

Three Discovery Tools: A Comparative Analysis of Retrieval Scope, Ranking Effectiveness, and Topic Diversity

Can Ekşi and Yurdagül Ünal*

Discovery tools facilitate access to large-scale academic collections, yet their retrieval performance varies. This study presents a comparative analysis of three discovery tools—EBSCO Discovery Services (EDS), EKUAL Discovery Services (EKUAL DS), and Piri Discovery Services (Piri DS)—evaluating retrieval scope, ranking quality, and topical diversity index. The iSearch test collection, derived from arXiv articles, was used with predefined search queries. To assess coverage, the full arXiv corpus was queried to identify indexing differences. A total of 63 queries were executed, and retrieved lists were analyzed for relevance and ranking distribution. Expert-evaluated relevant articles were used to assess retrieval accuracy. Ranking was measured using Discounted Cumulative Gain (DCG) and Normalized DCG (NDCG), and topical diversity was evaluated using the Shannon Diversity Index. EDS and EKUAL DS retrieved identical results, while Piri DS retrieved fewer records, affecting its retrieval completeness. Piri DS ranked relevant articles higher, but with broader distribution. While all tools exhibited comparable ranking performance, EDS and EKUAL DS demonstrated greater topical diversity. These findings offer empirical insight into the strengths and limitations of discovery tools and support libraries in improving search efficiency and retrieval strategies.

Introduction

The discovery of library resources is a concept independent of the size and scope of collections. Libraries have a responsibility to enhance the discoverability of their collections, to facilitate users' access to the information they need. With the advancement of computer and internet technologies, catalogs and indexes, which played an important role in the discovery of library resources in the past, have been replaced by web-based online library catalogs. The proliferation of electronic resources has led to the need for the development of new systems which would make the process of accessing information more efficient. In addition to online catalogs, discovery systems have emerged that make libraries' local collections, and licensed memberships discoverable. As a result, catalogs have been transformed into discovery services, and

*Can Ekşi is a librarian at Istanbul Bilgi University, email: can.eksi@bilgi.edu.tr. Yurdagül Ünal is an Associate Professor in the Department of Information Management at Hacettepe University, email: yurdagul@hacettepe.edu.tr. ©2026 Can Ekşi and Yurdagül Ünal, Attribution-NonCommercial (<https://creativecommons.org/licenses/by-nc/4.0/>) CC BY-NC.

bibliographic information of electronic publications has been integrated into these systems. Publishers providing subscriptions to electronic resources initially developed centralized search services by integrating their own platforms (e.g., EBSCOhost). However, over time, centralized indexes that also included other resources available in libraries were introduced (Breeding, 2015, p. 24). As a result, web-based discovery tools have emerged, not merely as a part of library catalogs, but also as comprehensive search services which encompass library catalogs as well.

The key factors influencing a library's choice of a discovery tool are the scope of resources covered by the discovery tool, integration capabilities, the number of results retrieved for queries, and the ranking of these results based on relevance, ease of access, interface features, and personalization options.

The effectiveness of discovery tools is directly linked to the up-to-dateness and comprehensiveness of the collections they index. Ensuring regular updates is the shared responsibility of both service providers and the library itself. Libraries must verify that all resources within their collections are fully indexed by the discovery tool, and provide feedback regarding any potential omissions, or necessary improvements. Hartman and Bowering Mullen (2008, p. 211) state that web-based academic search engines serve as portals for open-access materials available on the internet and in institutional repositories. Users may prefer discovery tools over traditional search engines to access open-access resources online. Therefore, it is crucial for discovery tools to include prominent open-access resources across different disciplines, to enhance their search capabilities.

Topic diversity is another critical factor that discovery tools should consider when ranking search results. Particularly in literature reviews, the diversity of topics among retrieved articles is considered essential (Akbulut, 2022, p. v). In searches using query terms spanning multiple disciplines, limited topic diversity in top-ranked results may restrict users' access to findings from a broader range of topic areas. Ensuring diversity in search results is also crucial for queries conducted using short or ambiguous terms. In such cases, search results should be ranked within a defined relevance framework, while also considering users' diverse information needs (Santos et al., 2015, p. 1529).

The placement of relevant results—as determined by users—in the top-tier of the rankings is one of the most critical indicators of a discovery tool's performance. A notable information-seeking behavior in long search result lists is that users tend to prioritize higher-ranked publications, often disregarding lower-ranked results as irrelevant. A study conducted by Nichols et al. (2014) highlights a phenomenon referred to as the “first result syndrome.” According to the study, users tended to assume that the most relevant result appears first, thus ignoring other results. This highlights the critical role of relevance-based ranking in search results. However, optimization strategies employed by discovery tools to expand result coverage can paradoxically lead to an overwhelming number of results, many of which may be contextually less relevant, thereby complicating the information retrieval process. In the context of electronic resources, viewing and full-text download statistics serve as key decision metrics for libraries when renewing electronic resource subscriptions, or evaluating alternative access models. Because the ranking of search results in discovery tools directly influences these statistics, ensuring that search results are appropriately ranked based on relevance is essential for usage statistics to accurately reflect actual user behavior.

Discovery tools, which Breeding (2005) defines as “centralized search,” vary in terms of their interface features, the richness of indexed collections, and the relevance of search results. Over time, features designed to facilitate research during the search process or to organize search results have been continuously updated. Enhancements such as improved visual design, relevance-based ranking, and the integration of user-generated reviews for resources have been incorporated into these systems (Breeding, 2010, p. 32). As a result, discovery tools have increasingly resembled internet search engines, evolving into a single search box model. However, many librarians argue that simple searches conducted with a single keyword may lack precision and could potentially mislead users (Chickering & Yang, 2014). Despite these concerns, there has been a growing trend toward expanding search scopes, with many systems prioritizing a broad, unified search option, such as “search across all fields,” or “search everything,” rather than allowing users to refine searches by specific access points, such as title, author, abstract, or keyword.

This study aims to analyze the scope, relevance, and topical diversity of search results generated by discovery tools in response to queries. The research focuses on evaluating the performance of widely used discovery tools in university libraries throughout Turkey. Performance assessment is based on search result retrieval, ranking quality, and topic diversity index. For the comparative analysis, discovery tools with the highest usage rates in Turkey—EBSCO Discovery Service (EDS), EKUAL Discovery Service (EKUAL DS), and Piri Discovery Service (Piri DS)—were selected.

Developed by EBSCO, EKUAL DS was among the first discovery tools adopted by university libraries in Turkey. It serves as an indexing tool for databases made accessible to universities under the EKUAL (National Academic License for Electronic Sources) framework. For institutions seeking to integrate licensed electronic publications and bibliographic records of physical resources into the discovery ecosystem, EDS emerged as a significant alternative. However, the most notable competitor of EDS and EKUAL DS for university libraries is Piri DS, introduced to the market in 2021. Developed by INSERES, Piri DS is a specialized discovery tool that integrates library catalogs and databases, employing modern search algorithms enhanced with artificial intelligence (INSERES, n.d.).

This study seeks answers to the following research questions:

- Is there a significant difference among EDS, EKUAL DS, and Piri DS tools in terms of the number of retrieved results, the ranking of results based on relevance, and topic diversity of retrieved results?
- Are there differences among the discovery tools analyzed in terms of functions, such as search fields, search options, and filtering options?

Literature Review

With their adoption rates steadily increasing, web-based discovery tools have become an integral part of library services. Connaway et al., (2020) conducted a study involving over 1,300 participants from 68 countries, revealing that 84% of libraries used at least one discovery tool. The most frequently preferred discovery tools were identified as WorldCat Discovery (WDS) (36%) and EDS (35%). It is evident that libraries are centralizing discovery tools in their information access processes, thereby reducing their dependence on multiple platforms.

