Making Search Engines Notice: An Exploratory Study on Discoverability of DSpace Metadata and PDF

Le Yang

Assistant Librarian

Texas Tech University

le.yang@ttu.edu

ABSTRACT

This study analyzed the search results of four main Internet search engines on three newly created digital collections at Texas Tech University Libraries, in order to assess the search engine indexing on metadata and PDF files of digital items that are published in the institutional repository. The exploratory study found that only one search engine discovered PDF files, while the others only discovered metadata. The result of the study also suggested that metadata and PDF files can supplement each other to facilitate discoverability of the digital collections on search engines, and revealed some findings that are contradictory to what have been concluded in previous research.

KEYWORDS: Metadata Indexing, PDF Indexing, Search Engine Discovery, Digital Collections, Internet Discoverability, Digital Repository, DSpace

INTRODUCTION

Conversations about discoverability and metadata have proliferated since the rapid growth of digital collections in the 1990s. As items are put into digital repositories, associated metadata records have to be effective not only for the system search but also for external indexing by search engines in order to be discovered. A review of the existing literature reveals studies evaluating the effectiveness of metadata, digital resources discovery, search engine indexing, and search engine optimization strategies for digital collections. Researchers did not really come to an agreement, however, on whether search engine crawlers prefer indexing metadata or full-text PDF files in a digital repository. Based on these scholarly debates, some librarians may think that they do not have to invest in metadata for any digital items because the associated files will be full-text indexed, while others insist that robust metadata is needed for Internet discovery purposes.

In order to help librarians and researchers at Texas Tech University set priorities for digital collection works, and to help them have reasonable expectations of what search engines will and will not do to the digital collections being published in the local institutional repository, the author designed a methodology to explore whether search engines discover metadata or PDF files in Texas Tech University's DSpace-based digital repository. The author created three digital collections and employed different implementation strategies of metadata and PDF files on the same digital items. The author then used the title and the abstract to test the search through four main Internet search engines for the target items. Findings of this exploratory study suggested that metadata and PDF files are both important and supplement each other to facilitate search engine discovery. PDF files without supplemented metadata are not crawled or discovered by any search engines, while metadata associated with a PDF has a better chance of getting the

content indexed and found. Through this study, the author also discovered some findings that are contradictory to some generalized conclusions made in previous research.

LITERATURE REVIEW

Internet search engines' dominance in information-seeking behaviors suggests that if digital repositories are not indexed by the search engines, the digital collections are most likely invisible and undiscoverable to users (comScore 2011; Arlitsch and O'Brien 2012). Coates (2014) emphasized the importance of having search engines index digital repository content, because according to her study on digital repository traffic, over two-third of visits are from search engines. Hagans (2005) described search engines as users with substantial constraints who cannot read text in images, Javascript or Applets, or many other multimedia formats. Therefore, in order to be indexed by search engines, digital repositories need to be receptive to search engine crawlers (Hagans 2005). Shoukat, Iftikhar, and Haseeb (2011) observed that it usually takes the search engine crawlers five to thirteen days to index the information after the newly published content become available on the Internet, and the indexing covers 65 percent to 75 percent of the content.

Dawson and Hamilton (2006) introduced the concept of adopting Google-friendly metadata in digital collections, but they also pointed out that the concept might not be able to completely address the indexing issues. A later study conducted by Drewry (2007) proved what Dawson and Hamilton (2006) suspected, demonstrating an inadequate use of metadata by the search engines Google and Google Scholar. Based on the specificity of the search queries that brought users to the digital collections, Coates (2014) advocated for file full-text indexing as

more helpful in discovering resources and directing users than metadata indexing alone. In a recent study of digital collections' metadata effectiveness in Internet discovery, Yang (2016) found that certain metadata fields, including title, description, and subject, are indexed more by Internet search engines than other fields. However, Yang (2016) also pointed out the study might be affected by PDF full-text indexing. In other words, the visit from search engines can be caused either by metadata indexing or PDF full-text indexing.

