ARTICLE

A Framework for Measuring Relevancy in Discovery Environments

Increasing Scalability and Reproducibility

Blake Galbreath, Alex Merrill, and Corey M. Johnson

ABSTRACT

Institutional discovery environments now serve as central resource databases for researchers in the academic environment. Over the last several decades, there have been numerous discovery layer research inquiries centering primarily on user satisfaction measures of discovery system effectiveness. This study focuses on the creation of a largely automated method for evaluating discovery layer quality, utilizing the bibliographic sources from student research projects. Building on past research, the current study replaces a semiautomated Excel Fuzzy Lookup Add-In process with a fully scripted R-based approach, which employs the stringdist R package and applies the Jaro-Winkler distance metric as the matching evaluator. The researchers consider the error rate incurred by relying solely on an automated matching metric. They also use Open Refine for normalization processes and package the tools together on an OSF site for other institutions to use. Since the R-based approach does not require special processing or time and can be reproduced with minimal effort, it will allow future studies and users of our method to capture larger sample sizes, boosting validity. While the assessment process has been streamlined and shows promise, there remain issues in establishing solid connections between research paper bibliographies and discovery layer use. Subsequent research will focus on creating alternatives to paper titles as search proxies that better resemble genuine information-seeking behavior and comparing undergraduate and graduate student interactions within discovery environments.

INTRODUCTION

There is no denying the ubiquitous nature and importance of discovery environments (DEs) to academic libraries. And further, "effective optimization of these search platforms should be one of the organization's core competences."[1] Uhl states it in the following way: "[T]he quality of the discovery layer is one of the most important elements in determining whether or not the library is successful in its mission to its users."[2] Whether or not libraries achieve their goals is complicated because libraries have lost control of information retrieval to "proprietary algorithms" now dictating how results are chosen and organized.[3] Our study and those of a similar focus, such as a recent project from five California State institutions, examine how different discovery environments address the important task of effective customization and how we should measure the overall quality of the DE.[4]

About the Authors

Blake Galbreath <blake.galbreath@wsu.edu> is Assistant Head of Systems and Technical Operations, Washington State University. **Alex Merrill** <merrilla@wsu.edu> is Head of Systems and Technical Operations, Washington State University. **Corey M. Johnson** <coreyj@wsu.edu> (corresponding author) is Instruction and Assessment Librarian, Washington State University.

Submitted: 18 August 2023. Accepted for Publication: 10 December 2023. Published 17 June 2024.

https://doi.org/10.5860/ital.v43i2.16915

BACKGROUND

University Common Requirements is Washington State University's (WSU) current general education program. It was launched fall term 2012 and asks students to take courses in 12 competency areas.[5] One such area features the only required course for all undergraduates, Roots of Contemporary Issues (RCI).[6] Courses in each competency area must address various combinations of the Seven Undergraduate Learning Goals.[7] A central learning outcome embedded in RCI is information literacy, which is defined as the ability to understand an information need, find and evaluate sources relevant to the need, and productively and ethically synthesize information to address the need.[8] RCI final research papers are the curricular content used in this study.

On the road to writing the RCI final paper, students engage with a set of scaffolded assignments which challenge them to develop their topics from general ideas to structured thesis statements, gather a set of topic-relevant sources (e.g., history monographs, history journal articles, newspaper articles, and primary sources), and learn about Chicago Style citation. The students, who are free to research the historical roots of topics of their choosing, frequently use WSU Libraries' discovery environment Primo (Ex Libris) as a central database of choice for any/all source needs.[9] The Libraries use the New User Interface version of Primo and its Central Discovery Index (CDI). In this study, the researchers evaluate the effectiveness of our locally customized version of Primo, using the titles of RCI papers as search queries, and final paper bibliography sources as a tool for measuring patron use and success with the discovery environment.

LITERATURE REVIEW

Whether referred to as discovery environments, discovery layers, discovery systems, or discovery services, these search tools have similar features and functions. OCLC's Lorcan Dempsey has described discovery layers as providing "a single point of access to the full library collection across bought, licensed, and digital materials."[10] Hoeppner writes that a discovery layer is "a user interface and search system for discovering, displaying, and interacting with the content in library systems, such as a WSD (web-scale discovery) central index."[11] While the implementation and use of discovery environments is now well established in the academic library sphere, there are concerns about their operability and performance. Discovery layer services vendors make many promises and tout improvements over time, but patrons often use other means to find sources in support of their research.[12] "Not only have discovery layers sometimes produced questionable results sets, but they have proved, in aggregate, somewhat difficult to configure."[13]

By their nature, discovery environments offer access to huge and diverse research materials, prompting deployment of sophisticated relevancy ranking and faceting processes. Marshall Breeding, renowned authority of library technology, includes relevancy ranking as a key feature of discovery environments.[14] Dempsey posits that discovery environments emphasize refining results through "narrowing mechanisms" such as pre- and post-search facets.[15] A host of other librarian authors confirm these points in their listings of typical discovery layer service components, "single search box (search engine feel) for the entire central index, tags and clouds, book art, suggestions, relevancy rankings, facets, customizability by the institution (e.g., cosmetic, search defaults), and user accounts."[16] Ex Libris Primo, a prominent discovery environment system, "allows administrators to customize much of the look, feel, and functionality of the system, including relevancy rankings."[17] Beyond the mere presence of relevance rankings and facets, Mussell reports that compared to all features of DEs, the "ability to limit to scholarly articles only"

[faceting] and the "ability to sort by relevance," are the top two most important among users.[18] This study centers on an expansion of relevancy ranking and faceting evaluation within a local iteration of Primo. While the specifics of Primo's algorithms are proprietary, they encompass "the degree to which an item matches a query, a value score representing an item's academic significance, and the publication date of an item."[19]

Discovery environments have been examined extensively. Bossaler conducted a meta-analysis of 80 DE studies concluding that the largest percentage of research ventures focus on the use and usability of DEs by patrons.[20] There are many signs that DEs are meeting user perceptions of their information needs. One key source of patrons' positive feeling toward DEs is their similarity to Google searching. Before DEs were widely introduced (beginning around 2010), information experts knew college students were largely (83%) starting their research at search engines.[21] Lippincott (2005), in an article examining the information-seeking behavior of Net Generation students (Millennials), notes that their preference for Google is tied to its simplistic and responsive design, and its speed, convenience, and reliability.[22] Even a decade later, "we know that users, particularly student populations, prefer to use general search tools [Google] rather than online databases [traditional discipline specific, subscription-based systems]."[23]

