

Searching for Books and Images in an OPAC: Effects of LCSH, TOC and Subject Domain

Ingrid Hsieh-Yee, Young Choi, Bill Kules
 hsiehyee@cua.edu; choiy@cua.edu; kules@cua.edu
 School of Library and Information Science, Catholic University of America
 American Society for Information Science & Technology Annual Meeting, Oct. 2007, Milwaukee, Wisconsin

Abstract

We examined user queries for subject searches of books and images to analyze the effects of the Library of Congress subject headings (LCSH), table of contents (TOC), and subject domains on search success. User queries were collected through a Web survey that included four target items for searching: a book on educating at-risk children, a book on energy, a photograph of immigrants arriving at Ellis Island, and a poster on breast cancer prevention. 556 queries for books and 492 queries for images were searched against the Library of Congress OPAC to assess search results, because it has provided access to LC's rich collections, including its Prints and Photographs collections.

We found the retrieval differences between LCSH and TOC to be statistically significant for books and for images, but the effects of subject domains (education vs. science; history vs. medicine) were not. When searched against LCSH, queries for books and for images resulted in many empty sets, but produced fewer empty sets when searched against TOC. Both LCSH and TOC had limited success in retrieving relevant items. Data indicated users' search terms were more likely to match terms in TOC than in LCSH, but neither TOC nor LCSH performed well for subject searches.

Background

- Subject searching vs Keyword searching - A long history of debate in information retrieval (Rowley, 1994)

Still relevant in the 21st century due to

- Challenges to library catalogues (OPACs)
- Users' preference for keyword searching
- Popularity of Search engines
- Continuing cataloging and indexing practices, though value of subject headings has been questioned (<http://lists.ala.org/www/info/headings>)
- Strong user interest in subject searches and the need to support subject searches
- Debate over the use of controlled vocabulary and natural language for OPAC subject searches

Objective

- To investigate how well table of contents and subject headings support subject searches by users who prefer to search by keywords and whether subject domains affect how users conduct subject searches

Research Questions

- How successful are searches by LCSH and TOC for users who look for books and images in various subject areas?

- To what extent do users' search queries match LCSH and TOC?

- Do user queries for the science item match terms in LCSH and TOC better than queries for the social science item?

- Do user queries for books match terms in LCSH and TOC better than queries for images?

- When user queries are searched against LCSH and TOC, to what extent do they retrieve items similar to the target items?

- Are user queries related to the subject domain and format of desired objects?

This project is sponsored by a research grant from the Catholic University of America

Methodology

A Web survey, hosted by SurveyMonkey, with four search scenarios was conducted in November 2006 to collect user queries.

Four types of items were included in the Web study

- A book on educating at risk children (see below)
- A book on the history of energy physics
- A photo of immigrants arriving at Ellis Island (see below)
- A poster on cancer prevention

#1) You are given an assignment to write a research paper on teaching children in school settings. Your professor gives you the following description of a useful publication to help you focus your research topic.

This work takes the perspective that young children are resilient and that general education teachers can make a difference with those who have unique needs for social, emotional, and academic support. It guides teachers in understanding early risk characteristics in young children who have not been diagnosed as having a disability, and provides effective intervention strategies, tools, and authentic examples to support learning and reduce chances for failure in school settings.

1a) Please provide 3 - 5 search statements you would use to look for materials on this topic. (You may enter up to 250 characters per search.)

Search 1:

Search 2:

Search 3:

Search 4:

Search 5:

1b) Please indicate your familiarity with this search topic.

Rating: Not at all familiar Somewhat familiar Very familiar



Participants provided up to five search queries for each search scenario

43 SLIS students and recent graduates took part in the survey. Most have limited or no cataloging or reference experience.

Each search queries was searched in the Library of Congress online catalog, because the catalog

- Contains records for text and images
- Has many table of contents
- Has LC subject headings
- Supports searching by subject headings and note fields
- Has description notes for nonprint materials

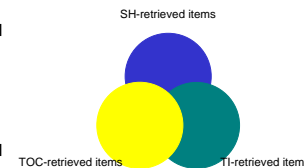
Analysis focuses on

- Items uniquely retrieved by LCSH =

all LCSH-retrieved items –
 TOC-retrieved items –
 Title-retrieved items

- Items uniquely retrieved by TOC =

all TOC-retrieved items –
 LCSH-retrieved items –
 Title-retrieved items



Sample Search Screen

The Library of Congress

LIBRARY OF CONGRESS ONLINE CATALOG

Help Search Search History Headings List Titles List Request an Item Account Status Other Databases Start Over

