2013

User-Centered Evaluation of a Discovery Layer System with Google Scholar

Tao Zhang
Purdue University, zhan1022@purdue.edu

Follow this and additional works at: http://docs.lib.purdue.edu/lib_fsdocs

Part of the Library and Information Science Commons

Recommended Citation

User-Centered Evaluation of a Discovery Layer System with Google Scholar

Tao Zhang

Purdue University Libraries
504 W. State Street, West Lafayette, IN 47907-2058, USA

zhan1022@purdue.edu

Abstract. Discovery layer systems allow library users to obtain search results from multiple library resources and view results in a consistent format. The implementation of a discovery layer is expected to simplify users' workflow of searching for scholarly information. Previous studies on discovery layer systems focused on functionality and content, but not quality of search results from the user's perspective. The objective of this study was to obtain users' assessment of search results of a discovery layer system (Ex Libris Primo®) and compare that with a widely used scholarly search tool (Google Scholar). Results showed that Primo's search results relevancy is comparable to Google Scholar, but it received significantly lower usability and preference ratings. A number of usability issues of Primo were also identified from the study. Results of the study are used to improve the interface of Primo and adjust relevancy ranking options. The empirical method of search results assessment and feedback collection used in this study can be extended to similar user-centered system implementation and evaluation efforts.

Keywords: Discovery layer, Google Scholar, search results, relevance, user-centered evaluation

1   Introduction

Discovery layer systems are web-based tools that search in a unified index of metadata from article databases, library catalogs, digital repositories, digital collections, and other scholarly information resources [1, 2]. Since the index is pre-harvested and centralized, the response time of discovery layer systems is significantly improved from previous tools built on federated search technology. Within a discovery layer's interface, search results are usually displayed in a consistent format. Users submit one search query and obtain search results from heterogeneous resources available from the library, which simplifies users' workflow and facilitate the utilization of those resources. Examples of discovery layer systems are Summon by Serials Solutions, WorldCat Local by OCLC, EBSCO Discovery Service by EBSCO, Primo by Ex Libris, and Encore Synergy by Innovative Interfaces.

Google Scholar was launched as a beta service in November 2004. Despite some skepticism, it has been widely accepted by researchers and students as an important scholarly search tool due to its easy to use interface, extensive content coverage, and highly relevant search results. A number of studies compared Google Scholar with traditional library databases in terms of coverage of discipline literature [3], the scholarliness of search results [4, 5], depth and breadth of coverage [6], and search performance [7]. By using link resolver technique, Google Scholar also has an impact on how users interact with the library to obtain materials and their perception of scholarly information [8]. The implementation of discovery layer systems is libraries' effort of responding to users' need of a search tool similar to Google Scholar in terms of user experience and covering only authoritative resources from the libraries (Google Scholar does not reveal its scope of source materials).

Implementation and evaluation of discovery layer systems have been reported in the recent literature. Becher and Schmidt [9] tested two discovery layer systems (WorldCat Local and Aquabrowser) with students and identified a list of features preferred by participants, including links to full-text articles using a link resolver, results incorporating both articles and books, and facets like date, format, and subject. Fagan et al. [2] conducted usability test with students and faculty members on EBSCO Discovery Service. They raised several questions related to discovery layer systems and libraries, such as helping users understand the scope and purpose of discovery layer in order to choose between a discovery layer or a subject-specific database, as well as enabling users navigate between the discovery layer and other library services and resources. Williams and Foster [1] conducted another usability study of EBSCO Discovery Service focusing on users' information seeking behavior with a discovery layer system. They found that participants mainly examined the first page of search results and relied heavily on the facets to distinguish between types of materials. They also suggested a need of instruction and documentation for users to better utilize the discovery layer. Ward, Shadle, and Mofjeld [10] tested WorldCat Local and identified usability issues for improvement.

These previous studies examined the functionalities and scholarly quality of search results of discovery layer systems as well as user preferences of features, content, and information display. Researchers also reported usability study results of discovery layers based on subjective feedback. However, the integration of empirical test and usability study has not been well established in a user-centered system implementation process. To fill this gap, this study compared users' subjective assessment of search results from a discovery layer system (Primo) and Google Scholar. Users performed searches on both systems and reported usability issues during the evaluation. The results not only provided quantitative data for system comparison, but also qualitative evidence for further improvement of the discovery layer.