Beel, Gipp, and Eilde (2010) supported the use of metadata for Internet discovery, stating that this use of metadata is especially important for academic search engines crawling the web. They explained that the only way for the search engine crawler to identify the PDF files is through the metadata with extracted information about the title and author (Beel, Gipp, and Eilde 2010). In 2010, Arlitsch, O'Brien, and McIntyre conducted research on Mountain West Digital Library and found that only 38 percent of digital items were in Google's index and none were in Google Image's index when searching by title. Worse, as the authors pointed out, was that only 15 percent of the clicks on the results pages of the Google search engine provided the users with direct links to the specific digital object, while most of them consisted of links directing back to a search results screen on the local digital repository (Arlitsch, O'Brien, and McIntyre 2010). The finding supported what Beel, Gipp, and Eilde (2010) insisted, reinforcing the importance of metadata in search engine indexing.

Because libraries have not developed a fully effective mechanism to aggregate, index, and search the institutional repositories, Google Scholar has become the best search engine for digital repository content (Arlitsch and O'Brien 2012). Google Scholar has a specific indexing focus, which is separated from its larger parent Google, limiting the results to peer-reviewed publications, patents, and legal cases (Assisi 2005; Arlitsch and O'Brien 2012). Google Scholar

integrates documents of different versions from different collections available on the Web, either from the commercial publishers or from the author's digital repository page, allowing visitor's selection of the appropriate versions (Lewandowski 2010). When comparing Google Scholar's indexing and abstracting service with other online search engines or library OPACs, many scholars did agree that Google Scholar's coverage has a bias, either towards the English language or towards a specific discipline. Despite this, Google Scholar is still considered the best search engine for abstracting and indexing services (Callicott and Vaughn 2005; Neuhaus et al. 2006; Mayr and Walter 2007; Walters 2007; Meier and Conkling 2008; Lewandowski 2010).

Beel, Gipp, and Eilde (2010) described Google Scholar as an "invitation based search engine," explaining that Google Scholar only indexes information from trusted sources. Once information is indexed by Google Scholar, Google Scholar will search the Web for corresponding PDF files, regardless of the place where the PDF files are published. If different PDF files of one article exist on the Internet, Google Scholar groups them to improve the article's ranking.

Arlitsch and O'Brien (2012) provided further explanation on Google Scholar's crawler, saying that the peculiarity of Google Scholar is that it indexes full-text documents and provides direct links to the PDF documents, avoiding metadata, institutional logos, and other information that are normally displayed to users. In their study, the authors proposed a more efficient method to address the issues of indexing, suggesting that transforming metadata schemas to more precise bibliographic information within the HTML page header tags can increase the indexing ratios of institutional repositories in Google Scholar. However, they also claimed that some other factors such as slow or misconfigured servers, crawler errors, inappropriate server response codes, etc., will affect the search engine indexing.

PROBLEM STATEMENT

In the current literature, researchers agreed that it is important for digital collections to be indexed by Internet search engines in order to be discovered. However, researchers did not agree on whether metadata indexing or file full-text indexing plays a more important role in search engine discovery. A limited number of studies discussed search engine optimization strategies for digital collections by improving metadata or transforming the schema. Some researchers evaluated and compared different search engines' indexing and discovery services for specific disciplines and stated that Google Scholar is by far the best search engine for digital repositories.

The question that remains is whether search engine crawlers have a preference for digital collection metadata or for attached files' full-text content. No literature was found comparing the indexing effects of digital item metadata and attached PDF file. The metadata librarians and digital services librarian at Texas Tech University felt the need to better understand different search engines' indexing effects on digital collection metadata and associated PDF files. As a result, the author employed a method to assess the indexing and discovering effects of four search engines on different digital collections that are customized specifically for the research. Findings from this exploratory study can help the librarians to set priorities for digitally publishing works.

