2013-06-19

Searching for known items: comparing search results for books in Millennium (III) and Primo at Boston University Libraries

Ammerman, Jack

http://hdl.handle.net/2144/16118
Boston University
Searching for known items: comparing search results for books in Millennium (III) and Primo at Boston University Libraries

Introduction

In August 2011, the BU Libraries deployed BU Libraries Search (Primo) as its primary discovery interface. In January 2013, the Millennium OPAC was taken offline. Most of the response to the Libraries’ introduction of BU Libraries Search (Primo) as the primary discovery interface for the Libraries’ resources is positive. Some report problems, however, when searching for “known items.” Other libraries have received similar reports when they deployed Primo. The clearly stated goal in the implementation of Primo is to provide an information discovery tool that satisfies the requirements of graduate level research for both high precision and contextual recall in searching and discovery. Diminishing the ability to discover known items would be a serious problem.

The Office of Digital Initiatives and Open Access, responsible for configuring and maintaining Primo, launched a study to determine the nature and extent of the problem and to use the gathered data to assist in resolving any configuration or metadata problems that might interfere with discovery of known items. The focus of this initial study is on non-serial literature.

Steve R. Thomas (University of Tennessee, Knoxville) presented a session at the Ex Libris Users of North America annual meeting in Salt Lake City (ELUNA 2012) titled “Discovery verses Precision: A Comparison of Known Item Searching in Primo and Aleph” in which he described a study he did of known item searching in Primo compared to Aleph. Aleph is a traditional integrated library system (ILS) that provides an online public access catalog (OPAC) for users to search. In addition to allowing users to search for books in Aleph, the UTK Libraries extract the bibliographic records in Aleph and add them to Primo to make them searchable in Primo. Thomas generated a random sample of bibliographic records from Aleph to compare searching in the Aleph OPAC and in Primo. He reports, “Given the confidence level of this sample, these results are essentially the same. In other words, the Primo Title Exact Match search and the Primo default search are about equal when compared to the Aleph Title Browse search.”

2
Searching for known items  
June 19, 2013  

Using a similar methodology this study attempts to compare the effectiveness of searching for “known items” in Primo and the Millennium OPAC. Like the UTK Libraries, Boston University Libraries provided both the Millennium OPAC and Primo for a period of 16 months. During that period bibliographic records were regularly harvested from Millennium to be loaded in Primo. Millennium has since been decommissioned, leaving Primo as the sole discovery interface.

While this study does not attempt to exactly replicate the methodology of the Thomas study, the results do add weight to Thomas’ conclusions that searching for known items in Primo is as effective as searching in a traditional OPAC like Aleph. In this study, Primo consistently outperformed Millennium. Emerging questions are: what is significantly different about the user experience that results in the perception that known item searching is more difficult in Primo? And, how can the Libraries address the issue?

Methodology

In order to achieve a 95% confidence level with a 4% confidence interval, a simple random sample of 600 bibliographic records was generated from Millennium using the Millennium Create List function. The total population from which the sample was drawn includes 1,050,176 bibliographic records for books held in the BU Libraries collections. One of the criteria for drawing the sample was the presence of the 001 field in the marc record. In the BU Millennium server, the 001 field carries a system number, most often the OCLC number. 20,811 records from the total population have no 001 tag, and were thus excluded, reducing the total population from which a sample was drawn to 1,029,365 bibliographic records. The excluded records are primarily of three types: a) brief bibliographic records created during the acquisition process; b) brief records created by the course reserves department; and c) vendor records that contain no 001 field.

In order to create a random sample of bibliographic records for the comparison searches, I looked for a randomly generated string of four digits as an embedded string in the Millennium 001 field. The four-digit random number (1330) was generated using the random number generator at: http://www.random.org/. The criteria in Table 1 were used to create the list of bibliographic records for books, each containing the four-digit string embedded in the 001 field.

