

Searching in the Wild: Observing Information-Seeking Behavior in a Discovery Tool

Heather A. Dalal, Amy K. Kimura, and Melissa A. Hofmann

Introduction

Knowing students' preference for single-input search engines such as Google, libraries have been turning to so-called "discovery tools," providing a Google-like search box for finding and accessing library content. Students, however, often lack understanding of the differences between Google and a discovery tool, not grasping that underneath the Google-like smoke-screen is a complex set of databases, indexes, catalogs, media, and journals where format matters, full text is not guaranteed, and results are returned by relevance instead of popularity. While "single search" discovery tools sound appealing to both students and librarians because they simultaneously search most library content, conducting academic research is far more complicated than constructing keyword searches. We wanted to find out how appropriate discovery tools are for students and novice researchers. Discovery tools are potentially a great place to start, but do they leave students overwhelmed with too many resources?

We connected with Professors Beth Bloom and Marta Deyrup at Seton Hall University Libraries. The pair have gained national attention with their Google-sponsored research entitled "The Truth is Out: How They Really Search."¹ They invited Rider University to participate with them on the discovery tool research following a very similar methodology. Our data has been collected concurrently and as a larger group we

will be conducting an ethnographic analysis of the data. The results presented in this paper are initial results from our slice of this larger project, specifically how they have impacted both our discovery tool administration and information literacy instruction.

As we set out to answer our questions, we discovered gaps in the students' knowledge that we did not expect. Upon voicing this to classroom faculty and library colleagues, we found they understood that college students in general had limited research skills but that many did not quite believe students at our own university would also exhibit such low levels of information literacy. We were not expecting to see so many problems, especially among our research participants, who as upperclassmen and graduate students were confident, experienced student researchers. Our observations verified the findings of others, such as the lack of understanding of constructing search queries and employing advanced searching techniques.² We also observed a strong inclination for favoring full-text articles and sacrificing of content for convenience, previously reported in the library literature.³

Methodology

Traditional usability studies of library discovery tools, geared toward gauging the effectiveness of the tool itself, involve a set of predefined questions or tasks.⁴ Our study, which allowed participants to "range free-

Heather A. Dalal is Assistant Professor-Librarian, Franklin F. Moore Library, Rider University, e-mail: hdalal@rider.edu; Amy K. Kimura is Assistant Professor-Librarian, Katharine Houk Talbott Library, Westminster Choir College of Rider University, e-mail: akimura@rider.edu; Melissa A. Hofmann is Associate Professor-Librarian, Franklin F. Moore Library, Rider University, e-mail: mhofmann@rider.edu

ly” as they made use of the discovery tool to pursue individualized research topics, is more ethnographic in nature.⁵ Using the methodology employed by Bloom and Deyrup, our student participants utilized OpenHallway, an unobtrusive web-based tracking tool that records voice and keystrokes.⁶ Participants logged in at their own convenience, in a venue of their choice, to record an hour of their research (in three 20-minute segments) for a course-assigned research project while providing verbal commentary. The typical goal of a usability study is to collect data to drive design of a tool. In our case, the aim was not so much to influence design but to improve the administration of our discovery tool and to develop more effective information literacy instruction. Typically in usability testing, only five participants are considered necessary for an effective study; more important, however, is the quality of the participants’ feedback.⁷ OpenHallway’s feature of recording narration allowed us to gather particularly insightful feedback from many of our participants.

Students were recruited via professors and campus mailing lists. They applied to participate in the project, providing the details of the research assignment they planned to use for this study. Upon approval, students were familiarized with using the OpenHallway software and instructed to start their research using the library’s discovery tool. In the fall of 2013, twenty-four students were trained, but only seven completed the project. Each was awarded a \$25 gift card incentive for completion of the project. Hoping to gain more participants for the spring of 2014, we increased the incentive to \$50, yielding nineteen trained applicants, but only four students completed the project.

Our group of participants was composed of juniors, seniors, and graduate students, all of whom have had experience searching for resources, information literacy instruction in either college or high school, and consultations with reference librarians. A survey conducted after the completion of their OpenHallway sessions revealed that most felt confident about their ability to find information in the library. Having such an experienced group initially seemed

like it might be a detriment to our research; we encountered some very sophisticated research problems that even non-library faculty might struggle to understand. Overall, however, the level of these students’ experience and capabilities revealed serious struggles with basic concepts of information literacy. Reviewing their recordings and sharing them with library and classroom faculty caused a considerable amount of surprise and dismay.

