Web of Science Citation Gaps: An Automatic Approach to Detect Indexed but Missing Citations

David Rodrigues
Iscte - University Institute of Lisbon, Portugal

António L. Lopes
Instituto de Telecomunicações, Iscte - University Institute of Lisbon, Portugal

Fernando Batista
Iscte - University Institute of Lisbon, Portugal
INESC-ID Lisbon, Portugal

Abstract

The number of citations a research paper receives is a crucial metric for both researchers and institutions. However, since citation databases have their own source lists, finding all the citations of a given paper can be a challenge. As a result, there may be missing citations that are not counted towards a paper's total citation count. To address this issue, we present an automated approach to find missing citations leveraging the use of multiple indexing databases. In this research, Web of Science (WoS) serves as a case study and OpenAlex is used as a reference point for comparison. For a given paper, we identify all citing papers found in both research databases. Then, for each citing paper we check if it is indexed in WoS, but not referred in WoS as a citing paper, in order to determine if it is a missing citation. In our experiments, from a set of 1539 papers indexed by WoS, we found 696 missing citations. This outcome proves the success of our approach, and reveals that WoS does not always consider the full list of citing papers of a given publication, even when these citing papers are indexed by WoS. We also found that WoS has a higher chance of missing information for more recent publications. These findings provide relevant insights about this indexing research database, and provide enough motivation for considering other research databases in our study, such as Scopus and Google Scholar, in order to improve the matching and querying algorithms, and to reduce false positives, towards providing a more comprehensive and accurate view of the citations of a paper.

2012 ACM Subject Classification Applied computing → Publishing; General and reference → Verification; Information systems → Digital libraries and archives; Information systems → Enterprise applications; Applied computing → Digital libraries and archives; Information systems → Data cleaning

Keywords and phrases Research Databases, Citations, Citation Databases, Web of Science, OpenAlex

Digital Object Identifier 10.4230/OASIcs.SLATE.2023.5

Funding This work is funded by FCT/MCTES through national funds and when applicable co-funded by FEDER – PT2020 partnership agreement under the scholarship reference Iscte_SIIC/01/2022, and projects UIDB/50021/2020 and UIDB/50008/2020.

1 Introduction

The number of citations on a research paper is of great value for the researcher and their work, since usually, the more citations a paper has, the higher chances of it being worth reading and having helpful information. Thus, the authors of said paper can also get prestige from the number of citations of their paper and will be more highly regarded. This is useful not only for the author to get recognition by his peers, but also for performance evaluation processes in the researchers' institutions that usually include publications' citations as one of the main metrics. This raises the importance of knowing the right amount of citations that any given paper has.

However, it is not always easy to keep track of citations, even with the existence of big citation databases, considering they too have their shortcomings. Each database has their own source list, which means that some papers that are indexed in a database, may not be found in another, therefore, there usually are unique citations for each database. But sometimes, the unique citations we find, in reality, are missing citations that one of the other databases did not find, because the citing paper is indexed in the database, but it is not tagged as citing the paper. That being the case, the database is missing information that another one has found, leaving us with incomplete data if only one database is used to check for the number of citations of any given paper. Therefore, it is very useful to devise a way to automatically find missing citations between databases, to get a closer depiction of the real number of citations in any given paper. The problem we aim to solve is to find these citations, where an article is indexed in multiple databases, but it does not count as a citation to the research paper we are studying in at least one of them. As a consequence of this problem, finding a way to automatically merge the information from multiple databases could get us a closer depiction of the real number of citations of a given work without having to go through the slow and tedious process of manually assessing these differences.

In this paper we present our first experiment, where we explore the potentially missing citations in Web of Science (WoS) using OpenAlex as a baseline and find some of the citing articles that WoS might be missing. There are two main reasons why we started by using OpenAlex as a baseline: first, it is a free tool that gives us all the information via an API, therefore it is easy to access and to gather information from it; secondly, OpenAlex gathers information from multiple other resources, which means it might have a more complete citation list for a given paper than other indexing databases. Consequently, OpenAlex allowed us to test our approach faster, and helped us make sure that we find missing citations in WoS, before moving on to Scopus and Google Scholar to complete our analysis.

