ilmedia

TECHNISCHE UNIVERSITÄT
ILMENAU

Markscheffel, Bernd; Mollenhauer, Sven:

A comprehensive comparison of arXiv and the Web of Science (WoS)

DOI: 10.22032/dbt.55570

URN: urn:nbn:de:gbv:ilm1-2023200076

Original published in:	2021 ICoASL : 7th International Conference of Asian Special Libraries ; November 24, 2021, National Library of Korea, Seoul, South Korea. – p. 55-69.
Conference:	International Conference of Asian Special Libraries (ICoASL) ; 7 (Seoul, South Korea) : 2021.11.24
ISBN:	978-89-965885-3-5
Original published:	2022-01
Original URL:	https://icoasl2021.mlive.kr/lib_paper/Proceedings%20of%20ICoASL%202021.pdf
[*Visited:*	2023-01-31]

© by the Authors, 2021

TU Ilmenau | Universitätsbibliothek | ilmedia, 2023
https://www.tu-ilmenau.de/ilmedia

A Comprehensive Comparison of arXiv and the Web of Science (WoS)

Bernd Markscheffel, Sven Mollenhauer
Chair of Information and Knowledge Management
Technische Universitat Ilmenau, Germany
bernd.markscheffel@tu-ilmenau.de, sven.mollenhauer@tu-ilmenau.de

Abstract

Scientific exchange is increasingly shifting to the Internet. Today, online literature and citation databases are important tools for scientific work, e.g. for exchanging information or investigating the current state of the art on a topic. Due to the large number of literature and citation databases that exist, and the limited amount of time available for a search, it is necessary to choose few or even only one database.

In this paper, we carry out a comprehensive comparison between Web of Science and arXiv. We compile a list of criteria for the comparison of these resources based on a literature analysis. Finally, 62 documents were found that dealt with comparisons between literature databases. Based on these comparisons, a concept matrix was created according to Webster & Watson (2002), in which the criteria for the comparison were summarized.

These criteria were then integrated into an adapted version of the criteria catalogue for the comparison of software packages from Jadhav & Sonar (2009) in order to provide a comprehensive picture, not only of content aspects, but also of functionality and usability issues. Based on these criteria, the Web of Science and arXiv databases were compared.

The main results can be summarized as follows: arXiv covers only a limited number of disciplines and has a strong focus on physics, mathematics and computer science. Web of Science covers significantly more subject areas and generally includes significantly more papers, which in contrast to arXiv all come from peer-reviewed journals. arXiv's biggest advantage is the topicality of the articles, since preprints are also accepted and thus the peer-review process can be bridged. Both databases are intuitive to use and have a similarly good simple search, but Web of Science's advanced search gives an experienced user much more possibilities to refine searches and to formulate distinctive queries. In general, Web of Science offers significantly more possibilities to conduct comprehensive literature searches due to the additionally stored citation data and corresponding analysis functions. arXiv, on the other hand, is

particularly well suited to learn about the latest state of the offered disciplines.

Keywords: Web of Science, arXiv, Comparison, Literature Databases

1. Introduction

Scientific exchange, and thus also scientific literature and citation databases, is increasingly shifting to the Internet. Today, online literature and citation databases are important tools for scientific work, e.g. for exchanging information or investigating the current state of the art on a topic (Falagas et al., 2008; Cronin & Blaise, 2001; Tsay, Tseng & Wu, 2019). Citation atabases are a special form of literature databases that store citation data in addition to bibliographic data (Jacso, 2004). Since there are a large number of online literature databases that differ, for example, in terms of scope, topicality or search functionalities, the question arises for many, especially junior researchers, which of these databases is the right one for which user and for which purpose. In order to answer these question(s), comparisons between different databases have already been carried out by several authors (Jacso, 2005; Falagas et al., 2008). So far, however, there is no comprehensive comparison between the literature databases Web of Science and arXiv.

In order to determine suitable criteria for a comparison between literature databases, a literature analysis, according to the methodology of Webster & Watson (2002), is carried out on already existing comparisons. The criteria identified in this process are then integrated into the criteria proposed by Jadhav & Sonar (2009) for comparing software packages adapted for this task. Based on these criteria, a comparison between Web of Science and arXiv is then performed. The paper is further structured in this sense: In the next section, we will first briefly introduce the objects of comparison (Web of Science and arXiv). In the following section we will go into detail about the methodology used for the comparison in order to identify the relevant criteria for the comparison. In the following section, this comparison is carried out and finally the findings are summarized and an outlook on further research potential is given.

