

# Relevancy Ranking Assessment Of A Discovery Tool

Satyabrata Sahoo<sup>1</sup>, Dr. Pijushkanti Panigrahi<sup>2</sup>

<sup>1</sup> Research Scholar, Department of Library and Information Science, University of Calcutta, Kolkata, India. ORCID: 0000-0002-8056-3692

<sup>2</sup>Professor, Department of Library and Information Science, University of Calcutta, Kolkata, India. ORCID: 0000 0002 5340 7512.

---

## Abstract

**Purpose-** The purpose of this paper is to assess relevancy ranking order provided against search term by a Discovery Tool (DT) and show relationship between the background of assessor and their assessment about the relevancy ranking order.

**Design/methodology/approach-** The relevant literature is reviewed minutely and the methodology has been designed in an apposite way. The search result found against a term assessed anonymously by hiding DT name and actual ranking order provided by the DT. Data collected in Likert Scale. Collected data analyzed by using statistical tool named SPSS. Respondents are the subject expert of the term used.

**Findings-** The relevance order provided by DT is mismatched with the calculated order found from the statistical analysis. On the other hand, it is found that there is a strong relationship between assessment and background of the assessor.

**Originality/ value-** Although a considerable amount of research has been done on evaluation of relevancy ranking order of DT but not so much done on Vu Find. The process used is unique and maintain a new way to assess the relevancy ranking of a DT to make the work anonymous and reduce biasness where assessor is representing the user group and investigator not involve in assessment.

**Keywords-** Discovery Tool (DT), Web-Scale Discovery (WSD), Relevancy Ranking (RR), Vu Find, Assessment of ranking order, Evaluation of ranking order, User Background, Relevancy Ranking Factors (RRF), OPAC, Next generation catalogue

**Paper type-** Research paper

## 1. Introduction

As the number of wild E-resources and library's local e-collections are growing day to day exponentially, an ardent need of effective single platform-based search facility like Web-Scale Discovery Tool to render exhaustive services to the users in simple, easy and fast way is felt. Discovery Tool (DT) is a next-generation catalogue that harvests data from OPACs and other sources, such as digital repositories, creating a single searchable index (Sadeh, 2008). The main aim of the tool is to enable library users to search and browse through all of the library's resources by replacing the traditional OPAC to include catalogue records, locally cached journals, digital library items, Institutional Repository (IR), Institutional Bibliography, other library collections and resources, etc. Discovery Tool displays results according to descending order of relevancy against input queries or search terms. But relevancy is perhaps a vague term as user preferences may vary across different contexts of time, understanding capability, purpose of use or need, academic background of users, research experience, subject knowledge, etc. Considering the fact that discovery systems are being implemented in libraries, one must know how well they perform in terms of relevancy ranking which depends on some pertinent factors. All these factors may be categorized into six different groups as (i) text statistics, (ii) popularity, (iii) freshness, (iv) locality and availability, (v) content properties and (vi) user background (Behnert and Lewandowski, 2015). There are so many standard methods of Information Retrieval (IR) evaluation approaches have been taken to evaluate library information systems proposed with respect to the web search context. Not so much of empirical studies are found in the relevancy ranking aspect in library and information system and the issues have not been evaluated as well with a greater perfection considering the user background. The present study is conducted towards exploring the effectiveness of a discovery tool in the relevancy ranking empirically. The actual user assesses the Library and Information System's results anonymously and tries to explore whether the user background has any relationship with relevancy ranking evaluation.