In the next sections, we present the literature review, followed by a step by step description of the approach to achieve the goal of automatically finding missing citations. Section 4 analyzes the results of an experiment performed with a set of more than 3000 articles. Section 5 describes a manual validation process, using a randomly chosen portion of the data we gathered, in order to make sure that our approach was working correctly. Finally, Section 6 presents the major conclusions and pinpoints possible directions for further work.

2 Literature Review

Since each citation database shows strength covering different areas, looking only at one database can be misleading [7] because a database may not have some of the articles that could potentially cite the paper which a researcher is looking at since there are unique citations for each database that the others do not have [5]. Moreover, missing publications from indexed journals can aggravate the problem, for example, in a study from 2006 to 2017 in the journal of enfermeria nefrologica, only 50.2% of the papers were indexed by Scopus [6]. Like Scopus, the other databases also show lack of coverage in some areas, but all of them are working on enhancing their coverage, and getting all the articles they can in their databases, showing some improvement over the years [7]. Nevertheless, neither database is perfect, and for better results, more than one should be consulted.

The importance of having an accurate citation count is partially shown in the research conducted by [9], where they correlated the citation count of a paper with the amount of times that a paper would be consulted, as well as the attention level that the readers had while examining it, showing that the higher citation count, the higher chance of a researcher to read the paper and also pay closer attention to it.

There are two reasons why a paper might be missing citations in a database: either it is because the citing paper is not indexed in the database, or because there was some error while processing the publication's citing data. [1] found that there are a variety of errors that can justify missing citations, and those errors can be committed by both the databases or the authors of the articles. There can be errors in the DOIs or in the references of a paper.

Exploring these errors, [2] tries to automatically find and correct wrong DOIs in the databases, that were wrongfully inserted either by the authors or by the databases themselves. Although no concrete numbers of the total of errors were provided, we can see there are a lot of errors found by the study. Ovid Technologies publisher alone had over 370,000 outgoing citation errors in 2 years. Besides wrongfully written DOIs, [4] reports that databases sometimes, mistakenly, give the same DOI to different articles, and since the DOI should be unique for each article, these errors can make a difference in bibliometric analysis.

A follow-up study reported by [10], tries to find what were the differences of the references lists in Scopus and WoS. While comparing the reference list of around 100.000 papers in both databases, using as a baseline the Elsevier ScienceDirect Article Retrieval API to get the references, they found that WoS had 77.2% of the papers with the same number of references, while 19.3% had fewer references. On the other hand, 96.4% of Scopus papers had the same amount of references. They manually analyzed random papers with different results and saw a variety of different reasons why this happened. Since the different databases themselves, can extract different references list of a paper, this can also explain why there can be missing citations in some databases, because if they don't extract a reference, or do it incorrectly, then the paper that it is being cited won't be found in the database as a cited paper and the citation will be missing.

Having now a bigger understanding of some of the problems we might face, and what causes them, we look for research that has been conducted where we can take inspiration from. The research conducted by [5], showed us that using queries to try to find articles through their authors' names, is most of the time a futile endeavor, and might not the best approach, since there are a lot of varieties of the same author name, as well as some databases extract the authors' names very poorly. Also, [3] identifies the missing citations of several articles in an automated way. Our work follows a similar methodology to theirs, the main difference being the way we identify if a possible missing citation paper is present in the database where it is missing from. While they check if the paper's journal or conference title is present in the list of indexed sources of the database in question, we aim to look for the paper itself in the database, making sure it really is indexed by it.

Finally, an advantageous tool we found is OpenAlex, which is a "fully-open index of scholarly works, authors, venues, institutions, and concepts" that allows us, using the DOI of a given paper, to get all the information that they gathered about that paper. OpenAlex gets all their data from multiple services, like MAG and Crossref being the more important ones, and also from ORCID, ROR, DOAJ, Unpaywall, Pubmed, Pubmed Central, The ISSN International Centre and Subject-area and institutional repositories from a plethora of platforms [8]. Using this API, it is possible to get easy access to the information from all these other platforms from a single query.

3 Proposed Method

Our main goal is to automatically find missing citations between different citation databases. In this project, we will be looking at Web of Science, using OpenAlex as a baseline reference. We first need to find all citations that OpenAlex found, but WoS did not. Afterwards, with these potentially missing citations, we have to search WoS to check if each citing paper is indexed in it or not. If it is, then it is considered to be a missing citation.

Table 1 Example of a report for paper A.