Results & Discussion

Reviewing the eleven hours of OpenHallway recordings, we identified three key areas as being problematic for students: the lack of an overall understanding of the interrelationship among the library catalog, databases, the discovery tool, and scholarly publications; failure to implement advanced search techniques (closely related to difficulties in defining topic and scope of research); and failure to locate the full text of articles.

Understanding of Research Tools and Scholarship

As we listened to our students describe their research and thought process, we were intrigued by how they characterized the discovery tool and comprehended scholarly communication and publication. One described the discovery tool as a large database and another just used it as the library catalog. Another student missed our instruction to start his research within the discovery tool: as he searched effectively in Medline, PsycInfo, and Academic Search Premier, he protested that he kept finding the same articles in the different databases and wished there was some way to combine searches to eliminate redundant results. It does not surprise us that he was unaware that his need would be met with the discovery tool, as he was in his final year at the university and had learned to research using subject specific databases before the implementation of the discovery tool in 2013.

Two students provided specific reasons why they disliked the discovery tool. One complained that it had “too many articles in it,” and another vocalized

that she liked the *idea* of the discovery tool but preferred the narrower results provided by individual databases. We agree that these students' preference for using a focused database is fair, as lack of precision is one of the commonly cited disadvantages of web-scale discovery.⁸ We know that their searches would have been more successful, however, if they had the knowledge and skills to take advantage of the various facets with which to limit the results. Overall, facets and limits were rarely employed by our participants. No student used a limiter beyond Full Text, Peer-Reviewed, English Language, and Date of Publication.

A graduate student articulated the most accurate perception of the discovery tool, comparing it to the popular airline fare comparison website Kayak.com. While this is not a perfect analogy, it suggests that she understood the relationship between the discovery tool and the resources it searches. This understanding was ultimately to her detriment, because when she found no results on her topic, she declared, "I cannot wait to tell my professor that ... there is literally no information on [my topic] ... in Rider databases." In reality, the library had plenty of articles on her topic, but her searching technique made it very challenging to find.

Beyond a lack of understanding of the interrelationship between the library catalog, databases, and search tools, our study also revealed a lack of understanding of the scholarly publishing model, or even the purpose of scholarly research at a fundamental level. One student complained, "Something I don't like about these databases is that they always have something that has to do with a certain area, one certain area, whether is Syria, Britain or Iraq. It just always has to do with one thing. And sometimes they are not even in English." This description gave us an impression that she did not understand the tools or even the point of scholarship. She was probably seeking a general, encyclopedia-style entry on her topic, and she did not understand why more specialized topics were appearing in her search results, or even see the value of scholarly communication.

Searching Behavior

Our observations revealed several problems with the students' search strategies. Students exhibited improper use of quotes and/or Boolean operators in their search strategies, yet only a few students either expressed confusion or frustration as to why the results did not change the way they expected. The others simply continued searching and changing their terms until they finally found content they deemed relevant enough. Students vocalized frustration when reviewing their results lists, claiming the library did not have anything on their topic, completely unaware that their search strategies were to blame for the mismatched or unsatisfying results.

Failure to Make Proper Use of Advanced Searching Techniques

While students seem to have remembered that librarians suggested using Boolean operators to narrow search results, not all students made proper use of them. We observed students using quotes around a single word (e.g., "*liberalism*") or placing quotes around search terms with Boolean operators and not expecting that their results would be narrower (e.g., "*music profanity and generations*," "*music genres +decades*"). Each time the search terms were changed with little explanation. Some participants employed the use of the ampersand or plus sign instead of *AND* in the discovery tool. Since these searches yielded some results, the students did not recognize their errors. For example, *e-commerce & target marketing* yielded six results while *e-commerce AND target marketing* would have yielded almost fifteen thousand results. One such student found nothing when searching *coal industry AND history AND failure in business*, and paused to think about her keywords. She then confidently stated, "maybe I'm supposed to do OR." She changed the *ANDs* to *ORs*, which broadened the search. While she was encouraged by the increase in the number of results, she did not realize that something might be wrong as she reviewed the titles and complained that items were on history topics, not business, and sighed "why is there nothing about coal?" Yet, she continued

to wade through the results and to use *OR* throughout her entire search session. Her search results finally became somewhat manageable when she attempted to search for *coal industry* within the title field, even with the addition of *OR companies OR United States OR coal*.

