Relevancy and the Creation of Shared Meaning

Samvera Virtual Connect
May 15, 2020
Getting Oriented - The Presenter

Max Kadel
Developer
Data Curation Experts

D C E
Developing Digital Repository Solutions
Getting Oriented - The Projects

curate.library.emory.edu
Curate - Staff-facing repository
Code name... Curate

Hyrax

digital.library.emory.edu
Emory Digital Collections - Public-facing discovery
Code name Lux
Blacklight
1. What is included in, *and excluded from*, search results

- What fields are being searched?
- What fields are being omitted?
- Completeness - precision vs. recall
  - Deposition? Reader Advisory?
- Defaulting to OR searching vs. AND searching
OR vs. AND

1. Barber and his dog, barber shop at corner of Fleet Street and Cornhill Street in Annapolis, Maryland, circa 1950s

2. A small fishing village, St. Lucia, B.W.I. : a barber at work : a street

3. Barber leaning on his barber chair

1. Barber and his dog, barber shop at corner of Fleet Street and Cornhill Street in Annapolis, Maryland, circa 1950s

2. A small fishing village, St. Lucia, B.W.I. : a barber at work : a street
2. How those search results are ranked

- Are certain fields given more weight? (e.g. title, author)
- Are others de-emphasized (e.g. abstract)
Relevancy Definition

3. Dependent on the specific user’s needs

- Impossible to make it fit every possible user’s needs
- Awareness of unconscious bias can help us make better decisions
What makes the meaning?
Search terms (User)

Search interface (Lux team)

Object description (Catalogers)

Solr index (Curate team & catalogers)

Solr query (Lux team)

Results & Relevancy
A Human Searches

- **Our Human**
  - Max

- **Has Identities**
  - White, educated, non-disabled, queer, trans man, from US, non-immigrant, documented...

- **Has Research Interests**
  - Music & representation, especially intersections of gender, masculinity, race, and imprisonment.

- **Search terms**
  - **Instruments**
  - **Leadbelly** (folk & blues musician)
  - **Max Roach** (bebop drummer)
  - **Prison**
Some Levers

- Relevancy test suite
  - Replicability, observability of results and ranking

- Search interface
  - Fielded search choices, clear options for limit & sort

- Solr query - Solr tuning, including query field boosting
  - E.g. in Blacklight catalog controller, boost the title and keywords:
    ```
    qf: "title_tesim creator_tesim keywords_tesim^2.1 subject_topics_tesim^2.5 abstract_tesim^0.4"
    ```

- Solr index
  - Make intentional choices about full-text indexing, more is not always better

- Object description
  - What is the “default” or unmarked subject? Should it be made explicit?
  - Titles for untitled objects
Relevancy Testing & Observation

- Starting point: direct interaction with system and collection
Observing tuning changes

Manual inspection of “civil rights” search: pamphlet gone from the front page. Running the test suite tells us more about the effects of that proposed ad-hoc change; the suspicious pamphlet is demoted, but Martin Luther King Jr. is also absent from the front page.
Further Reading

- [https://journal.code4lib.org/articles/10425](https://journal.code4lib.org/articles/10425) Identity and relevancy
- bell hooks’ writings