Paper A Analysis		
Citing Articles	WoS	OpenAlex
1	✓	✓
2	Missing	✓
3	✗	✓
4	✓	✓
5	✗	✓
6	✓	✗
Recorded Citations	3	5
Expected Citations	4	5
Missing Citations	1	0
Missing Percentage	25%	0%

An example of the information we want to find out is shown on table 1. If we analyze paper A, and WoS say that paper A is being cited by the articles 1, 4 and 6, and OpenAlex says that paper A is being cited by the articles 1, 2, 3, 4 and 5, there are 3 citing articles that OpenAlex found and WoS did not, which are 2, 3 and 5. The next step is to look for these 3 articles in WoS, in order to confirm if they are indexed by the database. And, for example, if only article 2 was found to be indexed by WoS, we can assert that article 2 is a missing citation of paper A on WoS.

In order to look for missing citations in an article, all we need is its DOI. The first step in the verification process is to clean the received data. Sometimes, the DOI can have extra information, such as the *https* link, the *doi.org* domain, or other invalid characters or blank spaces, so we standardize the DOI, so it is easier to work with it through the remaining steps.

Afterwards, we ask the institution's Current Research Information System (CRIS) API[1] for more information about this paper. From there, we can get the article's title, authors, year, the type of the article (journal, book, conference, etc.), as well as some information about the WoS representation of this paper, such as if the article is indexed in WoS, the url for the list of citing publications of our article and the unique identifier of WoS (accession number) for this article (WoS ID = WOS:xxxxxxxxxx).

The next step is to query OpenAlex for their information about this article. Once again, the query is done via the DOI, and all the information gathered about the article from the CRIS is compared with the one gathered from OpenAlex. With OpenAlex we can also get their list of citing publications, which comes with all the information they have about each of these citing publications, namely the DOI, title, publication year and the type.

After retrieving the list of citing publications from OpenAlex, we need the corresponding WoS list of citing publications to compare them both. Since the WoS API only provides the link for the page where the information is, we had to extract the information of each citing paper ourselves. We use the link provided by the CRIS in order to retrieve this information, but it would also be possible to get to the web page with the list of citing publications through the WoS unique identifier, since with that, we could build an URL that takes us directly to the WoS page about the paper, and from there we are only one click away from the list of citing publications.

[1] `https://ciencia.iscte-iul.pt/api/v2_6/doc`

Furthermore, we could also try to find the publication via the DOI and other information we got, but that method is more prone to errors, therefore, if possible, using the link to the page or the WoS ID is the better solution.

With the list of citing publications from both OpenAlex and WoS, we can look for possible missing citations in WoS. For this effect, we first try to match the OpenAlex publications with the WoS publications, and if both of these databases have the same publication tagged as citing our article, we don't need to do any further work for this article. If instead, the process does not find a match, then this publication is flagged as a possible missing citing publication for the article in WoS.

Having now all of these publications that are flagged as potentially missing citing publications in WoS, we need to check if they are in fact indexed in WoS. And only if they are, then they are considered missing citations. So once again, we must consult WoS to extract this information. This step is harder than the previous one, since before we already had a link to the page of paper we were looking for, while this time, we don't have any WoS information about this paper. In order to find the paper in WoS, we execute two different queries. First, we query using only the DOI of the publication '(query: DO=(doi))'. If this query does not provide results, we also try to find the publication via the title and publication year '(query: TI=(title) AND PY=(year))'. If any of the queries provides results, we extract the results from WoS the same way we extract the list of citing publications, and we double-check if the DOI is the same, or if the title, year and publication type are the same. The reason we don't use only the DOI to perform this check is because some publications don't have DOIs (or the database might be missing that information), but also because sometimes there are errors in the assignment of DOIs in databases, as seen earlier in the literature review. Hence, we might not find the correct publication via DOI, and do so through the title. If the queries' results are a match, then we are in the presence of a missing citation, and we tag it as such in the report we send back to the client. With every missing citation, we also provide the link of the publication we found that is supposed to be citing our article, so that the client can easily double check manually any missing citations they might want to check. These queries are very strict, and specially in the case of the title, if there is an error on a word, or any difference, they will not match, in the future we wish to improve these queries, using similarity techniques that will make it possible to find a match even in the presence of small errors. We would also like to try to expand these queries and use other information we might have on the paper to try to find it in the new database, although this must be carefully tested in order to limit the number of false positives that our application finds.