Comparing discovery tools with other academic platforms is crucial for understanding system usage trends. Wang et al., (2018) analyzed DOI link referrals and, based on data from

the Chronograph project (2010–2018), found that most DOI accesses were obtained through ProQuest (Summon, Primo, and other ProQuest databases) and Web of Science. Google (including Google Scholar and Google Search) ranked third, followed by Scopus and EBSCO (EDS), while WorldCat (WorldCat Discovery and WorldCat Local) exhibited a lower usage rate. These findings indicate that while discovery tools play a significant role in academic information access, users still tend to favor general platforms such as Google and Google Scholar.

User information-seeking behavior plays a critical role in the development of discovery tools. A study conducted by Ndumbaro (2023) using data from the University of Dar es Salaam Library catalog revealed that, on average, 1.9 terms were used for 5,018 queries, while the number of terms increased to 2.66 for 5,456 reformulated queries. The fact that 95.92% of a total of 30,474 queries contained three or fewer terms indicates that users predominantly prefer short and simple queries. These findings highlight the need for optimizing discovery tools to effectively accommodate and respond to short queries.

Comparative studies assessing the performance of discovery tools have established various criteria for measuring system effectiveness. Lee and Chung (2016) conducted a study comparing EDS with the ERIC, ERC, LISA, and LISTA databases, developing a formula to assess search result relevance. The study evaluated the top 10 search results by assigning relevance scores and comparing their impact levels. Search result evaluation was based on the degree of alignment between retrieved items and the search query, with a composite score calculated from the total assigned points. The study concluded that while EDS retrieves a broad range of results, its ranking algorithms require improvement to enhance result relevance and ordering.

A similar comparative analysis was conducted by Pulikowski and Matysek (2021) for Google, Google Scholar, EDS, and LISA. In this study, nine queries were performed under three topic categories within the field of library and information science, and the top 10 retrieved results were analyzed. The findings indicated that Google provided the best results for simple searches, although it was noted that Google does not eliminate duplicate results. Google Scholar demonstrated a performance similar to Google, whereas EDS fell below expectations.

Similarly, Hanneke and O'Brien (2016) compared EDS, Summon, and Primo OneSearch in terms of the number of results retrieved, and their relevance in the field of medicine and health sciences. Their findings suggested that EDS retrieved more relevant results than the other discovery tools. However, it was emphasized that the study was based on a limited dataset, making it insufficient for providing a general recommendation.

While these studies identified limitations in the ranking effectiveness of discovery tools, Akbulut and Tonta (2022) specifically examined ranking algorithms themselves. Their study evaluated commonly used ranking methods and proposed an alternative approach utilizing pennant access techniques to incrementally enhance relevance rankings. The findings suggest that this method could be implemented across various information systems, including discovery tools.

Further studies have also compared the effectiveness of various discovery tools and academic search platforms. Ciccone and Vickery (2015) compared Summon, EDS, and Google Scholar based on user queries. In this study, relevance assessment was conducted solely based on whether the query term appeared in the title or abstract. The results revealed no statistically significant difference among the three tools when searching for

a known item, while all three returned a comparable number of relevant results for topic searches. Similarly, Trujillo's (2025) study compared Primo, EDS, WorldCat Discovery, and Summon in terms of their performance in known-item searches. Focusing on the retrieval of popular books, the study found that Google and Amazon outperformed the library discovery tools analyzed. It was also observed that the library discovery tools' algorithms tended to emphasize certain factors, such as citation counts and the number of editions.

Walters (2009) compared Google Scholar with 11 bibliographic databases and found that, based on recall and precision values, Google Scholar outperformed most of the databases. Singh et al., (2023) conducted a comparative study of Web of Science, Scopus, and Dimensions, analyzing various attributes such as altmetrics, bibliographic matching, and abstract texts, along with a relevance assessment. According to participants' relevance scores, Web of Science outperformed the other databases in three of five queries, while Web of Science and Scopus performed equally well in one.

While these studies primarily focused on the technical performance and ranking effectiveness of discovery tools, user search behaviors and interface usability were equally critical factors in assessing their overall effectiveness. Asher et al. (2013) analyzed the search behaviors of users from Bucknell University and Illinois Wesleyan University across Summon, EDS, Google Scholar, and traditional library databases to evaluate the effectiveness of these tools based on retrieved results. The findings indicated that EDS was more effective in providing access to academic sources, and guided users more efficiently. However, Summon and Google Scholar were also preferred, particularly for their ease of use and user familiarity. In a comparable study, AlHamad (2025) compared abstracting and indexing (A&I) databases with discovery tools, using a survey conducted with 69 academic library staff. The study found that discovery tools were perceived as effective in terms of usability and broad access, but A&I databases were considered indispensable for conducting comprehensive and in-depth academic research.

The usability of discovery tools is another key variable influencing their effectiveness. Hamlett and Georgas (2019) conducted a study measuring the ease users experience with discovery tool interfaces and functionalities. Their study analyzed user interactions with Primo OneSearch, revealing that participants found the interface complex and overwhelming. Additionally, 23.3% reported difficulties in accessing full texts, while 40% struggled to locate the citation function.

A comparative study by Niu et al. (2014) examined Primo OneSearch and VuFind, assessing their prominent features through log data analysis. The results indicated that Primo OneSearch was preferred for retrieving articles, whereas VuFind was more frequently used for books and media sources.

Similarly, Tonyan and Piper (2019) investigated user opinions and experiences with Summon at the University of Colorado, concluding that, although Summon retrieved a high number of results, participants spent more time navigating these results. Nichols et al. (2014) also conducted a study on the Primo OneSearch discovery tool. Participants who used filtering completed the assigned tasks with ease; however, they struggled with sorting and refining long result lists.

Beyond individual usability assessments, broader trends and perceptions toward discovery tools in academic libraries have also been explored. Nichols et al. (2017) investigated trends, approaches, and librarian attitudes toward discovery tools. Aharony and Prebor (2015)

assessed key features, such as conceptual evaluation, satisfaction levels, attitudes, and user experiences. Wong (2024) conducted a study focusing on the organizational placement and management of discovery tools within academic libraries. Based on survey data collected from library staff, the study identified the specific roles and responsibilities held by staff members in the administration of discovery systems. It concluded that increased collaboration is needed among departments involved in the management process.

Data Sources and Methodology

Test collections used in the performance evaluation of information retrieval systems typically consist of three fundamental components: a document collection, search scenarios, and relevance assessments (Carevic & Schaer, 2014). This study was conducted using the iSearch test collection, which incorporates all three components. Developed within the framework of Lykke et al. (2010), the iSearch test collection comprises 434,813 physics articles from arXiv. A total of 65 scenarios were created by academics and graduate students in the field of physics. For each scenario, an average of 200 articles was selected from the iSearch test collection, and relevance assessments were performed. Participants rated the articles on a scale of 0 (irrelevant) to 3 (high relevance).

Data Collection and Query Process

To determine the coverage rate of the discovery tools in the arXiv collection, the entire corpus was queried, and the total number of results was determined by selecting the arXiv collection. When the results from 2009 and earlier, which were covered by the articles in the iSearch test collection, were filtered, EDS and EKUAL DS returned 579,000 results, while Piri DS returned 442,000 results. This difference indicates that arXiv records are missing in Piri DS's index.

