



Understanding Efficacy of Literature Retrieval on Robo-advisory in Finance Sector: Exploring Performance Metrics

Dr. Neha Gulati¹ & Payal Gupta²

¹Assistant Professor University Business School Panjab University, Chandigarh, India

²Research Scholar University Business School Panjab University, Chandigarh, India

ARTICLE INFO

Key Words: Information Retrieval, Precision, Recall, Robo-advisory, Systematic Literature Review (SLR)

ABSTRACT

Objective: Aim is to evaluate and compare performance of Scopus and Web of Science database in retrieving literature for Robo-advisory in finance sector.

Methodology: Five systematic literature reviews and bibliometric analysis on the theme Robo-advisory were selected. References of these 5 SLR were considered and a corpus of 137 most relevant documents were identified. From titles of 137 documents, most commonly used keywords were identified and search query “Robo-adv*” was formulated. Precision, Recall and F1 measure were calculated after executing the query on Scopus and Web of Science databases.

Results: Higher recall of 75.2% was exhibited for the query by Scopus as compared to 34.31% by Web of Science. Thus, Scopus is more effective in capturing relevant literature on the theme. The precision of query executed on Scopus was 65.71% as compared to 61.98% in Web of Science. Thus, implying that a large proportion of information retrieved from Scopus is relevant to search query thereby indicating a higher level of accuracy by Scopus. From the results of F1 score, Scopus has a better balance between precision and recall. Thereby concluding that Scopus is more effective in information retrieval as it retrieves lesser number of irrelevant documents.

Contribution: It offers valuable insights into the effectiveness of information retrieval from these databases on the theme under study. Researchers can make more informed decisions about selecting database for literature review and bibliometric analysis.

Received 15.12.2023; Accepted 02.02.2024

DOI: [10.48165/gmj.2023.18.2.5](https://doi.org/10.48165/gmj.2023.18.2.5)

Copyright @ Gyan Management Journal (acspublisher.com/journals/index.php/gmj)

Introduction

Information Retrieval is a systematic process used to extract relevant documents from the web according to users' requirements specified using keywords and operators through query. Systematic literature reviews (SLRs) and Bibliometric Analysis (BA) have gained prominence among researchers in recent years. SLRs offer a thorough and objective synthesis of existing literature and BA assists in understanding significance of research conducted by using quantitative methods to evaluate bibliographic data. **Researchers often face the challenge of choosing appropriate bibliographic DBs for conducting SLRs and BA.** Scopus and Web of Science are the most popular DBs used by researchers to retrieve literature (Pranckutė, 2021). Scopus was launched by Elsevier in 2004 and Web of Science was launched in 1997 by Clarivate. Both these DBs offer extensive coverage of scientific literature, including journals, conference proceedings, book chapters etc. across various disciplines. **Another challenge faced by researchers is the retrieval of relevant information.** Researchers face difficulty in effectively and precisely identifying relevant literature for conducting SLRs and BA. Success of SLRs and BA depends largely on efficiency and accuracy of the information retrieved from these databases (DBs). Plethora of studies have been conducted to compare these DBs. A comparative analysis of DBs- AGRICOLA, AGRIS, BIOSIS, CAB Direct, FSTA, Google Scholar, Scopus, and Web of Science has been performed on the theme agricultural information by Ritchie et al. (2019). A comparison of DBs- Cinhal, PubMed, Web of Science and Scopus in the retrieval of documents on the theme telemedicine has been conducted by Ahmadi et al. (2015). Some studies have focussed on comparison of journal coverage of Scopus and Web of Science (Mongeon & Paul-Hus, 2016; Singh et al., 2021) namely for research evaluation. Most bibliometric analyses have in common their data sources: Thomson Reuters' Web of Science (WoS). But there is dearth of research on comparison of DBs on the basis of information retrieval performance metrics- Precision and Recall and F1 score.

