DISCIPLINING SEARCH / SEARCHING DISCIPLINES:

Perspectives from academic communities on metasearch quality indicators

Katherine Skinner: Digital Programs Team Leader, Emory University

Rohit Chopra: MetaScholar Fellow, Woodruff Library, Emory University
Presentation Overview

- Project context: history, background, and goals
- How do we gather and understand user quality metrics?
- Project goals:
  - to gather foundational data concerning user quality metrics for metasearch retrieval
  - to discover the best way to present retrieval results based on user expectations
  - apply these findings to a working prototype system
  - evaluate the effectiveness of this system relative to standard or typically available alternatives
MetaScholar Initiative

- Digital library research program at Emory University to study applications of new technology for scholarly communication
- $2.6 M in funding for eight projects 2001-2005
- Collaborations with libraries, archives, museums, and other cultural memory organizations.
MetaScholar Research Areas

- Metadata Sharing
  - Programming services
  - Open source tools and systems

- Search and Discovery Systems
  - Combined searching and automated classification
  - Crafting and organizing coherent subject domains of related resources

- Digital Publishing
  - Evaluation of new tools for scholarly discourse / content creation
Project Context

- Frustration with current search and discovery systems for scholarly needs
  - Work with Southern Studies faculty group, our Advisory Board
  - AmericanSouth.org, Music of Social Change, MetaCombine
Current Search Systems

- Distributed approach:
  - Journals (ProjectMuse, JSTOR)
  - Books, Movies, Music (EUCLID)
  - Interlibrary Loans (WorldCat)
  - Archival Resources (OAIster, AmericanSouth.org, individual archives)
  - Web-based Resources (Google, Yahoo, etc)
Federated Approach

- Vision for a new, open source toolkit
  - Searching across various informational types
  - Rank according to user profiles and researched “quality metrics” as identified by particular user groups
  - Re-deployable at different institutions for their particular user communities
Staff

- Martin Halbert, Director of Digital Programs and Systems
- Aaron Krowne, Head of Digital Library Research
- Urvashi Gadi, Analyst and Programmer
- Rohit Chopra, MetaScholar Fellow
- Katherine Skinner, Digital Programs Team Leader
- Carrie Finnegan, Graduate Student Associate
Hypothesis

- “End-users of digital library metasearch services think in terms of implicit quality metrics that can be respected in retrieval ranking, improving results over the current text-query relevance and even hub-authority based methods. If effective metasearch systems are to be informed by empirical studies of actual users, this hypothesis must be tested” (Halbert and Krowne 2004)
Methodology

- Qualitative (Emory): Focus groups with Emory faculty and graduate students in the humanities, social sciences and sciences
- Quantitative (Virginia Tech): Survey instrument to record students responses to tasks on metasearch retrieval
Focus Groups Structure

- 9 focus groups between Oct. and Dec. 05
- User communities: Emory faculty and graduate students (in the humanities and sciences)
- 22 participants (11 humanities faculty, 3 sciences faculty, 8 humanities graduate students)
Individual Focus Group Design

- Introduction (objectives, participant rights)
- Hypothetical scenario:
  - Mock-ups of metasearch (for “globalization,” “civil rights movement”)
  - Highlighting metadata attributes (format, no. of citations, substantive quality rankings)
  - Sequence of mock-ups to show increasing levels of metadata / quality metrics
- Comparison with existing metasearch systems (eg., A9.com)
<table>
<thead>
<tr>
<th>Version 2: Screen a</th>
</tr>
</thead>
</table>

**MetaScholar Search Engine Version 1**

<table>
<thead>
<tr>
<th>civil rights movement</th>
<th>search</th>
</tr>
</thead>
</table>

Results sorted by: [ ] RELEVANCE AND/OR [ ] DOMAIN (peer-reviewed journals, books, archival records, websites) AND/OR [ ] NUMBER OF TIMES CITED

1. **Upon this Rock: The Black Church, Civil Rights and the Non-Violence Movement** (100%)
   - **Author:** Alston C. Galloway
   - **Source:** Political Science and Politics (Volume 11, no. 2, July 1978, pp. 181-194)
   - **Matches:** 223 (see all instances) | **Citations:** 121 articles cite this article (see all instances)
   - **Average rating:** ★★★★★

2. **Origins of the Civil Rights Movement** (100%)
   - **Author:** Alston D. Morris
   - **Source:** Libraries Information Database
   - **Matches:** 111 (see all instances) | **Citations:** 266 books cite this book (see all instances)
   - **Average rating:** ★★★★★

3. **Narrative of the Life of Frederick Douglass: An American Slave** (100%)
   - **Author:** Frederick Douglass
   - **Source:** James and Nancy
   - **Matches:** 246 (see all instances) | **Citations:** 117 books cite this book (see all instances)
   - **Average rating:** ★★★★★

4. **The Modern Civil Rights Movement: Federal Records in the National Archives and Records Administration** (100%)
   - **Author:** Compiled by Amanda Byrdonski, Glenn Lonsdale, and Martin Tuohy
   - **Source:** National Archives (NARA)
   - **Matches:** 155 (see all instances) | **Citations:** 213 articles cite this collection (see all instances)
   - **Average rating:** ★★★★★

