

Exploratory Subject Searching in Library Catalogs: Reclaiming the Vision

Julia Bauder and
Emma Lange

ABSTRACT

Librarians have had innovative ideas for ways to use subject and classification data to provide an improved online search experience for decades, yet after thirty-plus years of improvements in our online catalogs, users continue to struggle with narrowing down their subject searches to provide manageable lists containing only relevant results. This article reports on one attempt to rectify that situation by radically reenvisioning the library catalog interface, enabling users to interact with and explore their search results in a profoundly different way. This new interface gives users the option of viewing a graphical overview of their results, grouped by discipline and subject. Results are depicted as a two-level treemap, which gives users a visual representation of the disciplinary perspectives (as represented by the main classes of the Library of Congress Classification) and topics (as represented by elements of the Library of Congress Subject Headings) included in the results.

INTRODUCTION

Reading library literature from the early days of the OPAC era is simultaneously inspiring and depressing. The enthusiasm that some librarians felt in those days about the new possibilities that were being opened by online catalogs is infectious. Elaine Svenonius envisioned a catalog that could interactively guide users from a broad single-word search to the specific topic in which they were really interested.¹ Pauline Cochrane conceived of a catalog that could group results on similar aspects of a given subject, showing the user a “systematic outline” of what was available on the subject and allowing the user to narrow their search easily.² Marcia Bates even pondered whether “any indexing/access apparatus that does not stimulate, intrigue, and give pleasure in the hunt is defective,” since “people enjoy exploring knowledge, particularly if they can pursue mental associations in the same way they do in their minds. . . . Should that not also carry over into enjoying exploring an apparatus that reflects knowledge, that suggests paths not thought of, and that shows relationships between topics that are surprising?”³ However, looking back thirty years later, it is dispiriting to consider how many of these visions have not yet been realized.

The following article reports on one attempt to rectify that situation by radically reenvisioning the library catalog interface, enabling users to interact with and explore their search results in a profoundly different way. The idea is to give users the option of viewing a graphical overview of their results, grouped by discipline and subject. This was achieved by modifying a VuFind-based

Julia Bauder (bauderj@grinnell.edu) is Social Studies and Data Services Librarian, and **Emma Lange** (langemm@grinnell.edu) is an undergraduate student and former library intern, Grinnell College, Grinnell, Iowa.



discovery layer to allow users to choose between a traditional, list-based view of their search results and a visualized view. In the visualized view, results are depicted as a two-level treemap, which gives users a visual representation of the disciplinary perspectives (as represented by the main classes of the Library of Congress Classification [LCC]) and topics (as represented by elements of the Library of Congress Subject Headings [LCSH]) included in the results. An example of this visualized view can be seen in figure 1.