DATABASE: Library of Congress Online Catalog

Basic Search Guided Search

Search: immigrants ellis island all of these As: Subject: All (S:SUB)

AND OR NOT

Search: immigrants ellis island all of these As: Contents Notes (N:CON)

25 records per page

Begin Search Clear Search Set Search Limits

FINDINGS

Mean number of items retrieved

| | Cases | Mean | Stan. Dev. |
|---------|-------|-------|------------|
| Ed TOC | 144 | 43.5 | 191.6 |
| Ed SH | 144 | 324.5 | 1208.8 |
| Sci TOC | 134 | 76.5 | 321.8 |
| Sci SH | 134 | 428.5 | 1813.7 |
| HisTOC | 126 | 26.0 | 102.3 |
| His SH | 126 | 282.8 | 1446.8 |
| Med TOC | 120 | 100.5 | 642.3 |
| Med SH | 120 | 386.1 | 1701.6 |

Book Searches

No interaction between search type (TOC vs. SH) and subject domain (ed. vs. sci.) ($F = 0.145$, $p = 0.7034$)

Retrieval differences between TOC and LCSH were significant for both subjects ($F = 11.537$ and $p = 0.0007$)

ANOVA Table for NewTSEP

| | DF | Sum of Squares | Mean Square | F Value | P Value | Lambda | Power |
|-------------------------|-----|----------------|--------------|---------|---------|--------|-------|
| Mod:Domain | 1 | 13908130.361 | 13908130.361 | 11.537 | .0007 | 11.537 | .943 |
| Discipline | 1 | 660337.851 | 660337.851 | .540 | .4629 | .540 | .110 |
| Mod:Domain * Discipline | 1 | 174921.526 | 174921.526 | .145 | .7034 | .145 | .066 |
| Residual | 552 | 665477498.655 | 1205575.179 | | | | |

Image Searches

No interaction between search type (TOC vs. SH) and subject domain (history vs. medicine) ($F = 0.016$, $p = 0.8994$)

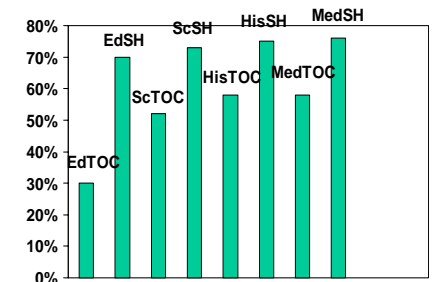
Retrieval differences between TOC and LCSH were significant for both subjects ($F = 6.749$, $p = 0.0097$)

Search Type vs. Format

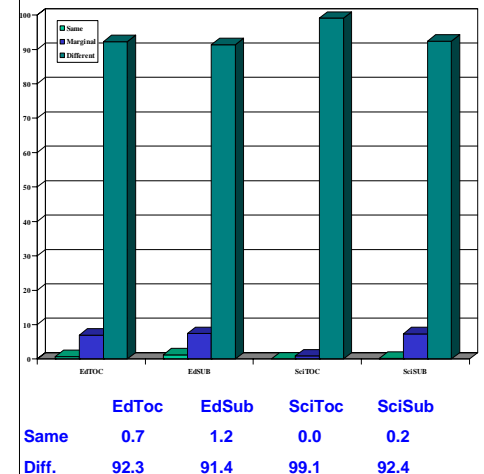
No interaction between search type (TOC vs. SH) and format (bk vs. image)

Retrieval differences between TOC and SH were statistically significant for both formats ($F = 17.677$, $p < 0.0001$)

SH searches resulted in more zero sets than TOC searches



Very few items retrieved by TOC and SH were similar to the book items used in the Web survey



CONCLUSIONS

- The retrieval differences between TOC and LCSH were found to be statistically significant for books, images, and works in different subject domains.

- Subject domain and format did not affect how user queries matched TOC or LCSH.

- Retrieval by LCSH resulted in 70% to 83.5% of empty set, compared to retrieval by TOC that resulted in 30% to 61% of empty set. Data suggested users' search terms had better chances of matching in TOC than in LCSH.

- However, 92% to 99% of the items retrieved by TOC were not similar to the two book items used in the survey, suggesting retrieval by TOC was still not effective (or precise enough) in finding items in the desired subject area.