Through all this process, one of our main goals is not to overload the APIs or repeat processes unnecessarily. With that in mind, we use delays between every request to avoid overloading the systems. Moreover, we have a local database which is used as a cache, so if we are asked to analyze the same publication in a short period of time (the time period of when the data becomes outdated varies from each type of data, and ranges from 1 day to 1 month), we don't have to gather all the information from external sources to provide an answer. Therefore, for the entire process, every time we need any data, we first query our local database, and only if we don't have the data stored locally we will do the process described above in order to get the necessary data.

Finally it is important to state that in order to create a report for any publication, the publication has to be indexed in both databases, so if we can't find it either in WoS or OpenAlex, no report is created. Also, if the publication has no citations in either database, there isn't anything to look for, and no report is created. A diagram with the representation of our system's design can be seen in Figure 1.

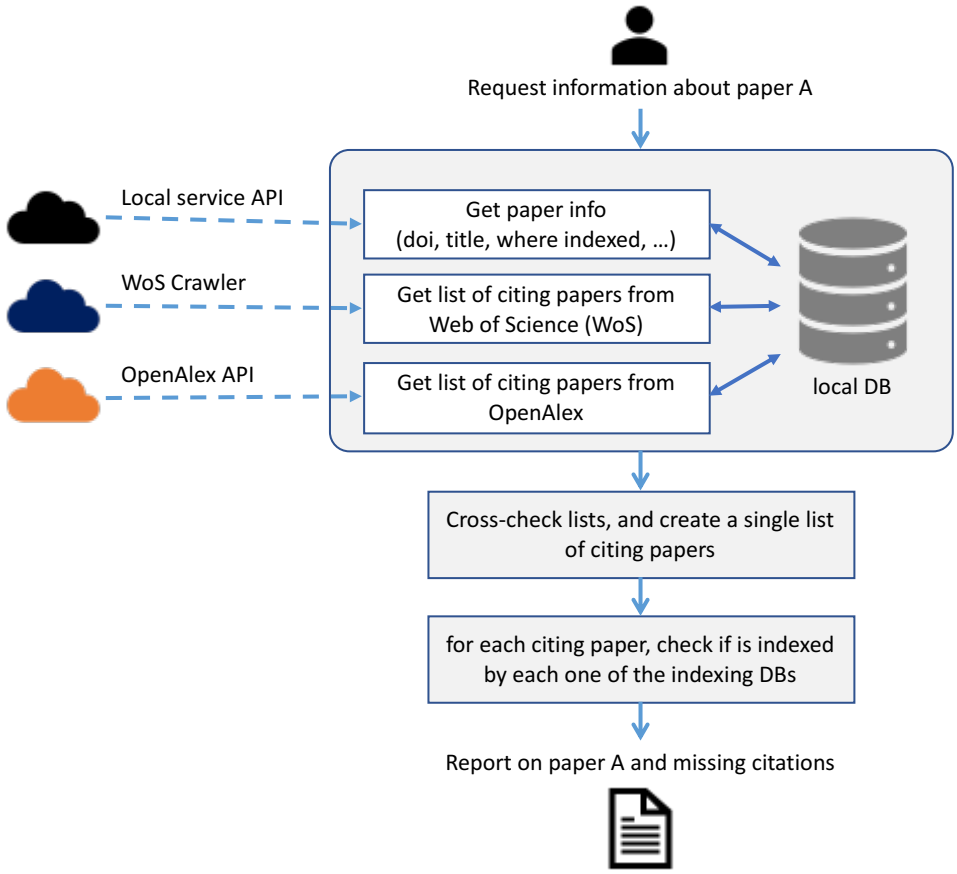

Figure 1 System's design.

4 Evaluation

In order to evaluate our proposed approach, we performed an experiment where we used the papers provided by our institution's CRIS as the dataset. From the 52.000 papers present in the dataset, we can only use 12.500 of them, since the others do not have a designated DOI.