## 2. Review of Literature

Implementation and evaluation of discovery systems have been reported in recent literature. Large & Beheshti (1997) reviewed OPAC studies published between 1990 and mid-1996 concerning different methods used for OPAC research including interviews and observations of users conducting search tasks, think-aloud techniques, and transaction log analysis. They pointed out "users (age, gender, knowledge, etc), search tasks (different search types demand different methods and metrics), the OPAC or system (features, interfaces), the kind of library and the type of data collection (researchers collect their data in different ways)" etc variables have to consider carefully. Gordon and Pathak (1999) proposed seven parameters for evaluation of discovery systems to ensure highly accurate and informative results. According to them "Relevance judgments should be made by the actual user, not by the investigators." Lau and Goh (2006) observed an important correlation with increasing query length and corresponding search failure. Olson (2007) describes a study of Aqua Browser, to investigate the effect of faceted browsing on the scholarly research users and suggested that "if faceted navigation can better match consumers to the goods that they specifically are interested [in], perhaps libraries can expect similar results in matching scholars to the research materials that best fit their specific needs"(p. 551) Firstly VuFind ordered search results according to date as

relevancy order was not so improved, in May 2010 field weighting based relevancy ranking and advanced search features added to Vu Find (Denton and Coysh, 2011). Zhang (2013) compared the search performance of Google Scholar with Primo using an experimental design: twelve participants rated the relevance of results for four search tasks on a scale from 1-7, as well as the usability of the systems. It seems, however, that the source of the results was not anonymous, so the participants knew the systems being evaluated which, in turn, may have influenced ratings and the search terms are not taken from controlled vocabulary. Asher, Duke and Wilson (2013) conducted a study to evaluate discovery tools in a laboratory situation. They collected data about user observation of discovery tools and their search process. But data have biasness and have questions about genuinity as those are collected in unnatural laboratory situation. Niu, Zhang and Chen (2014) studied search activities on two discovery systems (Vu Find and Primo) concerning transaction log analysis. They are unable to find out why some results were selected by the information seeker, and to what degree the selected information was able to satisfy the seeker's information need. Ciccone and Vickery (2015) also evaluated the search performance of Google Scholar in comparison to the discovery systems- Summon and EBSCO Discovery service. They used a sample of 225 search queries extracted from log files, but, again, the sources had not been made anonymous. Deodato (2015) has done an evaluation of LIS search tools where "results were recorded in the exact order retrieved and ranked on a scale of 0-3". In this procedure, an order effect of search results may impress the assessor. Behnert and Lewandowski (2017) proposed a framework for conducting retrieval effectiveness studies in the library and information system context.

We, in our studies, have followed their principles and select assessors cum users accordingly. This study has been carried out anonymously in every aspect where assessor doesn't know which Discovery Tool's result they are evaluating and what the actual relevancy order of the articles provided by the said DT.

### **3. Objectives of the Study**

The study has been planned to get the answers of the research questions and the research gap found in the literature review with many objectives. The main objectives of the study are mentioned herewith.

- a. To find out the relevancy ranking effectiveness of a Discovery Tool.
- b. To explore whether does any relation exist in user background and relevancy assessment of a Discovery Tool

### **4. Methodology**

The study was conducted very precisely maintaining all measures minutely. As we have approached to assess search performance of Vu find Discovery Tools (DT), first of all we find out which Library's search system designed using Vu Find. So many libraries found those who are using Vu Find but we have chosed the institution where from it has been developed and maintained. A standard search term was taken from Library of Congress Subject Headings (41<sup>st</sup> edition) and put that term to search inbox of Falvay Memorial Library of Villanova University, New York, USA. Other specification selected as peer reviewed journal article and English language. The Discovery System provided a result of 20 articles in a page according to its

relevancy against the search term “Library metadata”. The investigators double clicks upon each and every title and printed the details of each article with author details, abstract, summary, publication data etc. From this printed copy investigators again produce a reprographic copy excluding the name of Discovery Tool, relevancy order number as well as other information as the respondent does not know where from the search results collected i.e., to make the study anonymous. One short questionnaire to get the demographic data and academic background data with the reprographic copy of articles is given to the respondents with instructions what to do and details of 7-point Likert scale (where 1=very irrelevant, 2=moderately irrelevant, 3=slightly irrelevant, 4=neutral, 5=slightly relevant, 6=moderately relevant and 7=very relevant). Collected Data analysed with the help of SPSS, version 20.0. The statistical tool SPSS is used to check the reliability of the collected data. Mean Score (MS) and standard Deviation (SD) for each article are also computed, leading to the finding of relevancy order. Furthermore, the t-test and ANOVA test are used to verify the consistency of the opinions among different group of respondents.