Precision is defined as proportion of relevant items retrieved to all items retrieved. It is a measure of

specificity implying that whether a database is capable of excluding non-relevant items (Buckland & Gey, 1994; Gusenbauer & Haddaway, 2020) Measuring precision involves judgement on the part of researcher to identify relevant documents. **Recall** is defined as proportion of number of relevant items retrieved to all relevant items. It is a measure of sensitivity (Gusenbauer & Haddaway, 2020) validity, and explanatory power. Yet, the search systems allowing access to this evidence provide varying levels of precision, recall, and reproducibility and also demand different levels of effort. To date, it remains unclear which search systems are most appropriate for evidence synthesis and why. Advice on which search engines and bibliographic databases to choose for systematic searches is limited and lacking systematic, empirical performance assessments. This study investigates and compares the systematic search qualities of 28 widely used academic search systems, including Google Scholar, PubMed, and Web of Science. A novel, query-based method tests how well users are able to interact and retrieve records with each system. The study is the first to show the extent to which search systems can effectively and efficiently perform (Boolean or effectiveness (Walters et al., 2009) of a database. **F1 Score** is defined as harmonic mean of precision and recall. This score combines precision and recall into a single value thereby balancing the two metrics.

Robo-advisors provide advice to investors through the use of algorithms. Robo-advisors are an innovative FinTech product. Various SLRs and BA are available on this theme but there is paucity of research on comparison of these DBs in context of Robo-advisory literature. The present study is an effort to perform comparative analysis of performance of Scopus and Web of Science DBs in relevance to the literature retrieval on theme: Robo-advisory in finance sector. Information retrieval metrics such as Precision, Recall and F1 score are calculated to compare search performance of Scopus and Web of Science.

Organisation of paper is as follows. Section 2 and 3 presents review of literature and objectives of the study. Section 4 highlights the methodology followed. Section 4 presents results of analysis. Conclusion, limitations and future research avenues are discussed in Section 5.

Literature Review

Literature on Comparison of DBs

A comparative analysis of three DBs- Scopus, Web of Science and Dimensions regarding their journal coverage was conducted by Singh et al. (2021). Results highlighted that there is a wide variation in the journal coverage of these three DBs. Dimension database has an exhaustive coverage as compared to other DBs. Though Scopus was launched much after the launch of Web of Science, but it has made significant strides and has 'proven itself to be more effective than Web of Science. Scopus facilitates ease of navigation to users (Pranckutė, 2021).

Zhu & Liu (2020) highlighted that majority of researchers use Scopus and Web of Science. Visser et al., (2021) posits that large number of documents that have significant number of citations and references are covered by Scopus but they are not available in Web of Science. **All journal articles covered by Web of Science are covered by Scopus.**

A study was carried to select most appropriate database for literature search on theme telemedicine by comparing the performance of DBs- Cinhal, PubMed, Web of Science and Scopus by Ahmadi et al. (2015). Researchers concluded that the best database to conduct research on telemedicine is PubMed followed by Scopus and Web of Science. Evaluation of google scholar's performance on

the theme later life migration was conducted by Walters et al. (2009). Results highlighted that Google Scholar has better precision and recall as compared to other DBs such as 'Academic Search Elite, AgeLine, ArticleFirst, EconLit, GEOBASE, MEDLINE, PAIS International, POPLINE, Social Sciences Abstracts, Social Sciences Citation Index (SSCI), and SocINDEX'

Literature on Systematic Literature Reviews and Bibliometric Analysis on Robo-advisors

Comprehensive SLRs and BA have been conducted to study the research contribution on the theme: Robo-advisory. Table 1 presents summary of SLRs and BA on theme Robo-advisory.

Objectives of the Study

1. To assess performance of Scopus and Web of Science database in retrieving literature for Robo-Advisory in finance sector.
2. To compare the efficacy of Scopus and Web of Science database in retrieving literature for Robo-Advisory in finance sector.
3. To suggest database for conducting literature review and bibliometric analysis on Robo-Advisory in the finance sector.