For our experiment we decided to take a year-based approach, which means we carried out the analysis of all papers of a predetermined set of years. We chose the years 2015, 2018 and 2021. We chose these years for a multitude of reasons. First, we wanted years that would be relevant for our motivation, and since our university only examines the papers from the last 10 years to evaluate their researchers' performances, we did not want to look at papers that were older than 10 years. We also wanted to analyze multiple years, not only so we would have more data, but also so we could compare the results from each year and see if we could notice anything changing or even improving over time. The year of 2018 was the first year we explored, since it was before the COVID pandemic that started in 2019, and that was an untypical year that probably would not give us a trustworthy result. The year of 2021 was chosen because we also wanted a more recent year, but not so recent that the publication would not have had time to impact the scientific community, since if from the moment a publication is published, it still has to be read by other members of the community, that have to write and publish their own work. Finally, the year of 2015 was chosen to keep the 3 year gap between 2018 and 2021. A summary of the results from these 3 years is present at Table 2.

Table 2 Results per year.

	Year		
	2015	2018	2021
Papers analysed	820	950	1304
Papers that met the criteria	446	497	596
Total OpenAlex Citations	10490	10653	4114
Total WoS Citations	8760	9277	3411
Papers with MC	128	140	130
Total MC	233	265	198
Average percentage MC	5,30%	5,10%	9,60%
Average percentage MC in papers with MC	18,50%	18,30%	44,17%
Maximum MC	12	25	25

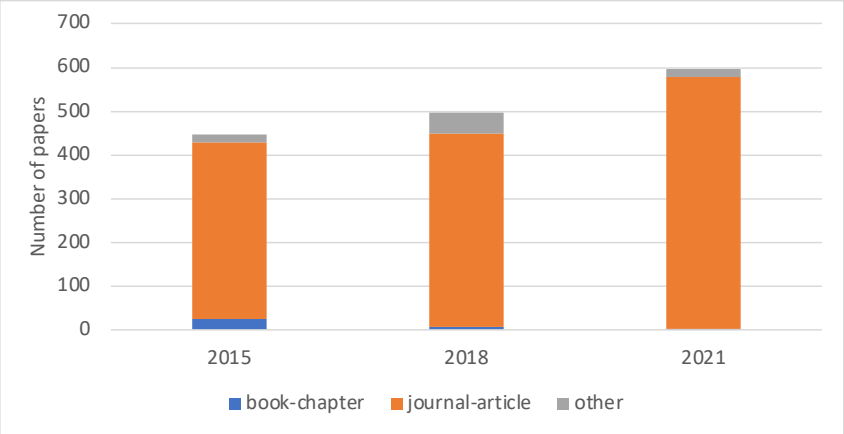

Figure 2 Type of publications being analysed.

Figure 2 shows the number of papers being considered in our experiments for each year, by publication type. It is interesting to observe that most of the papers are of the type "journal-papers", a tendency that is increasing over the years.

In total, 3074 papers were analyzed by our system, where 820, 950 and 1304 are from the year 2015, 2018 and 2021, respectively. For all the analyzed papers, a report is only created if the paper is indexed in both databases (WoS and OpenAlex) and if the paper has at least one citation in either of them, therefore, for each of the years, only 446, 497 and 596 (respectively) reports were created, giving us a total of 1539 reports of different papers. We can see that the values are increasing over time, not only in the number of papers, but also in the number of reports created, especially in 2021, where there was an increase in over 350 papers and almost 100 more reports when compared to 2018, whereas if we compare 2018 with 2015, the increase was only of 120 papers and 51 reports. While the number of reports did not increase as much as the number of papers, that is probably due to the fact that since 2021 papers are much more recent and maybe some of them are not indexed in both databases yet or they do not have any citations. If we did the same analysis in one year, the data from 2021 would be the one that would change the most.

All the data we are presenting next comes from the papers where a report was created. The total number of citations in WoS for each year were 8760, 9277 and 3411 respectively. The huge drop in the number of citations in 2021, comes from what has been said before, which is that since the papers have had less time to be cited, the number of citations is gonna be lower. The low number of citations, when compared to the other years, also greatly influences the percentages that are presented in Table 2.

