Research on Retrieval Strategies for United Nations Sustainable Development Goals Publications: Postprint

Authors: Yu Xi

Date: 2025-03-20T00:00:00+00:00

Abstract

[Purpose/Significance] United Nations Sustainable Development Goals (SDGs) publications constitute the bibliometric foundation for measuring the contribution levels of countries, regions, and institutions to SDG-related research; studying and optimizing publication retrieval strategies can provide methodological references for undertaking related work. [Method/Process] Based on the Elsevier and Clarivate retrieval system platforms, this study takes SDG-related papers as the retrieval object, comparatively analyzes the differences in search results between the two databases and the reasons underlying these differences, and proposes improvement recommendations by addressing the strengths and weaknesses of the two query methods. [Results/Conclusion] Compared with traditional disciplinary and literature classification systems, the United Nations Sustainable Development Goals publication classification is a system without participation from authors and editors, making it currently difficult to establish a precise mapping relationship between publications and retrieved literature. Therefore, it is necessary to fully recognize the classification differences among different retrieval system platforms, optimize SDG publication retrieval strategies, fully exploit and utilize library digital resources, and provide literature resource support for the Sustainable Development Goals in teaching and research.

Full Text

Preamble

Journal of Literature and Data Science
Volume 7, Issue 1, March 2025

Study on Retrieval Strategies for United Nations Sustainable Development Goals Publications

Yu Xi
(The Library of Tianjin Normal University, Tianjin 300387, China)

Abstract:
[Purpose/Significance] United Nations Sustainable Development Goals (SDGs) publications serve as the foundational literature for measuring the research contributions of countries, regions, and institutions toward sustainable development goals. Optimizing retrieval strategies for these publications can provide methodological references for related work. [Method/Process] Based on the Elsevier and Clarivate retrieval system platforms, this study examines SDG-related papers as retrieval objects, compares and analyzes differences in retrieval results between the two databases and the reasons for these differences, and proposes improvement recommendations based on the strengths and weaknesses of both query methods. [Result/Conclusion] Compared with traditional disciplinary and literature classification systems, the UN SDGs publications classification is a system that does not involve authors or editors. Currently, it is difficult to establish a precise mapping relationship between publications and retrieved literature. Therefore, it is necessary to fully understand the classification differences across different retrieval platforms, optimize SDG publication retrieval strategies, fully exploit and utilize library digital resources, and provide literature resource support oriented toward sustainable development goals for teaching and research.

Keywords: Sustainable Development Goals (SDGs) publications; retrieval strategy; Elsevier; Clarivate
Classification Number: G252.6
DOI: 10.31193/SSAP.J.ISSN.2096-6695.2025.01.11

In September 2015, the United Nations adopted the *2030 Agenda for Sustainable Development*, which first proposed the 17 "Sustainable Development Goals" (SDGs). The UN SDGs represent a global action initiative aimed at eradicating poverty, protecting the environment, and ensuring prosperity and peace, providing a global framework and indicator system for addressing the most critical social, economic, and environmental issues. These goals decompose sustainable development into more granular levels and serve as benchmarks for measuring global progress toward sustainable development.

The proposal of the UN SDGs has drawn attention from governments and research institutions worldwide, all seeking to contribute to solving global challenges. The U.S. Agency for International Development (USAID) is committed to advancing progress toward the SDGs. Australia emphasized in its 2017 Foreign Policy White Paper that it has a responsibility to contribute to global poverty reduction, alleviating suffering, and promoting sustainable development. The Australian government has joined forces with universities, businesses, and civil society organizations to address the risks and opportunities presented by the 2030 Agenda. China has incorporated the UN SDGs into its national 13th Five-Year Plan, issued the *China' s National Plan for Implementing the 2030*

Agenda for Sustainable Development, and established the International Research Center of Big Data for Sustainable Development. At the university level, the University of Auckland has developed a series of undergraduate and graduate courses focused on the SDGs, while Zhejiang University has offered a Global Summer School program to address environmental governance issues under the SDGs, and Fudan University has classified its teaching courses, research achievements, and practical activities according to the 17 SDGs and formulated an SDG action report. These university courses cover a wide range of disciplinary topics, yet traditional literature organization and retrieval strategies based on conventional classification systems are essentially unable to meet related research needs. It should be noted that scientific publications in this paper refer to a narrow concept, limited to academic papers in journals.