Table 1: Summary of SLRs and BA on Robo-advisory

Author	Objective	Database Used	Query executed on
Wagner (2024)	To find determinants of conventional and digital advisory.	Scopus IEEE SSRN	March 2022
Fahruri et al. (2024)	To understand research contribution on adoption of Robo-advisors	Scopus	Jan 25, 2023
Rico-Pérez et al. (2022)	To perform bibliometric analysis on Robo-advisors	Scopus Web of Science	May 21, 2022
Darskuviene & Lisauskiene (2021)	To study impact of Robo-advisors on behavioural biases of investors.	Google Scholar, Web of Science, Science Direct, Springer Link and Taylor&- Francis databases	2021
Manaf et al. (2023)	To analyse research on adoption of Robo-advisors.	Scopus Web of Science Science Direct	Feb, 2023

Source: Author's own compilation from extant literature

Methodology

The current study aims to compare the performance of two DBs- Scopus and Web of Science in relevance to availability of literature on theme of Robo-advisory in finance sector. Following steps have been performed:

Step 1: Formation of Corpus: 5 SLRs and BA on Robo-advisors were found from the literature. From the references of these 5 SLRs and BA, 137 documents relevant to the theme were identified. These 137 documents formed the corpus of the study.

Step 2: Identifying keywords: Most commonly used keywords were identified from title of documents available in the corpus. Table 2 presents the keywords and their frequency. It is evident from Table 2 that keyword “Robo-advisors” was found in 60 titles (44% of the total documents) followed by the keyword ‘Robo-advisors” that appears 27 times. It is worth nothing that all these keywords are slight variation of the keyword Robo-advisors. **Using these keywords, a search query “Robo-adv*” is formulated** (Wild card character asterisk (*) represents one or more characters in a search query. Hence inclusion of * in the search query, retrieves relevant documents closely associated with the theme Robo-advisory).

Table 2: Number and Percentage of appearance of keywords among 137 documents

Keywords	Number of times appeared	Percentage of appearance of keywords among 137 documents
Robo-advisors	60	43.79%
Robo-advisor	27	19.70%
Robo-advisory	15	10.94%
Robo-advice	09	6.56%
Robo-advising	7	5.10%

Source: Author’s own compilation

Step 3a: Execution of query on Scopus

The query “robo-adv*” was executed on Scopus on 09/03/2024 to search from Title, Abstract and Keywords. A total of 385 documents were retrieved. Figure 1 represents the snapshot of 385 results retrieved from the query executed on Scopus.

Step 3b: Executing the query on Web of Science

The query “robo-adv*” was executed on Web of Science on 09/03/2024. A total of 121 documents were retrieved. Figure 2 represents the snapshot of 121 results retrieved from the query executed on Web of Science.