A total of 63 queries¹ were run on EDS, EKUAL DS and Piri DS systems, and the result lists were analyzed. For example, when Query 1 (manipulation, nano spheres, peptides, immobilization) was run, EDS and EKUAL DS returned 13,234 results, while Piri DS returned 10,000 results.² In response to this query, it was verified whether the articles in the iSearch test collection were included in the result lists. It was found that only one of the nine articles with an interest score of between 1 to 3 was included in the lists. All 63 queries were run, and the ranking values of the accessible articles were entered into the dataset.

In the study, the query terms in the iSearch test collection were combined with Boolean operators to form composite queries. The queries were realized by using the "OR" operator in the Title and Abstract fields. After each query was run, the results were filtered by selecting the arXiv collection.

¹The iSearch test collection also includes records of printed materials from the Royal Library of Denmark (Lykke et al., 2010, p. 628). However, since these records do not contain arXiv articles, they were excluded from the study. Additionally, 1,163 records with incomplete data were removed from the dataset. The 5th query was excluded as it contained only records outside the study's scope. Similarly, for the 17th query, all evaluated articles received a relevance score of 0, indicating that they were deemed irrelevant. Following these exclusions and adjustments, the dataset included 63 queries with at least one article assigned a relevance score of 1 to 3, totaling 2,403 articles.

²In EDS and EKUAL DS, the first 25,000 results could be exported, whereas in Piri DS, only the first 10,000 results were accessible for export. Therefore, the cutoff point for each query was set at 10,000.

Data Analysis and Performance Evaluation

After completing the queries, all articles with positive relevance scores in the iSearch test collection were searched within the result lists. The rank positions of the retrieved articles and the search result depth were recorded in the dataset. Queries that did not return any results were excluded from the analysis.

As EDS and EKUAL DS were found to return exactly the same number and order of results, the results of these two discovery tools were shown in a single list. Rank positions, total number of results, and the formulas to be used in the calculation of metrics were organized as two separate data sets: EDS / EKUAL DS, and Piri DS.

The articles subjected to relevance assessment constituted only 2% of the collection. However, articles that were not assessed for relevance, but which were related to the query terms, were also included among the results retrieved by the discovery tools. The extent to which these unassessed articles influenced the ranking of other articles deemed highly relevant by experts remains unknown. The presence of unassessed articles and their potential impact on result rankings were among the limitations of this study.

Measures and Calculations

DCG (Discounted Cumulative Gain) is a metric used to measure the ranking performance of search systems based on the relevance of the results (Cossock & Zhang, 2008). DCG is calculated based on the relevance score and rank position of the results returned for a query. In this study, the DCG value was calculated using the following formula:

$$DCG = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)}$$

The i value in the formula represented the ranking, p represented the total number of results, and rel represented the relevance value (Singh et al., 2023).

The DCG value varies depending on the search result depth. Therefore, it is necessary to normalize the data, recalculate DCG, and obtain a valid value for each ranking. NDCG (Normalized Discounted Cumulative Gain) evaluates ranking performance regardless of the number of retrieved results (Brama et al., 2022). For the normalization process, the IDCG (Ideal DCG) was computed, and the NDCG (Normalized Discounted Cumulative Gain) value was obtained. IDCG represents the recalculated DCG value based on the optimal ranking order (Wang et al., 2013). The formula is as follows:

$$IDCG = \sum_{i=1}^{|REL_p|} \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

In this context, the NDCG value is calculated as the ratio of DCG to IDCG.

The NDCG value varies between 0 and 1. A value closer to 1 indicates that the ranking is closer to the ideal ranking.

The Shannon Diversity Index was used to measure the diversity of topics in the study. This Index evaluates the diversity of information provided by the system by measuring the distribution of different topics within a defined range. The Shannon Diversity Index was calculated according to the following formula:

$$H(i) = -\sum p_i \log_2 p_i$$

In the formula, p_i denotes the probability of a particular topic category being found in the total results (Han & Kobayashi, 2002). In the literature, it is stated that users are generally interested in the results on the first page, but users searching for specific topics examine a wider set of results (Wu & Kelly, 2014). Therefore, in this study, the first 20 results were set as the threshold value in the diversity analysis.

Findings

When the entire arXiv corpus was selected, EDS and EKUAL DS returned 2.371 million results, while Piri DS returned 2.228 million results. When all publications from 2009 and earlier which constituted the iSearch test collection were retrieved, EDS and EKUAL DS returned 579,000 results, while Piri DS returned 442,000. The analysis showed that the difference between the number of results retrieved by the discovery tools was due to the fact that the year information of the results retrieved by Piri DS was largely inaccurate. It was found that there were problems in accessing articles dating back to 2007. Articles with incorrect year information in Piri DS were eliminated from the search results when the year filter was applied.

In the iSearch test collection, a total of 9,905 articles were evaluated for relevance and assigned a score. Of these, 7,502 articles received a relevance score of "0;" 1,603 articles were scored "1;" 524 were scored "2;" and 276 were scored "3." In total, 2,403 articles were categorized as low, adequate, or high interest.

In this study, 63 queries from the iSearch test collection were run on the discovery tools. No discovery tool returned results for queries 49 and 62. Therefore, the number of queries that returned results in at least one discovery tool was 61, and the total number of articles in these queries was 2,356.

The number of queries in which at least one article appeared within the top 10,000 results was 40 (65%) for EDS and EKUAL DS, with a total of 524 articles (22%) listed within these results. In contrast, for Piri DS, the number of queries with at least one article within the top 10,000 results was 47 (77%) with a total of 832 articles (35%) included in these queries. These findings indicate that Piri DS retrieved articles from the dataset at a higher rate within the top 10,000 results compared to EDS and EKUAL DS. Additionally, an examination of the results ranked beyond the top 10,000 in EDS and EKUAL DS revealed that a total of 207 articles (8%) were ranked between 10,000 and 25,000.

The queries with the highest percentage of results within the top 10,000 were identified as Query 20 and Query 41 for EDS and EKUAL DS. In Query 20, 39 out of 62 articles (63%) were ranked in the result lists, while 83 out of 145 articles (57%) in Query 41 appeared within the top 10,000. For Piri DS, the queries with the highest retrieval percentage were Query 27 (N = 63, 72%), Query 29 (N = 102, 88%), and Query 41 (N = 77, 53%). An analysis of queries that retrieved a high number of results in EDS and EKUAL DS revealed that they frequently contained multiword terms and compound expressions (e.g., "far-zone calculations"). Similarly, Query 27, which yielded a high number of results in Piri DS, also included compound terms, such as "single-photon indistinguishability." For Query 32, which contained special characters (e.g., "N = 4 SYM"), all three discovery tools returned the same number of results (N = 26, 56%). In contrast, when examining queries for which none of the discovery tools retrieved any results (e.g., Query 5, Query 19), no differences were found in the number of terms, nor was there any use of punctuation, or special characters. These findings indicated that the number of terms, punctuation marks, and special characters did not have a direct determining effect on search performance.

The limited presence of corpus articles in the results retrieved by the discovery tools may be attributed to the prevalence of articles with a relevance score of 1. Therefore, it is essential to conduct a detailed examination of the distribution of articles with relevance scores of 2 and 3 within the result lists. The distribution of retrieved results based on relevance scores is presented in Table 1.

TABLE 1					
Retrieval Performance of Discovery Tools for iSearch Test Collection Articles					
Relevance Score	iSearch	Retrieval of Relevance-Scored iSearch			
		Articles in Discovery Tools			
	Total	EDS / EKUAL DS		Piri DS	
	N	N	%	N	%
1	1,573	296	18.8	479	30.5
2	512	150	29.3	234	45.7
3	271	78	28.8	119	43.9
Total	2,356	524	22.2	832	35.3

EDS and EKUAL DS retrieved 78 out of 271 articles with a relevance score of 3, and 150 out of 512 articles with a relevance score of 2; whereas, Piri DS retrieved 119 and 234 articles, respectively. It was observed that Piri DS covered approximately half of the articles with relevance scores of 2 and 3; this proportion remained lower in EDS and EKUAL DS. Additionally, EDS and EKUAL DS retrieved only 228 out of a total of 783 articles with relevance scores of 2 and 3 within the top 10,000 results. This finding indicates a significant limitation of these systems.

