
Indicators of **Urban Sprawl**. Prepared by Oregon's Department of Land Conservation and Development May 1992. * From 1970 to 1990, the density of **urban** ...
darkwing.uoregon.edu/~pppm/landuse/sprawl.html

NCPA - BA #287 - The Truth about Urban Sprawl
Urban sprawl has sparked a national debate over land-use policy. ... The Clinton administration has

Idea & Link Collector
User ID: 203
System: Kittery
Sequence: 0
Topic: WA

Idea/example:

Link:

Ideas & Links

art theft and high value personal proper
http://localhost:8080/Service2/src/Portsmou

mystery/crime writer real-life inspiration
http://www.lapdonline.org/get_involved/stole

international laws protecting cultural pr
http://portal.unesco.org/culture/en/ev.php-UF

Da Vinci Code and FBI anti-art theft effo
http://www.fbi.gov/hq/cid/arttheft/aboutus.htm

BBC televeison series on art crime artic
http://www.bbc.co.uk/bbcfour/documentaries

museum/collection security and art thef
http://www.museum-security.org/artcrime.htr

law enforcement and international art cr
http://www.interpol.int/Public/WorkOfArt/Def.

Wal-Mart and ageing workforce (greeter
http://www.infoshop.org/news/article.php?st

online resources for older workers (near

Fig. 3. The interface used by participants was comprised of the system under test (left) and the Collector form (right).

Participants used a laptop with a 15 inch, 1280 × 1024 pixel display, an external keyboard and mouse, and a desktop microphone. Camtasia Studio 3 was used to capture screen video and audio. The SERVICE search system was configured to log all pages visited, plus detailed data on category and result list clicks, mouse movements, and scrolling.

3.6. Procedure

Sessions were individually conducted in an office on the university campus. After participants signed the informed consent form, they completed the online entry questionnaire and viewed the training video appropriate to the first interface condition. Following the video, the scenario and task were described, and they practiced with the topic "urban sprawl." A training checklist ensured that they used the basic system features on their own or with prompting. During both the training and measured tasks, they were encouraged to use a think-aloud protocol (Ericsson & Simon, 1984).

They were then presented with the first topic. They completed the online pre-search questionnaire, performed the timed search and completed the post-search questionnaire. This was repeated for the second topic. After a short break, they were shown the video for the second interface and given practice time to become comfortable with it. The remaining two searches were then completed. The session concluded with a semi-structured exit interview. All sessions, including training, searches and interviews were recorded and participants were instructed to think out loud while they searched. This provided a total of about 100 minutes of audio and video per session.

Materials and procedures were pilot tested with 12 participants. Based on the pilot tests, the practice time was extended to permit participants to work until they felt comfortable with both the systems and the task. The scenario and task descriptions were edited to clarify the task in response to questions from pilot testers. The final pilot tests confirmed that the session duration was about 2 h and 15 min, including about 30 min for the semi-structured exit interview.

3.7. Analysis methodology

The quantitative data sets were analyzed using the null hypothesis that there was no difference between the groups. A p-value of 0.05 was used to reject that hypothesis. For ANOVA analyses, when the raw data did not follow a normal distribution, it was transformed using a logarithmic transform (Jaccard, 1983). For all significant ANOVA results, the normal Quantile–Quantile (Q–Q) Plots were examined to confirm that the residuals were distributed normally.

A limited qualitative analysis was conducted on responses to three open-ended questions from the interview:

1. Did the categorized overview change the way you searched? Can you describe an example?
2. Can you describe an example where the categorized overview [helped; OR hindered, frustrated or mislead – whichever not indicated in previous question]?
3. Did you notice any difference in how you used the categorized overview each time? Can you describe an example?

These questions required introspection and reflection. Introspection and reflection can allow the investigator to gain access to thoughts that are "mediated by knowledge structures or artifacts that we design and use" (Nielsen, Clemmensen, & Yssing, 2002). Categorized overviews are designed expressly to expose specific knowledge structures, thus this form of analysis is appropriate. To minimize known problems with verbal reports (Ericsson & Simon, 1984, pp. 19–30), the questions were constructed to elicit specific examples and concrete details.

Responses for each question were transcribed into an Access database and an inductive approach was used to develop and assign an initial code list. Particular attention was paid to how participants articulated their thoughts about search tactics, actions, and outcomes. Each response was reviewed by one researcher, who assigned a short label to sets of related comments. After 12 responses were coded, the codes were reviewed.

Obvious duplicates were merged before coding the remaining responses. A second full pass was conducted to review the initial assignments and assign a small number of new codes. The codes were divided into five groups to organize the subsequent analysis.

This analysis represents a principled approach to answering the research questions, drawing on the naturalistic inquiry paradigm (Guba & Lincoln, 1982). It complements the quantitative analysis, which seeks to identify commonalities across search experiences, by illuminating differences in search experiences. The use of a single researcher is an acknowledged limitation of this study; however, the analysis and results were peer-reviewed.

4. Results

These 24 sophisticated users coping with challenging search tasks over a two-hour period produced a wealth of data. The quantitative data show some differences in behavior and strong preferences. They do not show objective differences in outcomes. The qualitative data include thoughtful comments indicating strengths and weaknesses of the categorized overviews.

4.1. Summary of quantitative results

4.1.1. Original location of viewed (clicked on) pages in search result list

Searchers viewed (clicked on) a total of 924 pages from the search results. The results of a 2 (system) × 4 (topic) factorial analysis indicated a significant difference by system $F(1,919) = 8.96$, $p < 0.01$ and by topic $F(3,919) = 5.73$, $p < 0.01$. Searchers viewed pages at a mean (median) depth of 28.4 (18) when using the categorized overview, whereas they viewed pages at a mean depth of 22.3 (12) with the baseline. The plot in Fig. 4 shows modest but noticeable differences in the distribution of viewed pages. Although the distribution is similar for the two systems, searchers viewed results from a broader portion of the result list with the categorized overview.

