Evaluation of the NSDL and Google for Obtaining Pedagogical Resources

Frank McCown, Johan Bollen, and Michael L. Nelson

Old Dominion University
Computer Science Department
Norfolk, VA 23529 USA
{fmccown,jbollen,mln}@cs.odu.edu

Abstract. We describe an experiment that measures the pedagogical usefulness of the results returned by the National Science Digital Library (NSDL) and Google. Eleven public school teachers from the state of Virginia (USA) were used to evaluate a set of 38 search terms and search results based on the Standards of Learning (SOL) for Virginia Public Schools. Evaluations of search results were obtained from the NSDL (572 evaluations) and Google (650 evaluations). In our experiments, teachers ranked the links returned by Google as more relevant to the SOL than the links returned by the NSDL. Furthermore, Google's ranking of educational material also showed some correlation with expert judgments.

1 Introduction

When the question "What is the most venomous snake in the world?" was posted to AskNSDL, the response was, "I did a search in google.com for 'poisonous snakes world' and came up with some good information..."[2]. A search for the term 'google' at ask.nsdl.org reveals that Google is frequently used to answer AskNSDL questions. Why are the questions posed to an NSDL related web site answered using Google instead of the NSDL? The NSDL only accepts educationally sound material into its library [17], so naturally it should produce more trusted results than would Google, which accepts any web-crawlable content.

The National Science Digital Library (NSDL) [24] is a U.S. government funded online library that provides pedagogical resources for science, technology, engineering, and mathematics. In an attempt to provide only relevant and highly educational material, the NSDL obtains its contents from on-line material that is submitted by the education community or from focused crawls [4, 5] or from specific, targeted collections [1, 14]. Because all NSDL content is available on-line, some of the same material is also indexed by Google. The number of resources incorporated within the NSDL is relatively small when compared to Google because the NSDL acquires content solely for the purpose of supporting science education. Figure 1 illustrates a theoretical view of all the useful educational content that is on the web and how it has been crawled by Google and/or federated into the NSDL.
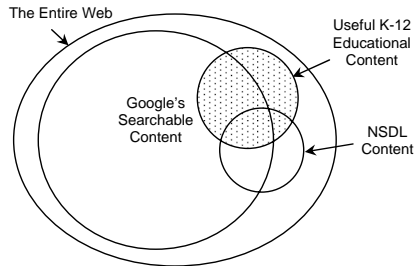
Fig. 1. Educational content on the Web

We wanted to measure the degree to which Google and the NSDL could locate relevant educational material. To do so, we designed a real-world scenario that showed how an educator might use the NSDL and Google to find educational content. This scenario provided an appropriate set of search terms that could be used to query the NSDL and Google. The Commonwealth of Virginia uses the state-wide standard called the Standards of Learning (SOL) for Virginia Public Schools [20] that mandates what should be taught in the public schools to children in grades K-12 (ages 5-18). The SOL lists detailed and specific topics which educators are required to teach and test for at each grade level. These topics can be researched using the web to retrieve background or additional information. Using the SOL for mathematics and the sciences, we devised a set of search terms and submitted them to the NSDL and Google. We used paid volunteer Virginia teachers to evaluate the search results in terms of their pedagogical usefulness. The results of the evaluation show that Google is able to produce more relevant educational material in its top 10 results than the NSDL.

2 Related Work

When evaluating web search engines, focus may be given to any number of factors: user interface, user effort, query response time, retrieval effectiveness, etc. Given that the primary goal of search engines is to locate relevant documents quickly, the most common measurements are precision and recall [6]. Precision, the ratio of retrieved relevant documents to the number of retrieved documents, is the primary measurement used by many web search engine comparisons as their primary evaluation measure [6, 8, 9, 11, 13, 15, 16, 21]. Recall is the ratio of retrieved documents to the total number of relevant documents available. Because the total number of relevant documents in a collection is rarely known, recall is much more difficult to measure due to the huge number of returned search results produced by search engines, and has therefore been largely ignored by web search engine comparisons.