Beyond mirroring the Google-like search experience to garner favor with young researchers, there are a host of other studies and reasons DEs are satisfying user expectations. At Linfield University, although library staff thought the transition to a DE fairly onerous, patrons said they generally found what they were seeking.[24] Whether librarian researchers are utilizing user surveys ("... more than 80 percent of participants across both studies responded that they felt 'Positive' or 'Very Positive' about the discovery system after completing the test"), System Usability Scales (OneSearch (Primo) scores well with the usability tool according to Perrin), questionnaires and focus groups ("ease of use" ratings were high for Summon at Ryerson University), or usability testing (25 University of Toledo students stated they felt positive about the DE, would use it again, and would recommend it to others), overall satisfaction with DEs seems very common.[25]

There are also signs and investigations showing DEs are not meeting, or at least not fully addressing, patron information needs. DEs offer a vast array of popular and scholarly library materials requiring students to exercise source evaluation skills which they often do not possess or which are underdeveloped. Students often do not look beyond the first page of results, so they are apt to use sources with lesser authority, currency, or relevance to their topics.[26] In Valentine's DE study, the researchers noted that although students were asked to find relevant articles for the topic, they logged the first results they received without employing any discernment strategies.[27]

Two other areas of concern related to patron problems with DEs are issues of low facet understanding/use and finding full text/interlibrary utilization. According to many studies, students largely focus on simple searches and rarely use/understand faceting, especially post-search faceting, when searching in DEs.[28] To provide one illustrative fact from Hanrath's work, "27 participants attempted four tasks each, and a facet was used in 26 of the resulting 108 opportunities."[29] Valentine discovered that students did not realize the list of post-search facets available depends on the varying characteristics of the items in the results list.[30] In terms of full-text discovery and interlibrary loan use, Perrin concluded users were only able to find the full text of an article about 38% of the time, and Jacobs reports that users have trouble understanding interlibrary loan.[31] In terms of finding the full text of articles, DE users tend to have problems with both link resolvers and the web interfaces of publishers or aggregators.[32]

Many studies report that DEs are not meeting their potential because they contain library jargon that users do not know. Students often are confused by what it means to limit to "scholarly" or "peer-reviewed" materials.[33] Other troublesome terms include "holdings," "citation," "reviews;" some are even baffled by the difference between the terms "article" and "journal."[34] Students do not know library location names and are stymied by the need to click on vendor names to get to the full text of articles.[35]

In addition to reasons why discovery environments are not meeting user needs, patrons often view subject-specific databases as more effective than DEs. When Mussell recently asked patrons "How helpful were the results you found for your most recent research assignment via the following sources?," publisher databases were cited as "helpful or essential" more often than Google, Google Scholar, and Summon.[36] Research subjects also rated challenges they typically face with searching for materials. The challenge most often classified as difficult was "becoming overwhelmed by the number of results in searches."[37] Beyond user perceptions, Dahlen's study finds the articles selected from indexing and abstracting databases were more authoritative than those from the DE, and Kennedy notes the quality of the metadata for DE records is not as high as indexing and abstracting services.[38] Perhaps Kennedy stated it best when writing "Simply having a large central index does not guarantee that resources will be discoverable."[39]

One of the aims of the current study is to maximize its reproducibility by decreasing manual intervention wherever possible. Bosker evaluated various forms of fuzzy string matching (approximate string matching) between target and response sentences within speech intelligibility studies.[40] Their study looked at Levenshtein distance, Jaro distance, and Token sort ratio as potential predictors of human-generated scoring which could then be used to automate the matching process and thereby reduce reliance on manual intervention.[41] Another objective of the current study is to find a quality proxy for actual student research queries. Fischer et. al. have proposed a transaction log analysis methodology using Google Analytics.[42] The researchers considered using the transaction log analysis provided by Ex Libris, but their supplied data only includes a list of the most common search queries and those resulting in zero returned records. The study explained in the pages below fills a gap in the literature; while most DE investigations evaluate system quality through user satisfaction or usability measures (Pierre and Walton being the most recent examples), the researchers aim to create a largely automated framework methodology for assessing DE effectiveness.[43]

METHODS

Research Questions

The desired outcome of this study was to refine the framework for testing the relevancy of results returned from Primo. In doing so, the authors attempted to answer the following questions:

1. Can the boundaries of the testing framework be altered to better align the source citations and the search results list?
2. Does the exclusion of newspaper articles, reference entries, and reviews help increase the matching success?
3. Does the positioning of the successful match tell the researchers anything about whether certain search queries are more/less successful?
4. Can the analysis of fuzzy string matches be further automated to improve scalability and reproducibility of the framework ? If so, what kind of error rate does that introduce?

Workflow Overview

To answer said research questions, the authors designed and used the following framework:

1. Collected student research papers.
2. Extracted citations from student research papers.
3. Determined whether or not extracted citations existed in the WSU Primo instance. Both local and remote records were used in this determination and without regard to full-text availability or entitlements.
4. Extracted titles from student research papers to use as model search queries in Primo Search API.
5. Harvested up to the first 0–50 results from each model search query.
6. Converted extracted citations and the harvested search API results into normalized strings.
7. Performed a fuzzy matching algorithm (using an R package and Jaro-Winkler distance metric) between normalized strings to determine matching success rates.

Data Collection

The authors used a sample of 197 randomly selected research papers that were submitted as part of the Roots of Contemporary Issues courses in fall 2020 (n=98) and spring 2021 (n=99). The bibliographic citations from these 197 research papers were harvested and their titles extracted for use as the target responses in our fuzzy matching algorithm.

During the summer of 2021, as part of data preprocessing, the researchers separated the paper citations that were available in Primo from those that were unavailable in Primo. The researchers use the term "available" here to mean that a record corresponding to one of the citations in a student paper existed in our instance of Primo (regardless of immediate full-text availability). The term "unavailable" means that no such corresponding record could be found in our instance of Primo (i.e., the student must have used a source other than Primo to find said citation). Of the 805 paper citations from fall 2020, 442 (55%) were present within Primo; for spring 2021, 463 (59%) of 780 paper citations were present within Primo. In this process, the authors noted that paper citations of type website/webpage comprised the largest portion of those that were unavailable: 40% (147/363) from fall and 48% (151/317) from spring. Newspaper articles were the next largest category that were unavailable: 35% (126/363) from fall and 30% (94/317) from spring. Paper citations of type magazine article, instructor lecture and notes, and those that could not be determined made up the remainder of those that were unavailable in Primo. (See fig. 1.)