Some students often opted for keyword searches, and most changed terms when a search when unsuccessful. Most often, their new choice for search terms seemed little more than a guess. None of the students observed or used the words found in abstracts or subject terms provided to change their search strategies. This is a disappointing oversight, as librarians model this strategy when teaching research instruction.

Searching in the Wild, with an Undeveloped Topic

While some students began the exercise with pre-defined research questions, others were still deciding upon a topic while participating in the study. Their choice of search terms hindered their ability to find useful resources, which in turn resulted in unfruitful hour-long research sessions.

A student searched for *l-commerce*, which she defined verbally as “location-based commerce.” Had she tried that phrase, she would have found plenty of results. When she failed to find results with *l-commerce*, she tried *m-commerce* (but never tried “mobile commerce”), finally broadening her search to *e-commerce*. She then attempted to narrow her results within e-commerce with the aforementioned ampersand Boolean operator (e.g., *e-commerce & psychology*). At the end of her session, she announced she was making an appointment with the business librarian.

Another student who appeared to have only a vague topic idea searched for *jfk* [sic] and then complained the articles were about the airport. This did not deter her from using the search term repeatedly (e.g., *jfk and realism*; *jfk and the missile gap*); she only once tried *John F. Kennedy* when it was presented as an option within the *New York Times* database. She also searched for *realism*, *liberalism*, *Cuban missile crisis*, and “*liberalism in international relations*.” Un-

fortunately at the end of the hour, we did not have a good idea of what her research topic was. It seemed like she really needed general information, and her switch to a Google search later in the recording was a good decision.

Searching in the Wild, with a Developed Topic

A student was interested in how music lyrics have changed from her parents’ generation to today. Having a relatively clear topic did not improve her searching in the discovery tool. Her search terms were *lyrics*; *music lyrics and generations*; *music profanity and generations*; *music genres +decades, 1950’s music, 1950’s rock and roll, 1950’s Elvis Presley, 1990’s popular music*. She would often find an article and have opinions about its usefulness, but she rarely retained any for reading later. By the end of the recording, she was exasperated when she tried a new search phrase: *change listening habits*. The results, naturally, were limited: three articles about hearing aids. Throughout the hour of searching, she did not get any closer to her research answers and lost confidence in her ability to get this assignment completed before she graduated.

One student in the study notified us she was nearly done with her research but was trying to find some primary data on underrepresented minorities in the healthcare industry to strengthen her conclusion. Her search terms in the discovery tool were *demographics US ethnicity* [sic]; *demographics US*; *physican demographics*; *US demographics*; *Language ability US*; *Gender US*; *Race/ethnicity provider*. Eventually with *us physicans facts* as her search phrase, she was able to find a study about minorities in emergency medicine and a book *Physican Characteristics and Distribution in the U.S.*

Not all the students in the study employed poor search terms. Several were very familiar with their topics and previously identified jargon or phrases that were very beneficial. For example, a student searching for *stereotype threat AND spillover* was very familiar with the authors in the field and displayed enthusiasm about some new findings. Another participant search-

ing for support for marijuana legislation seemed to have no issue locating relevant useful articles. However, all students had issues with accessing the full-text articles given a citation, as discussed below.

Issues of Access

Our observations verified others' findings that immediate, full-text access is of paramount importance to students. Few in our observed group, in fact, exhibited any willingness to consider sources available by other means.⁹ Reasons for this, when stated, generally related to the lack of time required to obtain other materials before the project's due date. Almost certainly another underlying reason is a lack of knowledge about how to find or acquire articles by other means. One student, addressing us directly during her narration, openly acknowledged what she perceived would be our disapproval of her last-minute approach: "This [article] looks great except that there's no link to the pdf. I know that's super lazy, and you probably hate me for it, but I don't have time to go searching for it."