## 5. Data Collection

To collect data we prepare a questionnaire for general information of the respondents (Gender, Designation, Organisation, Subject area, articles and books published, h-index, etc). The reprographic copies with questionnaire send to 89 respondents out of which 72 (80.9%) responses are received in due time, which is considered as on excellent response. For our analysis, we set occupation into four mutually exclusive groups: 15.28% LIS Professors (Gr.A), 29.17% LIS Professionals (Gr.B), 22.22% LIS Researchers (Gr.C), and 33.33% LIS Students (Gr.D). The four groups were categorised according to gender. The 52.78% (38) male was suggested as Category 1 respondents, and the 47.22% (34) female was named as Category 2 respondents.

Table-I: Group and category wise general information

Gender	Gr. A	Gr.B	Gr. C	Gr. D	Total
Male	7	15	8	8	38 (52.78%)
Female	4	6	8	16	34 (47.22%)
Total	11 (15.28%)	21 (29.17%)	16 (22.22%)	24 (33.33%)	72 (100%)

## 6. Data Reliability

Reliability and validity are concepts used to evaluate the research quality. It indicates how well a method, technique or test measure something. Validity indicates about the accuracy and reliability intends to show the degree of consistency of a measure. Data reliability considers data source and requires knowledge of the respondent’s understanding and the completed

questionnaire. Reliability checking method are of various types. The data reliability of the survey is verified using Cronbach's alpha ( $\alpha$ ) with the help of statistical tool SPSS, which returned value 0.82; therefore, the value of the questionnaire survey revealed satisfactory reliability.

## 7. Data Analysis

The study aims to assess the efficiency and accuracy of relevancy ranking of a discovery tool with the comparison of users' opinion about the relevancy ranking provided against the results of a particular search term as well as find out if there is any relationship between users' background and their capacity of assessing articles' relevancy in concern with a particular search term. The primary data obtained from 72 respondents are analyzed. We computed the mean score (MS) and standard deviation (SD) of each article to demonstrate its level of relevancy. If two or more articles happened to have the same MS, we considered the one with lower SD to be more relevant. The MS, SD, and ranking of the articles are shown in the Table-II. The actual order of the articles provided by the DT are mentioned in the table II. A1 represents first article, A2 represents second article, A3 represents third article and so on up to 20<sup>th</sup> articles (A20). From the Table-II it is clear that the Discovery Tool provided relevancy ranking mismatched with the respondents' opinion-based ranking. Only one rank is same in both cases, other 19 cases defer from DT ranking. Column 6 of the Table-II has shown the opinion-based ranking clearly. It is clearly showing that only 5% i.e., one article's (A18) order is same in both cases and rest 95% i.e., 19 articles' order are not matched with user opinion-based ranking.

Table-II: Article wise MS, SD and opinion-based ranking

Article (DT Rank)	Min	Max	MS	SD	Rank (opinion based)	Match (Y/N)
A1	1	7	5.43	1.265	7	N
A2	1	7	4.04	1.378	20	N
A3	2	7	5.37	1.261	8	N
A4	2	7	5.51	1.473	6	N
A5	1	7	5.93	1.117	1	N
A6	1	7	5.26	1.695	11	N
A7	1	7	5.72	1.247	3	N
A8	3	7	5.72	1.153	2	N
A9	3	7	5.67	1.384	4	N

A10	1	7	5.53	1.453	5	N
A11	3	7	5.17	1.021	13	N
A12	1	7	5.13	1.310	14	N
A13	1	7	4.31	1.526	19	N
A14	2	7	4.96	1.657	16	N
A15	1	7	5.15	1.676	12	N
A16	1	7	5.35	1.416	9	N
A17	2	7	5.28	1.746	10	N
A18	1	7	4.60	1.667	18	Y
A19	1	7	5.07	1.541	15	N
A20	1	7	4.92	1.676	17	N