Figure 1. Unavailable versus available citations in Primo.

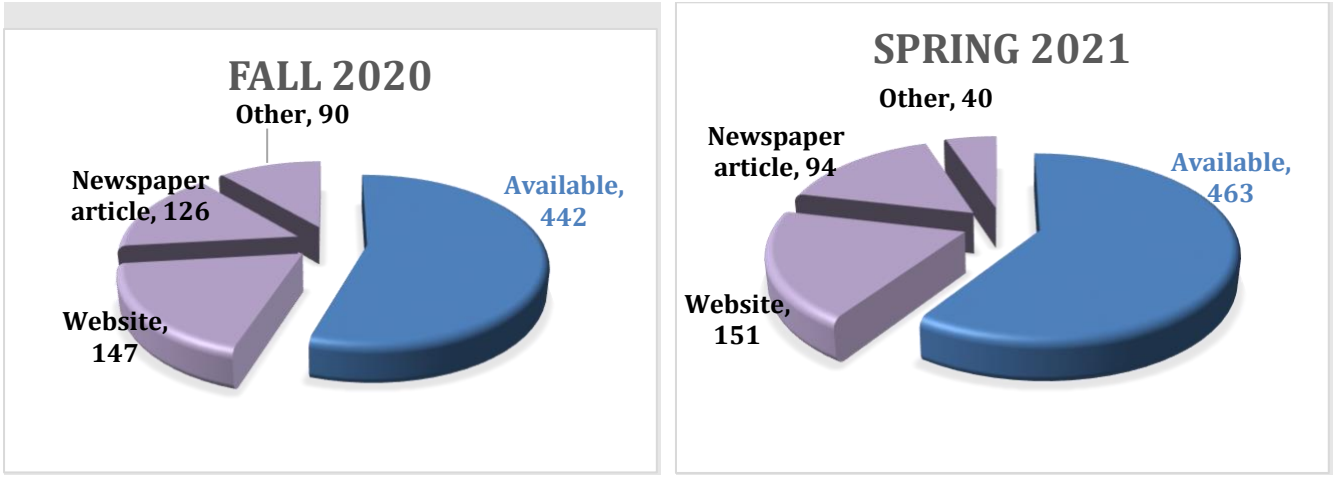

Table 1 is a breakdown of the paper citations that were present in Primo and their associated resource types (for full definitions of resource types in Primo, please see the Ex Libris document).[44] Journal articles and books comprised the vast majority of available source citations, indicating that Primo would have been a useful tool for finding these scholarly materials. Comparatively speaking, the other materials cited by Washington State University students were relatively absent from Primo, indicating that students would have had to have looked elsewhere.

Table 1. Source citations by resource type available in Primo for fall 2020 and spring 2021 terms

Resource type	Fall 2020 (% of total)	Spring 2021 (% of total)
Journal article	202 (45.70%)	235 (50.76%)
Books (ebooks/print)	180 (40.72%)	194 (41.90%)
Newspaper article	28 (6.33%)	20 (4.32%)
Book chapter	17 (3.85%)	3 (.65%)
Reference entry	6 (1.36%)	5 (1.08%)
Videos (evideos/DVD)	3 (.68%)	2 (.43%)
Journal	2 (.45%)	0 (0%)
Text resource	2 (.45%)	1 (.22%)
Report	1 (.23%)	1 (.22%)
Review	1 (.23%)	2 (.43%)
Semester citation count	442 (100%)	463 (100%)
Total citation count	905	

Search Query Creation

Building off previous work that indicated a natural-language-based query performs as well as or better than a machine-generated keyword search based on a supplied text corpus, the researchers used the original paper titles supplied by the student as a proxy for model Primo search queries.[45] Examples of paper titles as query include *Water Is Life: Standing Rock and the Repercussions of the Native Experience* (from fall 2020) and *Transgressions of Historical Racist Immigration Policies Reborn* (from spring 2021).

Search Results

Using the paper-title-as-query methodology, the authors constructed searches using the original paper titles and ran them against the Ex Libris Primo Search API endpoint. The basic structure of the endpoint used is

```
https://api-
    na.hosted.ExLibrisgroup.com/primo/v1/search?vid={vid}&tab={tab}&scope={scope}&q=any,conta
    ins," + $query + $facets + $date +
    "&lang=eng&offset=0&limit=50&sort=rank&pcAvailability=true&getMore=0&conVoc=true&inst={in
    st}&apikey={apiKey}.
```

The research paper titles were encoded as UTF-8 strings and stored as variable `$query`. The `$facets` variable stored querystring parameters *qInclude* and *multiFacets*, both of which were used to filter on the resource type facet category. The `$date` variable stored an additional *qInclude* querystring parameter, which was used to filter on the search creation date facet category (*facet_searchcreationdate*, currently undocumented on the Ex Libris Developer Network).

For fall 2020, the search creation date was set to range 1000–2020, while for spring 2021, the search creation date was set to range 1000–2021.

Searches were run on June 14, 2022, via the Primo New User Interface (NUI) using PowerShell and outgoing strings exported to CSV with columns Query (original title of student paper), Results (number of results returned from search), Title (Primo record title returned from search), Type (resource type of Primo record), and CreateDate (publication date of Primo record). Table 2 provides an example of exported CSV file for API results returned from fall 2020 with no facets applied.

Table 2. Example of exported CSV file for API results returned from fall 2020 with no facets applied

Query	Results	Titles returned	Type	CreateDate
CLIMATE REFUGEES. THE NEXT GREAT MIGRATION	15193	Global climate change, population displacement, and public health : the next wave of migration	book	2020
CLIMATE REFUGEES. THE NEXT GREAT MIGRATION	15193	Climate Migration at the Height and End of the Great Mexican Emigration Era	article	2018
CLIMATE REFUGEES. THE NEXT GREAT MIGRATION	15193	Does climate change influence people's migration decisions in Maldives?	article	2019

In addition to search results from queries 1) using no facets, the Primo Search API was used to retrieve search results from queries that 2) included only ebooks, print books, and book chapters; 3) included only articles, and 4) excluded newspaper articles, reference entries, and reviews. All told there were four search-query constructions (one query type by four faceting modes) for fall 2020 and spring 2021 each, for a total of eight CSV files.