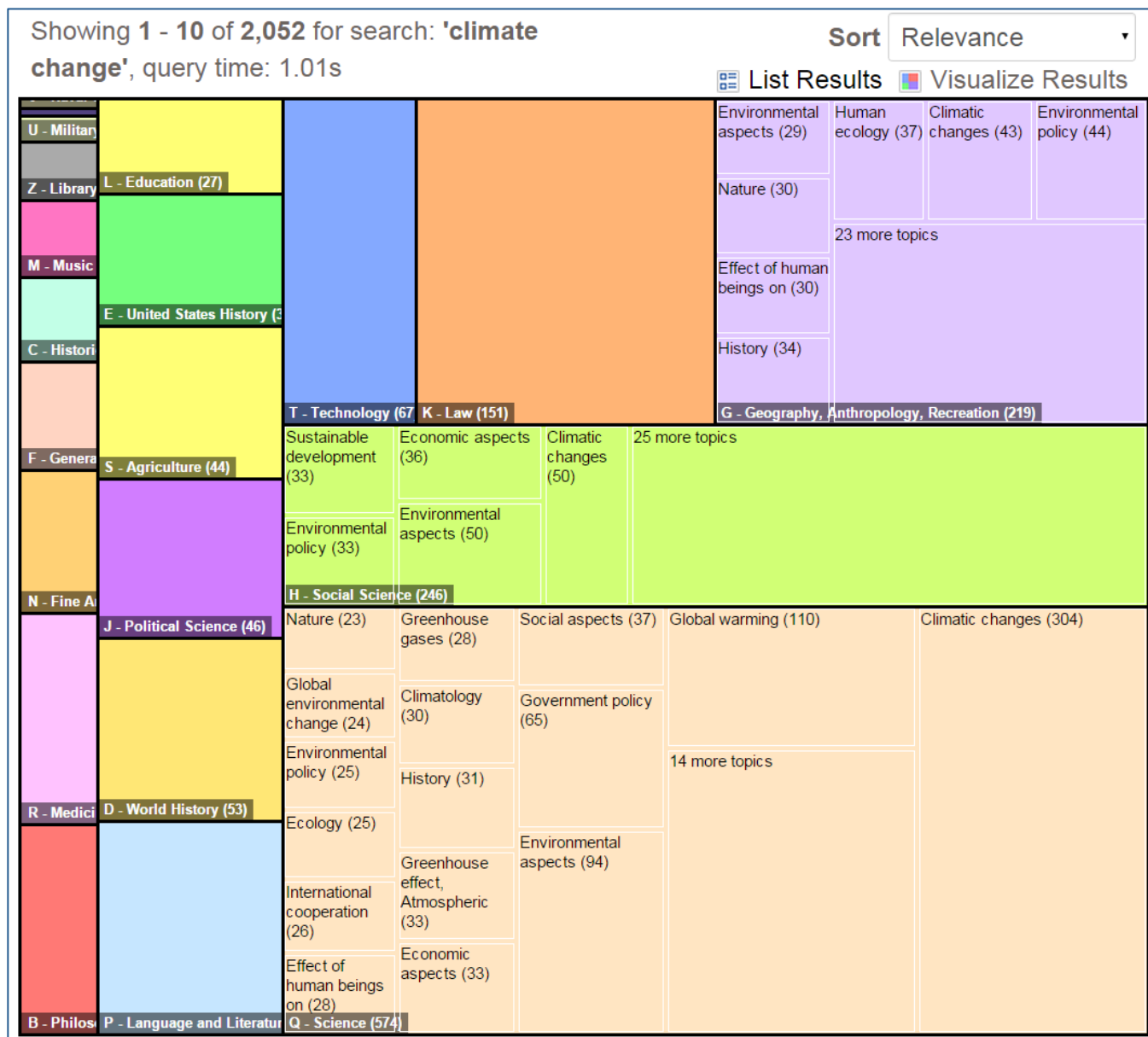


Figure 1. Visualization of the Results for a Search for “Climate Change.”

Subsequent sections of this paper summarize the library-science and computer-science literature that provides the theoretical justification this project, explain how the visualizations are created, and report on the results of usability testing of the visual interface with faculty, academic staff, and undergraduate students.

LITERATURE REVIEW

Exploratory Subject Searching in Library Catalogs

Since Charles Ammi Cutter published his *Rules for a Printed Dictionary Catalogue* in 1876, most library catalogs have been premised on the idea that users have a very good idea of what they are looking for before they begin to interact with the catalog.⁴ In this classic view, users are either conducting known-item searches—they know the titles or the author of the books they want to find—or they know the exact subject on which they are interested in finding books. Yet research has shown that known-item searches are only about half of catalog searches,⁵ and that users often have a very difficult time expressing their information needs with enough detail to construct a specific subject search. Instead, much of the time, users approach the catalog with only a vaguely formulated information need and an even vaguer sense of what words to type into the catalog to get the resources that would solve their information need.⁶

Even in the earliest days of the OPAC era, librarians were aware of this problem. Some of them, including Elaine Svenonius and Pauline Cochrane, speculated about better use of subject and classification data to try to help users who enter too-short, overly broad searches focus their results on the information that they truly want. One of Cochrane's many ideas on this topic was to use subject and classification data "to present a systematic outline of a subject," which would let users see all of the different aspects of that subject, as reflected in the library's classification system and subject headings, and the various locations where those materials could be found in the library.⁷ Svenonius suggested using library classifications to help narrow users' searches to appropriate areas of the catalog. For example, she suggests, if a user enters "freedom" as a search term, the system might be programmed to present to the user contexts in which "freedom" is used in the Dewey Decimal Classification, such as "freedom of choice" or "freedom of the press." Once the user selects a one of these phrases, Svenonius continued, the system could present the user with additional contextual information, again allow the user to specify which context is desired, and then guide the user to the exact call number range for information on the topic. She concluded, "Thus by contextualizing vague words, such as *freedom*, within perspective hierarchies, the computer might guide a user from an ineptly or imprecisely articulated search request to one that is quite specific."⁸