4.1.2. Proportion of pages collected from categorized facets

Not all pages were categorized in the available facets, and we were interested in whether searchers were more likely to collect categorized pages when using the categorized overview (see Table 2). Searchers collected a total of 679 pages. The proportion of categorized pages differed significantly by System, χ^2 (1,

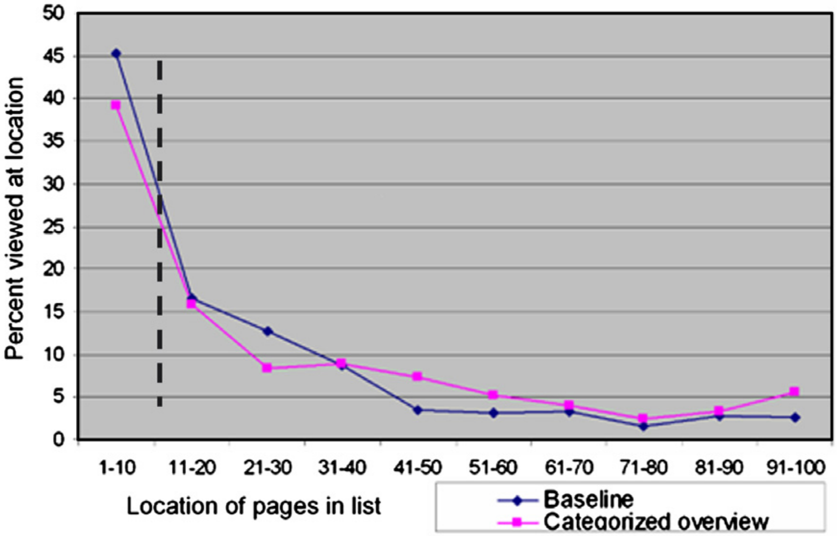


Fig. 4. Percent of pages viewed by original location of page within search results, for each system. The interface displayed approximately 10 results per screen. The dashed line shows the initial screen break.

Table 2
There was a significant difference between systems in the percentage of collected pages that had been categorized, i.e. appeared in at least one category

System	Percent categorized (%)
Baseline	75.4
Categorized overview	82.7

$N = 679) = 5.11$, $p < .05$, and Topic, χ^2 (1, $N = 679) = 18.00$, $p < .001$. The difference for the System factor (7.5 percentage points) has implications that are discussed in Section 5.1.

4.1.3. Number of queries issued during searches

Searchers conducted a total of 96 searches. All subjects except one issued at most 10 queries. One subject issued 15 queries during a search, and that outlier is removed from the following analysis. The results of a 2 (system) × 4 (topic) factorial analysis indicated a significant difference by system $F(1,87) = 7.15$, $p < 0.01$ and by topic $F(3,87) = 3.63$, $p < 0.05$. The mean (median) number of queries per search was 3.0 (2) for the categorized overview system and 3.5 (3) for the baseline.

4.1.4. Perceived organization of search results

Subjects were asked to rate agreement with the statement, "The system helped me organize my search results," (1 = strongly disagree, 9 = strongly agree). The results of a 2 (system) × 4 (topic) factorial analysis indicated a significant difference by system $F(1,88) = 42.11$, $p < 0.001$ and no significant difference by topic. The mean agreement for the categorized overview system was 7.4, and the mean agreement for the baseline system was 4.9. The corresponding medians were 7 and 5.

4.1.5. Agreement that system helped assess results and decide what to do next

Subjects were asked to rate agreement with the statement, "The system helped me assess the results of my queries to decide what to do next," (1 = strongly disagree, 9 = strongly agree). The results of a 2 (system) × 4 (topic) factorial analysis indicated a significant difference by system $F(1,88) = 13.63$, $p < 0.001$ and no significant difference by topic. The mean agreement for the categorized overview system was 6.5, and the mean agreement for the baseline system was 5.3. The corresponding medians were 7 and 5.

4.1.6. Adjectives to describe system

Subjects were asked to rate eight aspects of the systems (see Fig. 5). The 2 (system) × 4 (topic) factorial analysis for each of the eight system adjectives (semantic differentials) identified three measures that showed significant differences by system: terrible/wonderful $F(1,88) = 7.05$, $p < 0.01$; dull/stimulating $F(1,88) = 13.73$, $p < 0.001$; and disorganized/organized $F(1,88) = 45.7$, $p < 0.001$. The analysis indicated marginally significant differences by system for the frustrating/satisfying measure $F(1,88) = 3.03$, $p < 0.10$. No significant differences by topic were identified.

4.1.7. Idea quality

The quality of the generated ideas was assessed blind by a single researcher. High quality ideas would pose a question or paradox, contain conflict and human interest elements, indicate the context of the idea, and reflect intangible elements such as "coolness." Other factors included timeliness and potential impact. Two passes were made through the ideas for each topic to gain familiarity before assigning a final quality rating.

Searchers generated a total of 679 ideas. Idea quality was generally low, perhaps in part because of the time limit, which several participants commented on. Although a nine-point scale was used (1 = poor, 9 = excellent), the highest rating assigned was 5. A Wilcoxon rank sum test did not detect a significant difference in Idea Quality by System. A Kruskal–Wallis test detected a marginally significant difference by Topic, $p < 0.10$.

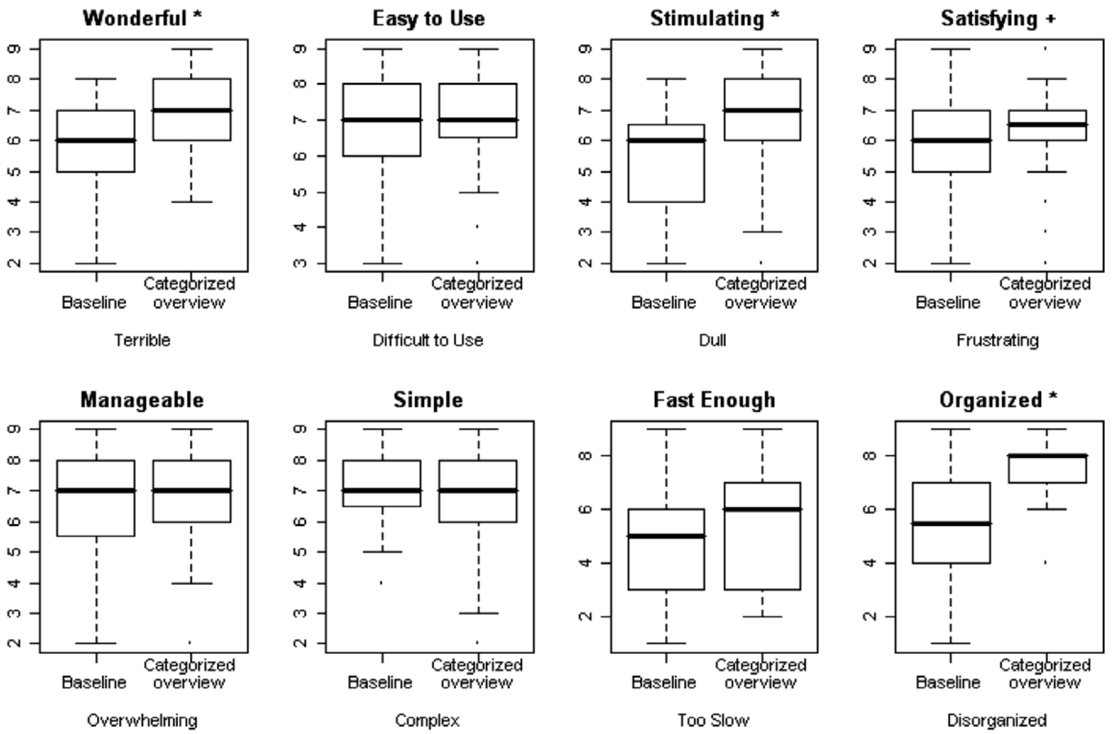


Fig. 5. Adjectives by System. Asterisk (*) indicates statistically significant differences ($p < 0.05$). Plus sign (+) indicates marginally significant differences ($p < 0.10$).