Even when readily available, however, students frequently failed to find the full text of an article. There were three primary reasons for this:

- Failure to fully examine or understand the article-level results page (even when containing links to article- or journal-level access)
- Failure of the pdf, link, or link resolver to take the student directly to the full text (and an unwillingness to follow links at all due to past failures)
- Subsequent failure to pursue an article via its citation, even in cases where full-text access would have been possible

These issues are discussed in more detail below. Essentially, and distressingly, *any* obstacle to one-click, immediate, full-text access ruled out that source for the majority of the students we observed.

Failure to Fully Examine or Understand the Article-Level Results Page

Depending on the database vendor from which the discovery tool is displaying results, the appearance

and placement of links can vary. Full-text access is often available via the brief record on the search result page, with the text "HTML Full Text" or "PDF Full Text" and a corresponding icon appearing directly below the citation; this is called native full text.¹⁰ On the full record view, these links appear in a menu to the left, with the HTML full text also automatically displaying below the citation. Native full-text access is the simplest and most frequently successful mode of access. Still, the fact that a results list might have some variation in appearance seems to present a challenge to students who are perhaps used to the standardized appearance of highly-frequented sites such as Google and Wikipedia; especially in their haste to complete assignments, the depth of familiarity with certain page layouts seems to prevent some students from absorbing the full contents of a search results page from the discovery tool. For example, one student viewing a citation failed to click on any of the three links presented to her directly below that citation (two "Full-Text Resources" and one "Journal Level Link"), claiming that "I just can't seem to find where it is... usually it will say it's at one of the libraries, but it's not popping up here. It's kind of frustrating. I'm actually probably not even going to bother. I'm sure I can find something else that's more convenient and easily accessible."

Failure of the Link or Link Resolver to Take the Student Directly to Full Text

Much greater difficulty arises when indirect links to articles come into play. These links appear in a variety of guises and behave in a number of different ways. For example, our discovery tool uses both custom links, set up with particular syntax to access a specific resource (such as JSTOR), and link resolver links. Link resolvers are software that interprets OpenURLs, connecting with a knowledge base of e-resource holdings to facilitate end user access to the large volume of publications typically included in an academic library's collection. Knowledge bases are extremely difficult to maintain due to the "immense, irregular, and quirky universe of published scholarly content."¹¹

Our links have different wording, depending on the resource to which it connects. This was a deliberate attempt on our part to be transparent as to the origin of the full-text provider, but the differences in terminology may be confusing to the student.

At the time this study was conducted, Rider's link resolver used the text "Get it @ Rider" for articles with full-text availability, similar to other institutions, with the intent of branding access to electronic resources. We observed that this terminology posed difficulties for several students. Students do not equate this as an option to finding the full text, a problem also identified by Imler and Eichelberger.¹² A number of them misunderstood it completely: one believed that it meant that the item was either physically located in the library ("I don't have time to go to the library, I need something now") or that it functioned like interlibrary loan ("I don't want to get it at Rider, I need things that Rider already has"). This was even the case with a student who had otherwise significantly above average information literacy skills. While scrolling down a list of results based on a decent search strategy, he muttered to himself "These are 'Get it at Rider,' so no, no, no..." Another student noted: "I always click on that and I don't understand where it takes me."

For users who did get as far as clicking on the link resolver, the experience was negative overall. Link resolver failure rates are relatively high (Trainor and Price reported a failure rate of 29%¹³), because they rely on a long chain of interdependent data, most of which is controlled by different entities. Incomplete or incorrect data anywhere along the linking chain will cause a failure, resulting in either false positives (linking to an item unowned by the library) or false negatives (failing to link to an owned item).¹⁴ From the end-user perspective, these failures are exceedingly frustrating, especially when encountered repeatedly. Two links on the same article (for example, the "JSTOR Full Text" icon, which is a custom link set-up in our discovery tool, and the "Get it @ Rider" link resolver), appearing next to each other, can take users to different destinations, some of which succeed and some of which fail. It can be confounding even for librarians

to determine where a link has taken them (why am I facing a paywall for a journal to which I know our library subscribes?), so it is no surprise that a student would have little patience for these perplexing results. One student in our study did not even bother to click on the link resolver because it failed so often; another student expressed surprise when it actually directed her where she expected. A third student avoided anything in JSTOR because she believed, based on previous experiences of navigating to that database via the discovery tool, that it "usually" required payment for articles. Additionally, the look and feel of these links changes over time as adjustments are made to the discovery tool and by content providers (which may not be controllable by the library), adding another layer of difficulty to which users are highly resistant. Trainor argues that false negatives damage the students' perception of the link resolver *and* the library, especially when students subsequently find the full text out on the open web.¹⁵ One of our students experienced just this when she Googled the pdf herself, thus scolding the library: "This is ridiculous, I don't understand why there's no pdf here [in the discovery tool]. I could easily upload it if I had access to the code."