**Table 1: Millennium Create List search criteria**

<table>
<thead>
<tr>
<th>Field</th>
<th>Operator</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAT TYPE</td>
<td>=</td>
<td>a</td>
</tr>
<tr>
<td>BIB LEVEL</td>
<td>=</td>
<td>m</td>
</tr>
<tr>
<td>BCODE3</td>
<td>!=</td>
<td>s</td>
</tr>
<tr>
<td>MARC Tag 001</td>
<td>has</td>
<td>1330</td>
</tr>
</tbody>
</table>

Jack Ammerman  
Digital Initiatives & Open Access  
Boston University Libraries
Searching for known items
June 19, 2013

This process generated a list of 757 records, the first 600 of which were exported from Millennium as the sample to be searched. It should be noted that BU Libraries’ implementation of Primo includes several million additional records for EBooks that were never loaded into Millennium. Generating the list from Millennium insured that all of the bib records would reside in both systems.

Following are the fields exported from Millennium into a tab-delimited file:

- Bibliographic record number
- Author (100$a)
- Main title (245$a)
- Subtitle (245$b)
- Statement of Responsibility (245$c)
- Publication Information (260$a, 260$b, 260$c)

The extracted records were imported into a spreadsheet where the data were cleaned to resolve problems in the export of records with diacritic characters and other characters that would impede the search. Where non-printable characters occurred, the title and/or author fields were copied into the field in the spreadsheet from the Millennium OPAC. The records were then manipulated to create four columns that would hold values to be used as the search strings for the searches:

- Main title
- Main title + author
- Shortened title
- Shortened title + author

The “Main title” column, consisting of 245$a, remained unedited. It included initial articles and punctuation contained in the original bibliographic records.

The “Main title + author” column was created by concatenating the “Main title” field with the last name of the author taken from the “Author” column.

Extracting the first words from the Main title string created the “Shortened title” column. A function was used to calculate the position of the first space occurring after the 20th character. The left portion of the Main Title was extracted up to that position. In the event the Main Title was less than 20 characters in length, the full main title was used. The function used to created this value is:

```
LEFT(<MainTitle>,Find(“ “,<MainTitle>,20))
```

The “Shortened title + author” column was created by concatenating the Shortened Title field with the author’s last name from the Author field.
Searching for known items
June 19, 2013

A small number of records had a single word title or a title not distinct enough to be a usable search term. When the Shortened title lacked sufficient words from the Main title to be a usable (distinctive) search string, additional words were added by modifying the function to begin looking for the position of the first space falling after the 30th character position. Single word main titles were occasionally supplemented with one or more words from the subtitle. All of the data cleaning and preparation was accomplished prior to beginning any search comparisons.

For each bibliographic record, four searches were executed in Millennium OPAC and replicated in Primo by cutting and pasting the search string from these four columns into the search box. All of the searches were simple keyword searches except the Millennium search for the Main Title. This search was executed on the Millennium title index. If the desired result was viewable within the first five results displayed on the screen (the visible screen without scrolling), the search was deemed successful and coded with a value: 1. This value was incremented (up to a maximum value of 5) for each refinement of the search required to navigate to the desired result. Refinements included

- Selecting a facet
- Selecting to see all versions of a FRBR group
- Scrolling the page to view records not visible in the first five results (incremented once for each five records)
- Navigating to the next or previous page
- Typing an additional value (such as a date) in order to refine the search
- Removing an initial article or punctuation that prevented the success of the search.

In a few cases, the desired record could not be located with the maximum four refinements to the search. In this case the search was coded with a value: ‘7’ to indicate a failed search.

**Analysis of the Data**

**Boston University Libraries**

The coded results were analyzed in Microsoft Excel to generate descriptive statistics. Primo generally outperformed Millennium by a small amount in all the searches. Millennium compared most favorably to Primo in the Short Title + Author searches. Chart 1 compares the successful searches in all four types of searches. Most searches of all four types succeeded without requiring any refinement of the search. The lowest being 87% for the Millennium Main Title search, and the highest being 97% for the Primo Main Title + Author search.
**Chart 1: Successful searches in Millennium and Primo**

When these “successful” searches are combined with those requiring a single refinement (Success + 1 Click), the percent of successful searches ranges from 93% for the Short Title search in Millennium to 99% achieved by Primo in all except the Short Title search.

**Chart 2: Successful Searches + 1 Click Refinement**
Searching for known items
June 19, 2013

The number of successful searches in Primo exceeded successful searches in Millennium for the Title, Title+Author, and Short Title searches. It was one less than Millennium for the Short Title + Author search. When successful searches were combined with those that required one refinement, Primo exceeded Millennium’s performance in all four types.