When evaluating precision, most studies use human evaluators who issue queries about some topic, browse the returned result set, and record a relevancy

score for each link. What qualifies as "relevant" is often specific to the study. Some studies have used a binary view of relevance [15] where search results are deemed either relevant or not relevant. Others have used a non-binary measurement [8, 15, 16] allowing a document to be semi-relevant. We have taken the later approach as will be discussed in Sect. 3. In all of these studies, evaluators were limited to viewing the first n returned results. Values of 10 [8, 16] and 20 [6, 11, 13, 15, 22] were common. This prevented the evaluators from being overwhelmed by the volume of search results. This methodology is in keeping with studies that show most users do not go beyond the first screen of search results anyway [18, 19]. Once the relevance scores were obtained for the first n results, a precision at n (P@n) was computed and used in statistical analysis. We also limited our evaluation to the first 10 search results from Google and the NSDL.

The difficulty of measuring precision is two-fold: the expense of time-consuming human evaluations, and the reliance on subjective human judgments as to the relevance of a resource. Some methods have been developed to compare search engines that remove the subjective human component from the process [3, 6, 7]. Although these methods have shown some promise, they are still somewhat new and need to be further validated. These automatic evaluation procedures could be applied to our findings in future research in order to further validate their methods. It is unclear though whether these approaches could be applied based on our definition of relevance as material that is educationally useful.

None of the studies in the literature have been performed against the NSDL. Furthermore, none of the studies addressed the educational usefulness of the search results. An in-depth look at how educators judged the quality of educational material on the web and in digital libraries was performed in [23]. They examined the mindset used by educators when deciding if a web resource was useful for augmenting the classroom experience. Participants of their study showed that educators expect educational digital libraries to save them time and effort when compared to web search engines because DLs filter their content. In the framework of our intended comparison of the NSDL and Google, it was thus expected that the NSDL would contain a collection of educational documents that were better filtered for quality and educational usefulness than Google and that these documents would be rated higher by educators. Although our study did not gauge the worthiness of using the resources that were found by the NSDL and Google in the classroom, several of our evaluators reported that many of the items they evaluated would make excellent classroom resources.

3 Experiment

We examined the SOL for math and scientific educational content that students would need to learn. In examining each standard, we came up with a list of search terms that a student or educator might type into the search facility of the NSDL or Google in order to find out more information about the topic. For example, a teacher might select the terms we have bolded in the following Biology SOL:

BIO.3 The student will investigate and understand the chemical and biochemical principles essential for life. Key concepts include a) **water chemistry** and its impact on **life processes**; b) the structure and function of macromolecules; c) the nature of enzymes; and d) the capture, storage, transformation, and flow of energy through the processes of photosynthesis and respiration.

We paid 11 Virginia public school teachers to judge the quality of the search terms that we devised based on the SOL. The teachers were assigned according to their specialty. In other words, biology teachers were assigned to judge the quality of search terms from the Biology SOL, chemistry teachers judged search terms from the Chemistry SOL, etc. The evaluators used a web-based evaluation system that we developed to perform their evaluations. A screen shot of the system is show in Fig. 2. After rating each search term, the evaluators were then presented a randomized list of at most 20 search results, combining the top 10 search results from the NSDL and Google. The evaluators were unaware of the source of the search results or rankings as Fig. 2 illustrates. Each search result was viewed by the evaluator, and then a score from 1 to 6 was assigned based on the quality of the search result. The score was based on how well the search result provided educational content (specific to the SOL) that was useful for student education. One was the best score, five was the worst score, and six indicated the web content was inaccessible (due to an old link, hidden content, login was required, etc.).
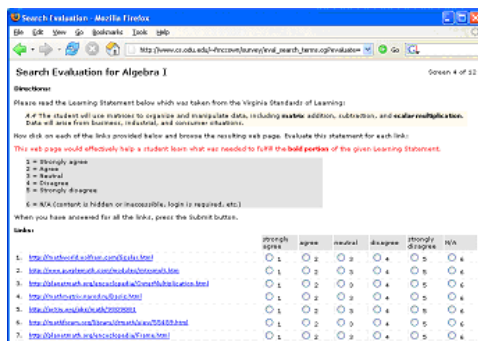

Fig. 2. Randomized and anonymous links in the evaluation system.