2   Method

2.1   Participants

The user evaluations were conducted with twelve graduate students (nine females and three males) at Purdue University who reported good experience of scholarly search with various systems. Participants ranged in age from 22 to 35 (*Mean* = 26.3, *SD* = 3.2). Descriptive statistics of participant background experience are listed in Table 1. Each item in Table 1 was measured in 5-point Likert scale from 1 (not at all) to 5 (very much).

Table 1. Summary of participant background experience.

Experience with …	*Min.*	*Max.*	*Mean*	*SD*
Finding library books on library website	3	5	4.4	0.65
Finding articles on library website	4	5	4.5	0.52
Finding course reserves on library website	2	5	3.5	0.97
Scholarly databases like Web of Science and Academic Search Premier	1	5	3.8	1.09
Google Scholar	3	5	4.6	0.65

2.2   The Ex Libris Primo® Discovery Layer

The interface of Primo is shown in Fig. 1. The interface consists of several functional areas: (1) top links to other library services; (2) search input box and options under the top links; (3) facets on the left column of the page for refining search results; and (4) list of search results showing title, author, and publication information. The detailed display of a search result is shown in Fig. 2. For each search result with "Full text available" label, the user can click on its title or the "View Online" link to open the full text article, or click on the "Details" link to view metadata including title, author, subjects, journal ("is part of"), abstract (description), and identifier (e.g., ISBN). The interface also highlights any matching keywords in the search results.

2.3   Tasks

Participants performed a set of four searches on both Primo and Google Scholar. They were given keywords as topics for course papers for the first three searches and created their own search keywords for the fourth search. The keywords given to participants for the first three searches were: classroom assessment techniques, food addition, and natural language processing. The keywords were chosen to represent typical topics in social science, life science, and engineering. For each search, participants examined the first ten results returned by Primo and Google Scholar and gave each result a relevance rating ranging from 1 (highly irrelevant) to 7 (highly relevant). Participants were also encouraged to ask questions and report anything they were not clear of.
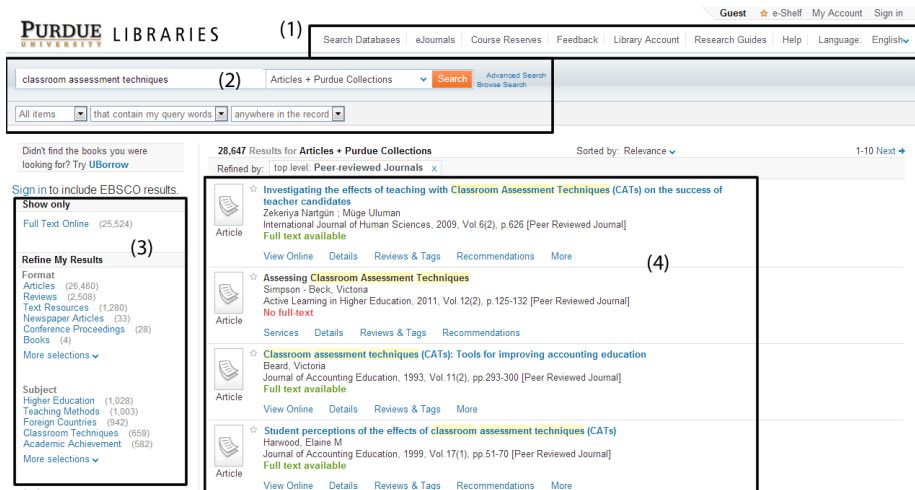
Fig. 1. Interface of Ex Libris Primo®.
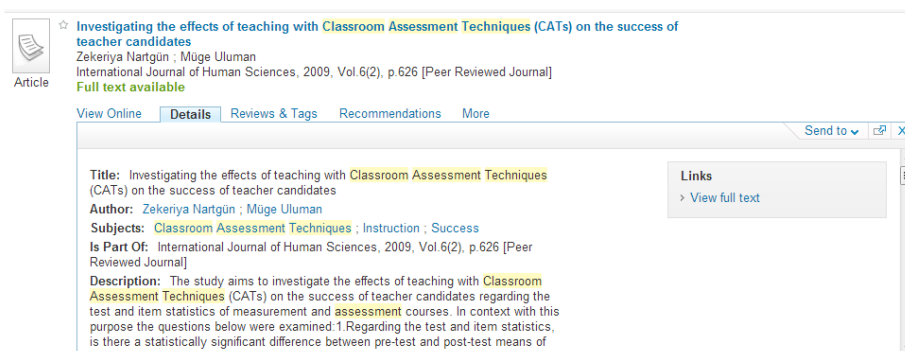
Fig. 2. Detailed display of a search result.