EDS and EKUAL DS retrieved fewer results within the top 10,000 rank positions, and the retrieved results were positioned at lower rankings compared to Piri DS. The rank distribution of results based on relevance scores is presented in Table 2. Within the top 100 results EDS and EKUAL DS ranked only 2.6% of the articles with a relevance score of 3 and 2.0% of the articles with a relevance score of 2. In contrast, in Piri DS, these percentages were 20.2% for articles with a relevance score of 3 and 13.7% for those with a relevance score of 2.

TABLE 2												
Ranking Distribution of Retrieved Results Based on Relevance Scores in Discovery Tools												
Distribution of Rankings	EDS / EKUAL DS						Piri DS					
	Relevance Score						Relevance Score					
	1		2		3		1		2		3	
	N	%	N	%	N	%	N	%	N	%	N	%
1–100	4	1.4	3	2.0	2	2.6	42	8.8	32	13.7	24	20.2
101–500	30	10.1	6	4.0	9	11.5	82	17.1	49	20.9	37	31.1
501–1,000	36	12.2	18	12.0	12	15.4	57	11.9	39	16.7	15	12.6
1,001–5,000	172	58.1	103	68.7	44	56.4	207	43.2	83	35.5	32	26.9
5,001–10,000	54	18.2	20	13.3	11	14.1	91	19.0	31	13.2	11	9.2
Total	296	100.0	150	100.0	78	100.0	479	100.0	234	100.0	119	100.0

An analysis of the lower-ranked results revealed that relevant articles in EDS and EKUAL DS fell outside the first 1,000 positions. EDS and EKUAL DS ranked 68.7% of the articles with an interest score of 2 between the 1,001st and 5,000th ranking positions; this ratio was 35.5% in Piri DS. For articles with an interest score of 3, the distribution was 56.4% for EDS and EKUAL DS, and 26.9% for Piri DS. These findings indicated that EDS and EKUAL DS tended to rank high-interest articles lower in the result lists, while Piri DS ranked these articles higher.

Within the top 500 results, it was found that 14.1% of articles with an interest score of 3 and 6.0% of those with an interest score of 2 were ranked within this range in EDS and EKUAL DS. In contrast, 51.3% of articles with an interest score of 3 and 34.6% of those with an interest score of 2 were within the top 500 in Piri DS. When the lower rank positions were examined, it was determined that 70.5% of articles with an interest score of 3 in EDS, and EKUAL DS were ranked outside the 1,000th rank position. Similarly, 82.0% of articles with an interest score of 2 were listed outside the 1,000th rank position. In Piri DS, these rates are comparatively lower than in EDS and EKUAL DS, with 48.7% for articles with an interest score of 3 and 36.1% for those with an interest score of 2. These findings indicate that EDS and EKUAL DS not only retrieved fewer relevant results but also tended to position them in lower rank positions. On the other hand, Piri DS provided a more balanced ranking distribution for relevant results, demonstrating a comparatively superior performance to EDS and EKUAL DS in this regard.

The values obtained in DCG calculations depend on the quality of the top-ranked results (Cossock & Zhang, 2008). In discovery tools, DCG values increase when relevant results are accessed in higher ranks (Akbulut, 2022, p. 56). For instance, in Query 42, the result list contained 44 articles with relevance scores of 1, 2, and 3. Among these, the article entitled “Flow Instabilities of Magnetic Flux Tubes—III. Toroidal Flux Tubes,” which had a relevance score of 3, held the highest DCG value (0.903) in Piri DS. The position of this article within the 10,000 retrieved results was ninth. Conversely, the lowest DCG value (0.076) belonged to the article listed first in the query results. This article had an interest score of 1 and was positioned in 8,808th place. The DCG values for each query in the discovery tools are illustrated in Figure 1.

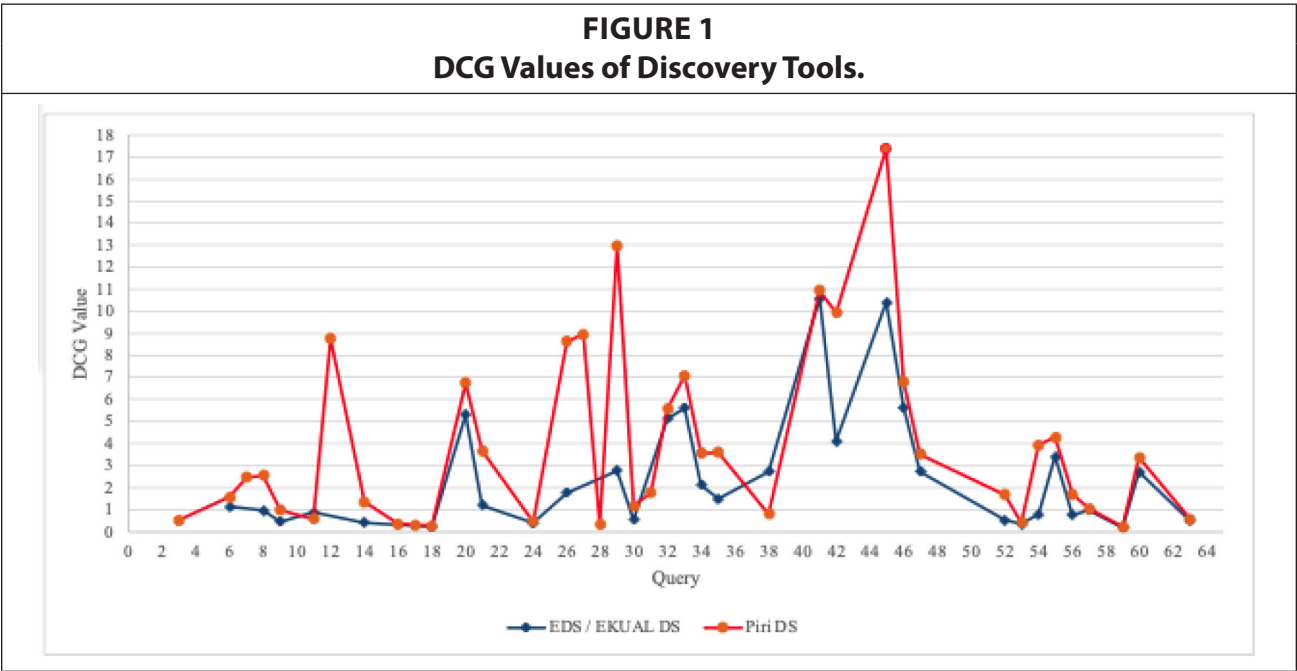


Table 3 presents the DCG and NDCG values for all queries where multiple articles were retrieved. Queries that retrieved only a single article were excluded from the analysis, as they did not provide a meaningful assessment of ranking performance. Because such queries inherently have an NDCG value of “1,” they tend to inflate the average values. According to the values presented in Table 3, the lowest NDCG value in EDS and EKUAL DS (0.780) was obtained from the ranking results of Query 29. In Piri DS, the lowest NDCG value (0.649) was calculated for the ranking results of Query 21.