2. Objects of the comparison

2.1 Web of Science

Web of Science (WoS) is a fee-based literature database for searching literature offered by Clarivate Analytics. It consists of several scientific literature databases from different subject areas, first and foremost the multidisciplinary database Web of Science Core Collection (Clarivate Analytics, 2020a) This consists of several indices, including the Science Citation Index Expanded(SCIE), probably the best-known index, which has existed since 1964, when it was still called the Science Citation Index, and is constantly being expanded (WoS Group, 2020a). The Web of Science Core Collection alone includes articles from over 21,100 peer-reviewed journals in 254 scientific disciplines (WoS Group, 2020b). An important feature of Web of Science is that not only bibliographic data are stored in the databases, but also data on citations (WoS Group, 2020c). Thus, it is possible to find related, directly or indirectly relevant, publications to a selected publication (Jacso, 2004). Thus, the Web of Science is one of the most important tools for a wide range of bibliometric studies at all levels of observation (Bauer & Bakkalbasi, 2005). The

deliberately limited number of sources included and evaluated by Web of Science, according to its own criteria, is a considered quality feature. The included sources are regularly reviewed and adjusted (Tunger, 2007). For example, if a journal in the SCIE no longer meets the required criteria, it will be removed from the SCIE and moved to the ESCI (Emerging Sources Citation Index). If it no longer meets the criteria for ESCI, the journal will be removed from the Web of Science Core Collection(see WoS Group (2020d) for the exact process and the evaluation criteria).

2.2 arXiv

arXiv (pronounced like the English word "archive," the X represents the Greek letter “chi”) is a free, scientific literature database, originally founded in 1991 by Paul Ginsparg, now maintained and operated by Cornell University (Steele, 2012). Unlike Web of Science, arXiv is limited to specific subject areas (physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering, and economics) and includes not only articles from peerreviewed journals, but also preprints, among others (Berg et al. 2016; arXiv, 2020a).

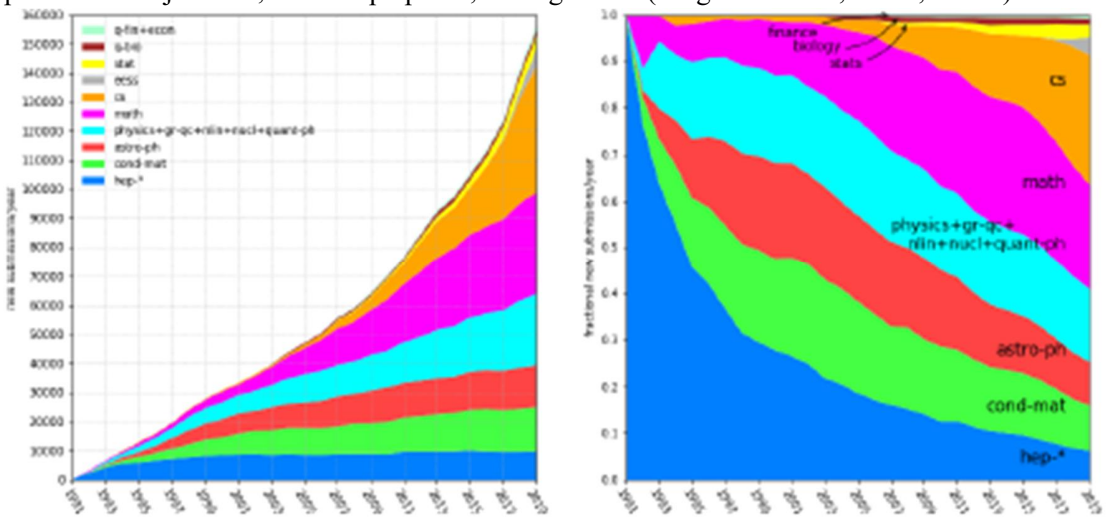


Figure 1: Annual submissions by field on arXiv ("hep" = High Energy Physics, "cond-mat" = Condensed Matter Physics, "astro-ph" = Astrophysics, "phys+gr-qc+nlin+nucl+quant-ph" = Other Physics, "math" = Mathematics, "cs" = Computer Science, "stats" = Statistics, "biology" = Quantitive Biology, "finance" = Quantitive Finance (arXiv, 2020b)

Although arXiv includes papers from several disciplines, Figure 1 shows a clear dominance of physics and mathematics - and in recent years an increasing share of publications from computer science. As already mentioned, arXiv, in contrast to WoS, also includes articles that have not yet been peer reviewed. Preprints play a major role in scientific exchange, especially in the fields of astronomy and physics or in those areas, as we all had to recognize just now in the Corona crisis, where the fastest possible communication of scientific findings is required. So, arXiv enables faster exchange than would be possible through peer-reviewed journals by bridging the time to publication (Lariviere et al., 2014).

All registered users can submit articles. This also distinguishes the moderation process, which does not check journals for quality, but mainly determines whether the submitted papers have a "scholarly value" (arXiv, 2020a) and assigns the papers to a corresponding subject area (arXiv,

2020c). Although arXiv, in contrast to Web of Science, does not store citation data, it is possible to use the extension Bibex (arXiv Bibliographic Explorer), a project that displays citation data for the respective articles from Semantic Scholar, ADS (astrophysics data system), Prohpy and Inspire HEP29 and has been integrated into the site (Bierbaum, 2020).

3. Methodology

In order to make a comparison between the two databases, it is first necessary to identify criteria for such a comparison. For this purpose, a literature analysis was carried out with the aim of identifying comparison criteria from existing comparisons between known literature databases.

Pairwise searches were conducted for comparisons between the Google Scholar, Scopus, and Web of Science databases. This resulted in 3 queries that were searched in the Web of Science database:

[comparison AND „Web of Science“ AND „Google Scholar“],
[comparison AND „Web of Science“ AND „Scopus“] and
[comparison AND „Google Scholar“ AND „Scopus“].