2.4 Experiment Design and Measures

This study used a within-subject experiment design, in which participants worked with both Primo and Google Scholar on the same set of four searches. Response measures included: (1) participant ratings of the relevancy of search results; (2) participant usability ratings of Primo and Google Scholar, measured by System Usability Scale [11]; (3) participant preference ratings of the two systems; (4) questions and comments from participants during the evaluation; and (5) participant responses to open questions in the final questionnaire (the most positive and negative aspects of Primo and Google Scholar).

2.5　Procedure

Participants first signed a consent form and completed a demographic survey regarding their experiences of scholarly search. They first conducted four searches using either Primo or Google Scholar and moved on to the other system. The order of testing Primo and Google Scholar was balanced across participants. The order of the four searches with a system was randomized. Participants used their own keywords for one of the four searches with each system. After participants finished all searches with a system, they rated that system's usability in SUS. After all search tasks, participants gave a preference rating for each system based on their overall experience and responded to the open questions in the final questionnaire. Each evaluation session lasted about 1.5 hours.

3　Results

3.1　Comparison between Primo and Google Scholar

The descriptive statistics of relevancy ratings of search results (ranged from 1 to 7) from the four searches are shown in Table 2, 3, 4, and 5. The SUS (measured by 5-point Lickert Scale) and system preference (ranged from 1 to 10) ratings are shown in Table 6.

Table 2. Descriptive statistics for relevancy ratings of search by keyword: classroom assessment techniques.

Article	Primo				Google Scholar			
	Mean	*SD*	*Min.*	*Max.*	*Mean*	*SD*	*Min.*	*Max.*
1	6.67	0.49	6	7	6.75	0.62	5	7
2	6.33	1.23	3	7	6.50	0.80	5	7
3	6.42	0.51	6	7	6.42	1.00	4	7
4	6.50	0.67	5	7	6.08	1.00	4	7
5	5.92	1.62	1	7	6.25	1.06	4	7
6	5.83	1.40	3	7	6.33	1.44	2	7
7	5.58	1.78	2	7	6.50	0.80	5	7
8	5.50	1.00	4	7	6.25	0.75	5	7
9*	4.67	1.72	2	7	6.25	1.22	3	7
10*	4.83	2.29	1	7	6.83	0.39	6	7
Average	5.83	0.72	4.5	6.9	6.42	0.45	5.7	7

An one-way ANOVA (Analysis of Variance) showed that participants gave significantly higher relevancy ratings of search results from Google Scholar than Primo, but only for one or two articles from each search (Article 9 and 10 for first search: $F(1, 11) = 8.09$, $p = 0.0160$ and $F(1,11) = 9.78$, $p = 0.0096$, respectively), Article 10 for second search ($F(1, 11) = 8.19$, $p = 0.0155$), Article 3 for third search ($F(1, 11) = 5.69$, $p = 0.0362$), and Article 3 for the fourth search ($F(1, 11) = 10.17$, $p = 0.0086$). The corresponding descriptive statistics are marked with star in Tables 2-5. There was

a significant difference of average relevancy ratings between the keyword-given search and participants' own search ($F(3, 77) = 9.17$, $p < 0.0001$). The post-hoc results showed that participants gave higher average relevancy ratings with the keyword-given search (*Mean* = 6.17 for first keyword search, *Mean* = 6.13 for second keyword search, *Mean* = 6.12 for third keyword search) than their own search (*Mean* = 5.31). Another ANOVA showed that participants gave significantly higher usability ratings and preference scores of Google Scholar than Primo ($F(1, 11) = 24.94$, $p = 0.0004$ and $F(1, 11) = 9.27$, $p = 0.0112$). Usability ratings were positively correlated with preference ratings ($r = 0.60$, $p = 0.0021$) and average relevancy ratings of search results ($r = 0.59$, $p = 0.0023$).