4.2. Qualitative findings

The relatively long (2 h) study time enabled participants to consider their tactics and produced insights into the search processes of sophisticated searchers coping with challenging tasks. This section is organized by the derived code groups: behavioral differences, cognitive and affective impacts, judgments of outcome, and facet and category usage.

4.2.1. Behavioral impacts

Table 3 provides an overview of the behaviors that participants commented upon.

Participants indicated that they used the overviews to filter, narrow, refine and explore their results. One participant was particularly effusive about the ease of narrowing her results, appreciating the immediacy of the interaction.

> I loved it. I was in love with that. I wish Google had that…With 3 clicks you have 5 pieces of information.

Two participants felt that they used fewer queries, which is consistent with the quantitative findings, and five felt that their queries were more general when they used the categorized overview. Two people commented that they used the overview when they were stuck. Most participants commented positively on the apparent reduction in work, but two expressed reservations about the change in their tactics.

Participants commented on interesting effects that the categorized overviews had on tactics (Table 4), including using it before looking at the result list, using it in an ancillary or backup role (e.g., when they felt stuck), and using it to understand the distribution of results across categories or to confirm interest in a particular result in the list. Participants were observed reading the subcategory pop-up windows, which provided a form of query preview (Tanin, Plaisant, & Shneiderman, 2000), before clicking on that category or moving the pointer to a different category. Twelve participants said (later in the interview) they used the pop-up sub-

Please cite this article in press as: Kules, B., & Shneiderman, B. , Users can change their web search tactics: Design guidelines ..., *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.07.014

Table 3
The six behavioral codes that participants commented upon

Description	Example	Count	Percent
Overview helped to filter or narrow list	"When I searched for work place allergies, which I am not that familiar with, the categories helped me narrow down the queries."	7	29.2
Issued more general queries	"I knew that if I did a broader word it could be divided by the categories; I did not necessarily have to be so specific."	5	20.8
Issued fewer queries	"Maybe it made me a little bit lazy. But I felt like I had to do less because it would do more... it did not take as much from me because they were gonna sort through them and organize them for me... I guess I changed by doing less."	2	8.3
Ping-ponged – alternated between using the overview and the list	"And then maybe it's just out of habit, but I just I looked at, first, at the different sites that came up. I did not even look at the categories over there. And then after I like looked at, like, the first page of this and clicked on the ones that looked the most helpful, then I thought to go to the categories."	2	8.3
Explore – used the overview to explore the results	"Yeah, rather than narrow down my search by adding additional search words, I found myself narrowing my search by exploring categories and subcategories."	1	4.2
Used the overview to refine search	"When I wanted to refine that search it came into play and it was helpful, so I guess it was kind of mixed for me."	1	4.2

The count is the number of participants who made each type of comment.

Table 4
Observation and participant comments suggested that categorized overviews enabled seven tactics

Tactic	Description	Benefit
Broad queries	Type broader queries in the search box, with few terms, then narrow results using the categorized overview	Reduced cognitive effort to generate the query
Organize examination by overview	Use the categorized overview to determine the order in which result subsets are examined	Helps monitor search to keep it on track and efficient
Overview as backup	Examine the top portion of the list first. If not satisfied, examine the overview to identify subsets to examine	May help when relevant documents are not at top of list
Preview before narrowing	Examine the subcategory information before narrowing results to that category	Avoids low relevance results. Improves confidence in expected results of action
Assess result set	Scan categorized overview to determine what categories are represented and how results are distributed across categories	Helps provide an overall understanding of the results of the query. May help assess the overall quality of the results and by implication the query
Probe using categorized overview	Select specific categories and examine the results to assess subsets of the results	Reduces effort compared to typing multiple queries
Ignore	Ignore the categorized overview	Avoids or simplifies decisions about actions to take

category list to assess what they would find in specific categories or decide whether to explore them. The categorized overview appeared to provide cues, similar to "information scent" (Pirolli & Card, 1995), that induced participants to explore categories instead of issuing a new query.

Participants spoke of the difficulty of changing established search tactics. During the session, we observed some searchers change their tactics rapidly, whereas others only started to change. During their first categorized overview search they often appeared to be exploring the interface and probing categories; comments like "let's see what this is" were frequent. By their second categorized overview search, all but two participants took advantage of the overview. Those two participants did not appear to change their tactics. They thought of specific ideas and searched for related pages instead of using the overview to guide their idea generation.

In one case, a participant avoided a page because it did not fall in the category he expected:

> I was shocked at the category that it was under, and I did not pursue it but, and I can't remember the specific... seemed like it was very strange that it would be under that category... I'm not going to that site. [laughs] I just kept moving, which is probably not the best thing to do because it might be worth investigating but that's what I did.

4.2.2. Cognitive and affective impacts

Table 5 provides an overview of the cognitive and affective impacts that participants commented upon.

The placement of pages within categories generated comments from 19 participants. This is not surprising because the interview questions specifically asked about problems. Eight participants commented on pages that did not belong within a category at all, judging them as incorrectly categorized. Eight others indicated that they found unexpected pages in a category. This occurred even though the instructions emphasized that it was typically the web sites that were categorized, not specific web pages. The prevalence of these concerns suggests that searchers may not understand the nature of the relationship entailed by category membership, as this comment suggests:

> In the human smuggling one, because that one has a lot to do with geography but I noticed that in the geography sections you'd click on Europe but it would not be about Europe, it'd be like... like I said, companies based in Europe talking about human smuggling anywhere, you know? It was not always exactly what you'd think it would be... yeah, it could be a BBC story talking about something in Asia but it still categorized as Europe... It would be hard to fix that... I don't think it was a big problem, you just have to know that something could sort of have a double meaning like a geographic location.

Seven participants commented on the structure or organization of a facet as being confusing or non-intuitive.