Effectively mapping scientific publications to corresponding UN SDGs is the literature foundation for conducting SDG-related research. Determining how to map scientific journal publications to SDG categories to retrieve relevant publications is therefore crucial. To implement the UN SDGs, governments have established sustainable development research institutions, and universities have set up SDG-related teaching courses and conducted academic research on SDGs. As supporting institutions for teaching and research, university libraries should provide literature resource support services for SDG-related courses and academic research. For example, university libraries can analyze which SDGs their institution' s teaching content and research directions primarily target, investigate the coverage of literature resources related to these SDGs, and conduct SDG-based collection development and evaluation. However, the SDG classification differs from any previous classification system and does not rely on any existing classification method. Establishing a correspondence between literature resources and SDGs and accurately retrieving publications is a prerequisite for libraries to carry out such services.

This study examines current methods and status for retrieving SDG publications, searches for a university' s SDG-related publications on both Elsevier and Clarivate platforms, compares the retrieval results from both platforms, identifies the reasons for differences, and proposes feasible improvement solutions for the methodological deficiencies in both platforms' SDG publication retrieval methods, aiming to provide references for SDG publication query research.

1. Retrieval Methods for UN Sustainable Development Goals Publications

Since the UN proposed the SDGs, scholars, information service institutions, and academic research organizations have conducted research on identifying and retrieving related publications. Currently, three main retrieval methods are commonly used for UN SDG-related publications.

The first method is Boolean logic querying. The prerequisite for Boolean logic querying is generating relevant subject terms, i.e., ontology construction. These

subject terms are then used individually or in combination, and the Boolean logic query conditions are applied to publication titles, abstracts, keywords, and other text fields. Armitage et al. applied Boolean logic methods to limit retrieval objects to publications that directly contribute to the UN SDGs while minimizing the impact of question formulation on Boolean logic to produce more targeted results. Bordignon limited the retrieval scope to all scientific journal classifications to narrow down keyword polysemy and used the CorTexT text mining tool to enrich the selected publications. The University of Auckland used text mining technology and n-gram analysis methods to extract global and local keywords from publication metadata, ranking them based on the number of publications containing the keywords and their TF-IDF values. They also created a localized version based on information queries from local researchers, considering more papers targeting Australian and New Zealand research topics. The Sustainable Development Solutions Network (SDSN) for Australia, New Zealand, and the Pacific region provided a "Compilation List of SDG Keywords" for SDG mapping, jointly developed by several universities in the region. Monash University developed the initial keyword list, which was modified based on input from SDSN and peer universities, resulting in a total of 915 SDG-specific keywords covering all 17 SDGs. Austrian researchers used datasets from internal research literature systems of 13 Austrian universities to map Austrian research publications to SDGs. Confraria et al. obtained SDG-specific terms from policy reports, publications, and forums, selected and processed these terms, and used them to identify citation-based publication sets.

The second method involves machine learning, using supervised (classification) or unsupervised (clustering) approaches to map relevant publications to UN SDGs. Supervised methods typically use the same publication queries to obtain labeled datasets for training models. Unsupervised methods usually employ publication text representation or citation mapping, with resulting clusters directly or via intermediate clusters (topics) mapped to corresponding SDGs. The UN used a topic modeling algorithm in unsupervised mode to build an SDG classifier for the Department of Economic and Social Affairs' publications. Hajikhani et al. utilized the interrelationship between science and technology to compile machine learning models through UN SDG publication classification, classifying SDG-related patent literature to reveal the performance of SDGs in patents. Zhang et al. applied four mapping studies to deep learning methods for SDG mapping. Digital Science extracted key terms based on the UN' s definitions of SDGs, defined keyword search strings for each goal, and its Dimensions platform created a training set for each SDG. The Dimensions platform applied natural language processing and machine learning methods to generate classification schemes for the 17 SDG training sets. Digital Science collaborated with Google Cloud to integrate Dimensions with Google Cloud' s BigQuery, enabling large-scale analysis of the Dimensions database and using the Dimensions SDG classifier for case analysis of the UK' s Research Excellence Framework (REF).

The third method combines Boolean logic querying and machine learning to identify and obtain relevant publications. The European University Alliance

used keyword combinations, Boolean and proximity operators to develop 169 target-level SDG query methods and built machine learning models mapping multilingual academic papers to UN publications. Elsevier, in collaboration with the University of Southern Denmark, the European University Alliance, and the University of Auckland, developed keyword query methods for mapping scientific publications to SDGs, supplemented by machine learning to greatly improve recall and accuracy. This method has been adopted by the *Times Higher Education* (THE) University Impact Rankings.

