The secrets behind Web of Science's DOI search

Junwen Zhu[1] · Fang Liu[2] · Weishu Liu[3]

Abstract

With the flourish of scientific literature, the Digital Object Identifier (DOI) is increasingly adopted in academia to uniquely identify research articles. By using Web of Science's DOI search, we find that millions of DOI names appear to begin with an alphabetic character which violates the naming rule of DOI. In this study, we try to uncover the secrets behind Web of Science's DOI search and explain this mysterious phenomenon. A similar investigation is also conducted by using the Scopus database. A few suggestions are provided at the end of this paper.

Keywords Digital Object Identifier · Article number · Web of Science · Scopus

Introduction

With the flourish of scientific literature, the Digital Object Identifier (namely DOI) is increasingly adopted in academia to uniquely identify research articles (Boudry and Chartron 2017; Gorraiz et al. 2016). Inspired by the above two studies, we try to explore how DOI is adopted in academic publishing by using the Web of Science database. However, during the initial data retrieval phase, we find that huge numbers of records can be retrieved by using the DOI search query "DO=A*" ("*" is the wildcard) occasionally.[1] Does it mean that thousands of records' DOI names begin with the letter "A"?

---

[1] The field tag "DO" stands for "DOI" and is designed to search in the DOI name field in Web of Science.

✉ Weishu Liu
wsliu08@163.com

Junwen Zhu
jwzhu@ed.ecnu.edu.cn

Fang Liu
liu_fang2014@163.com

[1] Faculty of Education, East China Normal University, Shanghai 200062, China

[2] School of Accounting, Zhejiang University of Finance and Economics, Hangzhou 310018, Zhejiang, China

[3] School of Information Management and Engineering, Zhejiang University of Finance and Economics, Hangzhou 310018, Zhejiang, China

More specifically, we find an article entitled "Single-stranded DNA and RNA origami" which was published in Science in 2017. This article's DOI name is "https://doi.org/10.1126/science.aao2648". It is strange to find that this article can be retrieved in Web of Science's advanced search platform by using the search query "DO=https://doi.org/10.1126/science.aao2648 AND DO=A*". However, none of this article's identity-related fields begin with the letter "A" as shown in Fig. 1.

How does this strange phenomenon happen? In this article, we try to uncover the secrets behind Web of Science's DOI search. For comparison, a similar investigation will also be conducted by using the Scopus database.

Analysis

An overview

In order to map the full picture of Web of Science's DOI search, we use the following queries in Table 1 to capture all records with DOI names beginning with any number between 0 and 9 or any letter between a and z (the DOI name is case-insensitive, more information about DOI please refer to http://www.doi.org/doi_handbook/2_Numbering.html). The citation index is limited to three journal indexes: Science Citation Index Expanded (SCI-EXPANDED, 1900–2017), Social Sciences Citation Index (SSCI, 1900–2017), and Arts and Humanities Citation Index (A&HCI, 1975–2017). Table 1 lists the results of the comprehensive DOI search by using Web of Science's online advanced search tool.

Since a legal DOI name should begin with the prefix "10.", it is reasonable that only the search query "DO=1*" can hit records with DOI names in Web of Science. However,

Single-stranded DNA and RNA origami

By: Han, DR (Han, Dongran)[1,2]; Qi, XD (Qi, Xiaodong)[3,4]; Myhrvold, C (Myhrvold, Cameron)[1,2]; Wang, B (Wang, Bei)[2,5]; Dai, MJ (Dai, Mingjie)[1,2]; Jiang, SX (Jiang, Shuoxing)[3,4]; Bates, M (Bates, Maxwell)[6]; Liu, Y (Liu, Yan)[3,4]; An, B (An, Byoungkwon)[6]; Zhang, F (Zhang, Fei)[3,4] ...More
View ResearcherID and ORCID

SCIENCE
Volume: 358 Issue: 6369
Article Number: eaao2648
DOI: 10.1126/science.aao2648
Published: DEC 15 2017
Document Type: Article
View Journal Impact

Document Information
Language: English
Accession Number: WOS:000417918500033
PubMed ID: 29242318
ISSN: 0036-8075
eISSN: 1095-9203