Each teacher evaluated at least 6 sets of search terms and search results. Each evaluation took about 1.5 hours to complete. The evaluators used common graphical web browsers like Internet Explorer and Netscape to perform the evaluations. The computers used to perform the evaluations had high-speed Internet connections so that our evaluators could work quickly and would not be prone to giving poor ratings to items that downloaded slower. The computers

were capable of displaying Adobe Acrobat files (PDF) and Microsoft PowerPoint files (PPT). PDF and PPT files made up 10% of the search results returned by Google.

The NSDL and Google queries were performed between Dec. 29, 2004 and Jan. 2, 2005. The evaluations (and web site accesses) were performed between Jan. 3 and Jan. 20, 2005. Our evaluation did not use the advanced search features of the NSDL and Google because research has shown that users rarely use advanced features [19]. Besides that, the NSDL lacks many of the advanced search features of Google. The NSDL does allow the user to limit the results to text, images, audio, video, interactive controls, or data, but it does not support boolean operators. In order to perform a fair evaluation of Google and the NSDL, we followed the guidelines suggested by [11, 13].

We used a Likert 5-point scale to determine relevancy. For each search result returned, the evaluators gave a score based on "This web page would effectively help a student learn what was needed to fulfill the bold portion of the given Learning Statement" where the Learning Statement was a section from the SOL. A score of 1 indicated they strongly agreed with the statement about the web resource, and 5 indicated they strongly disagreed.

Our definition of "relevant" was based not only on if the search terms produced an item that pertained to the SOL, but also on how educationally useful the web resource was that was being evaluated. Because our evaluators were determining whether or not a resource would help a student learn about a particular subject, a search result might match the search terms quite well and be considered relevant in the traditional sense, but the evaluator may have given the search result a much lower score if the resource didn't help teach a student at a particular grade level. This could be because the page content was presented at a lower or higher level, making the content too trivial or difficult to understand.

The eleven teachers used as evaluators in our experiment were asked several questions in regards to their teaching experience, familiarity with the SOL, and familiarity with using the Internet as a teaching tool. The teachers were all from the Norfolk and Virginia Beach school systems. They have taught on average 5.1 years with 3.85 of those years teaching in the fields they evaluated. They were all very familiar with the SOL and using the Internet to augment classroom activities (average scores of 1.15 and 2, respectively, on a five point scale). Most teachers were unfamiliar with the NSDL (4.5 on a five point scale). We asked the teachers after the evaluation about their knowledge of the NSDL to avoid biasing the evaluations.

4 Evaluation Results

The eleven evaluators each spent an average of 1.5 hours to evaluate at least six sets of search terms and the roughly 20 search results generated by each query. Table 1 shows a total of 38 queries were produced from the SOL for all subjects producing a total of 334 (318 unique) links from NSDL and 380 (376 unique) links from Google. The evaluators produced 65 ratings for search

terms and evaluations for 572 NSDL-returned resources and 650 Google-returned resources.

Table 1. Summary of NSDL and Google search responses and evaluations

Subject	Evaluators	Search Queries	Responses		Evaluated Srch Terms	Eval. Resp.	
			NSDL	Google		NSDL	Google
Algebra I	1	6	52	60	6	56	60
Algebra II	1	7	62	70	7	56	70
Biology	3	6	52	60	15	134	150
Chemistry	1	6	52	60	6	52	60
Earth Sci.	4	7	62	70	25	218	250
Physical Sci.	1	6	56	60	6	56	60
Totals	11	38	334	380	65	572	650

Each search query contained anywhere from 1 to 5 words and averaged 3.28 words per query. This length seems reasonable since it is slightly smaller than the average query length used in some evaluations [6] and slightly larger than the average seen by some popular web search engines [19]. The search terms that we chose to find educational content from the SOL were given an average rating of 2.08 (median=2) by our evaluators indicating agreement that the search terms validly reflected the relevant SOL learning statements. When each query was performed, we took only the top 10 search results. Of the top 10 search results, the NSDL averaged 8.8 hits per query, and Google always produced 10. The lower number of NSDL hits may be due to the following reasons: the NSDL is querying from a smaller pool than Google; the NSDL performs an ANDing of each search term (hurting recall in favor of precision); and the NSDL shows search results for which no link is available (we did not use archived versions from either NSDL or Google).