UN SDG-specific publication retrieval methods mainly include three categories: "keyword querying," "machine learning models," and "keyword querying + machine learning models" (Table 1).

2. Case Analysis of Elsevier and Clarivate Retrieval Platforms

Based on the current research status of SDG publication retrieval methods, this study conducts case analysis and comparative research on retrieval methods for obtaining SDG-related publications using two different retrieval platforms: Elsevier and Clarivate.

2.1 Selection of Retrieval Platforms and Cases

This study selects Elsevier and Clarivate as the analysis platforms for SDG publication retrieval. Elsevier is the world's largest publishing group, and its Scopus and SciVal platforms are dedicated to providing literature retrieval, scientific measurement analysis, and evaluation services for scholars, research institutions, and management departments. Clarivate' s InCites platform provides scientific measurement indicators and citation analysis services.

Both Elsevier and Clarivate platforms have relatively comprehensive literature coverage, high quality, and powerful statistical analysis functions, making them suitable as analysis tools for scientific research resource assurance and disciplinary data and performance evaluation. Both Elsevier and Clarivate have developed mapping methods between UN SDGs and scientific publications, including mapping methods for SDGs from Category 1 to Category 16, but neither has conducted mapping research for Category 17. Therefore, conducting comparative research on the identification of SDG scientific publications by Elsevier and Clarivate is practically significant.

This study selects a university' s related publications as the case analysis object. The university is a comprehensive "Double First-Class" undergraduate institution covering multiple disciplines including science, engineering, economics, management, literature, law, medicine, education, and arts, with broad disciplinary coverage and numerous research achievements, meeting the sample analysis requirements. The study retrieves SDG-related papers published by the university between 2018 and 2022 that are indexed in both Elsevier and

Clarivate databases. The retrieval date was December 4, 2023, the document types were research papers and review papers, and the retrieval platforms were Elsevier' s SciVal platform and Clarivate' s InCites platform.

2.2 Comparison of Retrieval Results on SciVal and InCites Platforms

The retrieval results show that between 2018 and 2022, the university had a total of 2,696 SDG papers indexed in both the SciVal and InCites platforms. The UN SDGs consist of 17 major categories (SDG1 through SDG17). However, the classification of these papers into SDG categories differs between the two platforms. The distribution across different categories shows cross-relationships between the two databases, with almost every major category showing incomplete overlap between the two platforms (Figure 1 [Figure 1: see original paper]).

Figure 1. Cross-distribution of SDG papers on SciVal and InCites platforms

When counting each paper, some papers involving multiple SDGs may be counted repeatedly, so the total number of related papers exceeds the actual number of papers. Among the 2,696 papers, SDG7 has the highest number of papers in both databases, indicating that the university' s SDG publications are primarily concentrated in the SDG7 area, "Affordable and Clean Energy." Therefore, the following analysis focuses on the university' s SDG7 papers to further examine differences in retrieval results between the two platforms. Additionally, the number of shared subject terms and cross-distribution of SDG7 papers on both platforms were analyzed (Figure 2 [Figure 2: see original paper]).

Figure 2. Distribution of SDG7 papers on SciVal and InCites platforms

2.2.1 Analysis of Shared SDG7 Paper Subject Terms on Both Platforms (1) Shared SDG7 Paper Subject Terms on Both Platforms

In reality, papers classified as SDG7 in both SciVal and InCites are not completely identical. According to retrieval statistics, 848 papers were classified as SDG7 in both databases (Figure 2), accounting for 51.02% of total SDG7 papers in SciVal and 85.83% of total SDG7 papers in InCites. These shared SDG7 papers likely relate to research topics contained in both databases.

The main subject terms for SDG7-related papers were identified from SciVal, and micro-topics for SDG7-related papers were identified from InCites. The results show that SciVal' s SDG7-related papers have 1,000 main subject terms, while InCites' SDG7-related papers correspond to only 77 micro-topics. By comparing SciVal' s SDG7 main subject terms with InCites' SDG7 micro-topics, 12 substantially identical subject terms were identified (Table 2). By examining the titles, abstracts, and keywords of the 848 shared SDG7 papers, multiple papers were found to be related to these shared subject terms.