Ideas such as these had little impact on the design of production library catalogs until the late 1990s, when a Dutch company, MediaLab Solutions, began developing AquaBrowser, which features a word cloud composed of synonyms and other words related to the search term and allows users to refocus their search by clicking on these words.⁹ AquaBrowser became available in the United States in the mid-2000s, shortly before North Carolina State University launched its Endeca-based catalog in 2006.¹⁰

While AquaBrowser's word cloud is certainly visually striking, the feature that these and most of the subsequent "next-generation" library catalogs implement that has had the most impact on search behavior is faceting. Facets, while not as sophisticated as the systems envisioned by

Svenonius and Cochrane, are partial solutions to the problems they lay out. Facets can serve to give users a high-level overview of what is available on a topic, based on classification, format, period, or other factors. They can also help guide a user from an impossibly broad search to a more focused one. Various studies have shown that faceted interfaces are effective at helping users narrow their searches, as well as helping them discover more relevant materials than they did when performing similar tasks on nonfaceted interfaces.¹¹ However, studies have also shown that users can become overwhelmed by the number and variety of facets available and the number of options shown under each facet.¹²

Visual Interfaces to Document Corpora

When librarians were pondering how to create a better online library catalog, computer scientists were investigating the broader problem of helping users to navigate and search large databases and collections of documents effectively. Visual interfaces have been one of the methods computer scientists have investigated for providing user-friendly navigation, with perhaps the most prominent early advocate for visual interfaces being Ben Shneiderman.¹³ In recent years, Shneiderman and other researchers have built and tested various types of experimental visual interfaces for different forms of information-seeking.¹⁴ However, with a few exceptions, most of these visual interfaces have remained in a laboratory rather than a production setting.¹⁵ With the exception of the “date slider,” a common interface feature that displays a bar graph showing dates related to the search results and allows users to slide handles to include or exclude times from their search results, few current document search systems present users with any kind of visual interface.

METHOD

The Grinnell College Libraries use VuFind, open-source software originally developed at Villanova University as a discovery layer to use over a traditional ILS. VuFind in turn makes use of Apache Solr, a powerful open-source indexing and search platform, and SolrMarc, code developed within the library community that facilitates indexing MARC records into Solr. Using SolrMarc, MARC fields and subfields are mapped to various fields in the Solr index; for example, the contents of MARC field 020, subfield a, and field 773, subfield z, are both mapped to a Solr index field called “isbn.” More than fifty Solr fields are populated in our index. Our visualization system was built on top of VuFind’s Solr index and visualizes data taken directly from the index.

The visualizations are created in Javascript using the D3.js visualization library, and they are designed to implement Shneiderman’s Visual Information Seeking Mantra: “Overview first, zoom and filter, then details-on-demand.”¹⁶ The goal was to give users the option of viewing a graphical overview of their results, grouped by disciplinary perspective and topic, and then allow them to zoom in on the results from specific perspectives or on specific topics. Once they have used the interactive visualization to narrow their search, they can choose to see a traditional list of results with full bibliographic details about the items. This would, ideally, provide a version of the

systematic outline that Cochrane envisioned. It should also support users as they attempt to narrow down their search results and focus on a specific aspect of their chosen subject without overwhelming them with long lists of results or of facets.

Currently, we are visualizing values of two fields, one containing the first letter of the items' Library of Congress Classification (LCC) numbers and the other containing elements of the items' Library of Congress Subject Headings (LCSH). This data is visualized as a two-level treemap.¹⁷ First, large boxes are drawn representing the number of items matching the search within each letter of the LCC. Within the largest of these boxes, smaller boxes are drawn showing the most common elements of the subject headings for items matching that search within that LCC main class. Less common subject heading elements are combined into an additional small box, labeled "X more topics"; clicking on that box zooms in so that users only see results from one LCC main class, and it displays all of the LCSH headings applied to items in that group. Similarly, users can click on any of the smaller LCC boxes, which do not contain LCSH boxes in the original visualization, to zoom in on that LCC main class and see the LCSH subject headings for it. Both the large and the small boxes are sized to represent what proportion of the results were in that LCC main class or had that LCSH subject heading.