The researchers designed the initial search to be open ended in order to establish a baseline for the search comparisons. That is, the study assumed that patrons most often use the default, basic search functionality, with no facets selected. Also, given the problematic nature of the newspaper resource type in discovery systems, the researchers excluded this resource type in faceted searches.[46] In a refinement of previous work, the researchers altered the search types to be Open-Ended, Books Only, Articles Only, and Constrained (Open-Ended minus newspaper articles, reviews, etc.).

Each Primo Search API returned titles for the top 50 results, moving a bit beyond users' usual first-page-only search behavior, in an effort to provide consistency to the framework (e.g., some search results lists were tens of thousands, others were hundreds of thousands) and retain the ability to place citation matches in context (where in a result set, 1–50, a citation appears).[47]

Data Cleaning

In a previous study, the authors found that small variations in the titles that were harvested from student citations and returned from the Primo API led to the researchers needing to perform a thorough quality assurance check on the fuzzy matches to ensure that a viable match was not missed because of small variations in the strings. These small variations in strings, like two spaces between words instead of one or differences in nonessential punctuation, led to matching scores needing a second human check to confirm title matches were not missed. For this round of

research, the titles were run through a more rigorous data normalization procedure. This data normalization procedure consisted of a search-and-replace function that utilized a regular expression in OpenRefine to normalize the titles completely. The regular expression or regex (`[^a-zA-Z0-9]`) removed every character that was not within the ISO basic Latin character set (`A-Z` or `a-z`) or a number 0–9. Researchers chose to do this in OpenRefine as opposed to within the R scripting environment as OpenRefine has a more approachable interface for quickly manipulating, normalizing, and reviewing the results of the normalization process than the RStudio scripting environment.

Matching Process

Previous work to verify citation matches relied on an Excel add-in called Fuzzy Lookup, and a fair bit of manual manipulation.[48] To reduce human intervention, increase the reproducibility of the process, and increase the configurability of the matching mechanism, the authors utilized an R-based approach, employing the *stringdist* R package and applying the Jaro-Winkler distance metric as the matching evaluator. For a full description of the process please see the referenced OSF site.[49] This investigation focused on results that had a score below 0.8, where 0 represents full overlap of the compared strings and 1.0 represents no overlap, which researchers reviewed and confirmed.[50] The Jaro-Winkler distance score was used to discard obvious nonmatches and the researchers manually confirmed matches using title and resource type as the main criteria.

Table 3. Sample matches and nonmatches between student paper citation titles and Primo Search API

	Normalized citation title	Citation resource type	Normalized results title	Result resource type	Confirmed match
0	Behaviorunbecomingacommunistjewishreligiouspracticeinsovietminsk	Article	behaviorunbecomingacommunistjewishreligiouspracticeinsovietminsk	Article	Yes
0.1872	runawaysrepertoiresandrepression	Article	runawaysrepertoiresandrepressionmarronnageandthehaitianrevolution17661791	Article	Yes
0.2353	incomeinequalityandeducation	Article	incomeinequalityandclassdividesinparentalinvestments	Article	No
0.2252	airpowerandtheenvironmenttheecologicalimplicationsofmodernairwarfare	ebook	theecologyofwarenvironmentalimpactsofweaponryandwarfare	Print book	No

RESULTS

Researchers attempted to match the available citations against the results returned from the API title search. For the fall 2020 research papers, the percentage of student citations that were matched using the API title search were as follows: Open-Ended, 2.04%; Articles Only, 2.97%; Books Only, 3.33%; and Constrained, 2.21%. The percentages for the spring 2021 research papers were higher across the board than in 2020 and were roughly proportional to the 2020 matches: Open-Ended, 5.40%; Articles Only, 6.81%; Books Only, 8.76%; and Constrained, 6.88%. These

results are consistent with the researchers' first study in that faceted searches resulted in higher matching success rates.[51] Also of note is the observation that the percentage matched via Books Only is highest in both terms. The results are summarized in table 4.

Table 4. Available source citations matched via API title search (matching success rate of available citations)

Search type	Fall 2020 (% of total)	Spring 2021 (% of total)
Open-Ended	9/442 (2.04%)	25/463 (5.40%)
Articles Only	6/202 (2.97%)	16/235 (6.81%)
Books Only	6/180 (3.33%)	17/194 (8.76%)
Constrained	9/407 (2.21%)	30/436 (6.88%)

In addition to calculating the number (and percentages) of student citations that were found using the API title searches, in other words, that appeared in the top 50 search results, the researchers also investigated potential trends concerning where in the top 50 the matches appeared. Across both academic terms and the four search types, there was at least one match in each group that appeared as the first result in the list (see low range numbers in table 5), while the matches appearing lowest in the list of 50 varied greatly between position 24 and 50 (see high range numbers in table 5). These results along with the mean average matching position appear in table 5.

Table 5. Positioning of matches within Primo search results lists

Matching positions for:		Fall 2020	Spring 2021
Open-Ended	Low range	1	1
	Average	11.89	16.00
	High range	35	43
Articles Only	Low range	1	1
	Average	17.83	12.93
	High range	47	46
Books Only	Low range	1	1
	Average	9.17	14.86
	High range	24	41
Constrained	Low range	1	1
	Average	16.78	15.54
	High range	50	47

DISCUSSION

Research Question #1: Can the boundaries of the testing framework be altered to better align the source citations and the search results list?

In the authors' previous study, all student citations were deemed viable regardless of whether the source citation was verified as available within Primo.[52] This led to the inclusion of citations such as lecture notes and other such materials that are not generally expected to appear in a discovery environment. For the current study, the researchers verified and included only those resources from the citation lists that were available in Primo (including both local and remote records and without regard to full-text availability or entitlements). Limiting the resources to only those that

are available in Primo increased the matching success rate, since it also decreased the denominator (see table 4). The researchers recognize that this step adds to the manual processing, but it is necessary to eliminate unmatchable items. The researchers also considered that the creation of a set of unavailable items could be useful for collection development purposes. For these two reasons, it would be advantageous to develop a more automated process to separate the available items from the unavailable. Recent developments from the discovery layer vendor may make this possible. For example, as of the May 2023 release, Ex Libris has made an exact phrase search possible for the title field.[53] If this advancement carries forward into the API structure, the researchers could then more easily automate a process that searches the exact title within Primo to establish the bibliography source's presence or absence.