The high success rate of searches naturally means that the percent of “failed” searches was very low, ranging from 0.17% for the Primo search of Shortened Titles + Author to 3.83% for the Millennium search of Shortened Titles. Primo had a lower number of failed searches than Millennium in each of the four categories.

The frequencies of occurrence for the results of the four types of searches are displayed in Table 2 below.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Main Title Score</th>
<th>Main Title + Author Score</th>
<th>Main Title + Author Score</th>
<th>Short Title Score</th>
<th>Short Title + Author Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Millennium</td>
<td>Primo</td>
<td>Millennium</td>
<td>Primo</td>
<td>Millennium</td>
</tr>
<tr>
<td>Failed</td>
<td>4</td>
<td>3</td>
<td>11</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>Success</td>
<td>521</td>
<td>566</td>
<td>572</td>
<td>579</td>
<td>526</td>
</tr>
<tr>
<td>1 Click</td>
<td>63</td>
<td>25</td>
<td>12</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>2 Clicks</td>
<td>11</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>3 Clicks</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>4 Clicks</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>n=</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>% Success</td>
<td>87%</td>
<td>94%</td>
<td>95%</td>
<td>97%</td>
<td>88%</td>
</tr>
<tr>
<td>% Success + 1 Click</td>
<td>97%</td>
<td>99%</td>
<td>97%</td>
<td>99%</td>
<td>93%</td>
</tr>
<tr>
<td>% Failed</td>
<td>0.67%</td>
<td>0.50%</td>
<td>1.83%</td>
<td>0.50%</td>
<td>3.83%</td>
</tr>
<tr>
<td>MEAN</td>
<td>1.148</td>
<td>1.062</td>
<td>1.039</td>
<td>1.035</td>
<td>1.132</td>
</tr>
<tr>
<td>STDEV</td>
<td>0.450</td>
<td>0.315</td>
<td>0.368</td>
<td>0.254</td>
<td>0.620</td>
</tr>
</tbody>
</table>

Jack Ammerman
Digital Initiatives & Open Access
Boston University Libraries
Searching for known items
June 19, 2013

Each of the pairs of searches was analyzed using a Paired-Sample T-Test. For Pair 1 (Title Search), there was a statistically significant difference favoring Primo. The mean score for Millennium was 1.19 (SD=6.32) and for Primo was 1.09 (SD=5.05). The mean difference between pairs was .095 (t=5.123; p = 0.000). For Pair 2 (Title + Author Search), there was a statistically significant difference favoring Primo. The mean score for Millennium was 1.15 (SD=0.837) and for Primo was 1.07 (SD=0.470). The mean difference between pairs was .083 (t =2.263, p = 0.024.) For Pair 3 (Shortened Title Search), there was a statistically significant difference favoring Primo. The mean score for Millennium was 1.36 (SD=1.220) and for Primo was 1.21 (SD=0.773). The mean difference between pairs was .152 (t =3.424, p = 0.001.) For Pair 4 (Shortened Title + Author Search), there was a statistically significant difference favoring Primo. The mean score for Millennium was 1.18 (SD=0.960) and for Primo was 1.07 (SD=0.361). The mean difference between pairs was .115 (t=2.749, p = 0.006.) The difference, however, is very slight.

These results suggest that though the difference between Millennium and Primo in successful searches is modest, it is statistically significant. Specifically, these data demonstrate that known item searching is more effective in Primo than in Millennium.

Table 3: Paired Samples Test

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>Lower</th>
<th>Upper</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1 TSM &amp; TSP</td>
<td>0.095</td>
<td>0.454</td>
<td>0.019</td>
<td>0.059</td>
<td>0.131</td>
<td>5.123</td>
<td>599</td>
<td>0.000</td>
</tr>
<tr>
<td>Pair 2 TAM &amp; TAP</td>
<td>0.083</td>
<td>0.902</td>
<td>0.037</td>
<td>0.011</td>
<td>0.156</td>
<td>2.262</td>
<td>599</td>
<td>0.024</td>
</tr>
<tr>
<td>Pair 3 STM &amp; STP</td>
<td>0.152</td>
<td>1.085</td>
<td>0.044</td>
<td>0.065</td>
<td>0.239</td>
<td>3.424</td>
<td>599</td>
<td>0.001</td>
</tr>
<tr>
<td>Pair 4 STAM &amp; STAP</td>
<td>0.115</td>
<td>1.025</td>
<td>0.042</td>
<td>0.033</td>
<td>0.197</td>
<td>2.749</td>
<td>599</td>
<td>0.006</td>
</tr>
</tbody>
</table>