RESEARCH DESIGN

Because of the availability of different repository systems and the varying levels of academic prestige of institutions that may or may not affect the indexing of search engines, it is extremely difficult, if not impossible, to carry out a research project that covers all options to

reach a generalized conclusion. Standards of selecting repository tools or search engine's preferences and emphasis on certain academically prestigious institutional repositories were not discussed in this exploratory study.

The author believes that carrying out an exploratory study with a small sample size can benefit Texas Tech University's digital and metadata librarians by better understanding how the digital collections on a local DSpace-based institutional repository are being indexed and searched by search engines. The author also believes that the sample size should not be the factor affecting the search engines' indexing preference on metadata or PDF full-text— a bigger sample size is not going to change the indexing results, metadata or PDFs, on the items being indexed, discovered, and linked.

The author retrieved ten copies of electronic theses and dissertations (ETDs) that had just been submitted in May 2015. Using the most recent ETDs guarantees these papers have never been indexed by the search engines. The author created three new digital collections, Collection 1, Collection 2, and Collection 3 in the digital repository DSpace, and published the ten papers to each of the three newly created digital collections within a two-hour timeframe.

The digital items in Collection 1 were implemented with complete metadata based on the information retrieved from the ETDs, including Dublin Core metadata fields Title, Author, Abstract, Date, Format, Language, Subject, Type, and Rights. Collection 1's digital items contained only metadata but no PDF files. In Collection 2, the digital items do not contain any other metadata fields except for the two required metadata fields Title and URI Identifier. For the Title metadata field, the author input "N/A" for each digital item. The author then uploaded the PDF files to each digital item. By doing this, each digital item in Collection 2 contained only

PDF files but no metadata information. In Collection 3, the author implemented complete metadata and uploaded the PDF files to each digital item. Collection 3 is the only collection containing both valid metadata information and full-text PDF files with real ETD content. A summary of the three collections is presented in Table 1.

[Place Table 1 here]

The Title and Abstract sections contained unique information for each digital item. Full title and full-text abstract were included in both PDF files and implemented metadata. Based on Shoukat, Iftikhar, and Haseeb (2011)'s findings that search engine crawlers need at most thirteen days to index the new information published on the Internet, the author waited for more than three months before moving forward to the next step, so that the search engine crawler had enough time to discover and index these newly published items,

Through Google Analytics, the author found that 99 percent of the search engine traffic coming to the digital repository was brought by four search engines (See Figure 1), where Google sent 82 percent, Google Scholar 15 percent, and Yahoo and Bing each had 1 percent. The author thus selected these four search engines for further test searches of each published ETD. In the research, the author did not use or combine any advanced operators when performing searches on these four search engines, neither did the author use targeted search "site: url", because this command sends a query over the DSpace built-in search tool and brings back corresponding URLs that navigate to the DSpace search results page, where the items have been indexed by DSpace internal indexing tools, not by the search engines. Moreover, when patrons perform Internet searches they wouldn't know in advance where the source site might be.

[Place Figure 1 here]

Instead, the author used the full text of the title and abstract of the ETDs as search queries, which contain exactly the same text and keywords, for search engines to discover matched results should these resources be indexed. Because the author was search accurate title text and abstract text, Google Scholar, Yahoo, and Bing brought back a limited amount of results, relevant and/or irrelevant, within only one page. However, Google brought back hundreds of results, mostly irrelevant, in several hundred results pages. According to Google (2015), Google machines search the index for matching pages and return ranked results that are the most relevant, which is determined by over 200 factors. Based on this, the author examined the first page, identified and visited the target items, and recorded which links on the results page directed to which of the three collections. Because the author used the exact title and abstract text for the search, it was assumed that if the most relevantly ranked results of targeted links did not show up in the first page, the results wouldn't show up in the rest of the hundreds of pages either.