### 7.1 t -test

t-test is based on t-distribution and is considered an appropriate test for judging the significance of a sample mean. A t-test can help us to decide whether the difference between the conditions or category is “real” or whether it is due merely to chance fluctuations from one time of testing to another. The t-test is used to ascertain if there is a significant difference between the mean scores of two independent groups. The significant value is set to be 0.05. The significance value is the main value that is used to evaluate the difference between two independent groups about their opinions of the relevancy of the articles. If the significant value is greater than 0.05, the two groups can be preserved as significantly the same, and no significant difference exists; otherwise, the two groups are not consistent with their perceptions. As there are distinct awareness and experience between the two categories on relevance ranking of discovery tools, the t-test is used in this study to check the opinion of the significance of relevancy ranking of articles whether the respondents in category 1 and category 2 have significant differences in the views. Table-III illustrates the results of the t-test performed, and compares the MS, SD, and the significant value of each article. Only 5 articles showing significantly different ( $p < 0.05$ ) in the opinion between category 1 and category 2. The articles A3( $p=0.001$ ), A7( $p=0.032$ ), A14( $p=0.002$ ), A15( $0.034$ ) and A19( $0.027$ ) are showing that category 2 has shown the opinion of most relevant than that of category 1. On the other hand, rest of the articles’ p- value are belonging within 95% level of confidence showing the significance value more than 0.05 ( $p > 0.05$ ). i.e. 75% opinion has not been different in comparison with one category to other.

Table-III: Independent sample t-Test results for Category 1 and Category 2

Article	Category 1		Category 2		Total		Sig	Sig. Diff (Y/N)
	MS	SD	MS	SD	MS	SD		
A1	5.53	1.224	5.32	1.319	5.43	1.265	0.612	N
A2	3.76	1.344	4.35	1.368	4.04	1.378	0.943	N
<b>A3</b>	<b>5.29</b>	<b>1.011</b>	<b>5.47</b>	<b>1.502</b>	<b>5.37</b>	<b>1.261</b>	<b>0.001</b>	<b>Y</b>
A4	5.53	1.538	5.50	1.420	5.51	1.473	0.777	N
A5	5.66	1.279	6.24	0.819	5.93	1.117	0.555	N
A6	5.13	1.679	5.41	1.725	5.26	1.695	0.272	N
<b>A7</b>	<b>5.39</b>	<b>1.443</b>	<b>6.09</b>	<b>0.866</b>	<b>5.72</b>	<b>1.247</b>	<b>0.032</b>	<b>Y</b>
A8	5.66	1.146	5.79	1.175	5.72	1.153	0.570	N
A9	5.63	1.239	5.71	1.548	5.67	1.384	0.058	N
A10	5.84	1.220	5.18	1.623	5.53	1.453	0.162	N
A11	5.37	1.076	4.94	0.919	5.17	1.021	0.139	N
A12	4.97	1.219	5.29	1.404	5.13	1.310	0.174	N
A13	4.37	1.567	4.24	1.499	4.31	1.526	0.757	N
<b>A14</b>	<b>4.45</b>	<b>1.781</b>	<b>5.53</b>	<b>1.308</b>	<b>4.96</b>	<b>1.657</b>	<b>0.002</b>	<b>Y</b>
<b>A15</b>	<b>4.89</b>	<b>1.842</b>	<b>5.44</b>	<b>1.440</b>	<b>5.15</b>	<b>1.676</b>	<b>0.034</b>	<b>Y</b>
A16	5.37	1.324	5.32	1.532	5.35	1.416	0.261	N
A17	4.89	1.857	5.71	1.528	5.28	1.746	0.411	N
A18	4.32	1.772	4.91	1.505	4.60	1.667	0.133	N
<b>A19</b>	<b>5.37</b>	<b>1.303</b>	<b>4.74</b>	<b>1.729</b>	<b>5.07</b>	<b>1.541</b>	<b>0.027</b>	<b>Y</b>
A20	4.68	1.662	5.18	1.678	4.92	1.676	0.895	N