The results were then narrowed down to the WoS categories "Information science library science", "Computer science interdisciplinary science", "Computer science information systems" and "Multidisciplinary sciences", since many of the results came from other scientific fields, such as medicine and biology. As a result, we retrieved 231 results. After removing the duplicates, the rest was analyzed for thematic relevance and further relevant results were identified via a backward and forward search based on these articles. As final result set 62 comparisons between literature databases were found.

4. Criteria Catalog and Comparison

As a basis for the comparison, the criteria proposed by Jadhav & Sonar (2009) for the comparison of software packages were used in order to compare not only, as most of the comparisons examined in the literature review, only key figures on literature databases, but to go one step further and also examine more software related criteria, such as user-friendliness, or criteria relating to the provider, so that the user can form a comprehensive picture of the databases examined. Table 1 shows the criteria catalog with the categories according to Jadhav & Sonar (2009) and as the result of the analysis of the literature analysis and their explanation.

Table 1 Citeria Catalog with its explanations (structure according to Jadhav & Sonar, 2009)

Criteria Category	Criteria: Explanation
Functionality	• Main aim: areas that the database is specifically focused on • Functions: all the features provided by the literature database • Interoperability: ability to use the database together with other tools or applications • Openness: to further developments or to external applications
Quality criteria	• User interface: how easy it is as a user to navigate and use the database • User types: ability of the database to be used by users of different types, such as beginners,

	advanced users, and professionals • Data presentation: how search results and any additional data are presented • Customizability of results: all the ways to refine search results after a query • Ease of use: how easy it is for a user to learn how to use the database and apply it accordingly. • Backup and recovery: options available to prevent data loss due to system crashes • Scope of articles: describes not only the number of all articles stored in the respective database, but also the period of publications covered (Bosman, 2006; Knackstedt & Winkelmann, 2006) • Completeness: articles published by one (or more) selected author(s) are all included in a database (Garcia-Perez, 2010; Tsay, Tseng & Wu, 2019; Goertzen, 2019). • Uniqueness: describes the proportion of articles that can only be found in one of the databases (Tsay, Tseng & Wu, 2019). • Quality of articles: is a statement about the control mechanisms of a database, which are intended to ensure the highest possible quality of the work by including only selected papers or journals (Cronin, 2001). • Scope of the citation data: quantity of all works for which citation data are available in addition to the bibliographic data. • Citation metrics: which of the numerous existing metrics are offered by the database (e.g.number of citation, h-index, or impact factor). • Actuality: how fast an article is included in the database. • Domain diversity: ability of the database to be used for different domains, not only different scientific disciplines, but also languages and regions of publication and types of content (Ball & Tunger, 2007).
Vendor criteria	• User manual: is there a manual for users and whether this manual contains relevant information and the main commands. • Tutorial: is a tutorial available to help users learn how to use the database. • Training: describes whether training is offered by the provider to explain the database and how to use it. • Consulting: provides the supplier of the database technical support and a consulting contact. • Communication, communication channels between user and provider. • Vendor experience: vendors experience as database provider. • Product reputation: reputation of the database in the market • Vendor reputation: reputation of the vendor in the market.
Costs	• License costs: costs of the database e.g. per user • Training costs: costs for the training for a user
Output	• Output: what options the database offers for exporting

4.1 Functional Criteria

Main aim: The main aims of the two databases examined are fundamentally different. Web of Science is especially designed for literature research and further bibliometric analysis. arXiv, on the other hand, emphasizes the fast provision of scientific documents for effective scientific communication by bridging the peer-review process. Functions: Web of Science has 5 different search functions:

- simple search,
- advanced search,
- author search,
- citation search and
- structure search.

For all search functions except the author search, it is possible to select which of the WoS databases or which indexes in this database are to be searched and in which time period. The

simple and *advanced search* support search operators - e.g. AND, OR, NOT-, wildcards, phrases and brackets to group compound Boolean operators (WoS Group, 2020e). The *advanced search* allows to formulate more complex searches by also including fields like title, author, organization etc. in the search term. The *citation search* should find all papers referencing one or more publications. In the first step, the publications are searched similar to the Basic Search and then, in the second step, the desired publications are selected from the search results. The *structure search* allows to draw chemical compounds or structures in an integrated tool and to search for articles on these compounds or structures. *Author search* allows to search for authors or, more precisely, to retrieve author profiles. These profiles contain information about an author as well as a list of all publications assigned to him. The search results are displayed as a list and can be analyzed using an extra button. As we can see in Fig. 2, the analysis shows how many papers from which subject area were found. Below this graphic, this data is also presented in full in tabular form. By clicking on one of the categories, the search results from that category can be displayed.

It is possible to create a citation report as can be seen in the upper right corner of Fig. 2. Figure 3 shows the citation report for the same example search. The h-index is calculated for the search results, and the total number of citations and citing articles is displayed. In addition, these numbers are also displayed without self-citations. Below this data, the citation data for each individual article is displayed in tabular form.