Table 2 . Shared SDG7 Subject Terms Between SciVal and InCites

SciVal Paper Subject Terms	InCites Paper Subject Terms
doubly fed induction generator	doubly fed induction generator
ground source heat pump	ground source heat pump
solar air heater	solar air heater
bio-energy	bioenergy
dye sensitized solar cells	dye-sensitized solar cells
hydrogen storage	hydrogen storage
inertial confinement fusion (icf)	inertial confinement fusion
organic solar cells (oscs)	organic solar cells
perovskite solar cells (pscs)	perovskite solar cells
renewable energy (re)	renewable energy
solid oxide fuel cell (sofc)	solid oxide fuel cell
the maximum power point tracking	the maximum power point tracking (mppt)

(2) Differences in SDG7 Paper Subject Terms Between Platforms

SciVal' s SDG7 main subject terms are numerous but contain many terms with different writing forms but identical meanings, such as "dye sensitized solar cell" (Table 3).

Table 3 . Different Writing Forms of the Same Term on the SciVal Platform

Original Term	Different Writing Forms
dye sensitized solar cell (染料敏化太阳能电池)	dye sensitized solar celldye sensitized solar cellsdye sensitized solar cells (dsscs)dye-sensitized solar cell (dssc)dye-sensitized solar cells (dscs)dye-sensitized solar cells (dssc)dye-sensitized solar cells (dsscs)dye-sensitized solar cells

This occurs because SciVal' s subject terms are extracted from literature using text mining technology, and different writing forms of the same term in the literature may all be extracted. To ensure comprehensive retrieval, all writing forms of subject terms are included in the search formula, but this method only works for existing literature and cannot encompass writing forms that appear in newly published literature. In contrast, InCites' SDG7 micro-topics are represented by a single term with broader content and more standardized writing forms.

After removing plurals, abbreviations, hyphens between words, and trailing periods from SciVal' s SDG7 main subject terms, only 609 subject terms remain. Among these 609 records, there are still numerous application forms for the same term, which is another main reason for the large number of subject terms in SciVal SDG7. Even after excluding differences caused by different writing forms and application forms, many differences remain in the subject term forms between SciVal SDG7 and InCites SDG7. Using "renewable energy" as an example, Table 4 lists its various application forms on the SciVal platform.

2.2.2 Analysis of SDG7 Papers Unique to SciVal As mentioned above, 814 SciVal SDG7 papers were not classified as SDG7 in InCites but were assigned to other SDG categories, with 518 assigned to SDG11 (Figure 2). It is speculated that SciVal SDG7 and InCites SDG11 share common subject terms. By examining SciVal SDG7 paper subject terms and InCites SDG11 micro-topics, the terms "lithium-sulfur batteries" and "lithium ion batteries (libs)" in SciVal SDG7 were found to be highly similar to "2.62.616 Lithium-Sulfur Batteries" and "2.62.138 Lithium-Ion Battery" in InCites SDG11. SciVal SDG7 subject terms such as "electric bus," "electric buses," and "hybrid electric vehicle (hev)" are similar in content to "4.18.788 Electric Vehicles" in InCites SDG11. Additionally, some papers include terms related to "lithium-sulfur batteries" and "lithium ion batteries" such as "electrode," "anode," "cathode," "battery," or "batteries." Examining these papers revealed that most are related to "lithium-sulfur batteries" or "lithium ion batteries," with a small portion about other battery types such as zinc-ion and sodium-ion batteries. These papers' micro-topic is "2.62.52 Supercapacitor," which is included in InCites SDG11. Searching for the aforementioned subject terms in paper titles, abstracts, and keywords found 456 papers containing these terms, with only 62 papers not containing them.

For the 62 papers not identified, screening their titles, keywords, and abstracts revealed that most contain SciVal SDG7 subject terms such as "solar power," "wind energy," and "geothermal system," as well as InCites SDG11 micro-topics like "aerosols," "distributed storage systems," "eddy current testing," "internet of things," and "supercapacitor."