This is easier to explain with a concrete example. Let's say a student were to search for "climate change" and click on the option to visualize the results. You can see what this looks like in figure 1. Instead of seeing a list of nearly two thousand books, the student now sees a visual representation of the disciplinary perspectives (as represented by the main classes of the LCC) and topics (as represented by elements of the LCSH) included in the results. Users could click to zoom in on any main class within the LCC to see all of the topics covered by books in that class, as in figure 2, where the student has zoomed in on "S – Agriculture." Or users could click on any topic facet to see a traditional results list of books with that topic facet in that main class. At any zoom level, users could choose to return to the traditional results list by clicking on the "List Results" option.¹⁸

We launched this feature in our catalog midway through the spring 2014 semester. Formal usability testing was completed with five advanced undergraduates, three staff, and two faculty members in the summer of 2014. (See appendix A for the outline of the usability test.) One first-year student completed usability testing in the fall 2014 semester. The usability study asked participants to complete a set list of nine specific, predetermined tasks. Some tasks involved the use of now-standard catalog features, such as saving results to a list and emailing results to oneself, while about half of the tasks involved navigation of the visualization tool, which was entirely new to the participants. Each participant received the same tasks and testing experience regardless of their status as a student, faculty, or staff, and each academic division was represented among the participants.

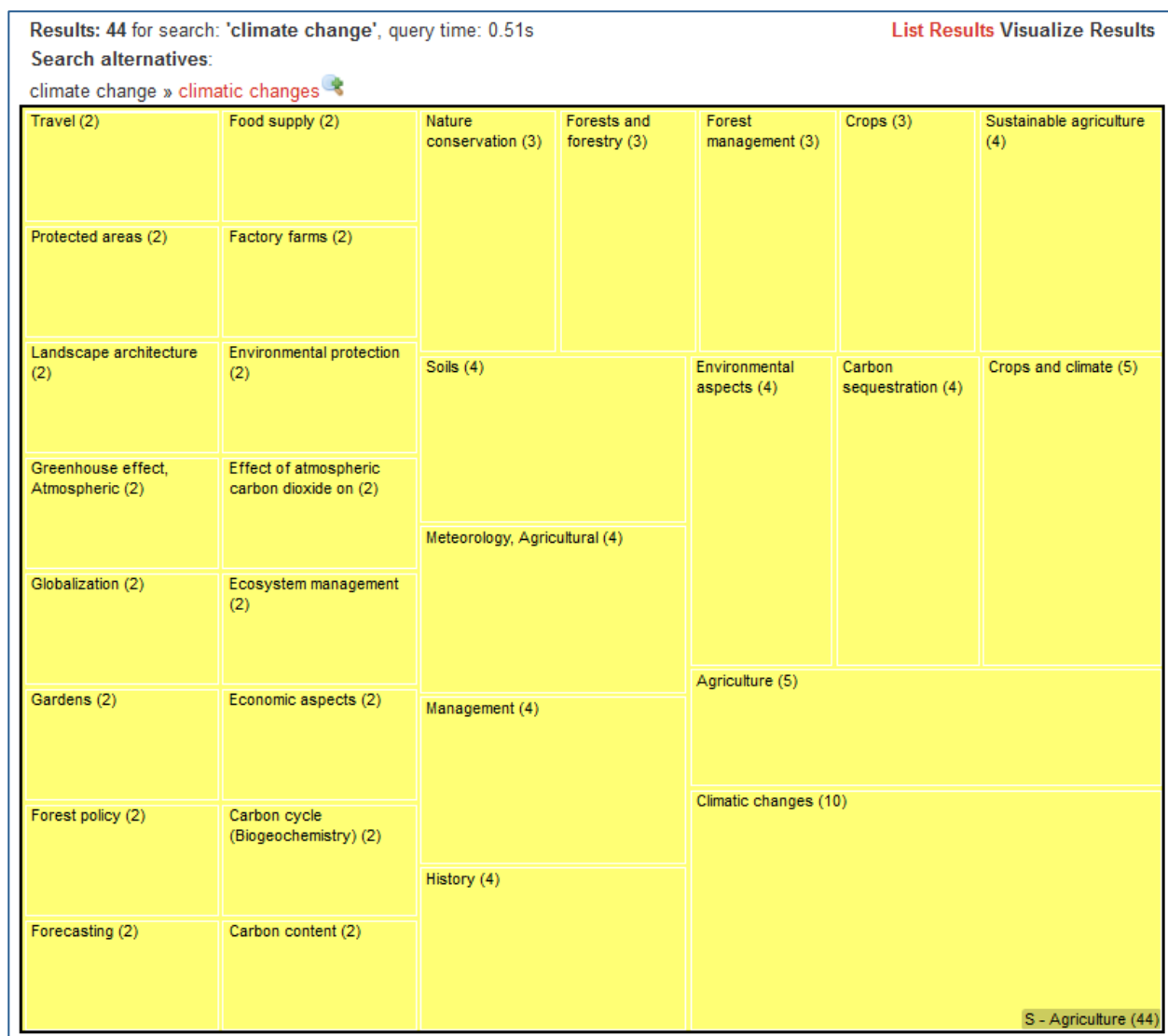


Figure 2. Visualization of the Results for a Search for “Climate Change,” Filtered to Show Only Results with Library of Congress Classification Numbers Starting with S.

RESULTS

Usability testing revealed no major obstacles in the way of users’ ability to navigate the visualization feature; the visualized search results were quickly deciphered by the participants with the assistance of the context set by the study’s outlined tasks. Familiarity with library catalogs in general, and the Grinnell College Libraries catalog in particular, showed no marked impact on users’ performance. No particular user group performed as an outlier in regards to users’ general ability to complete tasks or the time required to do so.

The most common issue to arise during the session concerned the visualization’s truncated text, which appears in the far left column of results when the descriptor text contains too many characters for the space allocated. (An example of this truncated text can be seen in figure 1.) The