In 2015, we found missing citations in 128 out of the 446 papers with a total of 233 Missing Citations (MC). This means that 28,70% of the papers had MC, averaging 1,82 per paper. The numbers in 2018 are very similar, with 140 papers out of 497 having MC (28,17%) and with a total of 265 MC, with 1,89 per paper, a slight decrease in the number of papers with MC but a slight increase of MC per paper. Once again the year of 2021 stands out with only 130 papers out of 596 with MC (21,81%) and with 198 MC, with an average of 1,52 per paper.

$$Average\_percentage\_of\_MC = MC/(MC + Citations)$$

To calculate the average percentage of MC we chose to use the formula presented above, which will tell us, out of all the citations that a paper should have (the ones the database caught and the ones it missed), how many are missing, giving us results from 0% to 100%. We also counted the average of MC for all the papers, and only for the papers that had MC. Out of all the papers, the percentage of missing citations both in 2015 and 2018 rounded the 5%, while in 2021 it was 9,6%. If we look only for the papers with MC in their reports, in 2015 and 2018 we had 18,5% and 18,3% respectively, while in 2021 we got an astounding 44,17%. In 2021 the number shoots up, especially because of the low number of citations in 2021. Once again we argue that these numbers in 2021 are higher because the papers did not have as much time to be cited yet, and the papers that already cited them are still being worked on by WoS, and because the papers are recent, the WoS process is still behind in referencing the citations on the papers. In Figure 3 we present the distribution of papers according to the percentage of missing citations. Although the average number of MC per paper is low (between 1,5 to 1,8), there were some cases where a high number MC were identified as it can be seen in Figure 4. In 2015 a paper had 12 MC, and in 2018 and 2021 both had a paper each with 25 MC each. Only 8% of the analyzed papers had over 3 MC, and 7 papers had 6 or more MC, showing us that most of them have 3 or less MC.

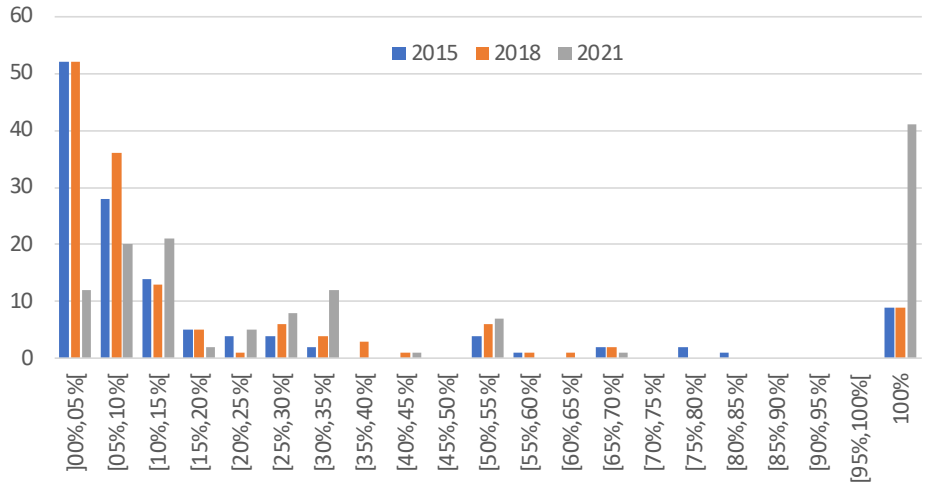

Figure 3 Distribution of papers according to missing citations percentage per year.

5 Results Validation

In order to have more trustworthy results, we manually checked a sample of the reports that we created. Part of that sample was retrieved at random, but another part was retrieved from reports that stand out because of unusual or unexpected results. We manually checked all

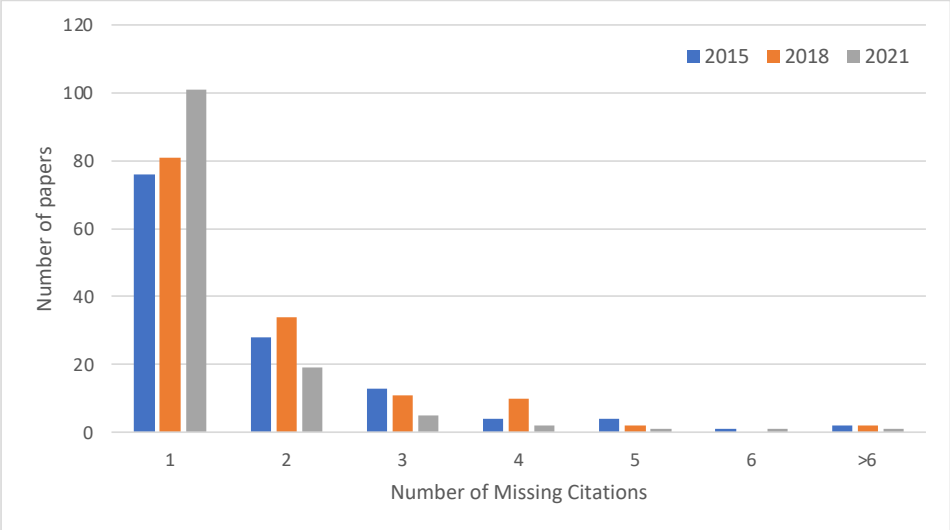

Figure 4 Distribution of papers according with the amount of missing citations per year.