University of Tennessee Knoxville Libraries

The data gathered in the University of Tennessee Libraries study were gathered in a similar fashion, though searches were done on the Aleph Title Browse Index, the Title Keyword Index, and General Keyword indexes to compare with searches in Primo using an Advanced Search for Title (exact), Title (contains), and the simple Keyword index. Thomas coded the UTK Libraries data differently. Using the same coding criteria used for the BU Libraries study, I recoded the UTK Libraries data. When the coding was normalized to be consistent with the data gathered from Boston University Libraries, the results were similar, though Aleph consistently outperformed Primo for “successful” searches, those requiring no refinement. (Chart 3) Like the data from Boston University, most searches were successful
without refinement with a range from 83% for Primo Keyword searches to 97% for Aleph Title Browse index searches.

**Chart 3: Successful Searches in Aleph and Primo**

Combining the “successful” searches with those requiring a single refinement (1 click), Primo outperformed Aleph in the Title Keyword and the General Keyword searches. Success rates ranged from 88% for the Aleph General Keyword search to 99% for the Aleph Title Browse search. (Chart 4)

**Chart 4: Successful Searches + 1 Click Refinement**
Searching for known items  
June 19, 2013

The frequencies of occurrence for the results of the three types of searches are displayed in Table 2 below.

Table 4: Descriptive Statistics for UTK Libraries

<table>
<thead>
<tr>
<th>Measure</th>
<th>Aleph Title Browse</th>
<th>Primo Title (exact)</th>
<th>Aleph Title Keyword</th>
<th>Primo Title</th>
<th>Aleph General Keyword</th>
<th>Primo Keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed</td>
<td>3</td>
<td>8</td>
<td>41</td>
<td>13</td>
<td>70</td>
<td>25</td>
</tr>
<tr>
<td>Success</td>
<td>636</td>
<td>585</td>
<td>582</td>
<td>566</td>
<td>550</td>
<td>541</td>
</tr>
<tr>
<td>1 Click</td>
<td>10</td>
<td>50</td>
<td>23</td>
<td>56</td>
<td>23</td>
<td>45</td>
</tr>
<tr>
<td>2 Clicks</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>3 Clicks</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>12</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>4 Clicks</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>n=</td>
<td>653</td>
<td>653</td>
<td>653</td>
<td>653</td>
<td>653</td>
<td>653</td>
</tr>
<tr>
<td>% Success</td>
<td>97%</td>
<td>90%</td>
<td>89%</td>
<td>87%</td>
<td>84%</td>
<td>83%</td>
</tr>
<tr>
<td>% Success + 1 Click</td>
<td>99%</td>
<td>97%</td>
<td>93%</td>
<td>95%</td>
<td>88%</td>
<td>90%</td>
</tr>
<tr>
<td>% Failed</td>
<td>0.46%</td>
<td>1.23%</td>
<td>6.28%</td>
<td>1.99%</td>
<td>10.72%</td>
<td>3.83%</td>
</tr>
<tr>
<td>MEAN</td>
<td>1.057</td>
<td>1.190</td>
<td>1.447</td>
<td>1.282</td>
<td>1.715</td>
<td>1.464</td>
</tr>
<tr>
<td>STDEV</td>
<td>0.458</td>
<td>0.770</td>
<td>1.515</td>
<td>0.972</td>
<td>1.866</td>
<td>1.300</td>
</tr>
</tbody>
</table>

Each of the pairs of searches was analyzed using a Paired-Sample T-Test. For Pair 1 (Title Index Search), there was a statistically significant difference favoring Aleph. The mean score for Aleph was 1.060 (SD=0.458) and for Primo was 1.19 (SD=0.770). The mean difference was -.133 (t=-5.041, p = 0.000). For Pair 2 (Title Keyword), there was a statistically significant difference favoring Primo. The mean score for Aleph was 1.45 (SD=1.515) and for Primo was 1.28 (SD=0.972). The mean difference was .165 (t=3.710, p = 0.000). For Pair 3 (General Keyword), there was a statistically significant difference favoring Primo. The mean score for Aleph was 1.72 (SD=1.866) and for Primo was 1.46 (SD=1.300). The mean difference was .251 (t=4.818, p = 0.000).