5. **Television News and the Civil Rights Struggle: The Views in Virginia and Mississippi** (65%)
   - **Author:** William G. Thomas III
   - **Source:** Southern Spaces (November 3, 2004, http://southernspaces.org/contents/williamthomas.htm)
   - **Matches:** 1 (see all instances) | **Citations:** 15 articles cite this article (see all instances)
   - **Average rating:** ★★★★★

6. **Voting Rights Act: General Records of the United States Government, Record Group 11, National Archives, August 8 1965, Public Law 89-100 (55%)**
   - **Source:** National Archives (NARA)
   - **Matches:** 2 (see all instances) | **Citations:** 117 sources cite this collection (see all instances)
   - **Average rating:** ★★★★★

7. **Abolitionism and Issues of Race and Gender** (68%)
   - **Author:** John F. McKelvey
   - **Source:** Library of Congress
   - **Matches:** 7 (see all instances) | **Citations:** 32 books cite this book (see all instances)
**Version 2: Screen c**

### MetaScholar Search Engine Version 1

**Results sorted by: RELEVANCE AND/OR DOMAIN (peer-reviewed journals, books, archival records, websites) AND/OR NUMBER OF TIMES CITED**

<table>
<thead>
<tr>
<th>Peer reviewed journals (see list)</th>
<th>Books</th>
<th>Archival Records</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cited more than 500 times</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Reading more about the Civil Rights Movement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The Civil Rights Movement: Struggle for Employment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **cited 100 to 500 times**        |       |                  |
| 1. Upon This Rock: The Black Church, Civil Rights, and Urban Utopianism |       |                  |
| 2. How the Civil Rights Movement Realized Labor |       |                  |

| **cited 1 to 100 times**          |       |                  |
| 1. To Pay for It: News and the Civil Rights Struggle |       |                  |
| 2. A Revolution in the All Rights Movement |       |                  |

---

### Books

- *Civil Rights Movement* by John D. Hicks (256 citations)
- *Litigation of Civil Rights Cases* by John D. Hicks (256 citations)

### Archival Records

- *The National Archives (NARA)*
  - Relevance: 100 percent
  - Matches: 135 (see all)
  - Average Rating: 4.5 stars
- *The National Archives (NARA)*
  - Relevance: 100 percent
  - Matches: 106 (see all)
  - Average Rating: 4.5 stars
Scope of Discussion

- Quality and ‘fitness’ metrics (substantive and descriptive attributes)
- Functionality / logic of retrieval and display
- Aesthetic / design issues
Findings

- Strong affirmation of need for good metadata:
  - at discipline and topical specialization level
  - at genre / domain level (archives, pamphlets, books, etc.)

- Understanding of quality / fitness contingent more on metadisciplinary and disciplinary orientation (sciences, humanities) than on research level (faculty or graduate students)

- Partial consensus across communities on abstract notions of academic quality

- Highly specific information requirements about quality / fitness of records, based on combination of task, discipline, and topical specialization
Demand for Transparency

- Transparency: Explanation of logic of ranking, meaning of each attribute (domain, format, number of citations), range of databases, sources searched etc.
- Dissatisfaction expressed about opacity of ranking criteria in current systems
- Transparency needs to be reinforced through institutional credibility
Empowerment to Evaluate Quality

- Non-quantifiability of academic value: Deep skepticism expressed about substantive quality indicators such as points system for ranking books and articles.

- Quality defined as the empowerment of academic communities to make their own judgment about quality.

- A ranking system for academic websites described as useful, if provided by legitimate associations e.g., the AHA for history websites.
Competing Notions of Quality

- Quality ranking requires balancing competing or contradictory imperatives
- Need to find “an equivalent of the academic peer-review process” in terms of credibility
- Need to avoid reinforcing “the very conservative base of peer review” while providing access to the broadest range of source materials for research
Discipline- and Task-specific Metrics

- Theology: divide records by “secular” and “theological” domains
- Interdisciplinary Studies: metadata on materials such as pamphlets and posters
- Mathematics: Search intelligence to be sensitive to mathematical terminology
- Biology: Identification of laboratory protocols in preliminary research for new methodology
- History, Literature: Archival metadata, full content of archives, and quality of archival holdings
Increased User Options as Quality

- Option to sort full text from summaries etc.
- Option to retain ‘memory’ of search sequence, if trying new options
- Concatenate results of same source from multiple access points (e.g., one periodical accessible from multiple databases)
- Option to sort results by word length, number of citations (e.g., Amazon.com)
- Option to switch sets of metrics midway through search
Areas for Developing Quality Metrics

- Digital images
- Archives: collection-level and item-level holdings
- Foreign-language content
- Pamphlets, posters, leaflets: cultural resources
- Websites (not counting online journal databases online etc.)
Conclusion: QM Project Phase II

- Creation of prototype system
- Specialized focus groups
  - Archives and archival holdings
  - Foreign language content
  - ‘Non-standard’ academic materials
- Reports