RESULTS

The author searched the titles of ten digital items on the Google search engine and browsed the result pages for the target items by identifying the URL sources with the digital repository. As shown in Table 2, five items from the metadata-only Collection 1 were successfully indexed and discovered by Google. The author was able to follow the links on the search results page to those five digital items. As for the metadata-and-PDF Collection 3, Google discovered five items' metadata as well as seven PDF files, as the author was directed to the PDF files directly and the browser triggered PDF downloads. The results also showed that Google

missed discovering two items' metadata in Collection 3, but successfully indexed and discovered their associated PDFs. Those items containing only PDF files but no valid metadata in Collection 2 were not indexed or discovered by Google at all.

[Place Table 2 here]

Google Scholar appeared to discover only titles from metadata but avoided searching or indexing titles from PDF files (see Table 3). Two items containing only metadata in Collection 1 were successfully discovered and linked to when performing the title search on Google Scholar. While in metadata-and-PDF Collection 3, five items were discovered by Google Scholar but the links directed only to the digital item pages instead of to PDF files. In PDF-only Collection 2, no items were indexed or discovered by Google Scholar.

[Place Table 3 here]

The Yahoo search engine discovered three items that belonged to metadata-only Collection 1 and four items that belonged to metadata-and-PDF Collection 3 (see Table 4). The Yahoo search engine did not discover any of the PDF files in PDF-only Collection 2, nor the PDF files in Collection 3 where metadata was supplemented. The Bing search engine discovered two items from metadata-only Collection 1 and two from metadata-and-PDF Collection 3 (see Table 5). Bing did not discover any PDF files in the PDF-only collection 2.

[Place Table 4 here]

[Place Table 5 here]

Besides discovering the results of title searching on the four different search engines, the author continued to test the abstract searching by employing the same search technique on the four search engines, trying to find the results of full-text indexing of the four search engines.

As shown in Table 6, four items' metadata from metadata-only Collection 1 was discovered by Google when searching the text of abstracts of digital items, and item pages were all linked to from the results page. Six items' metadata and six PDF files in metadata-and-PDF Collection 3 were also indexed by Google when performing the abstract searching. The author also noticed that although the same number of item's metadata and PDF files in Collection 3 were discovered by Google, only three of them were the same items. This means that in Collection 3, Google discovered three items' metadata but ignored indexing the associated PDF files, while it discovered another three items' PDF files but avoided their supplemental metadata. Google did not index or discover any items in PDF-only Collection 2.

[Place Table 6 here]

Table 7 shows that when searching abstracts on Google Scholar, three digital items' metadata in metadata-and-PDF Collection 3 were found. Google Scholar did not index any other metadata from metadata-only Collection 1, nor any PDF files from Collection 2 and Collection 3.

[Place Table 7 here]

Tables 8 and 9 show the results of abstract searching on Yahoo and Bing. Yahoo's search engine discovered one item's abstract metadata from metadata-only Collection 1 and two items' abstract metadata from metadata-and-PDF Collection 3. Yahoo's search engine did not discover any abstracts in PDFs that belong to Collection 2 or Collection 3. Bing discovered only one

abstract's metadata from Collection 3. Other than that, Bing did not discover any abstracts in metadata or PDFs.

[Place Table 8 here]

[Place Table 9 here]

Table 10 summarizes the results and discovering rates from this exploratory study on Texas Tech University's DSpace-based Institutional Repository. For Collection 1, the metadata-only collection, Google had a 50 percent disvoering rate on the title metadata and 40 percent discovering rate on the abstract metadata. Yahoo had a 30 percent discovering rate on the title metadata and 10 percent discovering rate on the abstract metadata. Google Scholar and Bing failed to discover the abstract metadata and had a 20 percent discovering rate on the title metadata. For Collection 2, which contained only PDF files but no valid metadata, all search engines failed to discover any information from this collection.