2.2.3 Analysis of SDG7 Papers Unique to InCites Among InCites SDG7 papers, 140 were not classified as SDG7 in SciVal but were assigned to other SDG categories, with 112 assigned to SDG13 (Figure 2). It is speculated that InCites SDG7 papers and SciVal SDG13 papers share common subject terms. By comparing InCites SDG7 micro-topics and SciVal SDG13 main subject terms, "2.62.2102 CO2 reduction" in InCites SDG7 was found to be identical to "CO2 reduction" in SciVal SDG13. Among the 112 papers belonging to SciVal SDG13, most contain "CO2 reduction" or "carbon dioxide reduction" in their titles, keywords, and abstracts. Other papers generally include terms such as "greenhouse gas emission," "global warming," or "CO2 electroreduction," all of which can be found in SciVal SDG13 paper subject terms, indicating that these papers may be classified as SDG13 in SciVal.

3. Discussion of Results and Cause Analysis

Excluding differences in literature coverage between Elsevier and Clarivate databases, this study retrieved UN SDG papers on their respective SciVal and InCites platforms and found significant differences in retrieval results. The specific reasons for these differences are roughly as follows.

(1) Different Classification Judgment Rules

Elsevier' s acquisition of SDG publications is primarily achieved by constructing relevant paper subject terms and corresponding search formulas. As early as 2015, when the UN adopted the *2030 Agenda for Sustainable Development*, Elsevier constructed the initial version of the SDG subject term list. Since then, it has continuously improved query strategies, updated the SDG subject term list, and refined SDG query methods. Retrieved papers are mapped to corresponding SDG categories and assigned values from 1 to 16 based on their SDG category. Elsevier has built training sets for 16 SDGs through machine learning models, generating relevant subject terms and their TF-IDF values for each SDG. Higher TF-IDF values indicate a closer relationship between the term and the SDG. To ensure readability, subject terms are not stemmed or lemmatized, so some subject terms may be duplicated, such as "recycle," "recycling," "industry," "industrial," etc. Machine learning models significantly increase the number of captured articles while maintaining precision above 80% by supplementing and improving SDG subject terms. Elsevier has loaded the SDG classification into its data analysis platform SciVal and updated publications for 16 SDGs. Researchers and institutions can track SDG achievements and progress through the SciVal platform.

Clarivate' s acquisition of SDG publications is primarily achieved through its Citation Topics classification system. Clarivate' s InCites platform includes multiple disciplinary classification methods, with Citation Topics being one of them. Citation Topics is a document-level literature topic classification method that breaks through journal limitations. It uses the Leiden algorithm to cluster literature based on direct citation relationships between papers, assigning a citation topic to each cluster. All literature in a cluster belongs to that topic, and each document is assigned to only one topic. Citation Topics includes three hierarchical structures: macro-topics, meso-topics, and micro-topics, currently comprising 10 macro-topics, 326 meso-topics, and 2,449 micro-topics. Each topic consists of a numerical identifier and topic name. This classification method was first launched on the InCites platform in December 2020 and updated twice in April 2023 and April 2024. Clarivate compares micro-topics with UN SDGs to construct mapping relationships between them. All literature under micro-topics mapped to each SDG is classified into the corresponding SDG category. In the InCites platform, each SDG paper involves multiple citation topics, and similarly, some citation topics appear in multiple SDG papers. Among InCites Citation Topics, 196 micro-topics are not exclusive to one SDG paper but may be distributed across 2 to 3 SDG papers, so some literature topics simultaneously

belong to multiple SDG papers.

Thus, Elsevier and Clarivate have different classification judgment rules for SDG paper topics. Elsevier' s retrieval system is based on document content, searching paper titles, keywords, and abstracts according to established SDG subject terms, so only papers containing SDG subject terms are considered SDG papers. Clarivate' s retrieval system is based on citation relationships between documents, determining whether a paper belongs to an SDG category by constructing mappings between micro-topics and SDGs and the paper' s assigned micro-topic.

(2) Classification Differences in Retrieval Results

Among the SDG papers retrieved on the SciVal platform, 375 papers could not be retrieved on the InCites platform, although these papers are indexed in the Web of Science Core Collection. The study found that among 2,437 micro-topics, only 1,898 micro-topics were mapped to SDGs by Clarivate. The micro-topics of these 375 papers were not within this mapping scope, so these papers are not considered SDG papers in InCites.

Among the SDG papers retrieved on the InCites platform, 3,190 papers are not considered SDG papers on the SciVal platform. The presumed reason is that the two platforms have different criteria for determining SDGs. These papers may not include the SDG subject terms provided by Elsevier, but their references or citing papers may contain numerous SDG papers. The number of these papers is relatively large, and searching and analyzing their references and citing papers is also challenging.