During this analysis, the researchers also observed that websites comprised a large portion of those citations that were unavailable in Primo, although this resource type represented a major category in the initial list of student citations. For example, web documents were approximately 20% of all citations in fall and spring (165/805 and 154/780, respectively). However, when we searched for citations in Primo, which could have retrieved any information type from the system, not a single web document was available. This is most likely because only a tiny fraction of online websites are indexed in Primo. Therefore, it could be fruitful to consider omitting this resource type from future iterations of the testing framework.

Another observation that surfaced during this analysis is related to the use of research paper titles as proxies for keyword searches. A potential issue here is that students are free to insert catchy or otherwise irrelevant words into their titles (e.g., plays on word and other poetic devices). Another possible issue is where a student might not include enough information in a title for it to sufficiently serve as a proxy for keyword search. The researchers deemed the following student paper titles to contain catchy or otherwise irrelevant information: *Great Leap Backward: Roots of Antibiotic Resistance in China*; *Too Many Mouths to Feed: Brazil, Amazon Deforestation, and Genetic Modification*; *Fada Beo An Réabhlóid 'Long Live the Revolution'*; *Bad Guys Wear Turbans: Examining 1,000 Years of Islamophobia in the West*; *le bon problème: Finding Balance in the Wine Industry*. Examples of titles with insufficient information included: *Sex Ed, Polarized*; *Disaster*; *Plagued*; and *Racial Tension*. The only one of these titles that produced a matching citation was *le bon problème: Finding Balance in the Wine Industry*. Overall, paper titles similar to the above are problematic. However, their occurrence in this study is not frequent (12/197), their analysis requires a high degree of subjectivity, and there are plenty of other titles that also did not result in matching records. The more central issue is that the use of paper titles as proxies for student searches did not create a reasonable matching success rate.

A significant amount of time was spent developing n-grams as keyword search queries in the previous investigation.[54] In order to focus more time on developing the framework further in the current paper, the researchers opted to streamline the process of search-query creation by using paper titles as the search query. In the end, the matching success rates were still not very high, but were higher than in the previous investigation. Overall, the researchers acknowledge that using a single search query to retrieve all relevant citations does not represent the information seeking process. In other words, research is iterative and involves a complex set of cognitive and affective variables.[55] This fact will be considered in subsequent investigations. Now that the framework is more stable, a new approach that incorporates multiple queries to gather citations should be formulated. This could be an additive approach that combines paper titles and n-grams from both investigations or one that relies more heavily on large language models, like ChatGPT, to reverse engineer queries from the research papers or citations. The researchers could also move away

from undergraduate assignments to explore using controlled vocabularies from articles and longer works such as dissertations and theses. This latter approach would then be relying on the key terminologies already established by the authors of each work.

Research Question #2: Does the exclusion of newspaper articles, reference entries, and reviews help increase our matching success?

The researchers considered the impact of including newspaper articles, reference entries, and review works in the open-ended searches. These resource types are large in number, not indexed very well, and often do not have descriptive titles. Reference entries also typically have very short titles and a significant portion of historical newspaper articles do not have titles at all. Newspaper articles are so numerous that Ex Libris has created a dedicated index called Newspaper Search that removes this resource type from the results lists and facets.[56] WSU has chosen not to enable Newspaper Search in its Primo instance yet, but perhaps should reconsider. Within the researchers' experiment, when compared to open-ended searches, the removal of these "noisy" resource types from the Primo results did increase the matching success rates, but only marginally (see table 4)—fall 20: Open-Ended = 2.04% vs. Constrained = 2.21%; fall 21: Open-Ended = 5.40% vs. Constrained = 6.88%.

Research Question #3: Does the positioning of the successful match tell us anything about whether certain search queries are more/less successful?

Another avenue of exploration was determining where in the results list a matched citation appears (i.e., somewhere between the first and fiftieth position in the results list), not just the binary positive or negative. It is notable that, across the two academic terms and the four types of searches, each set of results contained at least one match that was in the first position in the results list. It is also valuable to relay that the numerical average of the result position across the eight term/search type combinations was 13.55. In other words, across the 50-position spread, the matches are concentrated at the top of the results lists. However, there were plenty of results scattered across the bottom half of the positions (between 25 and 50). If the matches had more strongly clustered at the top of the results lists, it would have pointed to a stronger connection between the use of the local Primo system and student discovery of the sources valuable and relevant enough to be utilized in their research papers.

Research Question #4: Can the analysis of fuzzy string matches be further automated to improve scalability and reproducibility of the framework? If so, what kind of error rate does that introduce?

In their previous study on developing a framework for judging discovery environment effectiveness, the authors needed to intervene manually in the process in several places: 1) collecting the source titles and citations; 2) preparing and formatting the source and Primo API title lists so that an Excel Fuzzy Lookup could be performed; and 3) providing quality assurance on the citation matches by manually confirming matches. Researchers checked matches through reviewing both the source citation and the Primo record for an item to confirm a positive match or to correct a nonmatch that was not captured by the automated process correctly (due to punctuation differences, added titles, or spelling conventions).[57]

This same process of quality assurance was followed in the initial phases of the current study to establish a baseline of true matches. An example from the current study of a nonmatch that was reversed by the review process is in table 3. The source citation *Runaways, Repertoires, and Repression* does not include the subtitle that is present in the Primo results (before normalization), *Runaways, Repertoires, and Repression: Marronnage and the Haitian Revolution,*

1766–1791, resulting in a poor matching score. Without human review, these differences between the strings would have resulted in a nonmatching citation.

To further automate and routinize the framework, and find and correct both false positives and negatives, the researchers prepared both the source citation title and Primo results title by running the title normalization routine described in the methods section. Normalizing the titles has the potential to completely remove the need for review and contributes to scalability. However, the normalization routine used does have its trade-offs, including: 1) titles with non-Latin characters were disproportionally impacted and 2) certain types of matches were missed. The researchers believe the added scalability and reproducibility provided by the title normalization outweigh the trade-offs. In this round of research using a Jaro-Winkler distance score of 0.0, the researchers would have recorded an overall error rate of 11.01% (see table 6).