Other Information
IDS Number: FP8TH
Cited References in Web of Science Core Collection: 65
Times Cited in Web of Science Core Collection: 21

Fig. 1 An example of retrieved record by DOI search

Table 1 The DOI search in Web of Science

Rank	Search query	Number of records	Search query	Number of records
1	DO=0*	17	DO=A*	2,006,606
2	DO=1*	36,523,303	DO=B*	0
3	DO=2*	0	DO=C*	0
4	DO=3*	0	DO=D*	9
5	DO=4*	1	DO=E*	1
6	DO=5*	1	DO=F*	0
7	DO=6*	1	DO=G*	0
8	DO=7*	1	DO=H*	0
9	DO=8*	0	DO=I*	1
10	DO=9*	0	DO=J*	2
11			DO=K*	0
12			DO=L*	0
13			DO=M*	0
14			DO=N*	0
15			DO=O*	0
16			DO=P*	314,716
17			DO=Q*	0
18			DO=R*	0
19			DO=S*	44
20			DO=T*	0
21			DO=U*	66,673
22			DO=V*	0
23			DO=W*	0
24			DO=X*	0
25			DO=Y*	0
26			DO=Z*	0

Indexes=SCI-EXPANDED, SSCI, A&HCI. Timespan=1900–2017. Accessed on 6th March, 2019 from the library of Xi'an Jiao Tong University

the results demonstrated in Table 1 are quite surprising. Besides some records that can be hit by using search queries such as "DO=0*", large numbers of records can be hit by the search queries such as "DO=A*", "DO=P*", and "DO=U*".

How does this strange phenomenon happen? Are there really a lot of DOI names beginning with the letter of A or P or U? In the following sections, we try to uncover the secrets behind Web of Science's DOI search.

The DO=A* case

The field tag DO is designed to search the DOI names in Web of Science's advanced search platform. Theoretically, the search query "DO=A*" can hit any records with DOI names beginning with the letter "a" or "A" (DOI name is case-insensitive). According to Table 1, over two million records have been hit by the search query "DO=A*". Although

Table 2 The DO = ARTN* case

Rank	Search query	Number of records
1	DO = A*	2,006,606
2	DO = AR*	2,006,606
3	DO = ART*	2,006,606
4	DO = ARTN*	2,006,606
5	DO = ARTN *	2,006,606

Indexes = SCI-EXPANDED, SSCI, A&HCI. Timespan = 1900–2017. Accessed on 6th March, 2019 from the library of Xi'an Jiao Tong University

a previous study has reported a limited number of illegal DOI names in Web of Science (Huang and Liu 2019), it is still unbelievable to find two million illegal DOI names in three core journal citation indexes of Web of Science.

After checking the bibliographic data of the top one thousand most cited papers, we find all of these top cited papers' DOI names are legal (beginning with the prefix "10.") and some records' DOI field are empty. However, none of the DOI names of these top cited papers begin with the letter "a" or "A".

In order to investigate the secrets behind Web of Science's DOI search, we obtain the results of Table 2 through hundreds rounds of trial and error. The search queries listed in Table 2 become stricter gradually. The result of a latter query should be a subset of the previous one. Since the numbers of records hit by these five queries are equal, we can conclude that the records hit by all these five queries are identical. That is to say, two million records can be hit by the search query "DO = ARTN *" (Please note a blank exist between "ARTN" and the wildcard "*"). We guess the abbreviation "ARTN" stands for "article number" (article number is an identifier of a document used by some journals in addition to, or instead of, page number). Besides, we hold the opinion that the field tag "DO" may not only search in the DOI name field but also in the "article number" field under the circumstance where no public available field tag has been provided to search in the "article number" field. Although the Clarivate Analytics (owner of Web of Science) has warned users not to confuse "DOI" with "article number" (https://support.clarivate.com/Scientific andAcademicResearch/s/article/Web-of-Science-Core-Collection-Article-number-used-instead-of-pagination?language=en_US), is it possible that Clarivate Analytics itself confuses these two fields?