Students expressed genuine anger toward the University and the Libraries for "teasing" them with supposed access to full-text material which turned out to be, as far as they could discern, unavailable. Abstracts, while appealing to some students, caused frustration for others who seemed to feel that the provision of minimal information was just another way to "trick" them into believing that the whole article would be available. Databases of abstracts (e.g. RILM Abstracts of Music Literature) further complicate the situation, because employing the "full text" limiter in the discovery tool does not eliminate the presentation of these abstracts in search results, regardless of availability of the full *article* text.

Failure to Pursue an Article from a Citation

Provided a citation but no direct "full text" link, most students lack either the patience or the knowledge (or both) to locate the article, in print or online, even

when it is freely available to them. Besides a simple lack of time or laziness, this recalls our observation that many students appear to lack a fundamental understanding of the relationships among articles, journals, databases and search tools, and of the underlying scholarly publishing model. There seems to be a notion among some students, for example, that scholarly articles reside *in* Google or Google Scholar, or that a discovery tool is simply a very large database; they do not distinguish between a database and a search engine. Link resolvers, when they work, make the process of locating an article opaque; they allow users to “cheat” by bypassing that knowledge. The cost of this opacity is that users who depend on single-click access to articles may be unable to find that information when a link resolver fails.

What are some of the takeaways regarding access? First, we need to help students understand citations (and the scholarly publishing model, including copyright issues, and the relationship between articles, journals, databases, and search tools) and how to pursue resources that interest them. This can be a challenge, even for experienced users; changing platforms, lapses in subscriptions, and embargoes on access complicate the task.¹⁶ Plainly, no amount of tweaking the technology to make access easier or more transparent will replace the need for effective teaching if we want our students to be information literate.

On Campus Impact Faculty Engagement

We presented this research twice formally on campus to faculty outside of the library in a session entitled “Shocking Secrets of the Student Researcher.” Within the presentation, we played excerpts of the participants’ OpenHallway recordings. One composition instructor, who has always been very engaged in information literacy, partnering with librarians each semester, told us that this changed everything she knew about the students’ information literacy skills. She admitted she expected upperclassmen to be much more experienced. In addition, an English professor wrote:

“I was virtually entranced by the video-based data...The misconceptions evident even among high-achieving seniors...are vivid and crucial evidence of where and how faculty need to work more intentionally to help students achieve comfort with basic research skills in an increasingly complex and potentially confusing information environment...This is exactly the sort of research that can enable more fruitful partnerships between classroom instructors and library faculty and thus have a direct effect on students’ learning.”¹⁷

As we continue to collect data of this sort and share results on campus, we anticipate classroom faculty will be convinced that our students need more training and experience with information resources. Today’s ease of finding copious sources of information is a double-edged sword: the ease of access means that our students are stuck trying to filter out relevant, high-quality information using flawed searching behaviors and a misunderstanding of scholarly communication.

Discovery Tool Administration

Sites of student confusion have informed the administration of the discovery tool. For example, parenthetical wording was immediately added to the “Get it” text to emphasize immediate access: “Get it @ Rider University Libraries (via our ONLINE or print subscriptions).” False links to publishers are a known problem and are fixed as soon as discovered. Higher impact changes will be considered at the end of the academic year and as we evaluate the discovery tool’s integration with our new library catalog, introduced in January.

What Has this Done for Our IL Instruction?

These videos have informed information literacy instruction at Rider University. Classroom faculty and librarians alike were also shocked by students’ search behaviors and have questioned our information literacy instruction. We need to teach more effectively:

although our classroom time is usually severely limited, here is more evidence that teaching the mechanics is necessary but not merely enough. We need to teach more slowly and with more repetition to ensure real mastery of even the most basic concepts. We feel obliged now to explain, for example, why databases contain only abstracts, or why full text may not be immediately available, and that we are not attempting to “bait and switch” our frustrated students. We must focus on the development of appropriate search terms, narrowing with facets and limiters, emphasizing critical thinking, and pursuit of the full article.