subject boxes appearing in the furthest left column contain the least results, and therefore receive the least space within the visualization. This limited space sometimes results in truncated text. The full-text can be viewed by hovering over the truncated text box, but few users discovered this capability. Another common concern involved a participant's ability to switch their search results from the default list view to the visualized view. All participants were capable of selecting the "Visualize These Results" button required to produce the visualization, but a handful of participants expressed that they feared they would not find that option if they were not prompted to do so.

Participants remarked that the visualization initially appeared daunting but then quickly became comfortable navigating the results. Most participants, including staff, stated that they found the tool useful and intended to use it in the future during the course of their typical work at the college.

CONCLUSION

Librarians have had innovative ideas for ways to use subject and classification data to provide an improved online search experience for decades, yet after thirty-plus years of improvements in online catalogs, users continue to struggle with narrowing down their searches to produce manageable lists containing only relevant results.¹⁹ Computer scientists have been advocating for interfaces to support visual information-seeking since the 1980s. Finally, hardware and software have improved to the point where many of these ideas can be implemented feasibly, even by relatively small libraries. Now is the time to put some of them into production and see how well they work for library users. The particular visualizations reported in this article may or may not be the best possible visualizations of bibliographic data, but we will never know which of these ideas might prove to be the revolution that library discovery interfaces need until we try them.

Appendix A. Usability Testing Instrument

Introductory Questions

Before we look at the site, I'd like to ask you just a few quick questions.

—Have you searched for materials using the Grinnell College libraries' website before? If so, what for and when? (For students only: Could you please estimate how many research projects you've done at Grinnell College using the library catalog?)

In the Grinnell College Libraries, we're testing out a new tool in our catalog that presents search results in a different way than you are used to. Now I'm going to read you a short explanation of why we created this tool and what we hope the tool will do for you before we start the test.

Research is a conversation: a scholar reads writings by other scholars in the field, then enters into dialogue with them in his or her own writing. Most of the time, these conversations happen within the boundaries of a single discipline, such as chemistry, sociology, or art history, even when many disciplines are discussing similar topics. But when you do a search in a library catalog, writings that are part of many different conversations are all jumbled together in the results. It's like being thrown into one big room where all of these scholars, from all of these different disciplines, are talking over each other all at once. Our new visualization tool aims to help you sort all of these writings into the separate conversations in which they originated.