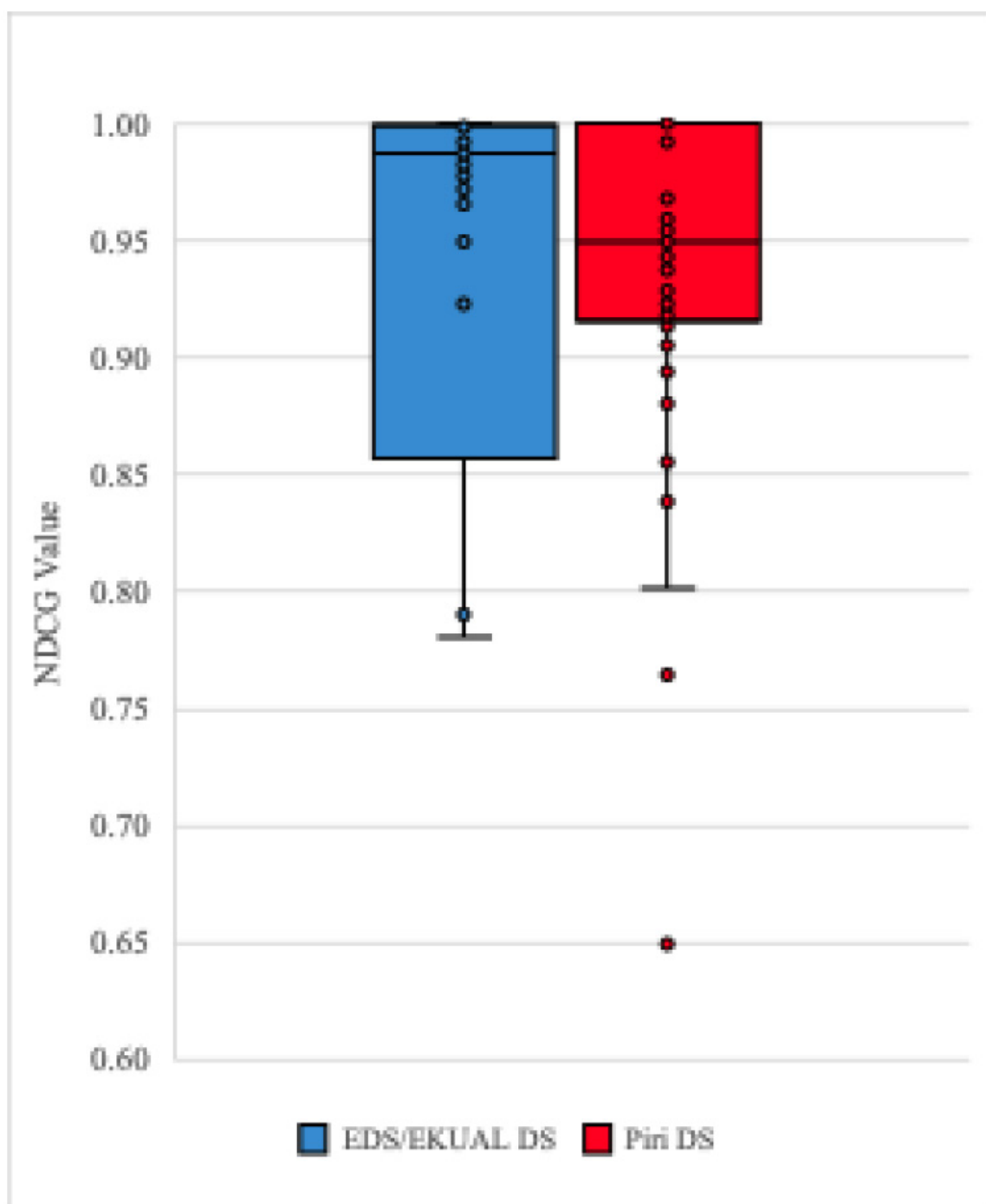
TABLE 3
DCG and NDCG Values of Discovery Tools

Query	EDS / EKUAL DS		Piri DS		Query	EDS / EKUAL DS		Piri DS	
	DCG	NDCG	DCG	NDCG		DCG	NDCG	DCG	NDCG
3	N/A	N/A	0.525	1.000	32	5.121	0.972	5.573	0.943
6	1.127	1.000	1.559	1.000	33	5.616	0.968	7.083	0.894
7	N/A	N/A	2.464	1.000	34	2.116	0.991	3.565	0.923
8	0.945	0.993	2.542	0.954	35	1.469	0.987	3.605	0.949
9	0.453	1.000	0.966	1.000	38	2.729	1.000	0.803	1.000
11	0.864	1.000	0.573	0.895	41	10.573	0.982	10.938	0.956
12	N/A	N/A	8.765	0.992	42	4.096	0.980	9.944	0.928
14	0.402	1.000	1.338	1.000	45	10.403	0.974	17.363	0.959
16	N/A	N/A	0.340	0.968	46	5.630	0.988	6.782	0.919
17	N/A	N/A	0.297	1.000	47	2.725	0.990	3.521	0.919
18	0.204	1.000	0.248	1.000	52	0.515	0.990	1.684	0.764
20	5.297	0.988	6.750	0.918	53	0.345	0.923	0.403	0.838
21	1.202	0.790	3.660	0.649	54	0.772	0.965	3.935	0.880
24	0.377	1.000	0.457	1.000	55	3.410	0.949	4.286	0.801
26	1.767	0.977	8.637	0.905	56	0.750	1.000	1.705	0.937
27	N/A	N/A	8.959	0.920	57	0.976	0.989	1.039	0.926
28	N/A	N/A	0.344	0.960	59	0.187	1.000	0.215	1.000
29	2.776	0.780	12.951	0.913	60	2.702	0.992	3.361	0.958
30	0.546	1.000	1.175	1.000	63	0.493	0.998	0.566	1.000
31	N/A	N/A	1.763	0.855					

When examining the average DCG³ values, EDS and EKUAL DS had an average of 2.470, while Piri DS had a higher average of 3.863. Although the higher DCG average in Piri DS suggested that it ranked results more effectively, differences in search result depth necessitated the calculation of Normalized DCG (NDCG). The NDCG average for EDS and EKUAL DS was calculated as 0.973; whereas, for Piri DS, it was 0.933. These values were quite close to each other, indicating that both discovery tools approached the ideal ranking (1.0). The

³ $DCG = \sum_{i=1}^p \frac{rel_i}{\log_2(i+1)}$

FIGURE 2
NDCG Values of Discovery Tools.



distribution of NDCG values is presented in Figure 2. In Piri DS, NDCG values predominantly ranged between 0.800 and 1. While Piri DS exhibited a broader and more balanced distribution, the values in EDS and EKUAL DS tended to cluster around the median (0.990) and the mean (0.973).