(3) Other Issues

Elsevier' s subject terms and Clarivate' s micro-topics overlap across different SDG literature, such as "CO2 reduction" existing in both InCites SDG7 and SciVal SDG13, and "lithium-sulfur batteries" and "lithium ion batteries" existing in both InCites SDG11 and SciVal SDG7. Thus, the same paper may be classified into different SDG categories in Elsevier and Clarivate.

Additionally, the same paper may contain subject terms from multiple SDGs, which may lead to classification attribution bias. Moreover, a paper' s research content and research objectives may involve different SDGs, thus belonging to different SDG categories.

4. Problems and Recommendations

4.1 Existing Problems

So far, no consensus has been reached on improving retrieval methods for SDG scientific publications, and there are still some ambiguous understandings at the theoretical level. First, the themes and concepts of SDGs are not clearly explained. The SDGs constitute a complex system with multiple definitions and interacting variables, and "sustainability" is difficult to define as it has different

interpretations across disciplines. Second, publications that contribute to SDG research are not easily defined, including: how publications should articulate their research themes to be considered SDG-related; whether SDG publications must make specific or direct contributions to SDGs; and whether publications making indirect contributions should also be counted. Third, there are structural differences in cultural traditions, language habits, economic development, industry characteristics, and research priorities across different countries and regions, but existing one-size-fits-all search queries lack sensitivity to these differences and fail to adequately address these nuances within the broad SDG framework.

At the technical level, both Boolean logic retrieval and machine learning algorithms have limitations. The selection, combination, and query structure of search terms in keyword-based Boolean logic methods all affect retrieval results. While keywords from official SDG documents are high-quality, they are often not comprehensive and cannot guarantee coverage of all SDG-related discourse. Challenges include whether selected keywords are relevant to SDG themes, whether they can explain publications' contributions to SDGs, and how to convert SDG thematic concepts into search query results to accurately identify relevant publications. Generally, supervised machine learning models are typically trained on relatively small and homogeneous corpora, making it difficult to handle out-of-sample cases. Moreover, models trained on different datasets with different parameters are almost impossible to integrate. Although unsupervised machine learning shows good promise, it is currently only applied to relatively small text corpora. In addition, machine learning models face other challenges.

4.2 Improvement Recommendations

As demonstrated above, both keyword-plus-machine learning retrieval strategies (such as the SciVal platform) and citation-plus-content mapping retrieval strategies (such as the InCites platform) have certain deficiencies when retrieving UN SDG publications. To address the limitations of these two retrieval strategies, this paper proposes the following improvement recommendations from a technical methodological perspective.

(1) Recommendations for Keyword-Plus-Machine Learning Retrieval Strategies

First, provide SDG-related terms and limit research topic areas to reduce subject term polysemy. By limiting subject terms to different disciplines or topics, the retrieval scope can be narrowed to exclude less relevant literature. Bordignon suggested limiting papers to Scopus subject areas when querying Elsevier.

Second, use regular expressions to unify the various writing forms of keywords appearing in text. For example, among the different variants of "dye sensitized solar cell" in Table 3, "dye-sensitized solar cells (dscs)" contains an incorrect writing form—the content in parentheses should be "dsscs." Therefore, regular

expressions can be constructed to cover all writing forms (including plurals, abbreviations, hyphens, and trailing periods). The regular expression is as follows:

```
"dye( |-)sensiti[a-z]{1,5}( |-)solar( |-)cell(s|)(\.|)(| )($|\(.*?\))"
```

Third, further verify the metadata of retrieved SDG literature to avoid mixing in non-target literature.

(2) Recommendations for Citation-Plus-Content Mapping Retrieval Strategies

In the InCites platform, some literature that has not been mapped or cited may contain SDG-related publications that are ignored during retrieval. These can be supplemented through keyword Boolean logic query methods.

InCites primarily identifies SDG literature based on citation and clustering, lacking semantic analysis of documents. Therefore, the SAO (Subject-Action-Object) semantic analysis method can be used to identify and analyze subjects, predicates, objects, and other elements in sentences and their semantic relationships, thereby obtaining main information and relationships in the text. The SAO semantic analysis method is commonly used for patent feature analysis and can also be applied to extract core content, features, and contextual information from paper titles and abstracts to determine and identify papers' SDG category attribution.