Scenarios

Now I am going to ask you to try doing some specific tasks using 3Search. You should read the instructions aloud for all tasks individually prior to beginning each. And again, as much as possible, it will help us if you can try to think out loud as you go along.

Please begin by reading the first scenario aloud and then begin the first scenario. If you are unsure whether you finished the task or not, please ask me. I can confirm if the task has been completed. Once you are done with Scenario 1, please continue onto Scenario 2 by reading it aloud and then beginning the task. Continue this process until all scenarios are finished. If you cannot complete a task, please be honest and try to explain briefly why you were unsuccessful and continue to the next.

1. Pretend that you are writing a paper about issues related to privacy and the Internet. Do a search in 3Search with the words "privacy Internet."
2. Please select the first WorldCat result and attempt to determine whether you have access to the full text of this book. If not, please indicate where you could request the full text through the InterLibrary Loan service.
3. Go back to your initial search results. Please choose "Explore these results" of the EBSCO database results. Choose an article. If you have unlimited texting, have the article's

information texted to your cell phone. Then, add the article to a new list for future reference throughout this project.

4. Go back to your initial search results. For Grinnell College's Collections results, click on the "Explore these results" link. Then click on the "Visualize Results" link to visualize the results. Which disciplines appear to have the greatest interest in this topic?
5. When privacy and the Internet are discussed in the context of law, what are some of the topics that are frequently covered in these discussions?
6. One specific topic you are considering is the legal issues around libel and slander on the Internet. How many resources do the libraries have on that specific topic?
7. Click on "Q – Science," to see the results authored by theoretical computer scientists. Based on these results, what are some of the topics that are frequently covered in their discussions when these computer scientists discuss privacy and the Internet?
8. Pretend that you are writing this paper for a computer science class and you are supposed to address your topic from a computer science perspective. Please narrow your results to only show results that are in the format of a book. Based on this new visualization, what might be some good topics to consider?
9. Add one of these books to the list you created in step 3. Please email all of the items on this list to yourself.

Debriefing

Thank you. That is it for the computer tasks. I have a few quick questions for you now that you have gotten a chance to use the site.

1. What do you think about 3Search? Is it something that you would use? Why or why not?
2. What is your favorite thing about 3Search?
3. What is your least favorite thing about 3Search?
4. Did you find the visualization function useful? Why or why not?
5. Do you have any recommendations for changes to the way this site looks or works?

REFERENCES

1. Elaine Svenonius, "Use of Classification in Online Retrieval," *Library Resources & Technical Services* 27, no. 1 (1983): 76–80, <http://alcts.ala.org/lrts/lrtsv25no1.pdf>.
2. Pauline A. Cochrane, "Subject Access—Free or Controlled? The Case of Papua New Guinea," in *Redesign of Catalogs and Indexes for Improved Online Subject Access: Selected Papers of Pauline A. Cochrane* (Phoenix: Oryx, 1985), 275. Previously published in *Online Public Access to Library Files: Conference Proceedings: The Proceedings of a Conference Held at the University of Bath, 3–5 September 1984* (Oxford: Elsevier, 1985).
3. Marcia Bates, "Subject Access in Online Catalogs: A Design Model," *Journal of the American Society for Information Science* 37, no. 6 (1986): 363, [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(198611\)37:6<357::AID-ASI1>3.0.CO;2-H](http://dx.doi.org/10.1002/(SICI)1097-4571(198611)37:6<357::AID-ASI1>3.0.CO;2-H)
4. Charles Ammi Cutter, *Rules for a Printed Dictionary Catalog* (Washington, DC: Government Printing Office, 1876).
5. David Ward, Jim Hahn, and Kirsten Feist, "Autocomplete as a Research Tool: A Study on Providing Search Suggestions," *Information Technology & Libraries* 31, no. 4 (2012): 6–19, <http://dx.doi.org/10.6017/ital.v31i4.1930>; Suzanne Chapman et al., "Manually Classifying User Search Queries on an Academic Library Web Site," *Journal of Web Librarianship* 7 (2013): 401–21, <http://dx.doi.org/10.1080/19322909.2013.842096>.
6. N. J. Belkin, R. N. Oddy, and H. M. Brooks, "ASK for Information Retrieval: Part I. Background and Theory," *Journal of Documentation* (1982): 61–71, <http://dx.doi.org/10.1108/eb026722>; Christine Borgman, "Why Are Online Catalogs Still Hard to Use?," *Journal of the American Society for Information Science* (1996): 493–503, [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199607\)47:7<493::AID-ASI3>3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1097-4571(199607)47:7<493::AID-ASI3>3.0.CO;2-P); Karen Markey, "The Online Library Catalog: Paradise Lost and Paradise Regained?," *D-Lib Magazine* 13, no. 1/2 (2007), <http://www.dlib.org/dlib/january07/markey/01markey.html>.
7. Cochrane, "Subject Access—Free or Controlled?," 275.
8. Svenonius, "Use of Classification in Online Retrieval," 78–79.
9. Jasper Kaizer and Anthony Hodge, "AquaBrowser Library: Search, Discover, Refine," *Library Hi Tech News* (December 2005): 9–12, <http://dx.doi.org/10.1108/07419050510644329>.
10. Kristen Antelman, Emily Lynema, and Andrew Pace, "Toward a Twenty-First Century Library Catalog," *Information Technology & Libraries* 25, no. 3 (2006): 128–39, <http://dx.doi.org/10.6017/ital.v25i3.3342>.
11. Tod Olson, "Utility of a Faceted Catalog for Scholarly Research," *Library Hi Tech* (2007): 550–61, <http://dx.doi.org/10.1108/07378830710840509>; Jody Condit Fagan, "Usability Studies of