Table 5
The 11 cognitive and affective codes that participants commented upon

Description	Example	Count	Percent
Categorization problems	"Why did they put News and Media under Computers? Publications under Shopping?"	19	79.2
Helped with task	"For the art crimes one, when I clicked on, I saw science and it was just, "What does that have to do with art crimes?" So that made me click on it and I found out that science can help solve art crimes. So that was something that I probably would not have picked up on if that subcategory had not been there."	6	25.0
Thought differently	"I think it kind of opened up my mind a little bit to investigate a little bit deeper."	5	20.8
Complex/overwhelming	"It was not helpful when you have, like, this broad topic, when you don't know really where to begin and you're just, like, overwhelmed and don't know where to start, so, like, having that extra column acted as a deterrent at times."	5	20.8
Less complex/ overwhelming	"I thought that was a lot easier than just having, like, a 100 laid out before you. I found it a lot less overwhelming, a lot less confusing."	5	20.8
Become more comfortable over time	"The more I used it the more comfortable I was with checking to see what was in the subcategories and also with remembering to go back out to the broader categories."	5	20.8
Frustration	"I tried to use it more on the first one and then got frustrated on the second one."	2	8.3
Helped when subject had idea in mind	"[Portsmouth] gave me, like, if I thought of an idea I could look for it in the subcategories or I could use the subcategories to give me an idea and it narrowed it down a lot better."	2	8.3
Misleading/distracting	"It might be a slight distraction."	2	8.3
Concerned they might miss something	"I don't know what the person was thinking when they categorized it. They might have been thinking about something that never occurred to me but that is perfectly relevant, so that might have hidden some information from me."	2	8.3
Cautious	"Well, then I was a little more cautious, especially since I knew that humans were the ones that were putting these sites into these categories, a little more careful."	1	4.2

The count is the number of participants who made each type of comment.

> Personal Finance under Home I guess that makes sense but it's not something I would go to intuitively. I might have gone to... Business if I was looking at finance, but Business is more like the corporate world and Home would be your personal world, so after viewing it I can see the logic but it would not have been there for me initially.

Two people found the topical categories too general and one person found them ambiguous.

Four people commented that the categories helped generate ideas. For three participants, the categories exposed them to different aspects of the topic:

> Without the categories I just saw a list and I just had this mentality that I did not want to go ahead and search through all of them but the categories made me think of different possibilities so I was more opted [sic] to search through a variety of different pages versus just looking for specific factors.

One person used the overview to get an overall sense of how results were distributed within or across top-level categories:

> It also changed how I originally took in the results. Rather than reading the titles and descriptions, I looked to see how they were divided up, what main categories there where, because I thought it would be faster way to see what I had in front of me, especially for this particular task where I'm looking for different angles within a larger topic. I wanted to see, "well, there's a social issue and a health issue and a business issue," so that lends itself very well to that.

One person was concerned about possibly missing useful pages in unexplored categories.

4.2.3. Judgments of outcomes

Table 6 provides an overview of participants judgments of their search outcomes.

Participant comments included judgments on the impact of the categorized overview on their searches. During their responses to the questions, ten participants indicated that the categorized overview was helpful. Three felt it was unhelpful and one commented that it was mixed overall. Eight participants commented that the problems they encountered were minor or did not hinder their search.

The qualitative data suggest that ideas were provoked by the categorized overviews. Six participants felt that they would not have generated specific ideas without the overviews. The data also suggest a possible negative outcome on the quality of ideas. One participant indicated concern that idea quality was negatively affected, indirectly, by changes in his search tactics due to the overview. His thoughtful comment indicated that he felt that he was not getting as many good results because he relied on exploring the categories instead of analyzing the results to identify new concepts and terms to refine his query. Training could help users decide when to use the categorized overview and when to submit a new query.

Most participants appreciated and used the overview, but there were no observed differences in the quality of the story ideas they generated with the overviews. This could indicate that the task was less dependent on gaining an overview than originally anticipated. When the categorized overview was not available, scanning the result lists and reformulating queries were reasonably effective tactics for generating article ideas. Participant comments suggest that the challenging nature of the experimental task, the tight time limit and the topic difficulty all contributed to the difficulty in making progress toward their goal and the generally low quality of ideas.

4.2.4. Facet and category usage

Table 7 provides an overview of participant comments on facet and category use.

Twenty-one participants commented on aspects of their use of the Topic Category facet. Seven commented on using the Geographic facets and four commented on using the US Government facet. Participants found that these facets helped narrow results and focus their search in ways that the Topic facet did not, as this participant commented:

Table 6
The nine judgment codes that participants commented upon

Description	Example	Count	Percent
Problems experienced (of any type) were not a hindrance	"Did it hinder searching at all? I would say generally no because I would go to the results here [indicates the list] first and then use this [indicates overview] as sort of a backup to reorder or filter again sort of thing. So it's a helpful tool."	4	16.7
Problems experienced (of any type) were a minor hindrance	"With it [the categorized overview] there I could either use it or not use it, so it's not really a problem that it's there. But one time I clicked on, I think, even though certain things are categorized under certain topics, things under US government might just mention US government. It might not be an actual government page."	4	16.7
Saw something that would not have been seen otherwise	"For the art crimes one, when I clicked on, I saw science and it was just, 'What does that have to do with art crimes?' So that made me click on it and I found out that science can help solve art crimes. So that was something that I probably would not have picked up on if that subcategory had not been there."	4	16.7
Search went faster	"Yeah, helped me organize the results better, and can make your search more efficient, save some more time."	3	12.5
Search went slower	"I found it more difficult than it was worth. It would take more time than to search through it myself."	1	4.2
Got more results from a new query	"I got a lot more when I actually did a separate search than when I just clicked on US Government and expected more stuff to be there, but it was good, because I still got a lot of good search results."	1	4.2
Search was more efficient	"Yeah, helped me organize the results better, and can make your search more efficient, save some more time."	1	4.2
Found poorer quality information	"I did not use as many queries, which is part of the reason why I did not get as good information."	1	4.2
Got side-tracked	"It led me down paths I did not need to go down because of the links on the side."	1	4.2

The count is the number of participants who made each type of comment.

Table 7
Specific mentions of Topic Category, Geography or US Government facet use

Description	Example	Count	Percent
Used "Topic Category" facet	"It was very helpful because you had the different categories and subcategories, which could help you define a topic better, like with the personal finance under home with the aging population search."	21	87.5
Used "Geography" facet	"I was getting a lot of stuff about the US, so I clicked on Europe and it gave me stuff about the UK."	7	29.2
Used US Government" facet	"The government sources are right there. Its just one click of the button and you have your government source. It's easier to cite it. You don't go looking for – like with Google – you'd go through what the US government has to say about workplace allergies. Here, it's in front of you, you know, Department of Health and Labor."	4	16.7

The count is the number of participants who commented on each facet.

> That really helps if you can narrow it down by geography, or if you're really looking for a credible source and you wish to go for government. The government sources are right there. It's just one click of the button and you have your government source. It's easier to cite it. You don't go looking for – like with Google – you'd go through what the US government has to say about workplace allergies. Here, it's in front of you, you know, Dept of Health and Labor.

4.3. Limitations

The experienced participants ($N = 24$) were primarily journalism students, so the scenario and task was appropriate for them, but they might not be representative of the needs of other exploratory searchers. A single, assigned scenario and task type was evaluated. Other exploratory search tasks may benefit more or less from the categorized overview. Post-search questionnaires were completed after each topic search, which could potentially influence search tactics for subsequent searches. The time available was substantial for each session, but not long enough for searchers to completely adapt their tactics. The time allocated to each task (12 minutes) was also short, which limited their ability to conduct more thorough searches and generate high quality ideas. A longitudinal or multi-day study could overcome this shortcoming by giving searchers time to adapt before conducting the assessed tasks.