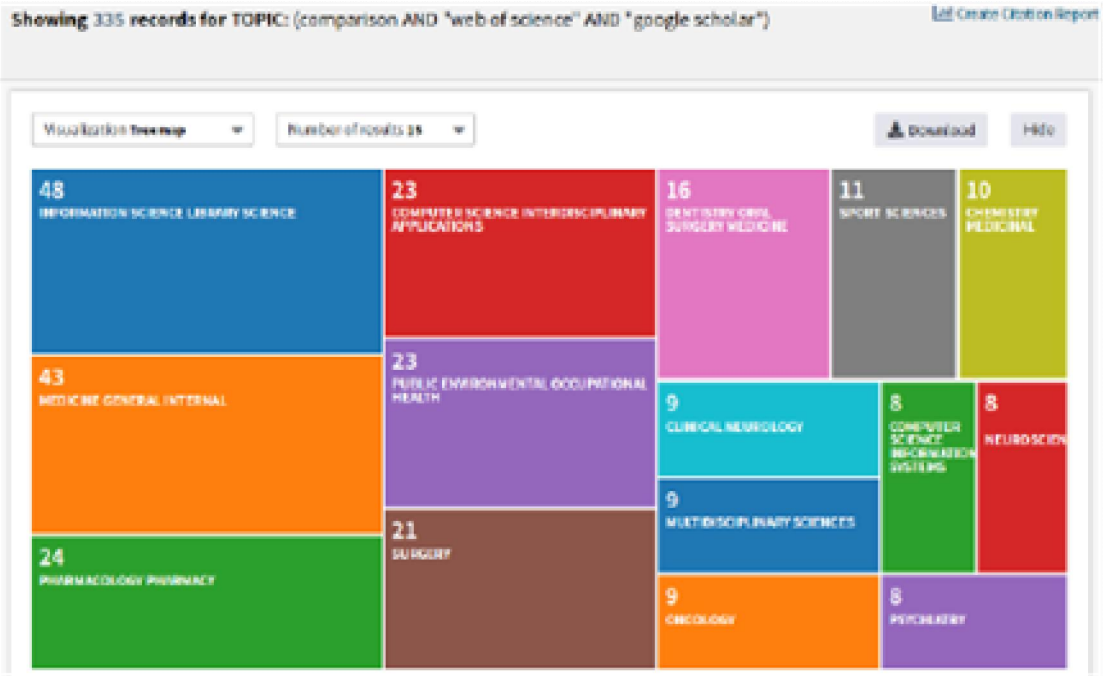


Figure 2: Analysis of the search results of a sample search

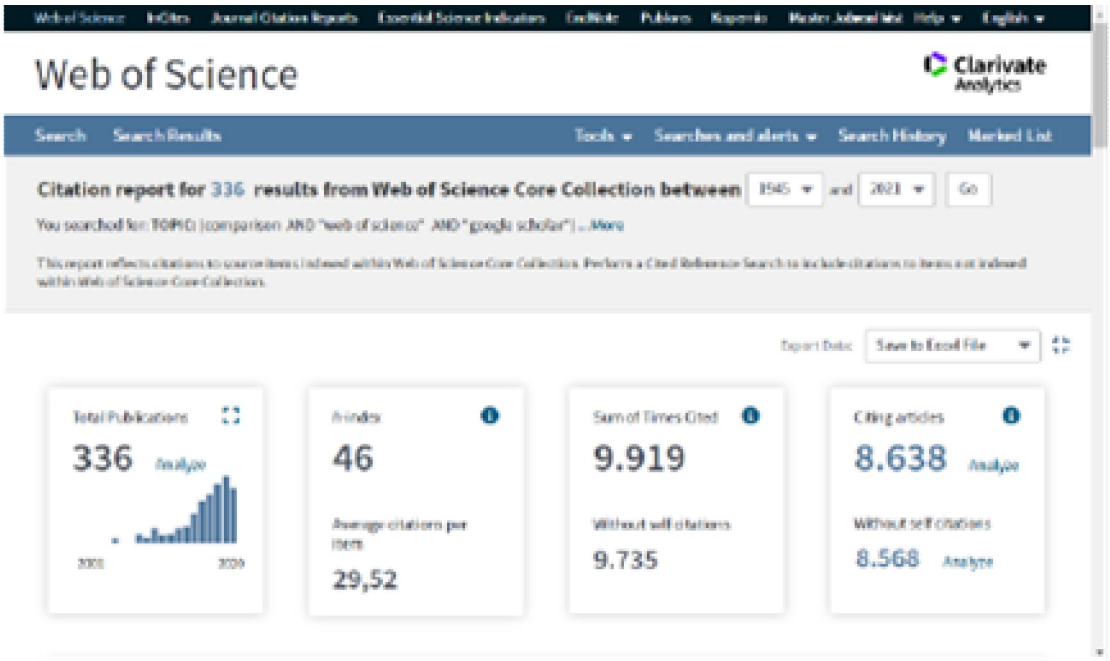


Figure 3: Citation report for a sample search

Search results can be further refined on the results page, e.g. a further search can be performed in the search results, or, for example, only certain document types or subject areas can be displayed. For each article included by Web of Science, the citation data is stored and can be displayed. Articles that cite the selected article can be displayed in a list like search results, filtered by additional criteria, and analyzed. The articles themselves are not stored on Web of Science, but a link to the article's page at the publisher is shown. The results of a search can be exported to various literature management programs, including EndNote, Citavi, RefWorks. etc. Export to Excel is also supported. Search queries that have been performed are automatically saved briefly in a search history but can also be saved manually. The search results are checked for new hits on a daily, weekly, or monthly basis, depending on the user's settings. If there are new matches, the user is notified by e-mail. Web of Science also offers a variety of other functions, such as a citation alarm, which, however, go beyond the search and analysis functionalities and are therefore not mentioned further here.