-
- Faceted Browsing: A Literature Review,” *Information Technology and Libraries* 29, no. 2 (2010): 58-66, <http://dx.doi.org/10.6017/ital.v29i2.3144>.
12. Kathleen Bauer, “Yale University Library VuFind Test—Undergraduates,” November 11, 2008, accessed September 9, 2014, http://www.library.yale.edu/usability/studies/summary_undergraduate.doc.
 13. See, for example, Ben Shneiderman, “The Future of Interactive Systems and the Emergence of Direct Manipulation,” *Behaviour & Information Technology* 1 (1982): 237–56, <http://dx.doi.org/10.1080/01449298208914450>; Ben Shneiderman, “Dynamic Queries for Visual Information Seeking,” *IEEE Software* 11 (1994): 70–77, <http://dx.doi.org/10.1109/52.329404>.
 14. See, for example, Aleks Aris et al., “Visual Overviews for Discovering Key Papers and Influences Across Research Fronts,” *Journal of the American Society for Information Science & Technology* 60 (2009): 2219–28, <http://dx.doi.org/10.1002/asi.v60:11>; Furu Wei et al., “TIARA: A Visual Exploratory Text Analytic System,” in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Washington, DC: ACM, 2010), 153–62, <http://dx.doi.org/10.1145/1835804.1835827>; Cody Dunne, Ben Shneiderman, Robert Gove, Judith Klavans, and Bonnie Dorr, “Rapid Understanding of Scientific Paper Collections: Integrating Statistics, Text Analysis, and Visualization,” *Journal of the American Society for Information Science & Technology* 63 (2012): 2351–69, <http://dx.doi.org/10.1002/asi.22652>.
 15. The most notable exception is Carrot² (<http://search.carrot2.org>), a search tool that will automatically cluster web search results and display visualizations of those clusters.
 16. Ben Shneiderman, “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations,” September 1996, accessed April 27, 2014, http://drum.lib.umd.edu/bitstream/1903/5784/1/TR_96-66.pdf.
 17. Ben Shneiderman, “Treemaps for Space-Constrained Visualization of Hierarchies: Including the History of Treemap Research at the University of Maryland,” Institute for Systems Research, accessed October 6, 2014, <http://www.cs.umd.edu/hcil/treemap-history>.
 18. To explore this feature in our catalog, go to <https://libweb.grinnell.edu/vufind/Search/Home>, do a search, and click on the “Visualize Results” link in the upper right.
 19. A recent Project Information Literacy report found that the two aspects of research that first-year students found most difficult were “coming up with keywords to narrow down searches” and “filtering and sorting through irrelevant results from online searches.” Alison J. Head, *Learning the Ropes: How Freshmen Conduct Course Research Once They Enter College* (Project Information Literacy, December 5, 2013), http://projectinfolit.org/images/pdfs/pil_2013_freshmenstudy_fullreport.pdf, 15.