The links retuned by the NSDL and Google had little in common. The 38 queries performed generated only 6 duplicate links in the top 10 results. The same queries produced only 9 duplicates in the top 20 results. One reason for the lack of commonality is because one quarter (78 of 318) of the unique search results produced by the NSDL were not indexed by Google. 31 out of the 334 NSDL search results (9.3%) were given a score of 6 (inaccessible content, login-required access, etc.) compared to 20 out of the 380 Google search results (5.3%). These are web pages that were inaccessible due to any number of reasons: stale links, temporarily off-line servers, intermittent Internet problems, login-only access, etc. In our experiment we had the evaluators group login-only search results with other inaccessible sites because if a user is trying to access on-line material quickly, they are more likely to move on to the next search result rather than going through the time-consuming process of registering for an account. Of the 31 NSDL results given a score of 6, 25 of them (80.6%) were from the Learning Online Network with CAPA at http://nsdl.lon-capa.org. LON-CAPA accounted

for 41 of the entire 334 NSDL search results (12.3%). 30 of the 41 LON-CAPA URLs (73.2%) were not indexed by Google.

Teachers rated each resource from 1 to 5 expressing, respectively, "strong agreement" or "strong disagreement" with the statement that the resources "would effectively help a student learn what was needed to fulfill the bold portion of the given Learning Statement." A rating of 6 indicated that the particular resource was inaccessible. Throughout the remainder of our evaluation, we will use the term "agree" to indicate *higher relevance* when discussing relevance of the search results. The term "disagree" will indicate *lower relevance*, and the term "neutral" will indicate *uncertainty* about relevance.

We compared the teacher ratings for all NSDL and Google search results to determine which on average performed best in terms of the SOL learning statements. The median rating for all NSDL search results was found to be 4 (disagreement), and its mean score 3.74 with a standard deviation of 1.48. The median rating for all Google search results, in comparison, was found to be 3 (neutral), and its mean 3.16 with a standard deviation of 1.57. These results are listed in Table 2.

Table 2. Descriptive statistics on NSDL and Google ratings

	Google	NSDL
Mean	3.16	3.74
Median	3	4
Std	1.57	1.48

Both search engines in the mean did not generate results that would satisfy the educational requirements expressed in the ratings evaluation statement. In other words, most raters did not feel the returned resources would help a student learn what was needed to fulfill a portion of the given Learning Statements (median for Google and NSDL was respectively 3 (neutral) and 4 (disagree)). However, even though both search engines performed poorly in the mean, the median and mean score for the NSDL seemed to suggest it performed worse than Google. In fact, a Wilcoxon signed rank test revealed a statistically significant difference ($p < 0.05$) between the NSDL and Google ratings. This pattern becomes more evident when we examine the actual distribution of ratings, e.g. "how often did a NSDL or Google search produce a satisfactory result indicated by a 1 or 2 score?"

We first determined the precision of NSDL and Google by counting the number of results rated adequate (any score less than or equal to 2), denoted R, compared to the total number of results generated by that search engine, denoted N. To avoid double counts we aggregated multiple ratings for the same search result by calculating the mean rating for that item. Google produced a

total of 380 search results for 38 queries, 145 of which were rated at a level ≤ 2 (agree to strongly agree). The precision for Google P_g was defined by:

$$P_g = R_g/N_g = 145/380 = 0.382 \ or \ 38.2\% \tag{1}$$

NSDL's precision (P_n) was defined in the same manner. 57 of the 334 NSDL search results were rated at a level ≤ 2 (agree to strongly agree), and P_n was determined as follows:

$$P_n = R_n/N_n = 57/334 = 0.171 \ or \ 17.1\% \tag{2}$$

Although both precision levels were relatively low, NSDL performance, according to this metric, was significantly worse than Google's. The fact that the NSDL produced an average of 8.8 results per query compared to Google's 10 results per query strengthens this conclusion.