## 7.2 ANOVA

The basic principle of Analysis of Variance (ANOVA) is to test for differences among the means of the populations by examining the amount of variation within each of the samples, relative to the amount of variation between the samples. an ANOVA is conducting to test

whether perception differences exist among more than two groups. We carried out a one-way ANOVA to test for significant differences in the respondents' views in responses from the four groups for each article. Moreover, ANOVA was used to inspect whether the opinions of the four different groups of respondents are consistent for each article. Analysis done by using SPSS version 20.0. If a significant value is less than 0.05, it recommends that a high degree of difference of opinion exists among the group; contrarily, if a significant value is greater than 0.05, then there are no differences in opinion among the groups. The result of the ANOVA presented in table-IV mentioning MS of each group for each article and significance value of all 20 articles. Table IV shows that the significance value (p-value) of 14 articles (A1, A2, A3, A4, A9, A10, A11, A12, A14, A15, A16, A17, A18, & A20) are less than 0.05 i.e., mean differences among the groups are significantly different. On the other hand, significance value of rest 6 articles (A5, A6, A7, A8, A13 & A19) are more than 0.05 (belonging within 95% of confidence) i.e., mean differences among the groups are not significant.

Table-IV: ANOVA test for Groups

Article	Mean Score				Total		Sig. Value	Sig. Diff. (Y/N)
	Gr. A	Gr. B	Gr. C	Gr. D	MS	SD		
A1	4.91	6.05	5.56	5.04	5.43	1.265	0.018	Y
A2	3.73	3.67	5.19	3.75	4.04	1.378	0.000	Y
A3	5.82	5.10	4.75	5.83	5.37	1.261	0.010	Y
A4	5.45	5.10	6.44	5.29	5.51	1.473	0.017	Y
A5	5.91	5.52	6.06	6.21	5.93	1.117	0.328	N
A6	6.09	4.95	5.19	5.21	5.26	1.695	0.093	N
A7	5.36	5.71	5.88	5.79	5.72	1.247	0.355	N
A8	5.36	5.43	5.94	6.00	5.72	1.153	0.253	N
A9	5.36	5.76	4.88	6.25	5.67	1.384	0.025	Y
A10	6.36	5.76	5.31	5.08	5.53	1.453	0.008	Y
A11	5.18	5.67	4.81	4.96	5.17	1.021	0.010	Y
A12	5.82	4.67	4.50	5.63	5.13	1.310	0.001	Y
A13	4.73	4.81	4.19	3.75	4.31	1.526	0.083	N
A14	4.45	4.29	5.38	5.50	4.96	1.657	0.034	Y
A15	6.27	3.95	5.13	5.71	5.15	1.676	0.001	Y



A16	6.27	5.00	6.19	4.67	5.35	1.416	0.000	Y
A17	3.73	5.33	6.50	5.13	5.28	1.746	0.000	Y
A18	3.82	4.24	4.00	5.67	4.60	1.667	0.000	Y
A19	4.91	5.00	5.56	4.88	5.07	1.541	0.549	N
A20	3.91	5.43	4.19	5.42	4.92	1.676	0.002	Y

## 8. Discussion and conclusion

The goal of the study was to find out whether relevancy ranking provided by the DT is matching with the users' opinion-based ranking as well as users' background as a factor to determine the relevancy order of search results. The effectiveness of relevancy ranking order of a Discovery tool is assessed based on the proposed seven features (Gordon and Pathak, 1999) for evaluation to ensure highly accurate and informative results. We have focused on the tasks (search term) based on genuine information needs and on the other hand we have judge the relevancy ranking order by the actual user instead of by the investigator to avoid biased and predefined assessment. From the study we can be able to understand all very carefully. Analysis shows clearly that 95% articles ranking has not been matched in order with the opinion-based ranking of the users which clearly shows that there is a need in the rethinking of the procedure of weighting method, factors for relevancy ranking and principles of relevancy ranking used by the Discovery Tool. The importance and weigh of the determining factors may be change from time to time as well as library to library to satisfy users. It is said that documents relevant to a particular user group are also relevant to an individual user with the same user background (Lewandoski, 2005). So, user background is an important factor to determine the ranking of the documents. T-test results support the consistency of the different categorically given opinion about the above inference. ANOVA test results showed that the assessors' opinion on the relevancy ranking of a discovery tool differ most of the cases. So, the academic background like qualification, tenure of experience in the field, publication portfolio and author metrics etc. have direct influence in the judgment or understanding of relevancy of documents against a search term.

