LINKING LIBRARY USERS TO INFORMATION THROUGH DATA

Rebecca Thompson and Heather Lea Moulaison
Missouri State University Library and University of Missouri School of Information Science and Learning Technologies

How do we improve the user's ability to find information in the online public access catalog (OPAC)? How can we expand the reach of the user to resources outside of the library catalog? Are linked data resources one solution?

Rationale

• Transaction log analyses have demonstrated that users do not query OPACs effectively
• Linked data initiatives are a way to link out to curated, trustworthy data on the web
• The connection between library content, user searches, and linked data is worth exploring

Research questions

1. Focusing on "author" entities (personal names and corporate bodies), how might we improve the patron's search experience through linked data?
2. How are patrons searching in the author index, and are they finding results?

What is linked data?

"Linked Data" refers to data published in accordance with principles designed to facilitate linkages among datasets, element sets, and vocabulary elements. (Library Linked Data Incubator Group, 2011)

Some linked data projects of interest to the library world

• Open Library: http://openlibrary.org/author allows users to merge records, points users directly back to libraries at the book-level
• NNDDB Mapper: http://mapper.nndb.com/ connections between 34,000 individuals are listed
• Library of Congress (LOC) Names: http://id.loc.gov/authorities/names Library of Congress name authority file as linked data
• Virtual International Authority File (VIAF): http://viaf.org/ "joint project of several national libraries plus selected regional and trans-national library agencies"
• DBpedia: http://dbpedia.org/ (SPARQL builder: http://dbpedia.org/sparql) contents of Wikipedia in linked data format
• Freebase: http://www.freebase.com/ "An entity graph of people, places and things, built by a community that loves open data." Entries may not be disambiguated, and link when possible to all other vocabularies mentioned here plus many more

Method

Analysis of author index (left-anchored) queries to Millennium III ILS at a medium-sized academic library in the Midwest from 28 March – 6 April, 2011.

• Transaction logs showed the characters input minus punctuation, time of the search, and date of the search
• No information about patron ISP, number of results returned, number of searches in session, etc. were recorded

Exactly 700 left-anchored author searches were carried out

• Because of the interest in "authors", 1-term queries (n=238, 34%) were not considered unless they reasonably represented an artist or association acronym (i.e. Beck, ACM, etc.)

Two-part analysis

1. The 462 retained queries were re-run against the OPAC to assess hits
   • Queries yielding zero hits in the OPAC were then manually relaunched in seven linked data repositories
2. To assess the quality of the retained set of OPAC queries, the search terms were analyzed for their context and construction

Results

1. OPAC queries and linked data
   • Of the retained OPAC queries (n=462), nearly two-thirds had some degree of success:
     - Over one third (34.4%, n=159) yielded zero hits in the OPAC (n=fail)
     - Of the failed searches (n=159), 72% (n=115) overall matched entries in two linked data repositories (see Figure 1)
   - Freebase
     • Other repositories used as source data for these two (e.g. DBpedia and LOC names) had similar, relevant content (see Figure 2)

2. Formulation of OPAC queries
   • Of the retained OPAC queries (n=462),
     - 16 were searching corporate bodies (3.47%), all had a plausible left-anchor
     - 438 were searching personal names (94.8%)
     - Of these, 99 had "incorrect" left-anchors
     - 7 were unknown or searching in the wrong index (1.5%)

Discussion & recommendations

Promote linked data in the OPAC when possible
• Provide access to Freebase-like lists of current resources in response to patron queries through linked data
• Help users disambiguate names based not only on library holdings, but on LD entries
• Take advantage of linked information in LD repositories to help patrons explore authors further (access to videos, MySpace pages, etc.)

Continue to work on providing a patron-centered OPAC
• To accommodate the over 20% of the retained author index queries that were incorrectly formulated, libraries might want to consider doing away with left-anchored searches
  - Avoid issues of searching the "right way" vs. the "natural way"
  - Make authority records and the whole MARC format more granular (i.e. implement RDA)
  - Link out to LD resources (that link back to libraries!) when faced with providing zero hits or to supplement information in the OPAC
• Libraries have been doing authority work well for a long time
  - All locally created authority records should be exported and shared to enrich the data available in general
  - Some of the records won’t be perfect, but that’s ok
  - Consider crowdsourcing authority work (see the example at Open Library and Freebase)

Research questions revisited

1. Focusing on "author" entities (personal names and corporate bodies), how might we improve the patron’s search experience through linked data?
A: Linked data sources, when it comes to personal names, may fill a void when it comes to library collections. Libraries should consider linking out to these resources as part of their mission to help patrons access resources.

2. How are patrons searching in the author index, and are they finding results?
A: 34% of author index searches were one-term queries. Of the author queries retained, over 20% were formulated incorrectly for searches in the OPAC, but were perfectly acceptable for searches in other systems (search engines, LD interfaces, etc.)