Visualizing Library Metadata for Discovery
Myung-Ja K. Han, Stephanie R. Baker, Peiyuan Zhao, Jiawei Li
{mhan3; srbaker; pzhao12; jiaweil3}@illinois.edu
University of Illinois at Urbana-Champaign Library

Research Questions
- What are the key set of information in catalog records that should be used for building entity relationships?
- What kind of entity relationships can be useful for users’ discovery activities?
- Will visualization help users’ discovery activities?

Sample Data
The data used for the project was extracted from a sample of 300,000 randomly selected library catalog records (from 7.4 million total bibliographic records and 11 million holdings records) as shown in the table below. The selected data was known as access points and most values are added with controlled vocabularies or codes, except titles.

<table>
<thead>
<tr>
<th>Data</th>
<th>MARC data fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name (Agent)</td>
<td>100, 110, 111, 700, 710, 711</td>
</tr>
<tr>
<td>Title (Work)</td>
<td>130, 245, 246</td>
</tr>
<tr>
<td>Bibliographic record identifier</td>
<td>001 (Local bibliographic record ID)</td>
</tr>
<tr>
<td>Subject</td>
<td>050, 082</td>
</tr>
<tr>
<td>Date</td>
<td>260 $c or 264 $c</td>
</tr>
<tr>
<td>Holdings record identifier</td>
<td>004 (Local holdings record ID)</td>
</tr>
</tbody>
</table>

User Interface
The new prototype discovery service supports only two simple search options, work and name, with the assumption that users will perform a ‘known item search,’ as found in Mischo, Schlembach, and Norman’s research (2015). The search results page displays related resources by visualizing relationships between entities.

Search Results Display
Work (Title) Search
If a user starts the search with a title, then the result page allows the user to browse related resources (works) by the same author, on the same topic, with the same publication date, or having the same holdings library(ies).

Name Search
If a user starts the search with a name, the search results page allows the user to browse by relationships associated with the name, such as works created by the name, close collaborators, and subject areas of all works associated with the name.

Visualization Considerations
Instead of displaying number of items with the same information as a list, as in faceted browsing service, the prototype visualization discovery service combines together the search results and displays them as one visualization. With such visualization, users can clearly see related items by subjects, collaborators by number of works, as well as works created by the name. This feature may help users to identify and select the resources better than in current search and discovery service interfaces.

Findings and Challenges
The prototype showed the benefits of using a selective set of data that is critical to discovery and visualization, as opposed to using all of the information included in an entire catalog record. However, this experimentation also showed that visualizing library catalog data is not that easy even with a small set of data from sample records. The challenges include:
- Inconsistent terms used in the records;
- Data quality;
- Granularity of data in certain data fields;
- Multiple data included in one data field.

Next Steps
As next steps, we will experiment with visualization of discovery in the following areas:
- Experiment with Library’s full catalog data to fully explore relationships between entities;
- Transform Library’s data to BibFrame 2.0 with entity reconciliation to test visualization;
- Identify an ideal set of bibliographic data to be used for the discovery services through user testing.

Building Relationships
Since visualizations work best for showing relationships between resources, we developed entity relationships between ‘work (title),’ ‘name’ and ‘subject.’ Those relationships are displayed through visualization that provide opportunities for users to understand, identify, and find related and similar resources in a more effective and organized manner.