As discovery tool provide single window to get results of his or her search as well as user always depend upon the default relevancy order and they only follow first page of the results so there is a need of minute principles to standardize the ranking strategy. More and more careful study about the factors, analysis of ranking strategy, new algorithm, new framework and model is inevitable to achieve the more or less satisfactory level by trial-and-error method.

## Reference

Asher, A. D., Duke, L. M., & Wilson, S. (2013), "Paths of discovery: Comparing the search effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and conventional library resources", *College & Research Libraries*, Vol. 74 No.5, pp. 464-488.

Babbie, E. (2007), *The Practice of Social Research*, 11<sup>th</sup> ed.; Thomson Wadsworth: Nashville, TN, USA.

Behnert, C and Lewandowski, D. (2017). "A framework for designing retrieval effectiveness studies of library information systems using human relevance assessments". *Journal of Documentation*. Doi: 10.1108/JD-08-2016-0099

Behnert, C and Lewandowski, D. (2015). "Ranking search results in library information systems- Considering ranking approaches adapted from web search engines". *The Journal of Academic Librarianship*, Vol.41 No.6, pp.725-735. <http://doi.org/10.1016/j.acalib.2015.07.010>

Canfora, G., & Cerulo, L. (2004), "A taxonomy of information retrieval models and tools". *Journal of Computing and Information Technology*, Vol. 12 No.3, pp. 175-194.

Ciccione, K., & Vickery, J. (2015), "Summon, EBSCO Discovery Service, and Google Scholar: A Comparison of Search Performance Using User Queries". *Evidence Based Library and Information Practice*, Vol.10 No.1, pp. 34. <https://doi.org/10.18438/B86G6Q>

Danjuma, S., Herawan, T., Ismail, M. A., Abubakar, A. I., Zeki, A. M., & Chiroma, H. (2017), "A review on Soft Set-based Parameter Reduction and Decision Making". *IEEE*.

Denton, W. & Coysh, S.J. (2011), "Usability testing of Vu Find at an academic library", *Library Hi Tech*, Vol.29 No. 2, pp. 301-319, <https://doi.org/10.1108/07378831111138189>.

Deodato, J. (2015), "Evaluating Web-Scale Discovery Services: A Step by Step Guide". *Information Technology and Libraries*, Vol.34No.2, pp.19-75. <https://doi.org/10.6017/ital.v34i2.5745>

Ellero, N. P. (2013). "An unexpected discovery: one library's experience with web-scale discovery service (WSDS) evaluation and assessment". *Journal of Library Administration*, vol. 53 No.5-6, pp. 323-343, doi: 10.1080/01930826.2013.876824.

Exner, N., Bollinger, S., & Sims, I. (2012), "Discovery in a hurry: fast transitions from federated search to article discovery". In M.P. Popp & D. Dallis (Eds.), *Planning and implementing resource discovery tools in academic libraries*, pp.351-365. Hershey, PA: Information Science Reference.

Gliem, J.A & Gleim, R. R. (2003), "Calculating, interpreting and reporting Cronbach's Alpha reliability coefficient for Likert-type scales". *Midwest Res. Pract. Conf. Adult Contin. Community Educ.*, pp. 82-88

Gordon, M., & Pathak, P. (1999), "Finding information on the World Wide Web: the retrieval effectiveness of search engines". *Information Processing & Management*, Vol. 35 No.2, pp. 141-180. [http://doi.org/10.1016/S0306-4573\(98\)00041-7](http://doi.org/10.1016/S0306-4573(98)00041-7)

Harari, Z. (2016), "Ranking user's comments by use of proposed weighting method". *Advances in Computer Science: an International Journal*, Vol. 5 No. 1, pp. 45-51.