Since Web of Science's online advanced search platform hasn't provided a field tag to search in the article number field yet. We download the bibliographic data of one thousand top cited papers in Web of Science to verify our guess. All these papers have article number values, however, none of these values begin with the prefix "ARTN". If we use both the article number and the accession number[2] to search in Web of Science (DO = "Article number accessed from Web of Science" AND UT = "Accession number"),[3] no record can be hit for these highly cited papers. However, after we add the prefix "ARTN" before the article number provided by the database and then use both the adjusted article number and the accession number to search, all these one thousand papers can be retrieved.

[2] All records in Web of Science have an accession number which can be used to uniquely identify a specific record in Web of Science.

[3] UT is the field tag designed for Web of Science to search in the accession number field.

Table 3 The DO=PII* case

Rank	Search query	Number of records
1	DO=P*	314,716
2	DO=PI*	313,992
3	DO=PII*	313,992
4	DO=PII *	313,992

Indexes=SCI-EXPANDED, SSCI, A&HCI. Timespan=1900–2017. Accessed on 6th March, 2019 from the library of Xi'an Jiao Tong University

Table 4 The DO=PMID* case

Rank	Search query	Number of records
1	DO=P*	314,716
2	DO=PM*	724
3	DO=PMI*	724
4	DO=PMID*	724
5	DO=PMID *	724

Indexes=SCI-EXPANDED, SSCI, A&HCI. Timespan=1900–2017. Accessed on 6th March, 2019 from the library of Xi'an Jiao Tong University

Based on the DO=A* case, we think that the field tag "DO" also searches in the article number field. This point has not been explicated by the database provider and may confuse many users. Besides, Web of Science needs users to add the prefix "ARTN" manually before many records' article numbers when searching by the article number.

The DO=P* case

Based on Table 1, about 0.31 million records can be captured by the search query "DO=P*". By trial and error, we obtain the results of Tables 3 and 4. Table 3 lists the search results of "DO=PII *" and Table 4 shows the search results of "DO=PMID *".

According to Table 3, over 0.31 million records can be retrieved by using the search query "DO=PII *". For the case of "DO=PII *", we download the bibliographic data of the top one thousand most cited papers. Different from the case of "DO=A*", most records' article numbers begin with the prefix "PII". Surprisingly, 12 records' article numbers do not begin with this prefix in our sample. The left 988 records can be retrieved by using the joint search of article number and accession number (that is DO="Article number accessed from Web of Science" AND UT="Accession number"). However, the left 12 records whose article numbers do not begin with the prefix "PII" cannot be retrieved by using the joint search of article number and accession number directly. More surprisingly, if we add the prefix "ARTN" before these 12 records' article numbers, the joint search of article number and accession number works. That is to say, for some records, the DOI search works for the article number field when there exists a prefix such as "ARTN" before an article number value. We manually collect these 12 records' article numbers from the corresponding publishers and find that all these records have article numbers beginning with the prefix "PII". The joint search of article number and accession number works by

Table 5 The DO = SICI* case

Rank	Search query	Number of records
1	DO = S*	44
2	DO = SI*	44
3	DO = SIC*	44
4	DO = SICI*	44
5	DO = SICI *	44

Indexes = SCI-EXPANDED, SSCI, A&HCI. Timespan = 1900–2017. Accessed on 6th March, 2019 from the library of Xi'an Jiao Tong University

Table 6 The DO = UNSP* case

Rank	Search query	Number of records
1	DO = U*	66,673
2	DO = UN*	66,673
3	DO = UNS*	66,673
4	DO = UNSP*	66,673
5	DO = UNSP *	66,673

Indexes = SCI-EXPANDED, SSCI, A&HCI. Timespan = 1900–2017. Accessed on 6th March, 2019 from the library of Xi'an Jiao Tong University

using the newly manually collected article numbers. That is to say, Web of Science has collected and stored more than one article number values for some records.

724 records can be retrieved by using the search query "DO = PMID *". Table 4 shows the detailed search results. We download the bibliographic data of all these records and import them into Excel. We find that over 90% of these records' article numbers begin with the prefix "PMID", however, 55 records' article numbers do not. That is to say, Web of Science has collected and stored more than one article number values for these 55 records.