arXiv has only two search functions. A simple search and an advanced search. For both search functions, arXiv also supports searching for phrases, Boolean expressions, and wildcards. Only in the advanced search can search results be narrowed by subject and time period. In addition, similar to the advanced search of Web of Science, several fields can be addressed in a search query in order to make the search more precise. arXiv displays the search results in a list, as in Web of Science, but unlike WoS, they cannot be further refined. The individual articles can be downloaded directly from arXiv as .pdf files. Although arXiv does not store citation data, it is possible to display citation data for an article through an integrated extension - Bibex. This citation data is retrieved by several third parties, e.g. Semantic Scholar, Prophy, ADS and Inspire HEP. arXiv has no builtin functionality to export search results to a literature management program and no possibility to save search histories. Furthermore, it is possible to search for

authors, but there are no author profiles like in the Web of Science and only articles of the respective author are displayed as search results.

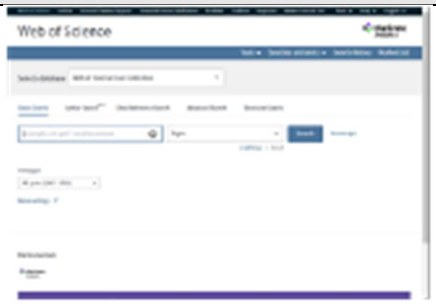
Interoperability: Web of Science offers interfaces for exporting to a variety of literature management programs (including EndNote, InCites, Refworks), as well as export functionality to Excel and a number of other data formats (such as HTML, BibTeX, plain text or CSV). arXiv does not offer export options by default.

Openness: Both providers offer an API that allows users to develop their own extensions and interfaces to the services of the. arXiv additionally maintains a list of projects via GitHub in which volunteers can participate.

4.2 Quality Criteria

In Table 2 we present the detailed comparison of the quality criteria section.

Table 2 Quality criteria comparison WoS arXiv

	WoS	arXiv
User Interface	 WoS Startscreen The interface is self-explanatory to a large extent. The buttons are labeled in an understandable way. When the dropdown menu for the search fields is opened, a short explanation for each of the options is displayed in addition to the choices.	 arXiv Startscreen The start page is much more overloaded compared to the start page of WoS, due to a huge amount of hyperlinks (on the screen-shot only about 50% of the hyperlinks are visible due to space limitations). The actual search function is placed in the upper right corner, very decentralized. The search interface itself is very organized, especially the simple search consists only of an input field for the search term, a dropdown menu for the examined fields and some explanations about the search options below.
User Types	Both databases are very easy to use for simple searches, even for beginners. For advanced users, both databases offer a more specific search functions in the form of the advanced search	
	Web of Science's simple search already includes all the functionalities that arXiv offers in its advanced search. The advanced search of Web of Science is suitable for advanced users, since the search results can only be limited by a self-written search expression. WoS offers (as written above) a lot more of analysis and report functions.	Advanced search in arXiv is done by selecting categories through a checkbox.

Data presentation	Both databases display the search results in the form of a list.	
	• Between10 and 50 results per page with: • title, author(s), journal title, date and number of citations • via hyperlink abstract and also usage data for each article is shown • via button WoS refers to publishers fulltext page • article data are supplemented with DOI, the author's address, information about the publisher, assigned subject areas, and a list of referenced works.	• Between 25 and 200 results per page with: • title, author(s), date, arXiv identifier, the DOI (if available) and abstract • arXiv allows you to download the work directly as a .pdf file
Customiz ability	For both literature databases, search results can be sorted according to certain criteria like:	
	• date of publication • number of citations, • number of uses (how often the full text of an article was accessed via Web of Science or how often the article was exported to a literature management system), • relevance, (indicates in WoS how many of the search terms were found in the fields Title, Abstract, Keywords, Keywords Plus. and Keywords Plus were found. Title and Keywords fields are weighted slightly higher), • last added to WoS, • number of uses in the last 180 days, • name of first named author, • source title, and • name of conference WoS allows to perform a further search within the search results or to filter by specific criteria, such as subject, date of publication, etc.	• announcement date, • submission date (for arXiv), and • relevance (without specifying how this is measured) arXiv does not allow to further refine the search results.
Ease of use	The basic functions of the Web of Science are very intuitive to use. Simple searches are no more complicated than a search on Google. For a user who has never used a literature database before, the amount of information and setting options of the search results might be a bit overwhelming, but it is not mandatory to use them. In order to use the advanced functions like the advanced search it is necessary to get familiar with Boolean expressions etc. first. Many functions such as the analysis of citations can be performed by simply pressing a button. Exporting to various literature management programs is also simple. However, if you want to export to literature management programs that are not directly supported, such as Citavi, you have to take the detour of exporting to a text file, which can then be imported into Citavi.Only the function for saving a search is unfortunate named. On the search results page you save a search with the button "Create an Alarm", because the functionality for alarms and saving a search in Web of Science belong together.	At arXiv, the search function is also very intuitive, and a simple search is no more complicated than a simple search on Google. The advanced search is, as already mentioned, more comparable to the simple search of the WoS and just as easy to use.