[Place Table 10 here]

For Collection 3, the metadata-and-PDF collection, Google was the only search engine that had successfully discovered titles (70 percent) and abstracts (60 percent) in PDF files. Google's discovering rates on the PDF files in Collection 3, supplemented with metadata, are also the highest discovering rates for the entire study. The other three search engines (Google Scholar, Yahoo, and Bing) did not discover PDF files at all, regardless of implemented metadata or not.

Regarding metadata discovery in Collection 3, Google was still the most successful search engine, having a 50 percent discovering rate on title metadata and a 60 percent rate on abstract metadata. Google Scholar came in second with a 50 percent discovering rate on titles

and a 30 percent rate on abstracts. Yahoo had a 40 percent discovering rate on titles and a 20 percent rate on abstracts. Bing was last, having a 20 percent discovering rate on titles and a 10 percent rate on abstracts.

DISCUSSION

Results of this exploratory study demonstrate some substantial findings that are contradictory to some generalized statements found in the existing literature. One of the most important findings is that in this case study Google is the only search engine that indexes, discovers, and links to PDF files, when the content is supplemented with metadata, and Google actually has a higher discovering rate on PDF files than on metadata. Google is also the only search engine that navigates the visits directly to PDF files instead of metadata pages. Google Scholar did not discover any PDF files published in these three collections at Texas Tech University's DSpace-based Institutional Repository, nor did Google Scholar provide direct links to any of the PDF files. Instead, Google Scholar discovered fewer metadata than Google from the titles and abstracts (See Table 10), and directed the visits to the digital item's pages. This finding contradicts what Arlitsch and O'Brien (2012) stated about how Google Scholar's crawlers index repositories and how Google Scholar (GS) "generally provides a link directly to the PDF document" in their article, noting that:[27]

> *"A peculiarity of GS's presentation of academic papers is that it generally provides a link directly to the PDF document… In other words, metadata, institutional logos, and other information normally displayed to users are lost unless they are inserted into the PDF itself."*

Results of this exploratory study also find that Google, compared to the other three search engines, has the best discovering rates on both metadata and PDFs that are published in Texas Tech University's DSpace-based digital repository. Google also discovered more PDF files than metadata in the same collection where both are available. These findings suggested that Google Scholar is not the best search engine for digital repositories, at least not in this case, unlike what was asserted by Arlitsch and O'Brien (2012):[28]

> *"Libraries have not developed a mechanism to aggregate and search IRs, and thus GS has become the best de factor search engine available for IR content."*

Metadata and PDF files are both important in digital collections. Indeed, they supplement each other to facilitate search engine indexing and the discoverability of digital items. Metadata plays a significant role in getting digital collections indexed by Internet search engines, because as this exploratory study's results show, search engines do not discover those digital items associated only with PDF files but which do not have any valid metadata. In contrast, digital items with metadata but without any PDF files still have a chance to be indexed and discovered by search engine crawlers. However, associating PDF files to digital items helps the digital collections have a better chance to be indexed by search engine crawlers. Implementing metadata and attaching PDF files to digital items appears to be the best strategy for making digital collections discoverable on the Internet.

In spite of the suggested best strategies, the author also noticed some exceptions in the study results. For example, the search engine discovered and provided a link to Item 5 in metadata-only Collection 1, but did not discover or provide a link to the same item in metadata-and-PDF Collection 3 (See Table 2 and Table 6). It is possible that search engines use some

mechanisms to eliminate links to items described with the same metadata. The search engine companies are by no means willing to share with the public or the researchers their algorithms, a specific-designed future research is thus needed to find out the answers to this question.

CONCLUSION

Digital collection discoverability on Internet search engines with a focus on metadata has been discussed for decades. Search engine optimization strategies have also been introduced to digital collections and institutional repositories, aiming to make digital content more search engine friendly. Researchers have discussed the transformation of metadata schemas and full-text PDF indexing in order to structure a correlation between search engines and digital repositories. However, researchers have not yet reached an agreement on whether metadata or full-text PDFs are more important for search engine indexing and discovering process. Many researchers have advocated for Google Scholar as the best search engine for journal articles and digital repositories since the Google Scholar crawler can avoid metadata to index PDF files and can compile results that include the discovered items along with direct links to the PDF files.