Table 6. Error rate for each search type

Search type	Error rate (%)
Open-Ended	5.88%
Articles Only	13.63%
Books Only	8.69%
Constrained	15.38%
Overall	11.01%

The authors observed that the error rate in spring 2021 was a result of missing subtitles in source citations as described above using the example from table 3. Moving forward, researchers will investigate methods to mitigate or control this impact so that, with a certain degree of confidence, they can scale the framework to draw more rigorous conclusions. One method to explore for controlling missing or incomplete added titles will be to refine and examine the Jaro-Winkler heuristic matching method that adds a penalty to mismatched characters in the first four characters of strings being compared.[58] Another potential control would be to extend the matching process to other parts of the citation in a secondary or even tertiary matching process. Performing a multistep matching process would allow for inconsistencies in title matches (e.g., missing subtitle matches) if the secondary/tertiary matching processes successfully match. For example, a matching publication date, format type, and/or author could be used to identify matches that would have been missed when only the title is being used (researchers are already confirming matches by visually comparing citation types so that an article is not erroneously matched against a book).

CONCLUSIONS AND NEXT STEPS

The most easily identifiable trend in the data is the low number of matches between the student paper sources and the first 50 results in each paper's Primo API searches. Whether the searches were open ended, default; constrained by eliminating newspaper articles, reference entries, and reviews; or were limited to books or articles only, the matching rates were small, ranging from just 2.04% to 8.76%. There are many possible explanations for this result. It might be the case that using the paper titles as the search query is not a quality proxy for the students' actual search queries (similar to what the authors discovered in the first paper, i.e., that n-grams and paper reader (human)–generated keywords did not produce higher matching rates).[59] Students simply

might be using different keywords and/or limiter combinations from what the researchers have constructed.

Another logical idea would be that students are largely not using Primo to find their research materials. This thought is furthered by the reality that during both academic terms featured in this study (fall 2020 and spring 2021), the physical libraries were closed due to the COVID pandemic and during this time the total number of Primo searches dropped by about 25%, according to Primo Analytics. On the other hand, one of this study's researchers has also investigated students taking Roots of Contemporary Issues during the pandemic closure (although not precisely the same students) concerning their use of Primo in finding books, journal articles, and primary sources for their research papers. From that research it has been discovered that the local Primo instance was the most frequently used database for finding monographs, and that for both journal articles and primary sources, Primo was second compared to all other databases.[60]

There are four other possible causes of the low matching rates. The first might be that students were looking beyond the first 50 results. Although this is possible, studies by Cmor, Kliewer, and Hamlett indicate that it is not likely.[61] The last three plausible explanations focus on the backend of Primo itself. The system is either dropping some of the titles students used (which seems highly unlikely, especially at high rates), or it is adding new sources fast enough that the sources the students used are getting pushed past the first 50 in the results lists. Across both phases of this research (2019–20 and 2020–21), the investigators found more matches in the latter (spring) term than the earlier (fall) term for nearly every search type. There is a connection in terms of the proximity to when the authors ran test searches and the time period under which students would have done their original searches. A last possible reason for the low matching rates is that underlying algorithms in Primo and CDI content changed, altering results lists. While all the searching done by students, and later by the researchers, occurs under the same version of the system, the researchers recognize that Primo and CDI monthly releases did occur in the interim and could have impacted the availability and placement of records within search results.

The framework being presented in this paper is reproducible with the data files offered in the Open Science Framework project. The framework could also be utilized for novel investigations by research communities at large with modifications for a local environment using the process outlined here and in more detail on the Open Science Framework project site, using the Primo API, Open Refine, and RStudio.[62] With the work completed thus far, the most human intensive aspect is collecting the appropriate source citations to be matched and some routinized data normalization performed in Open Refine to prepare the titles to be matched. The R matching procedure is expressed in three separate scripts and presented in an R Markdown Notebook, a simple formatting syntax that allows for authoring interactive HTML, PDF, and MS word documents, which can be opened and utilized in the open-source R integrated development environment RStudio with little knowledge of R or programming.[63]

The researchers remain determined to find a way to utilize patron research output as a tool for evaluating discovery environment quality. In doing so, the researchers migrated the framework to R to increase the scalability and reproducibility for future studies. A portion of the next round of research will be dedicated to exploring differences between utilizing undergraduate versus graduate student paper citation sources for potential matches to API search results. Future work could also bring in a mixed methods approach to reflect the information search process and information seeking behaviors of researchers and learners more accurately. The authors could augment the current quantitative approach with the addition of documenting the information

search process for a discrete number of subjects to get a more complete picture of where and how search refinement happens, which may inform steps that the researchers can take to capture the multistep search process. Finally, next steps will involve using ChatGPT to summarize paper content into search terms, which will hopefully produce higher source matching rates. This work is important because academic librarians understand "a frustrating or unsuccessful encounter with the discovery layer can bounce users away, possibly never to return" and there is nothing more paramount than delivery of relevant content to researchers.[64]

ENDNOTES

[1] Kim Durante and Zheng Wang, "Creating an Actionable Assessment Framework for Discovery Services in Academic Libraries," *College & Research Libraries* 19, no. 2–4 (2012): 217, https://doi.org/10.1080/10691316.2012.693358.

[2] Scott Uhl, "Applying User-Centered Design to Discovery Layer Evaluation in the Law Library," *Legal Reference Services Quarterly* 38, no. 1–2 (2019): 32, https://doi.org/10.1080/0270319X.2019.1614373.

[3] Uhl, "Applying User-Centered Design," 31.

[4] W. Jacobs, Mike Demars, and J. M. Kimmitt, "A Multi-Campus Usability Testing Study of the New Primo Interface," *College & Undergraduate Libraries* 27, no. 1 (2020): 1–16, https://doi.org/10.1080/10671316.2019.1695161.

[5] "The UCORE Curriculum," Washington State University Common Requirements, 2018, https://ucore.wsu.edu/faculty/curriculum/.

[6] "Welcome to the Roots of Contemporary Issues," Washington State University Department of History, 2017, https://ucore.wsu.edu/faculty/curriculum/root/.

[7] "Washington State University Learning Goals," Washington State University Common Requirements, 2018, https://ucore.wsu.edu/about/learning-goals.

[8] "Washington State University Learning Goals."