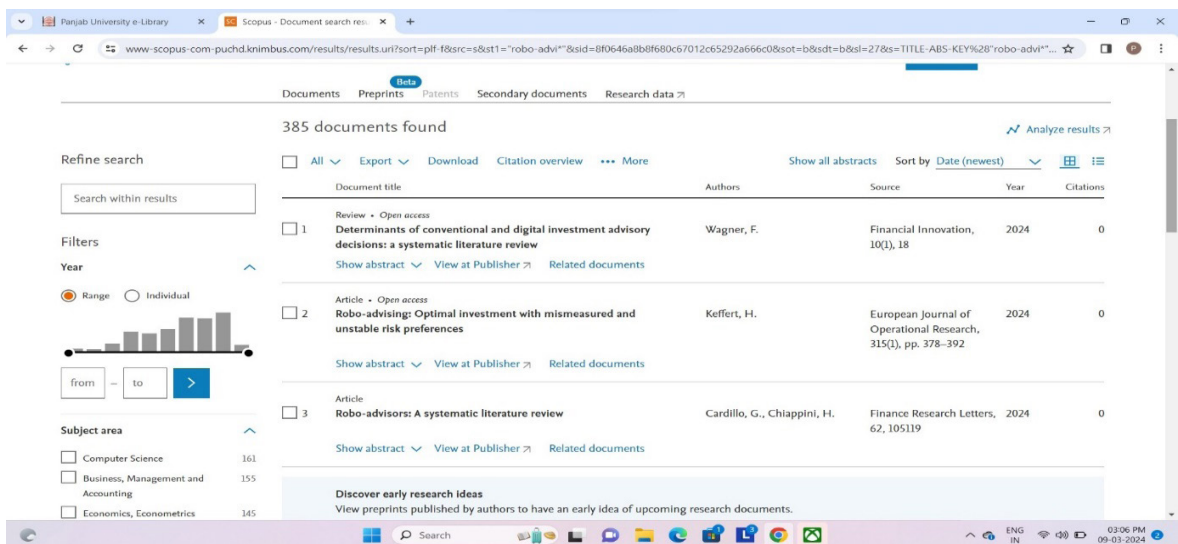


Fig. 1. Snapshot of 385 results retrieved from the query executed on Scopus.

Source: Scopus

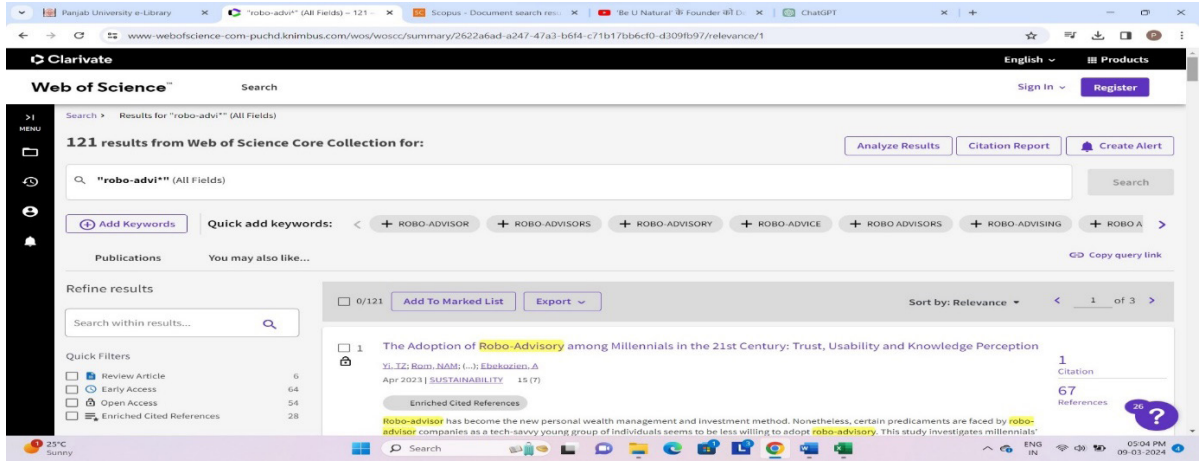


Fig. 2. Snapshot of 121 results retrieved from the query executed on Web of Science

Source: Web of Science

Table 3: Definition of Performance metrics for Information Retrieval

Performance metrics for Information Retrieval	Definition	Source
Precision	Precision is defined as proportion of relevant items retrieved to all items retrieved. It is a measure of specificity implying that whether a database is capable of excluding non-relevant items.	(Buckland & Gey, 1994; Gusenbauer & Haddaway, 2020)
Recall	Recall is defined as proportion of number of relevant items retrieved to all relevant items. It is a measure of sensitivity.	(Gusenbauer & Haddaway, 2020; Walters et al., 2009)
F1 Score	F1 Score is defined as harmonic mean of precision and recall.	(McSherry & Najork, 2009)

Source: Author’s own compilation from Literature

Step 4: Calculation of Precision, Recall and F1 Score

To compare the performance of Scopus and Web of Science for information retrieval on the theme “Robo-advisory”, the query was executed. From the results retrieved, the performance metrics-Precision, Recall and F1 Score were calculated. Table 3 presents definition of these performance metrics.

documents retrieved by Scopus is 3.18 times more than that of Web of Science. From a corpus of 137 documents, Scopus retrieves 103 documents and 47 were retrieved by Web of Science. Thus, Scopus retrieved larger proportion of documents as compared to Web of Science.