These results suggest that though the difference between Aleph and Primo in successful searches is modest, it is statistically significant. Specifically, these data demonstrate that known item searching is more effective in Aleph than in Primo when using the Aleph Title Browse Index. With both keyword searches, however, Primo excelled.
### Table 3: Paired Samples Test

<table>
<thead>
<tr>
<th>Paired Samples Test</th>
<th>Paired Differences</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Deviation</td>
</tr>
<tr>
<td>Pair 1 ATB &amp; PTE</td>
<td>0.133</td>
<td>0.675</td>
</tr>
<tr>
<td>Pair 2 ATK &amp; PT</td>
<td>0.165</td>
<td>1.139</td>
</tr>
<tr>
<td>Pair 3 AGK &amp; PK</td>
<td>0.251</td>
<td>1.332</td>
</tr>
</tbody>
</table>

### Discussion

**Comparing BU Libraries to UTK Libraries**

It is clear from the two studies that Primo performs well when compared to a traditional OPAC, though clearly Aleph Title Index searching outperformed Primo. The methodologies from the two studies are not identical making close comparison impossible. The need for recoding the data from the UTK Libraries study may have introduced problems as well. Though recoding was done consistently, it is seems problematic to do direct comparison of the two data sets.

In addition to the different methodologies, two important factors may explain why the BU Primo implementation fared better than the UTK Primo implementation. The two studies were done approximately a year apart. The UTK Libraries study was done when they were using Primo version 3.x. The BU Libraries study was done using Primo version 4.2. During the year, Ex Libris was working to improve performance in Primo including relevancy ranking. For known item searching, it is not clear how significant this might be. It would be interesting to perform the study on two Primo servers running the same version using the same from each study.

The normalization rules and configuration for the UTK Libraries’ and the BU Libraries’ Primo servers are not the same. One of the valued features of Primo is the ability to customize the normalization rules to best serve the discovery needs of the local library community. BU Libraries spent several months fine-tuning the normalization rules to work with the Millennium records and has since adjusted the normalization rules to work with Ex Libris Alma records. The ability to customize these normalization rules introduces a complicating factor in making a direct comparison.
Differing Search Strategies

Both studies do point to Primo’s strong performance in known item searching. Perhaps less obvious from the results is that the search strategies for searching a traditional OPAC and for searching Primo are different. Traditional OPAC searching assumes that one has already limited the search to those items (often physical but not exclusively) held in the library. The Millennium OPAC included 1,050,176 bibliographic records for books held by the libraries. The librarians have for years instructed students and faculty in effectively structuring searches, using defined fields, to effectively utilize the very good indexing provided by OPACs.

Direct comparison with Primo is difficult, but a search on the keyword “history” alone returned 896,935 books. In addition, it returned items of other resource types displayed in Table 4. The number of results for each material type appears in parentheses. This is obviously not a full count of all of the records in Primo, only the result of a single search.

Table 4: Results by Resource Type

<table>
<thead>
<tr>
<th>Resource Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articles</td>
<td>3,538,727</td>
</tr>
<tr>
<td>Reviews</td>
<td>1,107,783</td>
</tr>
<tr>
<td>Books</td>
<td>896,935</td>
</tr>
<tr>
<td>Text Resources</td>
<td>398,120</td>
</tr>
<tr>
<td>Journals</td>
<td>257,511</td>
</tr>
<tr>
<td>Conference Proceedings</td>
<td>163,648</td>
</tr>
<tr>
<td>Images</td>
<td>20,195</td>
</tr>
<tr>
<td>Newspaper Articles</td>
<td>19,165</td>
</tr>
<tr>
<td>Reference Entries</td>
<td>18,825</td>
</tr>
<tr>
<td>Dissertations</td>
<td>12,556</td>
</tr>
<tr>
<td>Maps</td>
<td>7,209</td>
</tr>
<tr>
<td>Other</td>
<td>2,028</td>
</tr>
<tr>
<td>Scores</td>
<td>1,045</td>
</tr>
<tr>
<td>Audio Visual</td>
<td>587</td>
</tr>
<tr>
<td>Legal Documents</td>
<td>453</td>
</tr>
<tr>
<td>Databases</td>
<td>437</td>
</tr>
<tr>
<td>Web Sites</td>
<td>3</td>
</tr>
</tbody>
</table>

The “simple search” box is a primary search interface for Primo. It is “Google-like” in its simplicity, requiring no understanding of the various fields in the bibliographic records or how they might be indexed. Rather, it relies on a relevancy-ranking algorithm to surface the most “relevant” records to the top of the search results list.
Searching for known items
June 19, 2013

In addition, a set of facets is displayed in the margin to allow the searcher to refine or filter the results set to develop a manageable sized set.