reports that had over 3 MC and where those 3 MC represented at least half of the citations of the paper. This is because if a paper had 4 MC out of 100, then it was quite plausible that WoS might be missing a few citations, while if there were 4 MC out of 5 total citations, then something might be wrong, since it means that WoS did not find most of the citations that it should have.

The manual verification was processed as such for any given report:

- We had the DOI of the paper that was initially provided to our system, as well as the OpenAlex and WoS web links for this paper. Both were double checked to make sure they related to the same paper.
- For each missing citation identified by the system, we checked the OpenAlex and WoS web links for the citing paper that our system matched, in order to make sure that the OpenAlex publication was the same we found in WoS.
- If the original paper being analyzed was the same in both databases, and all the missing citations were correctly matched with a paper indexed in WoS, then it was a missing citation.

We manually verified 40 random papers, from which 36 had correctly identified their 51 MC, while the other 4 reports had some errors.

In one of these reports with errors, out of 7 MC found, 2 of them were wrongly matched, meaning that the OpenAlex citing publication was not the same as the WoS publication we found through our queries. Upon closer inspection we found that the problem was that these 2 publications had the same DOIs in the different databases, and after a search for the DOI in google, it looked to us that WoS had the publication with the wrong DOI.

We also found 2 cases, where one of the missing citations was due to OpenAlex saying that the publication was citing itself, and since the publication was indexed in WoS, it was tagged as a missing citation. Therefore we have a total of 2 wrong MC in these 2 reports.

Finally the remaining report was wrong not due to the fact that it incorrectly matched citing papers, but because of an error in the data received by the CRIS, where the DOI given to the publication, provided wrong information about the publication, which led to the comparison of two different articles, in WoS and OpenAlex.

In this manual search, from the 40 analyzed papers, only 10% had some kind of mistake, one of the cases was due to WoS having assigned the wrong DOI to a publication, and the other one was due to the fact that we tagged the publication as citing itself. The latter is an easy fix we can do in our application so it won't happen again, while the former is a mistake that we cannot work around that easily, since it is a problem in WoS side. Finally the mistake from wrong data provided by the CRIS is out of our reach to fix, since the information in this database is provided mostly by the authors of the publications and it is prone to have some errors.

Afterwards we confirmed the reports that showed over 3 MC which represent at least 50% of the citations of the report itself. Out of 14 total reports, 13 of them had no errors, correctly identifying 82 MC. On the other hand, the remaining report was wrong, due to the fact that the publication we got from OpenAlex, was different from the one in WoS, making our system compare citation lists from two different papers. Therefore this report incorrectly gave us 25 MC. This error came from the CRIS database, which had 2 different DOIs for the same publication, and the DOI being analyzed did not match the information provided by the CRIS.

In general, we are satisfied with the system's performance as it has only encountered a few errors, primarily caused by inaccurate data. Although we can address some of these errors, there are others that are beyond the scope of our approach. As a result, there is a slight chance of misclassifying a Missing Citation as such. To mitigate the it's impact, our reports include links to both databases for any matched papers, allowing for manual verification. This process significantly reduces the effort required compared to starting from scratch, thereby saving researchers a considerable amount of time.

6 Conclusions and Future Work

This paper proposes an automated approach that leverages two indexing databases to locate missing citations in one of them. We focus on the case study of Web of Science (WoS) as the primary citation database, while utilizing OpenAlex as a reference point for comparison. Our method involves identifying all citing papers associated with a given target paper from both research databases. Subsequently, we examine whether each citing paper is indexed in WoS but not referenced in WoS as a citation, indicating the presence of a missing citation. By conducting experiments on a dataset comprising 1539 papers indexed by WoS, we successfully uncovered 696 missing citations using our approach. Although a small part of these missing citations might be wrongfully identified, as shown above, the data suggests that most of them are indeed missing citations, which proves the validity of our approach and the fact that WoS has missing information for citing publications, specially if we take into consideration that only OpenAlex was used as a baseline, while later on, we intend to add both Scopus and Google Scholar, and we expect to detect a higher amount of missing citations in WoS, as well as finding missing citations in Scopus and Google Scholar themselves.