Already, a lot has been gained by using videos of *our* students using *our* own library. With continued analysis and research, we hope to teach more effectively, both at the reference desk and in the classroom, and to foster more faculty engagement in information literacy instruction and research support, because students need more than one guide, out here in the wild.

Notes

- Beth Bloom and Marta Deyrup, “The Truth Is Out: How Students REALLY Search,” *Proceedings of the Charleston Library Conference* (2012): 203-208, doi: 10.5703/1288284315103
- Teri Oaks Galloway and Mary Finnan Hines, “Competitive Usability and the Catalogue: A Process for Justification and Selection of a Next-Generation Catalogue or Web-Scale Discovery System,” *Library Trends* 61, no. 1 (2012): 173-185, doi: 10.1353/lib.2012.0027; Kelly Meadow and James Meadow, “Search Query Quality and Web-Scale Discovery: A Qualitative and Quantitative Analysis,” *College & Undergraduate Libraries* 19, no. 2-4 (2012): 163-175, doi:10.1080/10691316.2012.693434
- Lynn Silipigni Connaway, Timothy J. Dickey, and Marie L. Radford, “‘If It Is Too Inconvenient I’m Not Going After It’: Convenience as a Critical Factor in Information-Seeking Behaviors,” *Library & Information Science Research* 33, no. 3 (2011): 179-190, doi:10.1016/j.lisr.2010.12.002; Chandra Prabha, Lynn Silipigni Connaway, Lawrence Olszewski, and Lillie R. Jenkins, “What Is Enough? Satisficing Information Needs,” *Journal of Documentation* 63, no. 1 (2007): 74-89, doi: 10.1108/00220410710723894
- Jody Condit Fagan, Meris A. Mandernach, Carl S. Nelson, Jonathan R. Paulo, and Grover Saunders, “Usability Test Results for a Discovery Tool in an Academic Library,” *Information Technology & Libraries* 31, no. 1 (2012):83-112, doi: 10.6017/ital.v31i1.1855; Megan Johnson, “Usability Test Results for Encore in an Academic Library,” *Information Technology and Libraries* 32, no. 3 (2013): 59-85, doi: 10.6017/ital.v32i3.4635.
- Jeffrey Rubin and Dana Chisnell, *Handbook of Usability Testing: How to Plan Design, and Conduct Effective Tests*, 2nd ed. (Indianapolis, IN: Wiley Publishing, 2008).
- Bloom and Deyrup, “The Truth,” 203.
- Jakob Nielsen, “How Many Test Users in a Usability Study?” Nielsen Norman Group, June 4, 2012, accessed February 18, 2015. <http://www.nngroup.com/articles/how-many-test-users/>
- Lisa M. Rose-Wiles and Melissa A. Hofmann, “Still Desperately Seeking Citations: Undergraduate Research in the Age of Web-Scale Discovery,” *Journal of Library Administration* 53, no. 2-3 (2013): 147-166, doi: 10.1080/01930826.2013.853493
- Connaway et al., “If It Is Too Inconvenient,”179; Prabha et al., “What Is Enough?,” 74.
- Cindy Trainor and Jason Price, “Rethinking Linking: Breaking New Life into OpenURL,” *Library Technology Reports*, 2010, <http://works.bepress.com/cinditrainor/2>
- Marshall Breeding, “Knowledge Base and Link Resolver Study: General Findings,” (Report commissioned the LIBRIS division of the National Library of Sweden.) May 1, 2012, http://www.kb.se/dokument/knowledgebase_linkresolver_study.pdf, p4.
- Bonnie Imler and Michelle Eichelberger, “Do They “Get it”? Student Usage of SFX Citation Linking Software,” *College & Research Libraries* 72, no. 5 (2011): 454-463. doi:10.5860/crl-141
- Trainor and Price, “Rethinking Linking,” p.5
- Ibid., p. 7.
- Ibid.
- Lydia Dixon, Lydia, Jody Condit Fagan, Meris Mandernach, and Stefanie E. Warlick, “Finding Articles and Journals via Google Scholar, Journal Portals, and Link Resolvers,” *Reference & User Services Quarterly* 50, no. 2 (2010): 170-181, doi:10.5860/rusq.50n2
- Arlene Wilner, personal communication, July 23, 2014.