Through this study, the author also identified some findings contradictory to what has been generally concluded in previous research. The findings from this case show that Google Scholar does not discover any PDF files published in the digital collections, nor does Google Scholar provide any direct links to PDF files in the digital collections. The results from this study show that Google has the best discovering rates on digital items for both metadata and PDF files, and Google is the only search engine that indexes the full-text of PDF files and provides direct

links to the associated files. Google discovers more PDF files than metadata in the collection where both are available to the search engine crawler.

As a result of these findings, it is recommended that librarians of digital resources conducted research on their local digital repositories, as the indexing effects and discovering rates on metadata or associated text files may be different case by case. Learning from the research on local digital collections can help develop a more customized metadata strategy for specific institutional repositories.

While Google Scholar may filter out non-scholarly materials and index only trusted institutional repositories, researchers using it should understand that coverage of relevant resources might not be as good in Google Scholar as it is in the general Google search, as this study shows. Researchers might also want to avoid using Yahoo and Bing for any academic materials, because based on this research those two search engines seem to be the most unreliable with low discovering rates. Academic libraries have invested much effort in publishing scholarly publications in their digital repositories and hope to make these valuable digital assets discoverable online. Librarians who work on institutional repositories should be implementing complete metadata and uploading PDFs online in order to give digital items the best chance of being searched.

There are some limitations in this research that can possibly affect the research results and the conclusion. The sample size of this exploratory study is relatively small, which could potentially be a reason why the findings differ from previous research. Sample size does not impact search engines' indexing preference, and when analyzing those already-indexed and already-discovered items on the search engine results pages, the sample size of the study should

not change the results of whether metadata or a PDF's full-text is indexed or discovered. However, future studies may be needed to demonstrate this principle in a wider range of contexts.

In addition, this study only looked at Texas Tech University's DSpace-based Institutional Repository. Other digital repository systems and other institutional repositories may react differently with search engines. Thus, the author is cautious about drawing generalizable conclusions and offers this case study as another viewpoint to help promote a better understanding of what metadata strategies might work best for our local repository and to encourage future research. The findings and conclusions in the article are reflections of what the author found based on this exploratory study and will be used by the author locally when developing digital collections.

Further studies are still needed to rectify the limitations of this research. A series of collaborative studies on different repository systems from different institutions are needed, as it is necessary to see a broader comparison on the indexing and discovering effects of metadata and PDF files. It is also noted that no search engine successfully indexed or discovered 100 percent of the three experimented collections; future research is needed to determine why crawlers index some content and ignore the rest. Search engines are commonly known for doing PDF full-text indexing, while the PDF-only collection without any metadata received no discovering at all. Conducting studies on other formats of associated files, such as PPT and RTF, would also enrich this particular aspect of literature.

As a metadata librarian working at a digital resources library, the author and colleagues wanted to establish a better understanding of search engines' indexing and discovering effects on digital content's metadata and associated PDF files. The goal was for the author and colleagues

at a digital resources library to be able to use this information to set priorities for digital collection strategies. In order to explore whether search engines have a preference in indexing Texas Tech University Institutional Repository's metadata or associated full-text PDFs, the author carried out this exploratory study by creating three digital collections and publishing digital items for testing the indexing feature of search engines. By comparing the test results on four search engines, Google, Google Scholar, Yahoo, and Bing, the author discovered that metadata and PDF files are indeed supplementing each other in digital collections. Search engines did not index PDF files in which no valid metadata were implemented; however, valid metadata helped more PDF files in the collection get indexed and discovered by search engines.

REFERENCE

Assisi, Francis C. 2005. "Anurag Acharya Helped Google's Scholarly Leap." Accessed January 16, 2016. http://www.indolink.com/SciTech/fr010305-075445.php/.