A frequency analysis was conducted of the NSDL's and Google's ratings to determine the performance differences at a lower level of granularity (Fig. 3). Figure 4 shows the NSDL and Google ratings according to the six different categories that were evaluated.
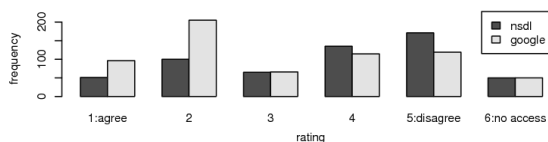

Fig. 3. Ratings for NSDL and Google search results (all domains)
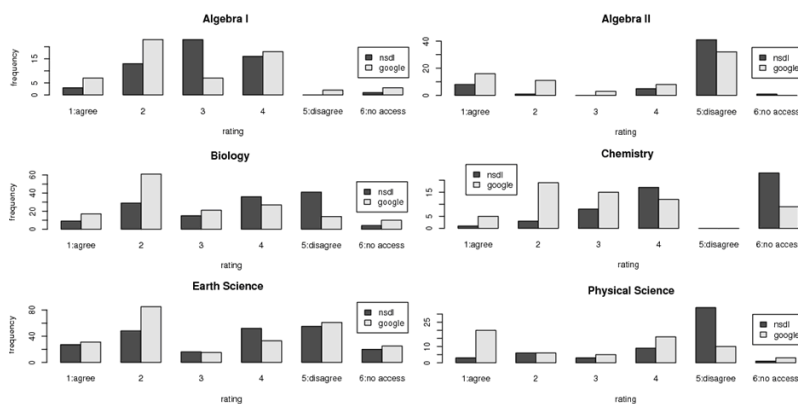

Fig. 4. Ratings for NSDL and Google search results (individual domains)

From Fig. 4 we can see that although NSDL and Google average ratings correspond to at best a neutral (3) or disagree (4), they differ considerably in their distributions. Google's ratings tend to cluster in the 2-"agree" range, but follow a bi-model distribution which peaks first at a rating of 2 and to a lesser degree at 4, indicating raters either approved or disliked the results, but tended to approve more than they disliked. The distribution of NSDL ratings is skewed to the right with the highest number of ratings observed in the 4-"disagree" and 5-"strongly disagree" range. Although the median of these distributions differs by only 1 point, these numbers hide strikingly different rating patterns favoring the Google results.

The bar plots in Fig. 4 indicate that although neither Google nor the NSDL perform exceptionally well, Google's ratings are consistently higher than the NSDL's for all subject domains. To verify this is indeed the case, we determined the median ratings for Google and the NSDL for the six mentioned domains. The results are listed in Table 3. For each comparison we ran a Wilcoxon signed rank test; p-values are listed in the third column.

Table 3. Median ratings split according to domain category

	Google	NSDL	p-value
Algebra I	2.5	3	0.437
Algebra II	4	5	0.001 **
Biology	2	4	< 0.001 **
Chemistry	3	4	< 0.001 **
Earth Sci.	3	4	0.209
Physical Sci.	3	5	< 0.001 **

P-values ≤ 0.001 (**) indicate statistically significant results. In all but two cases, namely Algebra I and Earth Science, the ratings indicate that Google significantly outperforms the NSDL. Furthermore, while only one of Google's median ratings slip below a "neutral" and two are actually rated "agree" (2), only one of NSDL's median ratings is a "neutral" (3), and all correspond to at least a "disagree" (4) rating, and two (Algebra II and Physical Science) correspond to a "strongly disagree" (5).

The effectiveness of a search engine is in great part determined by its ability to rank retrieved materials according to their relevance to the user query. In this case, however, our definition of relevance deviates from the traditional term-based model. We are not looking to match a query consisting of search terms to the resource that best matches those terms, but the resource which best matches the query in terms of its educational effectiveness and merit. We therefore need to determine the extent to which the ranking of search results from Google and NSDL match the obtained ratings that are based on the educational appropriateness and effectiveness of the returned results.

To investigate the degree to which the search result rankings of the NSDL and Google match expert ratings, we determined the Spearman Rank correlation between rankings and ratings. A small but statistically significant correlation was found between Google's search results rankings and our expert ratings (rho=0.125, p=0.001). In the NSDL's case, no significant correlation was found (rho=0.057, p=0.173). Google's search result ranking to some degree approximates how teachers would rate a set of search results, but NSDL's does not.