The study was limited by several factors related to the categories. Only three facets were used: Topic, Geography, and US Government. A modest proportion of pages was categorized (40–80%). The facets and categories enabled a pragmatic assessment of categorizing web search results with human-edited databases. Incorporating automated classification techniques might improve categorization rates.

The research was conducted in a laboratory setting rather than the participants' own workplaces. Participants did, however, show an awareness of these differences, and they commented on the similarity of the research setting to their workplace. More importantly, a single researcher analyzed and interpreted the raw qualitative data, including idea quality. The study does make modest use of triangulation with the quantitative data. The interpretations were closely tied to the raw data, often using the same language that participants used. Additional studies are needed to confirm these results.

5. Discussion

5.1. Differences in search behavior

The categorized overviews changed searcher behavior. With the overviews, participants explored significantly more deeply within the result list. This is consistent with previous studies (Käki, 2005).

When using the overviews, participants collected more pages that were categorized (i.e., they collected fewer uncategorized pages). Thus the categorized overview biased participants toward pages that were found in at least one category. Whether this bias is positive or negative depends on the context of search, the number of uncategorized pages, the value of the uncategorized pages, and the impact of not viewing the uncategorized pages. This bias reinforces the concern, expressed by one participant, that searchers might overlook important information by using the categories. Searchers should understand when they are narrowing their search to categorized results, whether it is important for them to view uncategorized results, and how to do so. This suggests a need for better training and/or clearer indications to searchers that their results are being filtered.

5.2. Categorization challenges

The use of the ODP categories resulted in several challenges. When page categories did not match searcher expectations, they experienced mild frustration, confusion or doubt. Three factors may have contributed to this: First, different kinds of relationships are encoded in the hierarchy. For example, pages from the British Broadcasting Corporation (BBC) were categorized under/TopicCategory/Arts/Television, which is closer to encoding an *is-a* relationship than an *about* relationship. Thus, when a BBC web page about a human smuggling story was found under Television, it was puzzling to many participants. It did not match their expectations. Second, participants commented on the generality or ambiguity of categories, particularly the topic categories. This could be attributed, at least in part, to the limited depth of the hierarchy (two levels) in the categorized overview. The ODP-assigned categories were frequently four or more levels deep. Truncating the categories removed detailed contextual information. Finally, the category structure was sometimes problematic. For example, some participants did not initially expect to find the Television category under Arts, and they found this troubling.

There are three implications for search interface designers: First, different relationships encoded in the hierarchy (e.g. *is-a* versus *part-of*) should be separated into separate top-level facets. Second, and more generally, parent-child (or broader-narrower) relationships that are clear when encountered while browsing a thesaurus or directory of web pages, will not always be clear when used in the context of a categorized overview of search results. The hierarchy may need to be changed, suggesting a new principle ("Use separate facets for each type of category") and refinement to the initial principle, "Visualize and clarify category structure." Practitioners should analyze at least the top two levels of a hierarchy, considering whether they need to be adjusted to provide the clearest overview. Third, facet analysis (Soergel, 1974) could yield more nearly orthogonal facets and identify additional facets. This might yield substantial improvements in the perceived accuracy of the category assignments. A lightweight tool could allow experienced indexers or "power searchers" with expertise in specific domains to customize hierarchies by splitting, merging, promoting, or hiding categories.

5.3. Cognitive impact of categorized overviews

These results suggest that the categorized overviews were no more difficult to use than the baseline overall. There was no significant difference in the overwhelming/manageable or complex/simple measures. This does not mean that complexity effects, which two participants commented on, should be ignored. Indeed, one participant specifically asked if he could hide the overview because it was distracting. But for this task most participants found the additional category information and the ability to preview and narrow results beneficial. This reinforces the value of providing searchers additional control over their search (Greene, Marchionini, Plaisant, & Shneiderman, 2000; Koenemann & Belkin, 1996; Shneiderman, Byrd, & Croft, 1998), including whether to display or hide the categorized overview.

Participants managed problems by relying on the stability of the categories during the second categorized overview task. This could be a benefit when compared to automatically clustered or dynamically generated categories, which will differ for each set of search results. The subjective measures lend support to this interpretation. Participants agreed that the categorized overview organized the results well and helped them assess their results and decide what to do next. Participants also found the categorized overview more generally appealing ("wonderful") and stimulating. The satisfaction ratings, which favored the categorized overview, were marginally significant.

5.4. Future research using longitudinal studies

The study highlighted an important consideration for future research: It takes time for searchers to reflect on their searches and refine their tactics. Researchers should consider using a longitudinal approach to investigate how web searchers adapt tactics when rich interfaces like categorized overviews are available (Kules, 2006a). Longitudinal studies have been used to examine changes in tactics and query terms in relation to changes in searchers' information problem stage while developing a research proposal (Vakkari, 2000). In-depth, longitudinal case studies have been used to evaluate information visualization interfaces and creativity support tools (Shneiderman et al., 2006; Shneiderman & Plaisant, 2006). These techniques integrate ethnographic and quantitative methods, using participant observation, surveys, interviews, and usage logs to study users performing complex tasks with individually defined goals. They present the opportunity to observe changes as searchers become familiar with an exploratory search system and tactics mature.

6. Design guidelines for categorized overviews

This study helped refine a set of design guidelines we are developing for categorized overview interfaces. They are particularly intended to support exploratory search. Eight design guidelines were suggested or refined by this study:

1. Provide overviews of large sets of results.
2. Organize overviews around meaningful categories.
3. Clarify and visualize category structure.
4. Tightly couple category labels to result list.
5. Ensure that the full category information is available.
6. Support multiple types of categories and visual presentations.
7. Use separate facets for each type of category.
8. Arrange text for scanning/skimming.

6.1. Provide overviews of large sets of results

During exploratory search, hundreds or thousands of results are potentially relevant. An effective overview helps searchers understand the contents of the result set and make decisions on which results to examine (Hearst, 1999b). This study showed that users did not always use the overview first, but providing one gives searchers the flexibility to use the best tool for them at the moment, either the overview or the list.[1] The ideal number of results will depend on many factors, including the task domain, topic, the quality and quantity of documents, and search engine capabilities. The fact that many pages viewed were ranked in the range of 50th-100th suggests that at least 100 results can be useful. The results of this study show that this can be done without introducing overwhelming complexity.

6.2. Organize overviews around meaningful categories

Gaining an overview of search results involves a number of cognitive subtasks, including interpretation of the results within the context of the searcher's internal mental model of the knowledge domain. Meaningful categories support learning, reflection, discovery, and information retrieval (Kwasnik, 1999; Soergel, 1999). This study suggests that categorized overviews based on topic, geography, and the US government supported beneficial search tactics. Categories based on document format, language, or Domain Name Service (DNS) domain may be useful (Kules et al., 2006). Numeric attributes such as date or size can be grouped into meaningful categories. Even abstract or computed attributes such as a journal impact factor (Garfield, 2005) can form the basis of meaningful, albeit controversial or limited, categories.