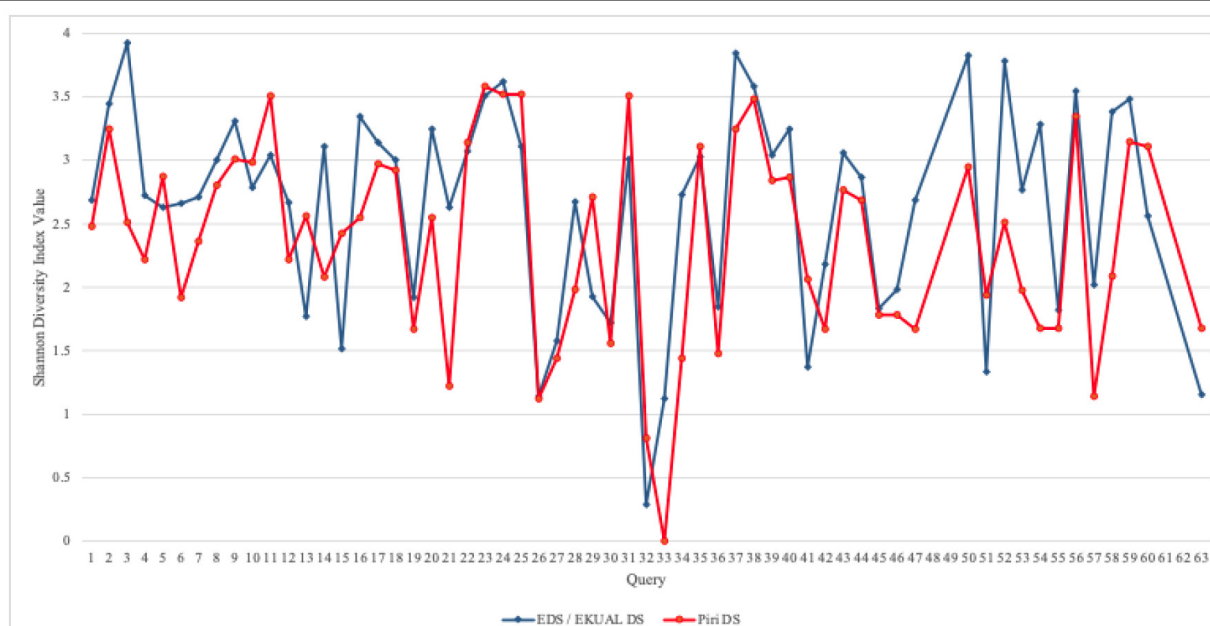
In 59 out of the 61 queries that returned results in the discovery tools, more than 20 results were retrieved, and the diversity index values for these queries were calculated.

According to the Shannon Diversity Index, as diversity decreases, the values approach 0. For example, in Query 32, because 19 out of the top 20 results retrieved by EDS pertained to the field of “high energy physics-theory,” the diversity index for this query was 0.286. Similarly, in Query 33, because all the top 20 results retrieved by Piri DS were related to “high energy physics-theory,” the diversity index for this query was 0. The diversity index values for all queries are presented in Table 4.

TABLE 4
Shannon Diversity Index Values of Discovery Tools

Query	Shannon Diversity Index		Query	Shannon Diversity Index	
	EDS / EKUAL DS	Piri DS		EDS / EKUAL DS	Piri DS
1	2.684	2.480	33	1.122	0.000
2	3.446	3.246	34	2.728	1.437
3	3.922	2.509	35	3.028	3.109
4	2.722	2.215	36	1.843	1.479
5	2.626	2.871	37	3.846	3.246
6	2.659	1.919	38	3.584	3.484
7	2.709	2.361	39	3.039	2.839
8	3.004	2.802	40	3.246	2.866
9	3.309	3.009	41	1.369	2.064
10	2.784	2.984	42	2.183	1.671
11	3.041	3.509	43	3.061	2.766
12	2.666	2.215	44	2.864	2.684
13	1.771	2.558	45	1.833	1.781
14	3.109	2.081	46	1.981	1.781
15	1.517	2.423	47	2.684	1.671
16	3.346	2.546	48	N/A	N/A
17	3.141	2.971	49	N/A	N/A
18	3.004	2.922	50	3.822	2.946
19	1.919	1.671	51	1.336	1.939
20	3.246	2.546	52	3.784	2.509
21	2.628	1.219	53	2.764	1.977
22	3.071	3.141	54	3.284	1.675
23	3.509	3.584	55	1.817	1.679
24	3.622	3.522	56	3.546	3.346
25	3.109	3.522	57	2.020	1.141
26	1.141	1.122	58	3.384	2.090
27	1.579	1.437	59	3.484	3.146
28	2.671	1.980	60	2.558	3.109
29	1.923	2.709	61	N/A	N/A
30	1.717	1.557	62	N/A	N/A
31	3.009	3.509	63	1.154	1.679
32	0.286	0.811			

FIGURE 3
Comparison of the Shannon Diversity Index of Discovery Tools.



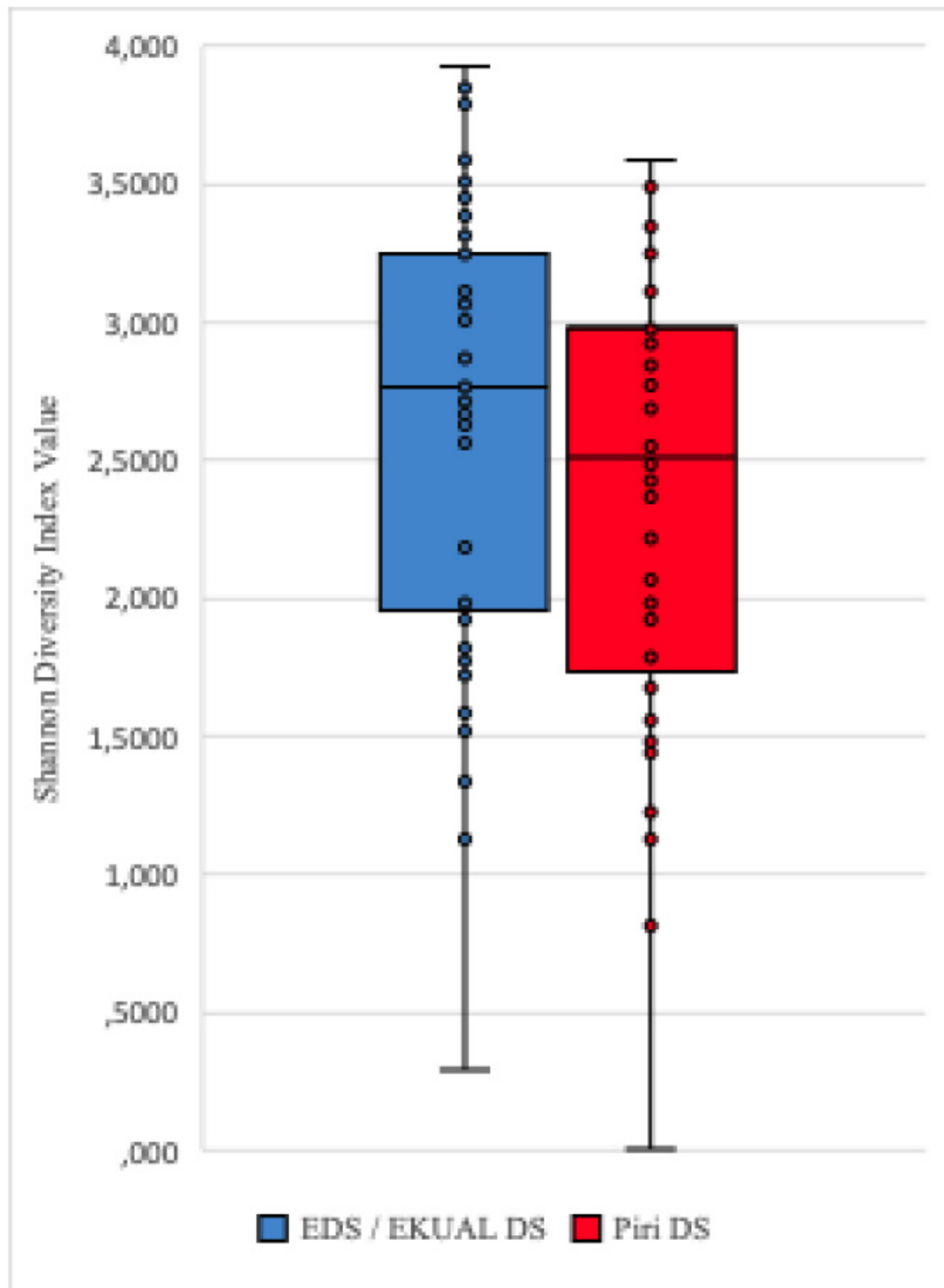
The greatest variation in diversity index values occurred in Query 3, Query 21, and Query 54. When the topic categories of the results of these queries were analyzed, it was found that EDS and EKUAL DS included a wider variety of topics in the top 20 results. The comparison of the Shannon Diversity Index between EDS / EKUAL DS, and Piri DS is presented in Figure 3.