Backup and Recover	In WoS search queries are automatically saved in a search history until the browser is closed. In addition, it is possible to save search queries permanently.	arXiv does not offer any possibilities to save search queries.
Scope of articles	The Web of Science includes (in all databases together) about 171 million papers (articles from journals, books and contributions from events), as well as over 90 million patents and 9.7 million records from studies. The Web of Science Core Collection includes articles from 1900 to the present; the entire Web of Science contains literature from 1800 to the present (Status as of 27.07.2020, Clarivate Analytics, 2020b).	arXiv contains about 1.8 million articles(arXiv, 2020f). The oldest article is from 1991(Status as of 16.09.2020). If one tries to search for articles from 1990 or earlier, you get an error message saying that it is not a valid date.
Uniqueness	Tsay, Tsen, and Wu (2019) also compared the uniqueness of papers between Web of Science and arXiv that were available from Nobel Prize winners in physics, from 2001 to 2013. They found that around 43% of all articles from the Web of Science could not be found on arXiv, and only 1.27% of all articles on arXiv could not be found on the Web of Science. However, this number may be heavily distorted, since it concerns the work of very specific, important(because Nobel Prize winners) scientists. Another, more comprehensive study showed that, across disciplines, only 64% of all arXiv articles can also be found on the Web of Science. This number varies greatly by discipline; for example, the share of articles from arXiv that can also be found in the WoS is 80% for highenergy physics, but less than 20% for computer science (Lariviere et al., 2014).	
Quality of articles	WoS decides on the inclusion of a journal after a four-part selection process. 1. it is examined whether the journal can be clearly identified. 2. it is checked whether a complete editorial evaluation of the journal is justified. This includes checking whether the journal contains mainly scientific material, whether English-language titles and abstracts are available for the articles, and whether bibliographic information is available. 3. editorial review for quality, looks at consistency between the journal's title, stated scope, editorial board composition and authorship, and published content. It also looks for evidence of editorial rigor and adherence to community standards. If a journal is rejected at this step,it cannot be reexamined for at least 2 years. 4. is to examine the impact of the journal. This is determined mainly on the basis of citations. Both the number of citations and their origin play a role. In addition, it is examined whether the journal is of importance and value to subscribers of WoS. Significance can take the form of a novel perspective, a regional focus, a unique specialization, or unusual content (WoS Group, 2020d). However, WoS has been criticized several times for including mainly English-language titles from North America and Western Europe (Meho, 2007; van Leuven, 2001).	The arXiv moderation process is much less rigorous. Submitted articles are rejected only if, in the eyes of the moderators, they do not contain original or substantive research (including undergraduate research, course projects, and research proposals), if they are insufficiently formatted (lacking references, presentations, etc.), or if they concern a subject area that is not offered by arXiv. In addition, duplicates will be rejected, as well as titles that have been submitted even though the submitter does not own the rights to the title. A continuous peer review process is not necessary (arXiv, 2020c). WoS arXiv Scope of the citation data Web of Science stores citations for all papers included in the Web of Science CoreCollection (1.6 billion citations in total) (Clarivate, 2020b).

Scope of the citation data	Web of Science shows the number of citations for each included article. Furthermore, it is possible to display the h-index.	The Bibex extension integrated in arXiv collects citation data from several sources: Prophy, Semantic Scholar, Inspire HEP and ADS. Prophy includes citations to over 90 million titles (Prophy, 2020), Semantic Scholar claims to include over 180 million titles (Scholar, 2020), Inspire HEP includes about 1.4 million titles Inspire, 2020), and ADS includes over 13.3 million titles (ADS, 2020). The extent of the overlap of titles among these different providers is not known.
Citation metrics	Web of Science shows the number of citations for each included article. Furthermore, it is possible to display the h-index.	In arXiv, the number of citations for individual articles can be displayed by the integrated extension Bibex. An h-index for authors or searches is not calculated.
Actuality	WoS databases are updated at different intervals, from daily to monthly. The Web of Science Core Collection, for example, is updated daily (Monday - Friday) (Clarivate 2020b).	arXiv is continuously updated because articles are submitted by the authors.
	A study by Lariviere et al. (2014) showed that the average time between publication of a preprint on arXiv and publication of the final paper in a journal (and thus inclusion in the Web of Science) depends strongly on the subject area. For physics, for example, this time span is comparatively low, averaging less than half a year, while for mathematics it is comparatively high, averaging more than a year.	
Domain diversity	Web of Science includes articles from all scientific disciplines, but according to its own statement, the coverage of natural sciences, health sciences, engineering, computer science and material sciences is the highest. There are regional indices for Korea, Latin America, Russia and China (Clarivate, 2020b). Although non-English papers are included, around 95% of all papers are in English. The majority of papers were published in North America and Western Europe, led by the United States and the United Kingdom (see Vera-Baceta et al.; 2019, Meho & Yang, 2006).	arXiv only accepts papers from the departments of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering, and economics (arXiv, 2020a). The focus is on articles from the departments of physics, mathematics, and computer science. Non-English language papers are also accepted as long as an Abstract in English is included. No details are available on the distribution of languages or the origin of papers (arXiv, 2020d).

