

Library OPAC Metadata Record Clusters for Fiction

Adam Trachtenberg
Columbia University

E-mail: adam@trachtenberg.com

ABSTRACT

We present a user interface for library online public access card catalogs (OPACs). This design improves author and title searching by grouping results by thematic categories. The HTML interface uses a three-column layout to provide relative context to a searcher for easier database browsing. Elements from physical catalogs are incorporated into this design, for greater comfort to non-computer savvy patrons.

KEYWORDS: online catalog design, clusters, web search, automatic clustering, metadata, bibliographic relationships

INTRODUCTION

Beginning in the 1970s, libraries converted their holding records from wood and paper card catalogs to electronic form. OPAC user interfaces were originally created for text terminals, but in the 1990s, as the World Wide Web gained popularity, libraries converted from ASCII to HTML.

However, OPAC design has changed relatively little. In many cases, the only usability benefit gained from the new medium is the hyperlinking of cross-references. Also, to simplify basic searching, some HTML front-ends eliminated features. For example, at Columbia University, the telnet interface allows searching by format and language, while the Web interface does not. So, simple tasks like finding a copy of *Faust* in the original German are very difficult. We present a method of displaying related records in clusters that allows this and related tasks without confusion.

RECORD CLUSTERING

Libraries often contain multiple records for a single work. However, 37% of readers have difficulty knowing what is in an online catalog and 28% have problems scanning through a long display [4]. One solution is to combine records into a set of smaller clusters. The Bradford OPAC 2 (BOPAC 2) does this by linking works by their related attributes. So, in a search for records titled *Ulysses* it groups all matches where the author is Joyce, James, regardless of publisher or edition. In contrast, most OPACs only combine identical records. Ridley suggests the BOPAC

“straightforward clustering techniques... dramatically improve the user interface.” [5]

However, BOPAC 2 fails to group using a record’s meta-characteristics, like format or language. Hearst et al. have shown novice users respond highly favorably to the use of hierarchical-faced metadata in a search context. [3] Using an unconstrained sorting task methodology, Carlyle reveals that readers organize fiction with regard to attributes like “physical format, audience, content description, pictorial elements, usage, and language.” [1] Further research by Carlyle and Summerlin proves the feasibility of mapping data to these clusters. Despite inconsistent bibliographic record standardization, or authority control, “preliminary research indicates that for some large works of fiction, as many as 94 percent of records... could be successfully clustered using automatic methods.” [2]

THREE-COLUMN LAYOUT

Carlyle does not provide any interface designs for this system. I propose a three columned layout, a left-to-right flow with continuing refinement of scope and richer detail in each column. The leftmost column contains a list of each of groups identified by Carlyle: illustrations, non-English language, amplifications, non-book, etc. Each top-level category shows the most popular subgroups for the search; for the “non-English” category, this might be French, Spanish, and German. Results are arranged in descending order from most to least popular and the number of hits is listed beside each item. We assume, as an approximation, that library holdings mirror user preferences. A similar design is used by eBay with success and we believe this interface should give superior results over traditional OPAC layouts while decreasing retrieval time.

Records appear not just in one category, but in every appropriate one. So, an illustrated edition of a book in Chinese with editorial commentary appears three times. Since it is unknown which path a reader will take to find a record, our interface provides every one we can determine.

Clicking on a category or subcategory displays a list of matching records in the middle column. Each record contains the canonical author name, book title and date of publication, and holding library names and call numbers. Ideally, these items would be ordered by popularity; here they are listed in reverse chronological order, under the assumption that current books are more preferable.