For all the 669 records with the article numbers beginning with the prefix "PMID", the joint search of article number and accession number can be used to retrieve these records. Similar to the previous finding, the joint search of article number and accession number doesn't work for article numbers not beginning with the prefix "PMID". However, if we add the prefix "ARTN" before these article numbers, the joint search takes effects again.

The DO = S* case

We also find some records can be retrieved by the search query "DO = S*". By hundreds of trial and error, we find the search query "DO = SICI *" returns the same result as "DO = S*". More details are demonstrated in Table 5.

44 records are retrieved by using these search queries. We download the bibliographic data and import them into Excel. The article numbers of these records all begin with the

prefix "SICI". All these records can be retrieved by the joint search of article number and accession number.

The DO = U* case

Similarly, many records can be retrieved by the search query "DO = U*". Through hundreds times of trial and error, we obtain the results of "DO = U*" search as demonstrated in Table 6. Both the search queries of "DO = U*" and "DO = UNSP *" retrieve the same results. We download the bibliographic data of the top one thousand most cited records among these 66,673 papers.

All these one thousand most cited records' article numbers begin with prefix. Most of them are with the prefix "UNSP", however, some records' article numbers are with the prefix "PII". That is to say, for some records, Web of Science has collected at least two article number values, one begins with "UNSP", the other with "PII". All these records can be retrieved by the joint search of article number and accession number, no matter whether the article number begins with the prefix of "UNSP" or "PII".

Other cases

Apart from the abovementioned cases, we still find that some other search queries can also retrieve limited number of records as demonstrated in Table 2. However, after a careful checking, these cases are different from the above four. All the left cases are due to the illegal naming of DOI names (not beginning with the prefix "10."). A more detailed discussion can be found in Huang and Liu (2019).

A check on Scopus

In order to make this research more comprehensive, we also investigate the DOI search tool in Scopus. Different from Web of Science, the Scopus online search platform provides the article number search tool by the field tag "artnum". Similar to the Table 1, we also retrieve related results by using the field tag "DOI" in Scopus.

Generally, only the search query "DOI(1*)" should obtain records. However, many other search queries, as demonstrated in Table 7, can also get many records. We download the bibliographic data of the records retrieved by these search queries, except the search query "DOI(1*)". Over 6500 records can be obtained. After a careful checking on the DOI names, we find that all these DOI names are illegal. That is to say, all the 6500 records' DOI names do not begin with the number "1", let alone the prefix "10.". Besides, according to a previous study, over 1500 records' DOI names in Scopus begin with "10" but not "10." (Huang and Liu 2019).

Conclusion and discussion

By using Web of Science's online advanced search platform, we try to uncover the secrets behind Web of Science's DOI search. We are surprised to find that millions of records can be retrieved by Web of Science's DOI search queries "DO = A*", "DO = P*", "DO = S*", and "DO = U*". Although some illegal DOI names existed violate the basic naming rule

Table 7 The DOI search in Scopus

Rank	Search query	Number of records	Search query	Number of records
1	DOI(0*)	272	DOI(A*)	519
2	DOI(1*)	50,404,202	DOI(B*)	5
3	DOI(2*)	50	DOI(C*)	211
4	DOI(3*)	34	DOI(D*)	298
5	DOI(4*)	5	DOI(E*)	69
6	DOI(5*)	8	DOI(F*)	10
7	DOI(6*)	0	DOI(G*)	62
8	DOI(7*)	3	DOI(H*)	110
9	DOI(8*)	0	DOI(I*)	294
10	DOI(9*)	51	DOI(J*)	436
11			DOI(K*)	320
12			DOI(L*)	75
13			DOI(M*)	61
14			DOI(N*)	28
15			DOI(O*)	1043
16			DOI(P*)	161
17			DOI(Q*)	0
18			DOI(R*)	753
19			DOI(S*)	1726
20			DOI(T*)	30
21			DOI(U*)	5
22			DOI(V*)	1
23			DOI(W*)	12
24			DOI(X*)	1
25			DOI(Y*)	2
26			DOI(Z*)	1

Data source: Scopus. Accessed on 6th March, 2019 from the library of Xi'an Jiao Tong University

that DOI names should start with the prefix "10.", the illegal DOI names could not explain millions of records hit by the above four search queries.