For papers whose research content and objectives involve different SDGs, citation context analysis can be used for judgment. In the Web of Science database, the location of target literature in all citing documents can be identified, and the value of target literature can be analyzed based on citation context. This method can also be used for SDG paper identification. For example, SDG-related subject terms can first be marked in the text, then classified according to the location of subject terms in the text (research background, research objectives, research methods, research conclusions, etc.), and finally, the SDG category attribution of related papers can be determined based on the application scenarios of subject terms.

Constructing retrieval strategies for UN SDG publications is a relatively complex process. While establishing mapping relationships between scientific publications and SDGs is relatively easy, evaluating the effectiveness of retrieval strategies is difficult. Current retrieval strategies all have certain defects, and the academic community has not yet established unified retrieval strategy standards. Compared with traditional disciplinary and literature classification systems, the UN SDG publications classification is a system that does not involve authors or editors, making it difficult to establish precise mapping relationships between paper publications and retrieved literature. Therefore, fully understanding the classification differences across different platforms as described above is necessary to maximize the remediation of these defects. Libraries should actively conduct optimization research on SDG publication retrieval strategies, fully exploit and utilize library digital resources, and provide literature resource support

oriented toward sustainable development goals for teaching and research.

Conclusion

Currently, international database retrieval platforms Elsevier and Clarivate have adopted different methods to construct mapping relationships between scientific publications and sustainable development goals. This study, after excluding data source coverage differences, retrieved SDG papers on both platforms, compared and analyzed the reasons for significant differences between them, and proposed improvement recommendations.

In fact, in addition to the paid platforms such as SciVal and InCites mentioned above, some open-source data platforms (such as Dimensions) also provide retrieval services for UN SDG-related publications. If retrieval research is conducted on these platforms' SDG-related publications and the results are compared with those from the above two platforms, it may further reveal differences between different retrieval strategies and can also serve as a useful supplement to SDG publications, making retrieval results more comprehensive.

References

[1] Analysis of the UN Sustainable Development Goals indicator system [EB/OL]. [2023-12-26]. https://sdg.js.org/.

[2] Take a holistic view of the sustainable development goals [EB/OL]. [2025-02-20]. https://2017-2020.usaid.gov/GlobalGoals.

[3] Key messages - Australia [EB/OL]. [2025-02-20]. https://sustainabledevelopment.un.org/memberstates/aust

[4] China releases the *China' s National Plan for Implementing the 2030 Agenda for Sustainable Development* [EB/OL]. [2023-11-21]. https://www.gov.cn/xinwen/2016-10/13/content_{5118514}.htm.

[5] International Research Center of Big Data for Sustainable Development [EB/OL]. [2023-11-21]. http://www.cbas.ac.cn/zxjs/zxgk/.

[6] Browse courses by SDG [EB/OL]. [2024-11-15]. https://www.auckland.ac.nz/en/about-us/about-the-university/the-university/sustainability-and-environment/studying-sustainability/browse-courses-by-sdg.html.

[7] SDG Global Summer School preview | Climate governance and sustainable development [EB/OL]. [2023-11-17]. https://mp.weixin.qq.com/s?___{biz}=MzA3NzUwOTcwNw==&mid=264

[8] Fudan University 2020 SDGs Action Report [EB/OL]. [2023-11-16]. https://www.op.fudan.edu.cn/_{upload}/article/files/09/d5/b70b34524eee925ec90526cb4c30/c1119570-461c-4ec0-9eef-61006f1721d3.pdf.

[9] Armitage C S, Lorenz M, Mikki S. Mapping scholarly publications related to the Sustainable Development Goals: Do independent bibliometric approaches get the same results? [J]. Quantitative Science Studies, 2020, 1(3): 1092-1108.

[10] Bordignon F. Dataset of search queries to map scientific publications to the UN sustainable development goals [J]. Data Brief, 2021(34): 1-12.

[11] Wang W W, Kang W H, Mu J W. Mapping research to the Sustainable Development Goals(SDGs) [EB/OL]. [2023-12-26]. https://www.researchgate.net/publication/368674683_{{{Ma

[12] The University of Auckland SDG keywords mapping [EB/OL]. [2023-12-26]. https://www.sdgmapping.auckland.ac.nz/.

[13] Universities and the SDGs [EB/OL]. [2023-12-26]. https://apunsdsn.org/regional-initiatives/universities-sdgs/.