5 Future Work and Recommendations

After performing our experiment and evaluating the results, we have noted several areas that could improve future experiments:

Improve the search terms by getting input from the evaluators. We wanted to limit the amount of work that the teachers needed to perform, so we picked search terms that we believed represented the SOLs. Although our search terms were shown to have been acceptable by the evaluators, some improvements could have been made by giving the evaluators the chance to modify the terms and come to a general consensus as to what the best search terms would be.

Use more evaluators. Volunteers, even compensated volunteers, are hard to enlist. If, for example, principals had required teachers to participate in the study, we would have more evaluators but the quality of their reviews may have suffered if they were coerced into participating.

Provide immediate procotor support duing the evaluations. We believe some of our evaluators may have been too harsh or too forgiving in assigning relevance because of a lack of guidance. Some may have not used the relevancy criteria that we explicitly asked them to evaluate. At least one of our evaluators gave consistently lower scores than other evaluators. A more tightly controlled, synchronous environment may have increased the quality of the reviews, but it would have restricted the number of possible participants.

In the course of our evaluation, we uncovered the following areas of improvement for the NSDL:

Rank search results according to quality. According to our analysis, the NSDL does not rank its search results according to perceived relevance. This has been suggested elsewhere [10].

Provide an advanced search capability. Although most users don't use the advanced searching features, those that do could benefit from producing more targeted queries. Originally the NSDL had just an advanced search page, then a usability study [12] suggested a simple search page. That was added, but the functionality of the advanced search page seemed to be reduced.

Provide the ability to target the grade level appropriateness of information. In our experiment, 16.5% of the NSDL results were from arXiv.org, an e-print archive focusing primarily on advanced research in physics, math, and computer science. Of these results, only one of them scored 2 or better. This suggests it may be useful for the NSDL to rate its content based on grade level appropriateness [23].

6   Conclusions

We have performed an experiment that demonstrates how educators might try to find educational material in the NSDL and Google. We created a real-life scenario whereby teachers needed to obtain web-based educational content based on educational requirements from the Virginia Standards of Learning. We queried the NSDL and Google using a variety of search terms and used paid volunteer teachers to evaluate the educational relevance of the search results.

Based on statistical analysis of the evaluation results, we found that Google tended to find more useful educational material than the NSDL; in 4 out of 6 subject areas, Google significantly outperformed the NSDL. Google's precision was found to be 38.2% compared to NSDL's 17.1%. Google's ranking of material outperformed the one applied by the NSDL search engine as indicated by its moderate correlation with expert ratings of resource relevance. Although the NSDL's collection of educational resources may be of higher quality and scope than what Google can offer, the latter's ranking will make those resources more easily and efficiently available. We analyzed the returned results from the NSDL and found that a significant portion of them required registration for accessing. About 1 in 4 NSDL resources were not indexed by Google. There was also very little overlap (6 duplicates in top 10 results from 38 queries) in the results returned by Google and NSDL. Finally we provided some guidelines for improving this type of evaluation in the future. We also provided some suggestions that could be used to improve the NSDL's search engine so that educators can use the NSDL to find relevant educational material more effectively than Google.

References

1. Arms, W., Dushay, N., Fulker, D., Lagoze, C.: A Case Study in Metadata Harvesting: The NSDL. Library HiTech. Vol. 21 **2** (2003) 228–237
2. AskNSDL. Response by Melodye Campbell. (March 20, 2004) https://ask.nsdl.org/default.aspx?from=srch&id=8337
3. Bar-Ilan, J. Methods for Measuring Search Engine Performance Over Time. Journal of the American Society for Information Science and Technology. Vol. 53 **4** (2002) 308–319
4. Bergmark, D.: Collection Synthesis. In Proceedings from the second ACM/IEEE-CS joint conference on Digital libraries (JCDL 2002) Portland, Oregon, USA (2002) 253–262
5. Bergmark, D., Lagoze, C., Sbityakov, A.: Focused Crawls, Tunneling, and Digital Libraries. In Proceedings of the European Conference on Digital Libraries (ECDL), Rome, Italy (September 2002) 91–106
6. Can, F., Nuray, R., Sevdik, A.: Automatic Performance of Web Search Engines. Information Processing and Management. Vol. 40 **3** (2004) 495–514
7. Chowdhury, A., Soboroff, A.: Automatic Evaluation of World Wide Web Search Services. In Proceedings of the ACM SIGIR conference. Vol. 25. (2002) 421–422
8. Chu, H., Rosenthal, M.: Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology. Proceedings of the ASIS annual meeting. **33** (1996) 127–135