For Query 3, it was found that the results retrieved by Piri DS were predominantly concentrated in the fields of quantitative methods and machine learning. By contrast, EDS and EKUAL DS included results not only from these fields, but also from diverse disciplines such as general economics, medical physics, and information theory, thereby offering a broader topical coverage. Similarly, in Query 21, while the results retrieved by Piri DS were largely focused on optics and quantum physics, EDS and EKUAL DS retrieved results from additional fields—such as signal processing and applied physics—contributing to greater topic diversity. Likewise, in Query 54, Piri Discovery primarily retrieved results related to mesoscale and nanoscale physics and materials science, whereas EDS and EKUAL DS also retrieved results from nuclear theory, applied physics, astrophysics, and optics, contributing to a broader range of topics.

The average topical Diversity Index was calculated as 2.648 for EDS and EKUAL DS and 2.374 for Piri DS. It was found that EDS and EKUAL DS offered relatively higher topical diversity than Piri DS. The distribution of Diversity Index values is presented in Figure 4.

In diversity indices, the maximum value is achieved when all categories are completely distinct from each other (Peet, 1975, p. 496). Similarly, in the topic diversity index, the maximum value is reached when the retrieved results contain an equal number of different topics up to the threshold value. The Diversity Index for a single-topic result in a query was calculated as 0.216. Because a query can contain results from a maximum of 20 different topics, the maximum possible Diversity Index was 4.321 (20×0.216). Accordingly, the highest Diversity Index was calculated as 3.922 (Query 3) for EDS and EKUAL DS, and 3.584 (Query 23) for Piri

FIGURE 4
Shannon Diversity Index Values of Discovery Tools.



DS. The lowest Diversity Index was 0.286 (Query 32) for EDS and EKUAL DS, while it was 0 (Query 33) for Piri DS. An analysis of the values indicated that EDS and EKUAL DS exhibited a higher number of queries with values closer to the maximum index, whereas Piri DS tended to cluster around the mean and median Diversity Index values (2.509). The median Diversity Index for EDS and EKUAL DS was 2.764. The findings suggest that while the diversity indices of the discovery tools were close to their median values, their average values remained below 3, indicating a significant gap from the maximum possible diversity. This suggests that the discovery tools did not provide sufficient topic diversity in their retrieved results.

Conclusions and Recommendations

This study presents a comparative analysis of the discovery tools EDS, EKUAL DS and Piri DS in terms of ranking accuracy, coverage rate, accessibility of high-interest articles, and topic diversity. The search terms used in the construction of the iSearch test collection were executed as queries in the discovery tools, and their retrieved results were examined to assess their coverage rate. The ranking quality of these tools was evaluated by determining the position of relevance-judged articles within the retrieved results. Additionally, the topic diversity index was calculated for the top 20 results retrieved by each query, providing a basis for comparing the diversity of topics across the discovery tools. The findings reveal significant differences in ranking performance, accessibility of high-relevance articles, and topic diversity.

EDS and EKUAL DS were found to be ineffective in ranking high-relevance articles in top positions, with a significant portion of relevant results appearing outside the first 1,000 rank positions. Specifically, EDS and EKUAL DS ranked 70.5% of articles with an interest score of 3, and 82.0% of articles with an interest score of 2, outside the first 1,000 results. While Piri DS performed better in this regard, its ability to retrieve only half of the relevant articles indicated similar limitations. This issue requires users to exert more effort in locating high-relevance results, often necessitating extensive filtering to refine result lists. To enhance ranking performance, discovery tools might improve their ranking algorithms to prioritize high-relevance articles. AI-driven ranking models might be implemented for complex queries, and relevance-based ranking algorithms might be developed to ensure that the most relevant results appear in higher positions.

EDS, EKUAL DS, and Piri DS exhibited similar performances in terms of overall coverage; however, when the entire arXiv collection was queried, Piri DS retrieved 172,000 fewer results. This indicated missing bibliographic records and suggested that certain publications had not been integrated into the discovery system. To ensure that discovery tool indices remain complete and up to date, closer collaboration between libraries and discovery service providers would be essential. Specifically, missing records might be regularly audited and incorporated into the discovery systems. Additionally, a significant portion of the publications retrieved from the arXiv collection in Piri DS contained incorrect publication year information. This metadata error rendered the year filter ineffective, which probably made it difficult for users to locate relevant publications. To resolve this issue, bibliographic records should be systematically verified, and erroneous metadata should be corrected. Moreover, libraries might conduct periodic audits of the collections included in their discovery tools, to ensure accuracy and completeness.

EDS and EKUAL DS provide greater topic diversity, compared with Piri DS. However, the topic diversity index values for both discovery tools remained significantly below the maximum threshold, indicating insufficient distribution of topics among retrieved results. To

address this limitation, ranking algorithms should be optimized to enhance topic diversity without compromising ranking quality. Specifically, ranking models that prioritize topic diversity up to a certain threshold could be developed to ensure a more balanced distribution of subjects in search results.

It was found that discovery tools differ in search fields, filtering options, and query processing mechanisms. Notably, variations in scope and functionality within the interfaces of EDS and EKUAL DS can create inconsistencies in user experience. To address this issue, it is recommended that discovery tools' user interfaces and functionalities be standardized. A single, unified interface for all libraries would ensure that all users have access to the same features and functions in a consistent manner.

Regarding discovery tools, libraries currently have only limited control of the management of their collections. To strengthen collection management, authorized accounts should be assigned to libraries, allowing them to manage collections, configure filtering options, and regulate access statistics. Additionally, some publishers do not fully integrate the collections they provide under subscription agreements into discovery tool indices, thereby restricting the scope of these tools. Stronger collaboration between discovery tool providers and publishers is essential to ensure the inclusion of missing collections in discovery indexes.

Acknowledgment

This article is based on the master's thesis of the first author, conducted under the supervision of the second author. The authors would like to thank the developers of the iSearch test collection for granting permission to use the dataset in this study. We also gratefully acknowledge Müge Akbulut for her valuable contributions to this research.

Declarations

Conflict of interest: The authors declare that there is no conflict in interest or financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Declaration of generative AI and AI-assisted technologies in the writing process: During the preparation of this work the authors, as non-native speakers of English, occasionally used ChatGPT to refine grammatical expression and terminology. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

- Aharony, N., and Prebor, G. (2015). Librarians' and information professionals' perspectives towards discovery tools: An exploratory study. *The Journal of Academic Librarianship*, 41(4), 429–440. <https://doi.org/10.1016/j.acalib.2015.05.003>
- Akbulut, M. (2022). Bilgi erişimde ilgi sıralamalarının artırımı olarak geliştirilmesi [Incremental refinement of relevance rankings in information retrieval], (Doctoral thesis, Hacettepe University).
- Akbulut, M., and Tonta, Y. (2022). İlgi sıralamalarının artırımı olarak geliştirilmesi: Pennant erişimle desteklenen yeni bir yöntem önerisi [Incremental refinement of relevance rankings: Introducing a new method supported with Pennant retrieval]. *Türk Kütüphaneciliği*, 36(2), 169–203. <https://doi.org/10.24146/tk.1062751>
- Asher, A. D., Duke, L. M., and Wilson, S. (2013). Paths of discovery: Comparing the search effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and conventional library resources. *College & Research Libraries*, 74(5), 464–488. <https://doi.org/10.5860/crl-374>
- AlHamad, M. M. (2025). Balancing precision and usability: Librarian perspectives on abstracting and indexing databases versus discovery services in academic libraries. *Internet Reference Services Quarterly*, 1–12. <https://doi.org/10.1080/10875301.2025.2472419>