[9] "Search It," Washington State University Libraries, 2020, https://searchit.libraries.wsu.edu/.

[10] Lorcan Dempsey, "Discovery Layers—Top Tech Trends 2," *LorcanDempsey.net*, 2012, http://orweblog.oclc.org/archives/002116.html.

[11] Athena Hoeppner, "The Ins and Outs of Evaluating Web-Scale Discovery Services," *Computers in Libraries* 32, no. 3 (2012), https://www.infotoday.com/cilmag/apr12/Hoeppner-Web-Scale-Discovery-Services.shtml.

[12] Sean P. Kennedy, "Uncovering Discovery Layer Services," *Public Services Quarterly* 10 (2014): 55, https://doi.org/10.1080/15228959.2014.875788; Marshall Breeding, "The Ongoing Challenges of Academic Library Discovery Services," *Computers in Libraries* 40, no. 1 (2020): 11, https://www.infotoday.com/cilmag/jan20/index.shtml.

[13] Uhl, "Applying User-Centered Design," 54.

[14] Marshall Breeding, "Major Discovery Product Profiles," *Library Technology Reports* 50, no. 1 (2014): 33–52, https://web.p.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=13d50467-13bb-465a-ab85-f9a0b41249f5%40redis.

[15] Lorcan Dempsey, "Thirteen Ways of Looking at Libraries, Discovery, and the Catalog: Scale, Workflow, Attention," *EDUCAUSE Review* (December 10, 2012), https://er.educause.edu/articles/2012/12/thirteen-ways-of-looking-at-libraries-discovery-and-the-catalog-scale-workflow-attention.

[16] Ellen Safley and Debbie Montgomery, "Oasis or Quicksand: Implementing a Catalog Discovery Layer to Maximize Access to Electronic Resources," *The Serials Librarian* 60, no. 1–4 (2011): 164–68, https://doi.org/10.1080/0361526X.2011.556028; Athena Hoeppner, "The Ins and Outs"; Kennedy, "Uncovering Discovery Layer Services," 55.

[17] Jacobs, Demars, and Kimmitt, "A Multi-Campus Usability Testing Study," 1.

[18] Jessica Mussell and Rosie Croft, "Discovery Layers and the Distance Student: Online Search Habits of Students," *Journal of Library & Information Services in Distance Learning* 7, no. 1–2 (2013): 26, https://doi.org/10.1080/1533290X.2012.705561.

[19] "Ranking: Deliver the Most Relevant Search Results," ExLibris Knowledge Center (Part of Clarivate), 2024, https://exlibrisgroup.com/products/primo-discovery-service/relevance-ranking/.

[20] Jenny S. Bossaller and Heather Moulaison Sandy, "Documenting the Conversation: A Systematic Review of Library Discovery Layers," *College & Research Libraries* 78, no. 5 (2017): 615–6, https://doi.org/10.5860/crl.78.5.602.

[21] Mussell and Croft, "Discovery Layers and the Distance Student," 19.

[22] J. K. Lippincott, "Net Generation Students & Libraries," *EDUCAUSE Review* 40, no. 2 (2005): 57, http://er.educause.edu/-/media/files/article-downloads/erm0523.pdf.

[23] Uhl, "Applying User-Centered Design," 53.

[24] Barbara Valentine and Beth West, "Improving Primo Usability and Teachability with Help from the Users," *Journal of Web Librarianship* 10, no. 3 (2016): 176–96, https://doi.org/10.1080/19322909.2016.1190678.

[25] Scott Hanrath and Miloche Kottman, "Use and Usability of a Discovery Tool in an Academic Library," *Journal of Web Librarianship* 9, no. 1 (2015): 17–18, https://doi.org/10.1080/19322909.2014.983259; Joy Marie Perrin et. al., "Usability Testing for Greater Impact: A Primo Case Study," *Information Technology and Libraries* 33, no. 4 (2014): 59, https://doi.org/10.6017/ital.v33i4.5174; Courtney Lundrigan, Kevin Manuel, and May Yan, "'Pretty Rad': Explorations in User Satisfaction with a Discovery Layer at Ryerson University," *College & Research Libraries* 76, no. 1 (2015): 47, https://doi.org/10.5860/crl.76.1.43; Christine Rigda, Margaret Hoogland, and Jessica Morales, "'But I Just Want a Book!' Is Your Discovery Layer Meeting Your Users' Needs?" *Journal of Web*

Librarianship 12, no. 4 (2018): 253, 254, 257, https://doi.org/10.1080/19322909.2018.1518183.

[26] Alexandra Hamlett and Helen Georgas, "In the Wake of Discovery: Student Perceptions, Integration, and Instructional Design," *Journal of Web Librarianship* 13, no. 3 (2019): 237, https://doi.org/10.1080/19322909.2019.1598919; Sarah P. C. Dahlen and Kathlene Hanson, "Preference vs. Authority: A Comparison of Student Searching in a Subject-Specific Indexing and Abstracting Database and a Customized Discovery Layer," *College & Research Libraries* 78, no. 7 (2017): 878–79, https://doi.org/10.5860/crl.78.7.878.

[27] Valentine and West, "Improving Primo Usability and Teachability," 185.

[28] Hamlett and Georgas, "In the Wake of Discovery," 232–33; Elena Azadbakht, John Blair, and Lisa Jones, "Everyone's Invited: A Website Usability Study Involving Multiple Library Stakeholders," *Information Technology and Libraries* 36, no. 4 (2017): 43, https://doi.org/10.6017/ital.v36i4.9959; Valentine and West, "Improving Primo Usability and Teachability," 187; Anita K. Foster and Jean B. MacDonald, "A Tale of Two Discoveries: Comparing the Usability of Summon and EBSCO Discovery Service," *Journal of Web Librarianship* 7, no. 1 (2013): 14, http://doi.org/10.1080/19322909.2013.757936; Kelsey Brett, Ashley Lierman, and Cherie Turner," Lessons Learned: A Primo Usability Study," *Information Technology and Libraries* 35, no. 1 (2016): 19, https://doi.org/10.6017/ital.v35i1.8965.

[29] Brett, Lierman, and Turner, "Lessons Learned: A Primo Usability Study," 11.

[30] Valentine and West, "Improving Primo Usability and Teachability," 187.

[31] Perrin et al., "Usability Testing for Greater Impact," 59; Jacobs, Demars, and Kimmitt, "A Multi-Campus Usability Testing Study," 10.