4.3 Vendor Criteria, Costs and Output

Both databases offer a quick guide to the most important search functions and terms, which can be found directly next to or below the search field. WoS offers a much more comprehensive guide, which can be accessed directly via a link next to the search button. Both databases also offer help files and a FAQ. Only WoS provides tutorials in the form of videos and self-guided training it also offers live training in several languages (WoS Group, 2020f). Only Web of Science offers customer support. Technical support is offered by both. WoS provides different contact options: a form on the website, by email, or by phone (Clarivate, 2020c). arXiv can only be contacted by email. Web of Sciences oldest index The Science Citation Index Expanded, (formerly known as the Science Citation Index) has been in operation since 1964. Web of Science was the first citation database and the only one of its kind for over 40 years (Li et al., 2010). Today, it is still considered one of the most important citation databases (Vera-Baceta et al., 2019). arXiv was founded in

1991 and has enjoyed increasing popularity ever since (arXiv, 2020a). In 2019, 243 libraries, institutions, and associations supported arXiv financially (arXiv, 2020e).

For the use of Web of Science, a license is required. The cost is not made public, as prices are always negotiated on an individual basis. A statement exists from the Texas A&M University Libraries that the cost of the Web of Science subscription in 2019 was $212,000 (Tabacaru, 2020). The cost of the video tutorials and live training is included in such a license. arXiv is completely free as a user

Web of Science supports a variety of export options: Single articles or multiple search results can be exported to End-Note Desktop or End-Note Online (both provided by ClarivateAnalytics), as well as to Excel, InCites, RefWorks and other file formats such as BibTeX, HTML and plain text. Furthermore, the search results can be sent by e-mail or printed (the last function is called "Fast 5k", because it can export up to 5,000 articles at once). However, with Fast 5k only author, title and source can be exported. With all other options, the abstract can also be included, or even the entire records of articles. However, it is not possible to export the original articles via Web of Science, because they are not stored by Web of Science. It is not possible to export more than 500 titles at once, except with Fast 5k. arXiv does not offer any functions to export to literature management programs or similar. Instead, it is possible to download any article directly from arXiv. Mostly in .pdf format, but some articles are also available in other file formats like Post-Script.

5. Summary and Outlook

In summary, both databases have different strengths. Web of Science is particularly well suited for basic literature searches due to the stricter quality controls and also due to the larger volume of articles. Web of Science is also better suited as a source for bibliometric analysis, as arXiv itself does not store citation data and does not offer any functions for analysis. arXiv, on the other hand, is particularly well suited to providing an overview of very recent research developments in one of the supported subject areas, because arXiv is a database that specializes primarily in preprints and provides highly current research results. Nonetheless, there is a potential loss of quality, as there appears to be no peer-review process. arXiv gains attractiveness that it can be used free of charge and all articles can be downloaded directly, also free of charge.

So far, the two databases have been compared exclusively from the point of view of content. However, data could not be found for all identified criteria, such as the costs of the Web of Science or the distribution of languages and countries of origin of the papers on arXiv. Some criteria, such as the completeness of the papers or the uniqueness of the papers, were only compared on the basis of data from the found literature, since an independent comparison would have exceeded the scope of this study. Furthermore, criteria such as ease of use and user interface have only been evaluated subjectively.

During the work on this topic, it became apparent that there are many uncovered areas that can still be explored. For example, the sources for citation data used by arXiv could be examined for

overlap or uniqueness of their data in order to get a complete picture of the citation data on arXiv, which in turn could be compared with the data of the Web of Science. Especially with regard to the criteria completeness and uniqueness of the articles, additional investigations could be carried out, as already mentioned in the critical appraisal, since the two investigations found on uniqueness, for example, resulted in different values. Moreover, the criteria catalog presented here could be used for comparisons between other literature databases.

6. References

[1] ADS (2020). About ADS. Retrieved from https://ui.adsabs.harvard.edu/about/

[2] arXiv (2020a). About arXiv | arXiv e-print repository. Retrieved from https://arxiv.org/about

[3] arXiv (2020b). arXiv submission rate statistics | arXiv e-print repository. Retrieved from https://arxiv.org/help/stats/2019_by_area/index

[4] arXiv (2020c). arXiv moderation. Retrieved from https://arxiv.org/help/moderation

[5] arXiv (2020d). Submissions in Languages Other Than English and Multiple-Language Submissions | arXiv e-print repository. Retrieved from https://arxiv.org/help/faq/multilang

[6] arXiv (2020e). arXiv Annual Update, January 2020 | arXiv e-print repository. Retrieved from https://arxiv.org/about/reports/2020_update

[7] arXiv (2020f). Monthly Submissions. Retrieved from https://arxiv.org/stats/monthly_submissions

[8] Ball, R., & Tunger, D. (2007). Science indicators revisited – Science Citation Index versus SCOPUS: A bibliometric comparison of both citation databases. *Information Services & Use*, *26*(4), 293–301. https://doi.org/10.3233/ISU-2006-26404

[9] Bauer, K., & Bakkalbasi, N. (2005). An Examination of Citation Counts in a New Scholarly Communication Environment. *D-Lib Magazine*, *11*(09), 1. https://doi.org/10.1045/september2005-bauer