Moreover, even if Web of Science has warned us not to confuse "DOI name" with "article number" explicitly (https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Core-Collection-Article-number-used-instead-of-pagination?language=en_US), we still find that Web of Science's DOI search tool also searches in the "article number" field. That is to say, Web of Science itself has mixed the "DOI search" with "article number search" without a clear reminder.

The majority of the records retrieved in this study do not belong to the incorrect DOI names as reported in previous studies (Franceschini et al. 2015; Zhu et al. 2019), but the mixture of "DOI search" and "article number search" found in this study will also confuse many users. Besides, for many records in Web of Science, users should manually add the prefix "ARTN" before the "article number" values in order to be retrieved by the DOI search. Comparatively, records with article numbers beginning with the prefix "PII",

"PMID", "SICI" and "UNSP" in Web of Science can be retrieved by the DOI search directly.

In order to deal with the situation found in this study, Web of Science should at least clarify that DOI search also retrieves the article number field explicitly. The users of Web of Science should also be informed about the manual adding of prefix "ARTN" before the article number for some records. Of course, the final solution is to create a new search tag to search in the article number field only and limit the DOI search in the DOI field exclusively (Scopus has already provided both the online DOI search and article number search). For academic users of Web of Science, the limitations of Web of Science's DOI search should be bore in mind. The search query of DOI search, especially a query with a wildcard, should be double checked.

Franceschini, Maisano, and Mastrogiacomo have conducted a series of studies to probe the errors in bibliometric databases (Franceschini et al. 2015, 2016a, b). The authors' team has also probed several types of errors or limitations of Web of Science regarding the funding acknowledgments information (Tang et al. 2017), language bias (Liu 2017), missing author address (Liu et al. 2018) and DOI errors (Zhu et al. 2019). We hope the database providers can work with the scientific community together to ensure a more reliable and user-friendly data source.

References

Boudry, C., & Chartron, G. (2017). Availability of digital object identifiers in publications archived by Pub-Med. *Scientometrics, 110*(3), 1453–1469. https://doi.org/10.1007/s11192-016-2225-6.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2015). Errors in DOI indexing by bibliometric databases. *Scientometrics, 102*(3), 2181–2186. https://doi.org/10.1007/s11192-014-1503-4.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016a). Empirical analysis and classification of database errors in Scopus and Web of Science. *Journal of Informetrics*, *10*(4), 933–953. https://doi.org/10.1016/j.joi.2016.07.003.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016b). The museum of errors/horrors in Scopus. *Journal of Informetrics*, *10*(1), 174–182. https://doi.org/10.1016/j.joi.2015.11.006.

Gorraiz, J., Melero-Fuentes, D., Gumpenberger, C., & Valderrama-Zurián, J. C. (2016). Availability of digital object identifiers (DOIs) in Web of Science and Scopus. *Journal of Informetrics, 10*(1), 98–109. https://doi.org/10.1016/j.joi.2015.11.008.

Huang, M., & Liu, W. (2019). Large numbers of easily identifiable illegal DOI names still exist in Scopus. Working paper.

Liu, W. (2017). The changing role of non-English papers in scholarly communication: Evidence from Web of Science's three journal citation indexes. *Learned Publishing*, *30*(2), 115–123. https://doi.org/10.1002/leap.1089.

Liu, W., Hu, G., & Tang, L. (2018). Missing author address information in Web of Science—An explorative study. *Journal of Informetrics*, *12*(3), 985–997. https://doi.org/10.1016/j.joi.2018.07.008.

Tang, L., Hu, G., & Liu, W. (2017). Funding acknowledgment analysis: Queries and caveats. *Journal of the Association for Information Science and Technology*, *68*(3), 790–794. https://doi.org/10.1002/asi.23713.

Zhu, J., Hu, G., & Liu, W. (2019). DOI errors and possible solutions for Web of Science. *Scientometrics, 118*(2), 709–718. https://doi.org/10.1007/s11192-018-2980-7.