[14] Körfgen A, Förster K, Glatz I, et al. It' s a hit! Mapping Austrian research contributions to the Sustainable Development Goals [J]. Sustainability/Molecular Diversity Preservation International (MDPI), 2018, 10(9): 1-13.

[15] Confraria H, Noyons E, Ciarli T. Countries' research priorities in relation to the Sustainable Development Goals [C]. 18th International Conference of the International Society for Scientometrics and Informetrics. Leuven, Belgium. 2021.

[16] South African SDG Hub [EB/OL]. [2023-12-29]. https://sasdghub.up.ac.za/home/.

[17] Wastl J, Porter S, Draux H, et al. Contextualizing sustainable development research [EB/OL]. [2024-01-05]. https://digitalscience.figshare.com/articles/report/Contextualizing_{{{Sustai

[18] Navigating the structure of research on Sustainable Development Goals [EB/OL]. [2024-01-05]. https://clarivate.com/lp/navigating-the-structure-of-research-on-sustainable-development-goals-2/.

[19] LaFleur M. Art is long, life is short: an SDG classification system for DESA publications [EB/OL]. [2023-12-26]. https://www.un.org/esa/desa/papers/2019/wp159_{2019}.pdf.

[20] Hajikhani A, Suominen A. Mapping the Sustainable Development Goals (SDG) in science, technology and innovation: application of machine learning in SDG-oriented artefact detection [J]. Scientometrics, 2022, 127: 6661-6693.

[21] Zhang R, Vignes M, Steiner U, et al. Matching research publications to the United Nations' Sustainable Development Goals by multi-label-learning with hierarchical categories [C]. IEEE 7th International Conference on Data Science and Advanced Analytics. Sydney, NSW, Australia. 2020: 516-525.

[22] Sustainable Development Goals classification [EB/OL]. [2023-12-29]. https://www.digital-science.com/resource/sustainable-development-goals-classification/.

[23] Fane B, Draux H, Wastl J. Using Digital Science' s Dimensions Database to track research with the UN Sustainable Development Goals [C]. 26th International Conference on Science, Technology and Innovation Indicators. Granada, Spain. 2022: 1-8.

[24] Schmidt F, Vanderfeesten M. Evaluation on accuracy of mapping science to the United Nations' Sustainable Development Goals (SDGs) of the Aurora

SDG queries [EB/OL]. [2023-12-29]. https://zenodo.org/records/4964606.

[25] Vanderfeesten M, Otten R, Spielberg E. Search queries for "Mapping Research Output to the Sustainable Development Goals (SDGs)" v5.0.2 [EB/OL]. [2024-12-29]. https://doi.org/10.5281/zenodo.4883250.

[26] Maurice V, Robert J, Lennart K. AI for mapping multi-lingual academic papers to the United Nations' Sustainable Development Goals (SDGs) [EB/OL]. [2024-12-29]. https://zenodo.org/records/6487606.

[27] Maxime R, Yury K, Alexandre B, et al. Improving the Scopus and Aurora Queries to identify research that supports the United Nations Sustainable Development Goals (SDGs)2021 [EB/OL]. [2023-12-29]. https://elsevier.digitalcommonsdata.com/datasets/9sxdykm8s4/4.

[28] Kashnitsky Y, Roberge G, Mu J, et al. Evaluating approaches to identifying research supporting the united Nations Sustainable Development Goals [EB/OL]. [2025-02-20]. https://arxiv.org/abs/2209.07285.

[29] SDG research mapping initiative [EB/OL]. [2023-11-16]. https://www.elsevier.com/about/sustainability/sc research-mapping-initiative.

[30] Traag V A, Waltman L, Van Eck N J. From louvain to leiden: guaranteeing well-connected communities [J]. Scientific Reports, 2019(9): 1-12.

[31] Citation topics [EB/OL]. [2024-11-19]. https://incites.zendesk.com/hc/en-gb/articles/22514077746961-Citation-Topics.

[32] Tay A. Mapping UN Sustainable Development Goals (SDG) to publications [EB/OL]. [2023-12-22]. https://library.smu.edu.sg/topics-insights/mapping-un-sustainable-development-goals-sdg-publications.

[33] Bordignon F. Search queries to map scientific publications to the UN Sustainable Development Goals [EB/OL]. [2023-12-22]. https://data.mendeley.com/datasets/xrx7ddbbb4/1.

Editor in charge: Ren Quan' e

Note: Figure translations are in progress. See original paper for figures.