Harman, D. (1986), "An experimental study of factors important in document ranking". *Research and development in information retrieval*, pp. 186-193.

Kothari, C. R. (2012), *Research Methodology: methods and techniques*, New Age International, New Delhi.

Large, A. & Beheshti, J. (1997), "OPACs: a research review", *Library and Information Science Research*, Vol. 19 No. 2, pp. 111-133, [https://doi.org/10.1016/S0740-8188\(97\)90039-6](https://doi.org/10.1016/S0740-8188(97)90039-6)

Lau, E.P., & Goh, D.H.L. (2006), "In search of query patterns: a case study of a university OPAC", *Information Processing & Management*, Vol.42 No. 5, pp. 1316-1329.

Lewandowski, D. (2009), "Ranking library materials", *Library Hi Tech*, Vol. 27 No.4, pp. 584-593, <http://dx.doi.org/10.1108/07378830911007682>.

Lewandowski, D. (2005), *Web Information Retrieval: Technologien zur Informations suche im Internet*, Frankfurt am Main: Dt. Ges. Fur Informations wiss. Und Informations praxis.

Niu, X., Zhang, T., & Chen, H. (2014), "Study of user search activities with two discovery tools at an academic library". *International Journal of Human – Computer Interaction*, Vol. 30 No.5, pp. 422-433. <https://doi.org/10.1080/10447318.2013.873281>.

Olson, T.A. (2007), "Utility of a faceted catalog for scholarly research", *Library Hi Tech*, Vol.25 No. 4, pp. 550-561

Plass meier, K., Borst, T., Behnert, C., & Lewandowski, D. (2015). "Evaluating popularity data for relevance ranking in library information systems". *American society for Information Science and Technology*, p. 66.

Sadeh, T. (2008), "User experience in the library: a case study", *New Library World*, Vol. 109 No. ½, pp. 7-24.

Zhang, T. (2013), "User-Centered Evaluation of a Discovery Layer System with Google Scholar". In A. Marcus (Ed.), *Design, User Experience, and Usability. Web, Mobile, and Product Design SE-34* (Vol. 8015, pp.313-322). Berlin; Heidelberg: Springer. [https://doi.org/10.1007/978-3-642-39253-5\\_34](https://doi.org/10.1007/978-3-642-39253-5_34)

### **Further reading**

Breeding, M. (2015). *The Future of Library Resource Discovery*. NISO.

Breeding, M. (2014). "Web-scale discovery services". *American Libraries*, Vol. 45 No.1/2, p. 25.

Breeding, M. (2007), "The birth of a new generation of library interfaces". *Computers in Libraries*, Vol. 27 No. 9, pp. 34-37.

Breeding, M. (2006), "Technology for the Next Generation". *Computers in Libraries*, Vol. 26 No. 10, pp. 28-30.

Fagan, J. (2010), "Usability studies of faceted browsing: a literature review", *Information Technology and Libraries*, Vol. 29 No. 2, p. 58

Lewandowski, D. (2015), "Evaluating the retrieval effectiveness of web search engines using a representative query sample", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 9, pp.1763-1775, <https://doi.org/10.1002/asi.23304>

Lewandowski, D. (2012), "A framework for evaluating the retrieval effectiveness of search engines", In C.Jouis, I. Biskri, J. G. Ganescia & M. Roux (Ed s.), *Next Generation Search Engines* , Hershey, PA: IGI Global, pp. 456-479 <https://doi.org/10.4018/978-1-4666-0330-1.ch020>

Vaughan, J. (2012), "Investigations into library web-scale discovery services", *Information Technology & Libraries*, Vol. 31 No. 1, pp. 32-82.