Searchers that use traditional structured search strategies effective in OPACs may be confused by the simple search approach used by Primo. Even though the data from these two studies indicates that Primo effectively displays the desired result for known items in the first screen, searchers accustomed to searching traditional OPACs may also be confused by the display of a large results set.

Conclusions & Questions for Investigation

It is clear that known item searches in Boston University's implementation of Primo are more likely to be successful than the same search in Millennium OPAC. The staff of the libraries has worked effectively to resolve metadata problems, adjust the normalization rules, and tweak the relevance ranking for local data sources. Though the results from the study at the University of Tennessee Knoxville Libraries do not quite as strongly favor searching in Primo, known item searching in Primo does compare well to searching in Aleph.

Despite data that indicate known item searching in Primo is more effective than the same searches performed in Millennium OPAC, the perception exists among some library staff and users that they are not able to easily find known items in Primo. The results of this study do not negate this perception. Discerning more clearly the issues that contribute to the perception is an area for continued investigation. Several possibilities might be explored.

1. Is there a correlation between search strategies (traditional vs. simple search) and the perception that one is finding or not finding known items in Primo? Do users perceive that they don't find known items more when they use the advanced search form with fielded search than when they use a simple search box?

2. The Advanced Search form allows the user to select a group of databases with a disciplinary focus (a Metalib Quickset) to search. Selecting one of these Quicksets includes results from only those databases. Are users that can't find known items selecting a Quickset and unknowingly filtering out the records with the item they seek?

3. When users perceive that they have not found the item for which they are searching, are they confused by the search interface? Are they confused by the display of results? How can these interfaces be improved?

4. Would additional training assist in reducing the number of users who perceive they have not been able to find the item for which they are looking? If so, what form of training would be most effective?

5. Are there particular user groups that experience difficulty in known item searching? If so, are there disciplinary differences? Age differences?
Searching for known items
June 19, 2013

Education level differences? Are they more or less successful when they have received library instruction in searching?

6. Though Primo outperformed Millennium based on this random sample, there may be particular kinds items that were not well represented in the sample that are difficult to discover. If so, what are they? Will improving the metadata make them more discoverable? Can discovery be improved for these types of items by modifying the Primo configuration or normalization rules?

Finally, there are several areas for future study. The BU Libraries recently implemented Primo’s new “Browse Search” functionality. It is now possible to search an author, title, subject, or call number browse index. Studying the effectiveness of this new search mode, particularly in comparison to the results from this study will be helpful.

This study did not explore the effectiveness of the Primo “Advanced Search.” This fielded search resembles the structured search strategies that are used in traditional OPACs. Numerous anecdotal reports indicate frustrating or unexpected results when using the Advanced Search form. A comparison study of the same random set of titles using the advanced search form would be helpful.

This study focused on known item searching. Subject and keyword searching is also used in traditional OPACs for discovery. A study that assessed the strengths and weaknesses of OPACs and discovery systems like Primo for discovery searching would be useful. The methodology for such a study would not be as straightforward as this one for known item searching. Judging what is missing from a result set can be difficult. This might be accomplished, though with the new Browse Search functionality.

NOTES

1 By “known items” we mean that the search is for a specific item, a book, for example, for which the user already has a full or partial citation. This is as opposed to what might be called a discovery search in which the user is searching for a subject or topic in order to discover sources that are as yet unknown to her.


3 The Millennium Main Title Search required refinement most often because of its uneven handling of initial articles and occasionally punctuation in the title. English language initial articles were generally handled without a problem. Initial articles in other languages often failed.
4 Thomas coded the position of the desired bibliographic record in the results list for Aleph and Primo. Based his recorded values, I recoded the UTK Libraries data to be comparable to the BU Libraries data.