[10] Berg, J. M., Bhalla, N., Bourne, P. E., Chalfie, M., Drubin, D. G., Fraser, J. S., . . .Wolberger, C.(2016). Scientific COMMUNITY. Preprints for the life sciences. *Science (New York, N.Y.)*, *352*(6288), 899–901. https://doi.org/10.1126/science.aaf9133

[11] Bierbaum, M. (2020). mattbierbaum/arxiv-bib-overlay. Retrieved from https://github.com/mattbierbaum/arxiv-bib-overlay

[12] Clarivate Analytics (2020a). LibGuides: Web of Science platform: Introduction. Retrieved from https://clarivate.libguides.com/webofscienceplatform

[13] Clarivate Analytics (2020b). Web of Science: Summary of Coverage. Retrieved from https://clarivate.libguides.com/webofscienceplatform/coverage

[14] Clarivate Analytics (2020c). Support - Clarivate. Retrieved from https://clarivate.com/contactus/support/

[15] Cronin, B. (2001). Bibliometrics and beyond: some thoughts on web-based citation analysis. *Journal of Information Science*, *27*(1), 1–7. https://doi.org/10.1177/016555150102700101

[16] Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *22*(2), 338–342. https://doi.org/10.1096/fj.07-9492LSF

[17] INSPIRE (2020). Literature Search - INSPIRE. Retrieved from

https://inspirehep.net/literature?sort=mostrecent&size=25&page=1&q=
[18] Jacso, P. (2005). As we may search - Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, *89*(9), 1537–1547.
[19] Jacso, P. (2004). Citation-enhanced indexing/abstracting databases. *Online Information Review*, *28*(3), 235–238. https://doi.org/10.1108/14684520410543689
[20] Jadhav, A. S., & Sonar, R. M. (2009). Evaluating and selecting software packages: A review. *Information and Software Technology*, *51*(3), 555–563. https://doi.org/10.1016/j.infsof.2008.09.003
[21] Lariviere, V., Sugimoto, C. R., Macaluso, B., Milojević, S., Cronin, B., & Thelwall, M. (2014). arXiv E-prints and the journal of record: An analysis of roles and relationships. *Journal of the Association for Information Science and Technology*, *65*(6), 1157–1169. https://doi.org/10.1002/asi.23044
[21] Li, J., Burnham, J. F., Lemley, T., & Britton, R. M. (2010). Citation Analysis: Comparison of Web of ScienceR, Scopus™, SciFinderR, and Google Scholar. *Journal of Electronic Resources in Medical Libraries*, *7*(3), 196–217. https://doi.org/10.1080/15424065.2010.505518
[22] Meho, L. I. (2007). The rise and rise of citation analysis. *Physics World*, *20*(1), 32–36. https://doi.org/10.1088/2058-7058/20/1/33
[23] Prophy (2020). Prophy Referee Finder. Retrieved from https://www.prophy.science/referee-finder
[24] Semantic Scholar (2020). Semantic Scholar | About Us. Retrieved from https://pages.semanticscholar.org/about-us
[25] Steele, B. (2012). Library-managed 'arXiv' spreads scientific advances rapidly and worldwide. *EZRA - Cornell University's Magazine*, *5*(1), 9. Retrieved from http://ezramagazine.cornell.edu/FALL12/ezra.pdf
[26] Tabacaru, S. (2019). *Web of Science versus Scopus: Journal Coverage Overlap Analysis*. Texas. Retrieved from https://oaktrust.library.tamu.edu/handle/1969.1/175137
[27] Tsay, M.-y., Tseng, Y.-w., & Wu, T.-l. (2019). Comprehensiveness and uniqueness of commercial databases and open access systems. *Scientometrics*, *121*(3), 1323–1338. https://doi.org/10.1007/s11192-019-03252-3
[28] Van Leeuwen, T. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & van Raan, A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics*, *51*(1), 335–346. https://doi.org/10.1023/a:1010549719484
[29] Vera-Baceta, M.-A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, *121*(3), 1803–1813. https://doi.org/10.1007/s11192-019-03264-z
[30] WoS Group (2020a). SCIE - Web of Science Group. Retrieved from https://clarivate.com/webofsciencegroup/solutions/webofscience-scie/
[31] WoS Group (2020b). Web of Science Core Collection - Web of Science Group. Retrieved from https://clarivate.com/webofsciencegroup/solutions/web-of-science-core-collection/
[32] WoS Group (2020c). Trusted publisher-independent citation database - Web of Science Group. Retrieved from https://clarivate.com/webofsciencegroup/solutions/web-of-science/

[33] WoS Group (2020d). Web of Science Journal Evaluation Process and Selection Criteria. Retrieved from https://clarivate.com/webofsciencegroup/journal-evaluation-process-and-selection-criteria/
[34] WoS Group (2020e). Search Rules. Retrieved from https://images.webofknowledge.com/WOKRS535R95/help/WOS/hs_search_rules.html
[35] WoS Group (2020f). Web of Science Learning Home - Web of Science Group. Retrieved from https://clarivate.com/webofsciencegroup/support/home/
[36] Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii. Retrieved from www.jstor.org/stable/4132319