Table 3. Descriptive statistcs for relevancy ratings of search by keyword: food addition.

Article	Primo				Google Scholar			
	Mean	*SD*	*Min.*	*Max.*	*Mean*	*SD*	*Min.*	*Max.*
1*	6.58	0.67	5	7	6.92	0.29	6	7
2	6.42	0.79	5	7	6.67	0.49	6	7
3	5.92	1.08	4	7	6.25	1.71	1	7
4	5.33	1.67	2	7	6.17	1.03	4	7
5	5.92	0.90	4	7	5.75	1.60	1	7
6	5.92	1.16	4	7	6.25	0.75	5	7
7	6.33	0.65	5	7	6.25	0.87	5	7
8	6.17	0.94	4	7	6.25	0.75	5	7
9	6.17	0.83	5	7	6.33	0.78	5	7
10*	5.25	1.71	2	7	6.58	0.67	5	7
Average	6.00	0.59	4.9	7	6.34	0.50	5.7	7

Table 4. Descriptive statistcs for relevancy ratings of search by keyword: natural language processing.

Article	Primo				Google Scholar			
	Mean	*SD*	*Min.*	*Max.*	*Mean*	*SD*	*Min.*	*Max.*
1	6.42	1.44	2	7	6.75	0.45	6	7
2	6.50	0.80	5	7	6.75	0.45	6	7
3*	5.42	1.68	1	7	6.67	0.65	5	7
4	6.00	1.41	2	7	6.83	0.39	6	7
5	5.58	1.88	2	7	6.50	0.80	5	7
6	6.58	0.79	5	7	6.42	0.79	5	7
7	6.08	1.16	3	7	6.00	1.48	2	7
8	4.58	2.27	1	7	5.92	1.16	3	7
9	6.00	1.21	3	7	6.50	0.52	6	7
10	5.50	1.68	2	7	5.50	2.02	1	7
Average	5.87	0.74	4.5	6.9	6.38	0.57	5.2	7

Participants answered open questions in the final questionnaire regarding the most positive and negative aspects of Primo and Google Scholar. Positive aspects of Primo mentioned by participants include: facets for refining results are very helpful; display

of search results is logical and easy to understand; highlight of search keywords; detailed information with abstract is easy to find; and search results are up to date. Negative aspects of Primo include: redundant search results from different sources; Primo does not show good results with highly specific keywords; it does not show preview of articles on the search results display; some article links are not available; participants have to click through the link resolver interface to see the full text article; and relevancy of results rely more on the title of materials than Google Scholar. For Google Scholar, the positive aspects mentioned by participants include: easy to use interface and search preferences similar to Google web search; easy to access full text articles; links for each search result include other articles citing this article, citation information, and importing citation to reference tools; the ability to filter results by date; and the results display is visually pleasing. For negative aspects of Google Scholar, participants mentioned less scholarly and less relevant articles in search results, inability to filter out books or refine results, inability to sort results by date or number of citations; and full text articles sometimes are not available.

Table 5. Descriptive statistcs for relevancy ratings of search by participants' own keywords.

	Primo				Google Scholar			
Article	*Mean*	*SD*	*Min.*	*Max.*	*Mean*	*SD*	*Min.*	*Max.*
1	6.25	1.06	4	7	6.50	1.17	3	7
2	5.42	1.56	3	7	6.08	1.16	4	7
3*	4.17	2.17	1	7	6.50	0.90	4	7
4	4.50	2.15	1	7	5.83	1.11	3	7
5	5.25	1.91	2	7	5.75	1.48	3	7
6	5.67	1.30	3	7	5.75	0.87	5	7
7	4.08	2.31	1	7	5.17	1.64	2	7
8	4.17	2.04	1	7	5.75	1.54	2	7
9	4.58	2.15	1	7	5.50	1.88	1	7
10	3.75	1.82	1	7	5.50	1.83	1	7
Average	6.25	1.06	4	7	5.83	0.67	4.5	6.9

Table 6. Descriptive statistics for usability and preference ratings.