9. Ding, W., Marchionini, G.: A Comparative Study of Web Search Service Performance. In Proceedings of the 59th Annual Meeting of the American Society for Information Science. Baltimore, MD, USA. **33** (1996) 136–142
10. Fulker, D.: Metadata Strategies to Address NSDL Objectives. In Proceedings of the 5th Russian Conference on Digital Libraries (RCDL 2003). St. Petersburg, Russia. (2003)
11. Gordon, M., Pathak, P.: Finding Information on the World Wide Web: The Retrieval Effectiveness of Search Engines. Information Processing and Management. Vol. 35 **2** (1999) 141–180
12. Hartson, H., Shivakumar, P., Pérez-Quiñones, M.: Usability Inspection of Digital Libraries: A Case Study. International Journal of Digital Libraries. Vol. 4 **2** (2003) 108–123
13. Hawking, D., Craswel, N., Bailey, P., Griffiths, K.: Measuring Search Engine Quality. Information Retrieval. Vol. 4 **1** (2001) 33–59
14. Lagoze, C., Hoehn, W., Millman, D., Arms, W., Gan, S., Hillmann, Ingram, C., Krafft, D., Marisa, R., Phipps, J., Saylor, J., Terrizzi, C., Allan, J., Guzman-Lara, S., Kalt, T.: Core Services in the Architecture of the National Science Digital Library (NSDL). In Proceedings from the second ACM/IEEE-CS joint conference on Digital libraries (JCDL 2002) Portland, Oregon, USA. (2002) 201–209
15. Leighton, H., Srivastave, J.: First 20 Precision Among World Wide Web Search Services (Search Engines). Journal of American Society for Information Science. Vol. 50 **10** (1999) 870–881
16. Ljosland, M.: Evaluation of Web Search Engines and the Search for Better Ranking Algorithms. SIGIR99 Workshop on Evaluation of Web Retrieval (1999)
17. NSDL Collection Development Policy Draft v.030715. Accessed 1-18-2005. http://content.comm.nsdl.org/doc_tracker/docs_download.php?id=452
18. Silverstein, C., Henzinger, M., Marais, J., Moricz, M.: Analysis of a Very Large Alta Vista Query Log. Technical Report 1998-014. COMPAQ Systems Research Center, Palo Alto, CA. (1998)
19. Spink, A., Wolfram, D., Jansen, B., Seracevic, T.: Searching the Web: The Public and Their Queries. Journal of the American Society for Information Science and Technology. Vol. 52 **3** (2001) 226–234
20. Standards of Learning for Virginia Public Schools, http://www.pen.k12.va.us/VDOE/Superintendent/Sols/home.shtml
21. Su, L.: A Comprehensive and Systematic Model of User Evaluation of Web Search Engines: I. Theory and Background. Journal of the American Society for Information Science and Technology. Vol. 54 **13** (2003) 1175–1192
22. Su. L., Chen, H., Dong, X.: Evaluation of Web-based Search Engines from an End-user's Perspective: A pilot study. Proceedings of the 61st Annual Meeting of the American Society for Information Science. Pittsburgh, PA, USA. (1998) 348–361
23. Sumner, T., Khoo, M., Recker, M., Marlino, M.: Understanding Educator Perceptions of "Quality" in Digital Libraries. In Proceedings of the Third ACM+IEEE Joint Conference on Digital Libraries. (JCDL 2003) Houston, Texas, USA. (2003) 269–279
24. Zia, L.: The NSF National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) Program: A Progress Report. D-Lib Magazine, 6(10), 2000.