This study also suggests that stable categories will allow searchers to reuse category knowledge on subsequent searches. Dynamic categories, such as those generated by automated clustering techniques, change with each query. Thus the learning benefits of stable categories may accrue less, but they may provide other benefits (Kules and Shneiderman, submitted for publication).

6.3. Clarify and visualize category structure

If categories are drawn from a classification, taxonomy, or ontology, the structure should be made visible. It provides context for individual category labels, shows relationships between concepts and allows users to focus on the portions of the concept space that are of most interest. The visual presentation must be disciplined to avoid overwhelming or disorienting searchers.

This study suggests that practitioners should review at least the top two levels of a hierarchy, considering whether they need to be adjusted to provide the clearest overview. In general, broad, shallow hierarchies are beneficial (Jacko & Salvendy, 1996; Kiger, 1984; Miller, 1981). This study showed that parent-child (or broader-narrower) relationships that are clear when encountered while browsing a thesaurus or directory of web pages are not always clear when used in the context of a categorized overview of search results. The structure of the hierarchy may need to be changed in these cases.

6.4. Tightly couple category labels to result list

Brushing and linking techniques tightly couple multiple views of data in an information visualization, so that an action in one view (brushing) is linked to an action in another view. This can be applied to search results to synchronize two views of the results, an overview and a detailed list (Klein, Reiterer, Müller, & Limbach, 2003). Judicious use of this technique can support richer interactions between category information and individual results. The SERVICE system provides two examples of tight coupling: clicking on category labels narrows or broadens the result list; and pausing the pointer over a result in the list highlights all the categories in the overview that contain the result. One benefit of tight coupling between the categories and results is that it allows searchers to very quickly see examples. Within a category, example results help to clarify the meaning

[1] We thank the anonymous reviewer for reminding us of this point.

of the categories and often provide indications of relevance, quality, etc. Even within well-known classifications, some category labels may be ambiguous or unfamiliar. Examples of individual pages can disambiguate category names (Dumais, Cutrell, & Chen, 2001).

When this capability is implemented, it is important to provide clear feedback indicating which categories are currently applied. Observations from this study suggest that participants occasionally forgot or overlooked the fact that they were viewing a subset of their original query.

6.5. Ensure that full category information is available

When using deep hierarchies, designers should ensure that full category information (the complete label or descriptor) is available to searchers. The category labels in the overview indicate which categories results are in, but this may be limited to the top few levels because of the limited display space. During this study, participants wondered aloud what specific category results were in. Participants were occasionally frustrated because only the top two category levels were visible in the overview (Section 5.2). Providing the full category label could alleviate this problem. Displaying category labels in each result can be helpful (Drori & Alon, 2003). However, when this was implemented in the SERVICE system, the individual results became too large because results often appeared in multiple categories. Therefore, it was disabled prior to the study. During development, we also experimented briefly with opening a pop-up window when the pointer moved over the result. A small hyperlink in each result may be an appropriate design compromise, although this was not implemented or evaluated. These alternatives should be investigated in future studies.

6.6. Support multiple types of categories and visual presentations

No single type of category is effective for all users, tasks, and domains. In her comparison of categories and clustering for organizing search results, Hearst (1999a) noted that neither categories nor automatically constructed clusters will always align with users' interests. Libraries provide subject, author, and title indexes and archives provide multiple finding aids for their holdings. GRiDL (Shneiderman, Feldman, Rose, & Grau, 2000), SuperTable (Klein, Müller, Reiterer, & Eibl, 2002), and the Clusty (www.clusty.com) and Exalead (www.exalead.com) search engines are examples of search result interfaces that permit users to reorganize results using alternate sets of categories. During previous studies, several participants noted that they would like to be able to select or define their own categories and re-arrange them for their own purposes (Kules and Shneiderman, submitted for publication). Likewise, no single presentation style is ideal for all situations and tasks (Risden, Czerwinski, Munzner, & Cook, 2000; Sebrechts, Vasilakis, Miller, Cugini, & Laskowski, 1999; Shneiderman & Plaisant, 2004; Swan & Allen, 1998). Exploratory searchers should be allowed to select a task-appropriate form of data display (Shneiderman, Byrd, & Croft, 1997). Alternatively, if that level of control and the corresponding increase in complexity is not appropriate for the intended users, designers should have a variety of categories and presentation styles to choose from, so they can choose appropriate categories and visual presentation styles. It may be useful to enable an experienced searcher to customize the overview and share it with others. Supporting multiple classifications and multiple visual presentations may enable users to view and explore search results from the perspectives most appropriate to their needs.

6.7. Use separate facets for each type of category

When a rich set of categories encodes multiple types of relationships, presenting them as separate visual facets can clarify meanings and relationships that might otherwise be ambiguous. For example, categories for *is-a*, *is-about*, and *part-of* relationships should be presented separately. Faceted classification organizes a domain into orthogonal sets of categories, which are ideally homogeneous, mutually exclusive, and represent a single characteristic of division (Vickery, 1960). It has been used to organize catalogs, classifications, and thesauri (Soergel, 1974; Vickery, 1960), information spaces on the Web (Louie, Maddox, & Washington, 2003), and non-web search interfaces (Yee et al., 2003). The importance of this principle was clarified during the development of the SERVICE system. During informal user tests, searchers experienced confusion when topical and geographic categories were used in the same facet. Separating geographic categories from topical categories in the final interface helped reduce this problem in this study. Other instances of categories that should have been separated out remained problematic. Therefore, hierarchies used in a categorized overview should be analyzed to determine whether they should be restructured into separate facets. The informal analysis performed during development yielded a noticeable improvement, suggesting that even a lightweight faceted analysis focused on the upper levels of a hierarchy could be beneficial.

6.8. Arrange text for scanning/skimming

At a perceptual level, users of search results attempt to rapidly ingest large amounts of text. We observed searchers scanning category labels, titles, URLs, and snippets of text to quickly select specific pages to view. They skimmed the pages and returned to the list to repeat this cycle. It could be argued that this is simply a result of the textual presentation format, but it also reflects more fundamentally that the source documents are inherently textual and are not easily presented graphically. Arranging these elements in a consistent manner (e.g. linear lists, columns, or matrices) (Teitelbaum & Granda, 1983) and ensuring that they are visible (rather than requiring interaction such as moving the pointer over an item) will support fast scanning and skimming. Aula (2004) found that presenting snippets as bulleted lists was 20% faster than the standard textual display. Appropriate use of font weights, styles, sizes, and colors will also help (Tullis, 1988).