	Primo				Google Scholar			
	Mean	*SD*	*Min.*	*Max.*	*Mean*	*SD*	*Min.*	*Max.*
Average SUS rating	3.59	0.43	2.74	4.21	4.27	0.29	3.74	4.79
Preference rating	7.08	1.38	5	9	8.33	0.65	7	9

3.2 Usability Issues of Primo

Comments from participants indicated that they had more usability concerns of Primo than Google Scholar. With Primo, participants had to go through the link resolver interface (Fig. 3) in order to access the article page, which is one more step compared to Google Scholar's workflow. Google Scholar allows users to click on the article title

in the search results and directly access the article page. This could partly contribute to higher usability ratings of Google Scholar than Primo. Because Primo indexes multiple databases and an article may be available from more than one database, it may be necessary to show a link resolver interface to let users know that the article can be accessed from multiple databases. However, comments from participants revealed that they did not fully understand the link resolver interface. Most participants would choose the article link from the database they know from the list of available sources, or randomly choose one link that works for them. The link resolver interface essentially forced participants to make an unnecessary choice they are not familiar with. Ideally, the Primo system should be able to handle multiple source links and choose the "best" link for users.

Title: Impact of food and predation on the snowshoe hare cycle.
Source: Science [0036-8075] Krebs, Charles yr:1995 vol:269 iss:5227 pg:1112

Electronic Full Text

📄 Full Text Online Gale Biography In Context

📄 Full Text Online JSTOR Life Sciences Collection

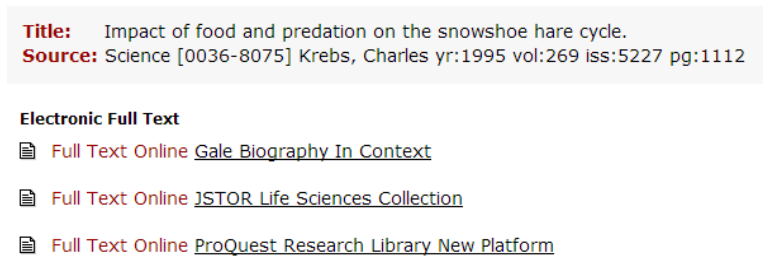📄 Full Text Online ProQuest Research Library New Platform

Fig. 3. The link resolver interface associated with Primo.

There were a few consistency issues of Primo based on participant comments and questions. In the search results display, Primo displays action links like "Request" for books, "View Online" for articles, and both links for journals. Participants initially did not understand this inconsistency caused by different types of materials. When a search result has multiple versions, its title is not clickable and Primo displays a link labeled as "Click here to view 2 versions" under the title. This was confusing for participants because they expected to click on the title to access a search result. Google Scholar displays an "All 2 versions" link at the bottom of the result and clicking on the title goes to one version of the article, which does not break the consistency and still gives users the option to view other versions. Another consistency issue is the use of text colors in Primo interface. Text label such as "Available at …" and "Full text available" is in green color and participants tend to see them as clickable links. In Primo, users can save the current search in the system or as a RSS feed, but the two links for saving query and RSS are located in the bottom left corner of the interface and thus difficult to find. Other search tools including Google Scholar usually place the save search link close to the search input box, which is much more prominent.

4   Discussion and Conclusion

The goal of this study was to gather data about user subjective assessment of search results from a discovery layer system (Ex Libris Primo) and identify usability issues from system implementation. The resulted task design did not support measuring task

completion time as part of response measures, since participants were encouraged to "think aloud" during the evaluation. Instead, the statistical comparison of relevancy ratings showed that the Primo's search effectiveness in terms of perceived relevancy is on par with Google Scholar. Participants gave higher average relevancy ratings in searches when they were given keywords than searches using their own keywords. This result revealed the individual differences of participants making relevancy judgment when they conducted their own searches. More importantly, usability issues of Primo has affected users' preference and perceived result relevancy, as the significant correlations between usability and preference and relevancy ratings suggested. Since Primo was being implemented at the time of the evaluation, this study did not examine the end-user features available in Primo for organizing and exporting results, user ratings and reviews, integration with library accounts, and sharing with external sites (e.g., Facebook). These features potentially could create a more engaging experience for users, which may affect users' preference ratings.