In our analysis we detected that in the later years, the percentage of missing citations skyrocketed, which makes sense since WoS might not have had the time to process the citations of more recent publications, while OpenAlex uses data from other platforms like DBLP and CrossRef, which might have faster algorithms to find missing citations, giving OpenAlex an advantage for more recent publications. Nevertheless, this can be a problem if an author is being evaluated on their more recent publications, making this process more critical for that circumstance. It could be interesting to follow up this research and see how the results have changed, and how long it takes for the WoS algorithm to catch up to the standards of older publications, since the publications analyzed from 2021, were between 1.5 and 2.5 years old when analyzed.

One aspect of our current work that we believe can be enhanced is the matching algorithms employed for publications. Merely augmenting query parameters is not sufficient, as the methodology outlined in this paper predominantly relies on the utilization of the DOI reference system, as well as basic searches that incorporate the title, authors, publication year, and publication type to establish connections between papers and their respective citations. In our forthcoming endeavors, we aim to introduce similarity metrics to refine paper searches, thereby potentially yielding more comprehensive and thorough outcomes.

The next step will be to add other indexing databases to our approach, such as Scopus and Google Scholar (GS), not only to try to find missing citations, but also because they might bring additional citing papers that WoS and OpenAlex might have missed. Especially in the case of GS, which has a broader indexing policy than WoS and Scopus. Moreover, a preliminary analysis has already shown us that there are cases where GS has more citations than OpenAlex for the same publication.

References

1 Robert A. Buchanan. Accuracy of cited references: The role of citation databases. *College & Research Libraries*, 67(4):292–303, 2006. doi:10.5860/crl.67.4.292.

2 Alessia Cioffi, Sara Coppini, Arcangelo Massari, Arianna Moretti, Silvio Peroni, Cristian Santini, and Nooshin Shahidzadeh Asadi. Identifying and correcting invalid citations due to doi errors in crossref data. *Scientometrics*, 127:3593–3612, 2022. doi:10.1007/s11192-022-04367-w.

3 Fiorenzo Franceschini, Domenico Maisano, and Luca Mastrogiacomo. A novel approach for estimating the omitted-citation rate of bibliometric databases with an application to the field of bibliometrics. *Journal of the American Society for Information Science and Technology*, 64(10):2149–2156, 2013. doi:10.1002/asi.22898.

4 Fiorenzo Franceschini, Domenico Maisano, and Luca Mastrogiacomo. Errors in doi indexing by bibliometric databases. *Scientometrics*, 102:2181–2186, 2015. doi:10.1007/s11192-014-1503-4.

5 Miguel A. García-Pérez. Accuracy and completeness of publication and citation records in the web of science, psycinfo, and google scholar: A case study for the computation of h indices in psychology. *Journal of the American Society for Information Science and Technology*, 61(10):2070–2085, 2010. doi:10.1002/asi.21372.

6 Erwin Krauskopf. Missing documents in scopus: the case of the journal enfermeria nefrologica. *Scientometrics*, 119:543–547, 2019. doi:10.1007/s11192-019-03040-z.

7 Henk F. Moed, Judit Bar-Ilan, and Gali Halevi. A new methodology for comparing google scholar and scopus. *Journal of Informetrics*, 10(2):533–551, 2016. doi:10.1016/j.joi.2016.04.017.

8 Jason Priem, Heather Piwowar, and Richard Orr. Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts, 2022. doi:10.48550/ARXIV.2205.01833.

9 Misha Teplitskiy, Eamon Duede, Michael Menietti, and Karim R. Lakhani. How status of research papers affects the way they are read and cited. *Research Policy*, 51(4):104484, 2022. doi:10.1016/j.respol.2022.104484.

10 Nees Jan van Eck and Ludo Waltman. Accuracy of citation data in web of science and scopus, 2019. doi:10.48550/ARXIV.1906.07011.