Performance of database regarding retrieval of relevant information is analysed using information retrieval performance metrics- Precision and Recall and F1 score.

Findings Documents Retrieved

Scopus retrieved total of 385 documents and Web of Science retrieved 121 documents. The number of

Calculation of Precision

Precision is defined as number of relevant items retrieved to total items retrieved. It is calculated using following formula:

$$\text{Precision} = \frac{\text{No. of relevant documents retrieved}}{\text{All documents retrieved}}$$

Precision of Scopus

No. of relevant items retrieved= 253

All documents retrieved= 385

Precision= $253/385 = 65.71\%$

Precision of Web of Science

No. of relevant items retrieved= 75

All documents retrieved= 121

Precision= $75/121 = 61.98\%$

INTERPRETATION

Results indicate that Scopus has a higher precision as compared to Web of Science. Thus, Scopus returned more relevant results for the search query on Robo-advisory.

5.3 Calculation of Recall

Recall is defined as number of relevant items retrieved to all relevant items in the corpus. It is calculated by following formula:

$$\text{Recall} = \frac{\text{No. of relevant documents retrieved}}{\text{All relevant documents}}$$

Recall of Scopus

Total records retrieved out of corpus of 137 documents is 103.

No. of relevant documents retrieved= 103

All relevant documents=137

Recall = $103/137 = 75.18\%$

Recall of Web of Science

Total records retrieved out of corpus of 137 documents is 47.

No. of relevant documents retrieved= 47

All relevant documents = 137

Recall= $47/137 = 34.30\%$

INTERPRETATION

Results indicate that Scopus has a relatively high recall rate for search query implying that Scopus is more effective at retrieving relevant documents from the corpus as compared to Web of Science. Scopus is able to capture a higher proportion of the relevant documents, making it potentially more comprehensive for the search query.

Calculation of F1 Score

F1 measure is defined as harmonic mean of precision and recall. This score combines precision and recall into a single value thereby balancing the two metrics. It gives an overall measure of effectiveness of search query. It is calculated by following formula:

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1 Score of Scopus

Precision of Scopus= 65.71 %

Recall of Scopus= 75.18%

$$\text{F1 Score (Scopus)} = 2 * \frac{0.6571 * 0.752}{0.6571 + 0.752} = 0.7011$$

F1 Score of Web of Science

Precision of Web of Science= 61.98%

Recall of Web of Science= 34.30%

$$\text{F1 Score (WOS)} = 2 * \frac{0.6198 * 0.3431}{0.6198 + 0.3431} = 0.4412$$

INTERPRETATION

Higher F1 score for Scopus indicates that it has a better balance between precision and recall as compared to Web of Science. This means that Scopus is more effective at retrieving relevant documents while also minimizing the number of irrelevant documents retrieved, compared to Web of Science.

Conclusions, Limitations and Future Research Directions

Information retrieval plays a crucial role in conducting SLRs and BA. The quality of the documents retrieved from DBs affects the effectiveness of SLRs and BA. It provides researchers with a comprehensive understanding of the existing literature on a particular field. The current study aims to perform a comparative analysis of performance of Scopus and Web of Science DBs in retrieving relevant documents on theme Robo-advisory financial services. Precision, Recall and F1 Score of Scopus is higher than Web of Science highlighting that Scopus outperforms Web of Science in terms of effectiveness in retrieval of relevant documents. These results suggest that researchers and practitioners in the field of Robo-advisory may benefit more from using Scopus for literature search in conducting SLRs and BA.