[32] Brett, Lierman, and Turner, "Lessons Learned: A Primo Usability Study," 17.

[33] Jacobs, Demars, and Kimmitt, "A Multi-Campus Usability Testing Study," 3–4; Brett, Lierman, and Turner, "Lessons Learned: A Primo Usability Study," 20.

[34] Valentine and West, "Improving Primo Usability and Teachability," 185; Brett, Lierman, and Turner, "Lessons Learned: A Primo Usability Study," 21; Jacobs, Demars, and Kimmitt, "A Multi-Campus Usability Testing Study," 3, 8.

[35] Valentine and West, "Improving Primo Usability and Teachability," 185.

[36] Mussell and Croft, "Discovery Layers and the Distance Student," 26.

[37] Mussell and Croft, "Discovery Layers and the Distance Student," 29.

[38] Dahlen and Hanson, "Preference vs. Authority," 878; Kennedy, "Uncovering Discovery Layer Services," 57.

[39] Kennedy, "Uncovering Discovery Layer Services," 59.

[40] Nimisha Singla and Deepak Garg, "String Matching Algorithms and Their Applicability in Various Applications," *International Journal of Soft Computing and Engineering* 1, no. 6 (2012): 218–22, https://www.ijsce.org/wp-content/uploads/papers/v1i6/F0304111611.pdf.

[41] Hans Rutger Bosker, "Using Fuzzy String Matching for Automated Assessment of Listener Transcripts in Speech Intelligibility Studies," *Behavior Research Methods* 53 (2021): 1945–53, https://doi.org/10.3758/s13428-021-01542-4.

[42] Rachel K. Fischer, Aubrey Iglesias, Alice L. Daugherty, and Zhehan Jiang, "A Transaction Log Analysis of EBSCO Discovery Service Using Google Analytics: The Methodology," *Library Hi Tech* 39, no. 1 (2021): 249–62, https://doi.org/10.1108/LHT-09-2019-0199.

[43] Jodi Pierre, "Discovery Services: Continuous Improvement with Ongoing Usability Testing," *Information Today* (March 2023): 4–8; Kerry Walton, Gary M. Childs, and Laurie Palumbo, "Testing Two Discovery Systems: A Usability Study Comparing Student Perceptions of EDS and Primo," *Journal of Web Librarianship* 16, no. 4 (2022): 200–221. https://doi.org/10.1080/19322909.2022.2125478.

[44] "Resource Types in CDI," ExLibris Knowledge Center (Part of Clarivate), 2024, https://knowledge.exlibrisgroup.com/Primo/Content_Corner/Central_Discovery_Index/Documentation_and_Training/Documentation_and_Training_(English)/CDI_-_The_Central_Discovery_Index/070Resource_Types_in_CDI.

[45] Blake L. Galbreath, Alex Merrill, and Corey M. Johnson, "A Framework for Measuring Relevancy in Discovery Environment," *Information Technology and Libraries* 40, no. 2 (2021): 11–12, https://doi.org/10.6017/ital.v40i2.12835.

[46] Galbreath, Merrill, and Johnson, "A Framework for Measuring Relevancy," 12.

[47] Greta Kliewer, Amalia Monroe-Gulick, Stephanie Gamble, and Erik Radio. "Using Primo for Undergraduate Research: A Usability Study," *Library Hi Tech* 34, no. 4 (2016): 572, https://doi.org/10.1108/LHT-05-2016-0052; Hamlett and Georgas, "In the Wake of Discovery," 237.

[48] Galbreath, Merrill, and Johnson, "A Framework for Measuring Relevancy," 8.

[49] Alex Merrill and Blake Galbreath, "A Framework for Measuring Relevancy in Discovery Environments," Open Science Framework (OSF), 2023, https://osf.io/wafbx?view_only=edf1715850e7474b90e6c521f7d82349.

[50] Mark P. J. Van Der Loo, "The Stringdist Package for Approximate String Matching," *The R Journal* 6, no. 1 (2014): 111–22, https://doi.org/10.32614/rj-2014-011.

[51] Galbreath, Merrill, and Johnson, "A Framework for Measuring Relevancy."

[52] Galbreath, Merrill, and Johnson, "A Framework for Measuring Relevancy."

[53] "Primo VE 2023 Release Notes," ExLibris Knowledge Center (Part of Clarivate), 2023, https://knowledge.exlibrisgroup.com/Primo/Release_Notes/002Primo_VE/2023/010Primo_VE_2023_Release_Notes.

[54] Galbreath, Merrill, and Johnson, "A Framework for Measuring Relevancy."

[55] "The Information Search Process," University of Illinois Libraries, 2023, https://www.library.illinois.edu/staff/wp-content/uploads/sites/24/2017/07/Information-Search-Process.pdf.

[56] "Configuring Newspaper Search for Primo VE," ExLibris Knowledge Center (Part of Clarivate), 2023, https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/130Configuring_Advanced_Search_Interfaces_for_Primo_VE/Configuring_Newspaper_Search_for_Primo_VE.

[57] Galbreath, Merrill, and Johnson, "A Framework for Measuring Relevancy."

[58] Van Der Loo, "The Stringdist Package for Approximate String Matching," 119.

[59] Galbreath, Merrill, and Johnson, "A Framework for Measuring Relevancy."

[60] Corey M. Johnson and Jennifer Saulnier Lange, "Source Seeking Database Usage Patterns for Roots of Contemporary Issues Students" (poster presentation at Washington State University Showcase), March 30, 2023, https://doi.org/10.7273/000004754.

[61] Diane Cmor and Xin Li, "Beyond Boolean, Towards Thinking: Discovery Systems and Information Literacy," 2012 IATUL Proceedings, paper 7, https://docs.lib.purdue.edu/iatul/2012/papers/7/; Kliewer et al., "Using Primo"; Hamlett and Georgas, "In the Wake of Discovery," 232–33.

[62] Merrill and Galbreath, "A Framework for Measuring Relevancy in Discovery Environment," Open Science Framework.

[63] Garrett Grolemund, "Introduction to R Markdown," *R Markdown from Rstudio*, July 16, 2014, https://rmarkdown.rstudio.com/articles_intro.html.

[64] Uhl, "Applying User-Centered Design," 29; Kennedy, "Uncovering Discovery Layer Services," 60.