6.9. Summary

These design guidelines for categorized overviews have been suggested or refined by the design and evaluation of the SERVICE system. They complement and extend general human-computer interaction, web design, information architecture, and information visualization principles. They will be useful for search interface designers because they provide guidance for the appropriate integration of visual overviews with search result lists, and particularly for the textual surrogates embedded in result lists. These guidelines are not exhaustive or comprehensive. Evaluations and analysis of other exploratory search interfaces will certainly suggest additional guidelines.

7. Conclusion

This study answers our research questions by revealing tactical and cognitive benefits of categorized overviews: (1) Searchers explored results more deeply; (2) they agreed that the categorized overviews helped them organize, explore and assess their results without being appreciably more complex than typical Google-like interfaces. No quantitive differences were detected in the quality of task outcomes. Our results describe benefits and limitations of categorized overviews and identify considerations for the structure of the categories. The study identified seven tactics that searchers used and revealed a bias in the choice of pages viewed: When categorized overviews were present, fewer uncategorized pages were viewed. Enthusiasm for categorized overviews was often voiced, e.g. "I loved it... I wish Google had that." Overall, the benefits of categorized overviews for many search tasks and users seem strong enough to warrant further research, refined designs, and more commercial implementations. Our results suggest that future studies may benefit from a longitudinal approach to analyze changes in tactics as searchers adapt to richer interfaces.

Finally, the study suggested or refined a set of eight design guidelines for categorized overview search interfaces. We believe these guidelines will be useful for digital library and web search designers, information architects, and web developers because they provide guidance for the appropriate integration of visual overviews with search result lists. The findings and guidelines must be tested, refined, and extended by additional studies that investigate the challenges of exploratory search.

Acknowledgements

The authors thank Young Choi, Allison Brochu, Renee Reighart and the anonymous reviewers for their thoughtful comments. This research was partially supported by an AOL Fellowship in Human-Computer

Interaction and National Science Foundation Digital Government Initiative grant (EIA 0129978) “Towards a Statistical Knowledge Network”.

Appendix

The scenario and task used for the study:

> Imagine that you are a reporter for a national newspaper. Due to some recent events, your editor has just asked you to generate a list of ideas for a series of articles on [the topic, e.g. urban sprawl]. There’s a meeting in an hour, so she does not need a lot of detail, but she wants a diverse list of 8–10 (or more) ideas for discussion. They should cover many different aspects of the topic, to appeal to a broad range of readers. Unusual or provocative ideas are good. You have about 10 minutes to conduct a short web search to find out what information is available and generate the ideas. Your results will be judged (by your imaginary editor) on the quality and diversity of ideas. For example, “public health impact” would be an okay idea and “obesity as a public health impact of urban sprawl” would be even better, because it is a bit more specific. As you use the search engine to explore and generate article ideas, enter them in the Collector form and include the web page that inspired your idea. It is important that you enter the ideas, not notes like “a good page.” Think of this list [point to the Collector] as a bullet list for the discussion.

Three interview questions were selected for coding and analysis:

1. Did the categorized overview change the way you searched? Can you describe an example?
2. Can you describe an example where the categorized overview [helped; OR hindered, frustrated or mislead – whichever not indicated in previous question]?
3. Did you notice any difference in how you used the categorized overview each time? Can you describe an example?

In question two, the object was to elicit feedback on whichever aspect (positive or negative) the participant did not mention when answering the first question.

References

Aula, A. (2004). *Enhancing the readability of search result summaries*. Paper presented at the Proceedings Volume 2 of the Conference HCI 2004: Design for Life, Leeds, UK. Retrieved May 30, 2007, from http://www.cs.uta.fi/~aula/aula_summary.pdf.

Bates, M. (1990). Where should the person stop and the information search interface start. *Information Processing and Management, 26*(5), 575–591.

Bates, M. J. (1979). Information search tactics. *Journal of the American Society for Information Science, 30*, 205–214.

Borlund, P. (2003). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3), paper no. 152.

Drori, O., & Alon, N. (2003). Using documents classification for displaying search results list. *Journal of Information Science, 29*(2), 97–106.

Dumais, S., Cutrell, E., & Chen, H. (2001). Optimizing search by showing results in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 277–284.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Fidel, R. (1985). Moves in online searching. *Online Review, 9*(1), 61–74.

Garcia, E., & Sicilia, M.-Á. (2003). User interface tactics in ontology-based information seeking. *Psychology Journal, 1*(3), 242–255.

Garfield, E. (2005). *The agony and the ecstasy-The history and meaning of the journal impact factor*. Paper presented at the International Congress on Peer Review and Biomedical Publication, Chicago, IL. Retrieved May 30, 2007, from http://garfield.library.upenn.edu/papers/jifchicago2005.pdf.

Golovchinsky, G. (1997). Queries? Links? Is there a difference? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA* (pp. 407–414). New York: ACM Press.

Greene, S., Marchionini, G., Plaisant, C., & Shneiderman, B. (2000). Previews and overviews in digital libraries: Designing surrogates to support visual information-seeking. *Journal of the American Society for Information Science, 51*(3), 380–393.

Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communication and Technology, 30*(4), 233–252.

Hearst, M. (1999a). The use of categories and clusters for organizing retrieval results. In T. Strzalkowski (Ed.), *Natural Language Information Retrieval* (pp. 333–373). Boston: Kluwer Academic Publishers.

Hearst, M. (1999b). User interfaces and visualization. In R. Baeza-Yates & B. Ribeiro-Neto (Eds.), *Modern Information Retrieval* (pp. 257–323). Reading, MA: Addison-Wesley.

Hearst, M., Elliot, A., English, J., Sinha, R., Swearingen, K., & Yee, P. (2002). Finding the flow in web site search. *Communications of the ACM, 45*(9), 42–49.

Jaccard, J. (1983). *Statistics for the Behavioral Sciences*. Belmont, CA: Wadsworth Publishing Company.

Jacko, J., & Salvendy, G. (1996). Hierarchical menu design: Breadth, depth, and task complexity. *Perceptual and Motor Skills, 82*, 1187–1201.

Janecek, P., & Pu, P. (2005). An evaluation of semantic fisheye views for opportunistic search in an annotated image collection. *Journal of Digital Libraries, 5*(1), 42–56.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management, 36*, 207–227.

Kabel, S., Hoog, R. d., Wielinga, B. J., & Anjewierden, A. (2004). The added value of task and ontology-based markup for information retrieval. *Journal of the American Society for Information Science and Technology, 55*(4), 348–362.

Käki, M. (2005). Findex: search result categories help users when document ranking fails. In *Proceeding of the SIGCHI Conference on Human Factors in Computing Systems, Portland, OR* (pp. 131–140). New York: ACM Press.

Kiger, J. (1984). The depth/breadth trade-off in the design of menu-driven user interfaces. *International Journal of Man-Machine Studies, 20*, 201–213.

Klein, P., Müller, F., Reiterer, H., & Eibl, M. (2002). Visual information retrieval with the SuperTable + Scatterplot. In *Proceedings of the Sixth International Conference on Information Visualisation (IV '02)* (pp. 70–75). New York: IEEE Computer Society.