Arlitsch, Kenning, Patrick S. O'Brien, and Sandra McIntyre. 2010. "Search Engine Optimization for Digital Collections." Accessed August 11, 2015. http://www.diglib.org/wp-content/uploads/2011/01/25DLFseo.pdf/.

Arlitsch, Kenning, and Patrick S. O'Brien. 2012. "Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar." *Library Hi Tech* 30 (1): 60-81. doi: 10.1108/07378831211213210.

Beel, Jöran, Bela Gipp, and Erik Wilde. 2009. "Academic Search Engine Optimization (aseo) Optimizing Scholarly Literature for Google Scholar and Co." *Journal of Scholarly Publishing* 41 (2): 176-190. doi: 10.3138/jsp.41.2.176.

Callicott, Burton, and Debbie Vaughn. 2005. "Google Scholar vs. Library Scholar: Testing the Performance of Schoogle." *Internet Reference Services Quarterly* 10 (3-4): 71-88. doi: 10.1300/J136v10n03_08.

Coates, Mildred. 2014. "Search Engine Queries Used to Locate Electronic Theses and Dissertations: Differences for Local and Non-local Users." *Library Hi Tech* 32 (4): 667-686. doi: 10.1108/LHT-02-2014-0022.

ComScore. 2011. "ComScore Releases September 2011 U.S. Search Engine Rankings." Accessed August 11, 2015. http://www.comscore.com/Insights/Press-Releases/2011/10/comScore-Releases-September-2011-US-Search-Engine-Rankings/.

Dawson, Alan, and Val Hamilton. 2006. "Optimising Metadata to Make High-value Content More Accessible to Google Users." *Journal of Documentation* 62 (3): 307-327. doi: 10.1108/00220410610666484.

Drewry, Josiah M. 2007. "Google Scholar, Windows Live Academic Search and beyond: A Study of New Tools and Changing Habits in ARL Libraries." Accessed January 16, 2016. https://ruby.ils.unc.edu/MSpapers/3310.pdf/.

Google (2015). "How Google Search Works." Accessed March 5, 2016. https://support.google.com/webmasters/answer/70897?hl=en.

Hagans, Andy. 2005. "High Accessibility is Effective Search Engine Optimization." Accessed August 11, 2015. http://www.alistapart.com/articles/accessibilityseo/.

Lewandowski, Dirk. 2010. "Google Scholar as a Tool for Discovering Journal Articles in Library and Information Science." *Online Information Review* 34 (2): 250-262. doi: 10.1108/14684521011036972.

Mayr, Philipp, and Anne-Kathrin Walter. 2007. "An Exploratory Study of Google Scholar." *Online information review* 31 (6): 814-830. doi: 10.1108/14684520710841784.

Meier, John J., and Thomas W. Conkling. 2008. "Google Scholar's Coverage of the Engineering Literature: An Empirical Study." *Journal of Academic Librarianship* 34 (3): 196-201. doi: 10.1016/j.acalib.2008.03.002.

Neuhaus, Chris, Ellen Neuhaus, Alan Asher, and Clint Wrede. 2006. "The Depth and Breadth of Google Scholar: An Empirical Study." *Portal: Libraries and the Academy* 6 (2): 127-141. doi: 10.1353/pla.2006.0026.

Shoukat, Ijaz Ali, Mohsin Iftikhar, and Abdul Haseeb. 2011. "Indexing Size Approximation of WWW Repository with Leading Information Retrieval and Web Filtering Robots." *International Journal of Advanced Research in Computer Science* 2 (3): 71-75.

Yang, Le. 2016. "Metadata Effectiveness in Internet Discovery: An Analysis of Digital Collection Metadata Elements and Internet Search Engine Keywords." *College and Research Libraries* 77 (1): 7-19. doi: 10.5860/crl.77.1.7.