However, the study suffers from some limitations. Since, the performance of database varies from one field to another (Walters et al., 2009), the limitation of this study is that these results are applicable in the context of extracting literature on the theme Robo-advisory and cannot be generalized to other themes. Another limitation is that search performance of only two DBs- Scopus and Web of Science has been analysed and compared. Future studies can focus on conducting a comparative analysis of other databases such as Google Scholar, Jstor, EBSCOhost etc. in retrieving literature on Robo-advisory financial services. The effectiveness of each database varies depending on the specific search query used. Future research can explore the impact of different search queries on database performance. The study is limited by timeframe of the literature included. Search query was executed on 09th March, 2024. Robo-advisory is a rapidly evolving field, and newer publications may have different indexing status that could impact database performance.

References

Ahmadi, M., Sarabi, R. E., Orak, R. J., & Bahaadinbeigy, K. (2015). Information retrieval in telemedicine: A comparative study on bibliographic databases. *Acta*

- Informatica Medica*, 23(3), 172–176. <https://doi.org/10.5455/aim.2015.23.172-176>
- Buckland, M., & Gey, F. (1994). The relationship between Recall and Precision. *Journal of the Association for Information Science and Technology*, 45(1), 12–19. [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1<12::AID-ASI2>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L)
- Darskuviene, V., & Lisauskiene, N. (2021). Linking the Robo-advisors Phenomenon and Behavioural Biases in Investment Management: An Interdisciplinary Literature Review and Research Agenda. *Organizations and Markets in Emerging Economies*, 12(2), 459–477. <https://doi.org/10.15388/omee.2021.12.65>
- Fahruri, A., Rusmanto, T., Warganegara, D. L., & Tjhin, V. U. (2024). Mapping the Research Landscape of Robo Advisor Adoption: A Bibliometric Analysis. *Journal of System and Management Sciences*, 14(1), 27–49. <https://doi.org/10.33168/JSMS.2024.0103>
- Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, 11(2), 181–217. <https://doi.org/10.1002/jrsm.1378>
- Manaf, S. M., Ismail, M. K. A., & Zakaria, S. (2023). Systematic Literature Review on Robo-Advisory Adoption towards Young People. *Environment-Behaviour Proceedings Journal*, 8(S115), 3–9. <https://doi.org/10.21834/e-bpj.v8isi15.5086>
- McSherry, F., & Najork, M. (2009). Computing Information Retrieval Performance Measures Efficiently in the Presence of Tied Scores. *Advances in Information Retrieval*, 11(3), 365–373. <https://doi.org/10.1163/156854109X446962>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Pranckutė, R. (2021). Web of Science (WoS) and Scopus: the titans of bibliographic information in today's academic world. *Publications*, 9(12). <https://doi.org/10.3390/publications9010012>
- Rico-Pérez, H., Arenas-Parra, M., & Quiroga-Garcia, R. (2022). Scientific Development of Robo-Advisor: A Bibliometric Analysis. *Review of Economics and Finance*, 20, 776–786. <https://doi.org/10.55365/1923.x2022.20.87>

- Ritchie, S. M., Banyas, K. M., & Sevin, C. (2019). A Comparison of selected bibliographic database search retrieval for agricultural information. *Issues in Science and Technology Librarianship*, 2019(93). <https://doi.org/10.29173/istl48>
- Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, 126(6), 5113–5142. <https://doi.org/10.1007/s11192-021-03948-5>
- Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, crossref, and microsoft academic. *Quantitative Science Studies*, 2(1), 20–41. https://doi.org/10.1162/qss_a_00112
- Wagner, F. (2024). Determinants of conventional and digital investment advisory decisions: a systematic literature review. *Financial Innovation*, 10(1). <https://doi.org/10.1186/s40854-023-00538-7>
- Walters, W.H., Walters, W.H., & Walters, W.H. (2009). Google Scholar Search Performance: Comparative Recall and Precision. *Libraries and the Academy*, 9(1), 5–24. <https://doi.org/https://doi.org/10.1353/pla.0.0034>
- Zhu, J., & Liu, W. (2020). A tale of two databases: the use of Web of Science and Scopus in academic papers. *Scientometrics*, 123(1), 321–335. <https://doi.org/10.1007/s11192-020-03387-8>