Klein, P., Reiterer, H., Müller, F., & Limbach, T. (2003). Metadata visualisation with VisMeB. In *Proceedings of the Seventh International Conference on Information Visualization (IV '03)* (pp. 600–605). New York: IEEE Computer Society.

Koenemann, J., & Belkin, N. J. (1996). A case for interaction: A study of interactive information retrieval behavior and effectiveness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground, Vancouver, British Columbia, Canada* (pp. 205–212). New York: ACM Press.

Kules, B. (2006a). *Methods for evaluating changes in search tactics induced by exploratory search systems*. Paper presented at the ACM SIGIR 2006 Workshop on Evaluating Exploratory Search Systems. Retrieved May 30, 2007, from http://www.takomasoftware.com/techreports/KulesESSEval-20060630b.pdf.

Kules, B. (2006b). *Supporting Exploratory Web Search with Meaningful and Stable Categorized Overviews* (Unpublished doctoral dissertation): University of Maryland, College Park. Retrieved May 30, 2007, from http://hcil.cs.umd.edu/trs/2006-14/2006-14.pdf.

Kules, B., Kustanowitz, J., & Shneiderman, B. (2006). Categorizing web search results into meaningful and stable categories using Fast-Feature techniques. In *Proceedings of the Sixth ACM/IEEE-CS Joint Conference on Digital Libraries, Chapel Hill, NC* (pp. 210–219). New York: ACM Press.

Kules, B., & Shneiderman, B. (submitted for publication). Using meaningful and stable categories to support exploratory web search: Two formative studies.

Kwasnik, B. H. (1999). The role of classification in knowledge representation and discovery. *Library Trends, 48*(1), 22–47.

Louie, A. J., Maddox, E. L., & Washington, W. (2003). *Using faceted classification to provide structure for information architecture*. Paper presented at the 62nd ASIS Annual Meeting, Washington, DC. Retrieved May 30, 2007, from http://depts.washington.edu/pettt/presentations/conf_2003/IASummit.pdf.

Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge University Press.

Miller, D. (1981). The depth/breadth trade-off in hierarchical computer menus. *Proceedings of the Human Factors Society*, 296–300.

Nielsen, J., Clemmensen, T., & Yssing, C. (2002). Getting access to what goes on in people's heads? – Reflections on the think-aloud technique. In *Proceedings of the Second Nordic Conference on Human-Computer Interaction, Aarhus, Denmark* (pp. 101–110). New York: ACM Press.

Pirolli, P., & Card, S. (1995). Information foraging in information access environments. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 51–58). New York: ACM Press.

Pirolli, P., Schank, P., Hearst, M., & Diehl, C. (1996). Scatter/gather browsing communicates the topic structure of a very large text collection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common ground, Vancouver, British Columbia, Canada* (pp. 213–220). New York: ACM Press.

Risden, K., Czerwinski, M., Munzner, T., & Cook, D. (2000). An initial examination of ease of use for 2D and 3D information visualizations of Web content. *International Journal of Human-Computer Studies, 53*(5), 695–714.

Sebrechts, M., Vasilakis, J., Miller, M., Cugini, J., & Laskowski, S. (1999). Visualization of search results: A comparative evaluation of text, 2D, and 3D interfaces. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 3–10). New York: ACM Press.

Shneiderman, B., Byrd, D., & Croft, W.B. (1997). Clarifying search: A user-interface framework for text searches. *D-Lib Magazine*.

Shneiderman, B., Byrd, D., & Croft, W. B. (1998). Sorting out searching: A user-interface framework for text searches. *Communications of the ACM, 41*(4), 95–98.

Shneiderman, B., Feldman, D., Rose, A., & Grau, X. F. (2000). Visualizing digital library search results with categorical and hierarchial axes. In *Proceedings of the Fifth ACM International Conference on Digital Libraries (San Antonio, TX, June 2–7, 2000)* (pp. 57–66). New York: ACM Press.

Please cite this article in press as: Kules, B., & Shneiderman, B. , Users can change their web search tactics: Design guidelines ..., *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.07.014

Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B., Candy, L., et al. (2006). Creativity support tools: Report from a US National Science Foundation sponsored workshop. *International Journal of Human-Computer Interaction, 20*(2), 61–77.

Shneiderman, B., & Plaisant, C. (2004). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (4th ed.). Boston: Pearson/Addison-Wesley.

Shneiderman, B., & Plaisant, C. (2006). *Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies*. Paper presented at the Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV '06): A Workshop of the AVI 2006 International Working Conference. Retrieved May 30, 2007, from http://hcil.cs.umd.edu/trs/2006-12/2006-12.pdf.

Soergel, D. (1974). *Construction and Maintenance of Indexing Languages and Thesauri*. New York: Wiley.

Soergel, D. (1999). The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science and Technology, 50*(12), 1119–1120.

Swan, R., & Allen, J. (1998). Aspect Windows, 3D visualizations, and indirect comparisons of information retrieval systems. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 173–181). New York: ACM Press.

Tanin, E., Plaisant, C., & Shneiderman, B. (2000). Browsing large online data with Query Previews. In *Proceedings of the Symposium on New Paradigms in Information Visualization and Manipulation (NPIVM) 2000*. Washington, DC: ACM Press.

Teitelbaum, R. C., & Granda, R. E. (1983). The effects of positional constancy on searching menus for information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA* (pp. 150–153). New York: ACM Press.

Toms, E. G., Freund, L., Kopak, R., & Bartlett, J. C. (2003). The effect of task domain on search. *Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative Research*, 303–312.

Tullis, T. (1988). Screen design. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 377–411). Amsterdam, The Netherlands: Elsevier Science Publishers.

Vakkari, P. (2000). *eCognition and changes of search terms and tactics during task performance: A longitudinal case study*. Paper presented at the Proceedings of the RIAO 2000 Conference. Retrieved May 30, 2007, from http://www.info.uta.fi/vakkari/Vakkari_Tactics_RIAO2000.pdf.

Vickery, B. C. (1960). *Faceted Classification: A Guide to Construction and Use of Special Schemes*. London: Aslib.

White, R., Muresan, G., & Marchionini, G. (2006). Evaluating Exploratory Search Systems – SIGIR 2006 Workshop Call for Papers. Retrieved May 30, 2007, from http://www.umiacs.umd.edu/~ryen/eess.

White, R. W., Kules, B., Drucker, S. M., & Schraefel, M. C. (2006). Supporting exploratory search. *Communications of the ACM, 49*(4), 36–39.

Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology, 55*(3), 246–258.

Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems, Ft. Lauderdale, FL* (pp. 401–408). New York: ACM Press.

Please cite this article in press as: Kules, B., & Shneiderman, B. , Users can change their web search tactics: Design guidelines ..., *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.07.014