Usability issues identified from the evaluation showed the importance of a smooth workflow for discovery layer systems. Because of the similar perceived relevancy of Primo and Google Scholar, participants paid more attention on the mechanism of accessing full text articles, instead of what is searched or covered by both systems. Although it is still important for libraries to help users understand the content differences of discovery layers and library databases, the integration of library tools and services (in this case the discovery layer and link resolver) is critical to ensure a satisfactory user experience leading to system acceptance. Inconsistency of user interface is a primary cause of usability issues identified in this study. A discovery layer's interface should conform to common design practices in other search tools (e.g., Google Scholar) so that users are able to transfer their experiences of other systems to the discovery layer. Since participants considered the facets of Primo as an important advantage, future studies of discovery layer systems could also focus on users' information seeking behavior with facets through empirical observation or transaction log analysis [12].

The pervasive use of Google Scholar provides challenges and opportunities for libraries to implement better search tools with easy to use interface and authoritative content coverage. Discovery layer systems appear promising to meet this goal because they integrate various information resources from the library and they provide a much more simplified interaction paradigm than previous generation of federate search tools. This study provided empirical evidence of the utility of the discovery layer in terms of both perceived relevancy of search results and qualitative user feedback. The comparison of the discovery layer system (Ex Libris Primo) with Google Scholar served as a benchmark test in realistic task settings, which is different from previous studies focusing on content coverage and search functionalities of discovery layers and Google Scholar. As discovery layer systems play an important role in bringing back users to library-centered scholarly search experience, the user-centered methodology in this study serves as a basis for similar system design and evaluations. The discovery layer system must be designed carefully to meet user needs and expectations, while maintaining a similar experience of other search tools within the library information systems.

References

1. Williams, S.C., Foster, A.K.: Promise Fulfilled? An EBSCO Discovery Service Usability Study. Journal of Web Librarianship. 5, 179–198 (2011).
2. Fagan, J.C., Mandernach, M., Nelson, C.S., Paulo, J.R., Saunders, G.: Usability Test Results for Discovery Tool in an Academic Library. Information Technology and Libraries. 31, 83–112 (2012).
3. Meier, J.J., Conkling, T.W.: Google Scholar's Coverage of the Engineering Literature: An Empirical Study. The Journal of Academic Librarianship. 34, 196–201 (2008).
4. Howland, J.L., Wright, T.C., Boughan, R.A., Roberts, B.C.: How Scholarly is Google Scholar? A Comparison to Library Databases. College & Research Libraries. 70, 227–234 (2009).
5. Callicott, B., Vaughn, D.: Google Scholar vs. Library Scholar: Testing the Performance of Schoogle. Internet Reference Services Quarterly. 10, 71–88 (2005).
6. Neuhaus, C., Neuhaus, E., Asher, A., Wrede, C.: The Depth and Breadth of Google Scholar: An Empirical Study. portal: Libraries and the Academy. 6, 127–141 (2006).
7. Walters, W.H.: Google Scholar Search Performance: Comparative Recall and Precision. portal: Libraries and the Academy. 9, 5–24 (2009).
8. Herrera, G.: Google Scholar Users and User Behaviors: An Exploratory Study. College & Research Libraries. 72, 316–330 (2011).
9. Becher, M., Schmidt, K.: Taking Discovery Systems for a Test Drive. Journal of Web Librarianship. 5, 199–219 (2011).
10. Ward, J.L., Shadle, S., Mofjeld, P.: User Experience, Feedback, and Testing. Library Technology Reports. 44, 17–23 (2008).
11. Brooke, J.: SUS - A Quick and Dirty Usability Scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., and McClelland, A.L. (eds.) Usability Evaluation in Industry. Taylor & Francis, London (1996).
12. Mussell, J., Croft, R.: Discovery Layers and the Distance Student: Online Search Habits of Students. Journal of Library & Information Services in Distance Learning. 7, 18–39 (2013).