- Brama, H., Dery, L., & Grinshpoun, T. (2022). Evaluation of neural networks defenses and attacks using NDCG and reciprocal rank metrics. *International Journal of Information Security*, 22(2), 525–540. <https://doi.org/10.1007/s10207-022-00652-0>
- Breeding, M. (2005). Plotting a new course for metasearch. *Computers in Libraries*, 25(2), 27–29.
- Breeding, M. (2010). The state of the art in library discovery 2010. *Computers in Libraries*, 30(1), 31–34.
- Breeding, M. (2015). Future of library discovery systems. *Information Standards Quarterly*, 27(1), 24–30. <https://doi.org/10.3789/isqv27no1.2015.04>
- Carevic, Z., and Schaer, P. (2014). On the connection between citation-based and topical relevance ranking: Results of a pretest using iSearch. *BIR@ECIR*. <https://ceur-ws.org/Vol-1143/paper5.pdf>
- Ciccone, K., and Vickery, J. (2015). Summon, EBSCO Discovery Service, and Google Scholar: A comparison of search performance using user queries. *Evidence Based Library and Information Practice*, 10(1), 34–49. <https://doi.org/10.18438/b86g6q>
- Chickering, F. W., and Yang, S. Q. (2014). Evaluation and comparison of discovery tools: An update. *Information Technology and Libraries*, 33(2), 5–30. <https://doi.org/10.6017/ital.v33i2.3471>
- Connaway, L. S., Cyr, C., and Gallagher, P. (2020). *Global Perspectives on Discovery and Fulfillment: Findings from the 2020 OCLC Global Council Survey*. OCLC Research. <https://doi.org/10.25333/0pmf-gy24>
- Cossock, D., and Zhang, T. (2008). Statistical analysis of Bayes Optimal Subset Ranking. *IEEE Transactions on Information Theory*, 54(11), 5140–5154. <https://doi.org/10.1109/tit.2008.929939>
- Hamlett, A., and Georgas, H. (2019). In the wake of discovery: Student perceptions, integration, and instructional design. *Journal of Web Librarianship*, 13(3), 230–245. <https://doi.org/10.1080/19322909.2019.1598919>
- Han, T. S., and Kobayashi, K. (2002). *Mathematics of information and coding* (Vol. 203). American Mathematical Society.
- Hanneke, R., and O'Brien, K. K. (2016). Comparison of three web-scale discovery services for health sciences research. *Journal of the Medical Library Association*, 104(2), 109–117. <https://doi.org/10.5195/jmla.2016.52>
- Hartman, K. A., and Bowering Mullen, L. (2008). Google Scholar and academic libraries: An update. *New Library World*, 109(5/6), 211–222. <https://doi.org/10.1108/03074800810873560>
- INSERES. (n.d.). *Piri Discovery Tool*. <https://inseres.com/en/product/piri.html>
- Lee, B., and Chung, E. (2016). An analysis of web-scale discovery services from the perspective of user's relevance judgment. *The Journal of Academic Librarianship*, 42(5), 529–534. <https://doi.org/10.1016/j.acalib.2016.06.016>
- Lykke, M., Larsen, B., Lund, H., and Ingwersen, P. (2010). Developing a test collection for the evaluation of integrated search. In: *Advances in Information Retrieval, 32nd European Conference on IR Research, ECIR 2010, Milton Keynes, UK, March 28-31, 2010. Proceedings (LNCS 5993)*, p. 627–630. Springer. https://doi.org/10.1007/978-3-642-12275-0_63
- Ndumbaro, F. (2023). Remote OPAC users' search query reformulation (SQR) patterns: A transaction log analysis. *Online Information Review*, 47(1), 162–176. <https://doi.org/10.1108/oir-09-2020-0389>
- Nichols, A., Billey, A., Spitzform, P., Stokes, A., and Tran, C. (2014). Kicking the tires: A usability study of the Primo discovery tool. *Journal of Web Librarianship*, 8(2), 172–195. <https://doi.org/10.1080/19322909.2014.903133>
- Nichols, A. F., Crist, E., Sherrieff, G., and Allison, M. (2017). What does it take to make discovery a success? A survey of discovery tool adoption, instruction, and evaluation among academic libraries. *Journal of Web Librarianship*, 11(2), 85–104. <https://doi.org/10.1080/19322909.2017.1284632>
- Niu, X., Zhang, T., and Chen, H. (2014). Study of user search activities with two discovery tools at an academic library. *International Journal of Human-Computer Interaction*, 30(5), 422–433. <https://doi.org/10.1080/10447318.2013.873281>
- Peet, R. K. (1975). Relative diversity indices. *Ecology*, 56(2), 496–498. <https://doi.org/10.2307/1934984>
- Pulikowski, A., and Matysek, A. (2021). Searching for LIS scholarly publications: A comparison of search results from Google, Google Scholar, EDS, and LISA. *The Journal of Academic Librarianship*, 47(5), 1–8. <https://doi.org/10.1016/j.acalib.2021.102417>
- Santos, R. L. T., Castells, P., Altungovde, I. S., and Can, F. (2015). Diversity and novelty on the web: Search, recommendation, and data streaming aspects. In *24th International Conference on World Wide Web* (pp. 1529–1530). ACM.
- Singh, P., Singh, V. K., and Piryani, R. (2023). Scholarly article retrieval from Web of Science, Scopus and Dimensions: A comparative analysis of retrieval quality. *Journal of Information Science*, 0(0). <https://doi.org/10.1177/01655515231191351>
- Tonyan, J., and Piper, C. (2019). Discovery tools in the classroom: A usability study and implications for information literacy instruction. *Journal of Web Librarianship*, 13(1), 1–19. <https://doi.org/10.1080/19322909.2018.1530161>
- Trujillo, N. (2025). Finding popular books in library discovery services: Known item searches in Primo, EDS, WorldCat Discovery, and Summon. *Journal of Electronic Resources Librarianship*, 1–13. <https://doi.org/10.1080/1941126X.2025.2455781>

- Walters, W. H. (2009). Google Scholar search performance: Comparative recall and precision. *Portal: Libraries and the Academy*, 9(1), 5–24. <https://doi.org/10.1353/pla.0.0034>
- Wang, X., Cui, Y., and Xu, S. (2018). Evaluating the impact of web-scale discovery services on scholarly content seeking. *The Journal of Academic Librarianship*, 44(5), 545–552. <https://doi.org/10.1016/j.acalib.2018.05.010>
- Wang, Y., Wang, L., Li, Y., He, D., Chen, W., and Liu, T.-Y. (2013). A theoretical analysis of NDCG ranking measures. In *JMLR: Workshop and Conference Proceedings* (pp. 1–30). Annual Conference Computational Learning Theory.
- Wong, S. (2024). Organizational structure around web-scale discovery services in Canadian academic libraries. *Partnership: The Canadian Journal of Library & Information Practice & Research*, 19(1), 1–18. <https://doi.org/10.21083/partnership.v19i1.7336>
- Wu, W.-C., and Kelly, D. (2014). Online search stopping behaviors: An investigation of query abandonment and task stopping. In *Proceedings of the American Society for Information Science and Technology* (pp. 1–10). American Society for Information Science and Technology.