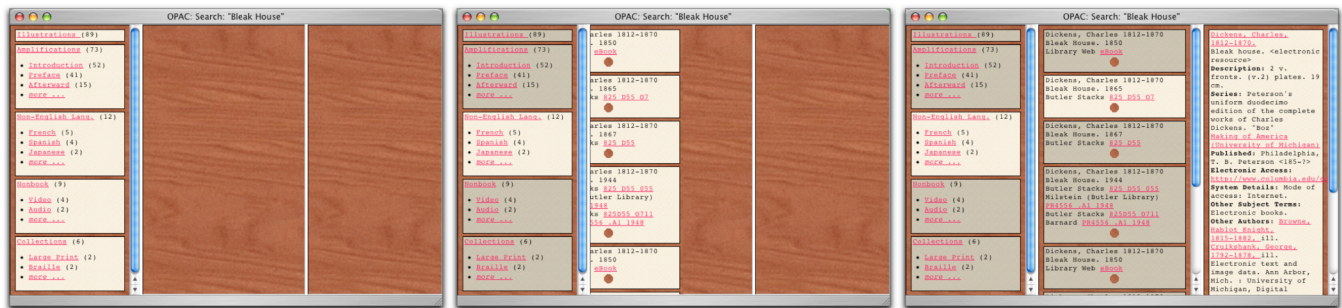


Figure 1: the search refinement process using a three-columned OPAC interface. The left window is the opening search results screen listing categories and subcategories; the middle window shows what occurs next: the chosen category is highlighted and a middle results panel begins to slide open; last, when a book is selected from the middle column, the rightmost panel of the right window reveals its complete bibliographic record.

At the same time, to improve flow and the reader's ability to scan the screen, the non-selected cards in the left panel are darkened. The contrast change pops out the important card and illuminates the search pathway. To further accentuate the left to right direction of information, the new panel smoothly slides out, as if the person opened a drawer.

When a searcher clicks on a card in the middle column the right panel displays that book's complete record. Besides the earlier fields this may include: other author names, a physical description, series data, additional publication information, and subject cross-references. Each subsequent record clicked causes the panel to include information about the next work and remove the old data. A similar transformation takes place between these two columns as in the previous two.

The traditional single-paneled OPAC design makes it extremely difficult to maintain context within a search. Searchers bounce back and forth between the brief views of multiple records and a detailed view of a single record. Upon returning to the index page of listings, they must repeatedly reorient themselves. A list in the middle column eases scanning for relevant records.

VISUAL DESIGN ELEMENTS

A major design consideration is incorporating elements from physical card catalogs. Many readers have strong affection for their wood shelves and find computers bewildering. Preserving the visual identity of the physical catalog is useful. It comforts the reader and distinguishes this interface from the dozens of generic interfaces he or she sees, allowing the reader to associate certain states of mind or search strategies with an environment that is a richer visual mnemonic.

The background texture is wood; records appear as "cards," manila-colored rectangular shapes with rounded top corners and a punched hole at the bottom. The text color, shades, and fonts are similarly derived from real library card ink and typography. This makes the experience less sterile and more tactile, like the texts people seek.

FURTHER DIRECTIONS

The current prototype is for gaining feedback and modeling flow. A more polished user interface will be converted from static HTML to dynamic code. I have identified a programming language, PHP, and extension toolkit, YAZ, for querying Z39.50 servers and early results project no implementation difficulties. Last, I will validate the design using a formal user testing program.

ACKNOWLEDGMENTS

Special thanks to: Beth Hondl for the original topic; W. Bradford Paley for design and project advice; Baruch Gabo and James Pak for ideas on flow and record contrast; Sarah Witte for knowledge on CLIO, past, present, and future; and the librarians at Columbia University and the New York Public Library for stories, suggestions, and ideas.

REFERENCES

1. Carlyle, Allyson. (1997). "Fulfilling the Second Objective in the Online Catalog: Schemes for Organizing Author and Work Records into Usable Displays." *Library Resources & Technical Services*. 41 (2): 9-100.
2. Carlyle, Allyson and Joel Summerlin. (2002). "Transforming Catalog Displays: Record Clustering for Works of Fiction." *Cataloging & Classification Quarterly*. 33, (3/4): 13-25.
3. Hearst, Elliott, English, Sinha, Swearingen, and Yee. (2002). "Finding the Flow in Web Site Search." *Communications of the ACM*. 45 (9): 42-49.
4. Matthews, J.R., et al (Eds.). (1983). *Using Online Catalogs: A Nationwide Survey*. New York: Neal-Schuman. p. 124.
5. Ridley, M.J.; Ayers, F.H.; Nielsen, L.P.S. (1998). *The Bradford OPAC 2 (BOPAC2): managing and displaying retrievals from a distributed search in Z39.50*.
<http://www.bopac2.comp.brad.ac.uk/~bopac2/report>