Science Communication: Research Productivity and Data Analytics using Open Source Software

Vanita Khanchandani

Central Library

IIT Delhi

Week: 03

Lecture: 09

Health Science Database: PubMed

Dear Learners, so welcome again to the course on Science Communication, Research Productivity and Data Analytics using Open Source Software.

Today we will learn about Health Science Database PubMed, how we can extract the data from PubMed, and how we can do the search in PubMed. So, one can reach PubMed using login Pubmed.ncbi.nlm.nih.gov.

PubMed is a platform that indexes journal articles and more back to 1947. It covers the areas of medicine, nursing, dentistry, veterinary medicine, healthcare systems, preclinical sciences and related areas. PubMed is developed and maintained by National Library of Medicine (NLM) and National Center for Biotechnology Information (NCBI), both at National Institute of Health (NIH) in Bethesda, Maryland.

As of December 2020, PubMed contains over 36 million records. PubMed is a platform that contains Medline, PubMed Central and additional PubMed records. So, one can sign in PubMed using eRA-Commons, Google account, ORCID, Login.gov, Microsoft, and NIH account. For document searching, if you write a keyword, it can give you the related keywords like for example, here we are writing heart and it is giving you the related keywords like heart failure, heart disease, heart rate, congenital heart. So, we will search using the keyword heart disease. One can save and email the records. So, we have different kind of NCBI filters, like you can refine document according to abstracts, full text, free full text and all, article attributes like associated data, article types like books and document clinical trials, meta-analysis, randomized control trials, reviews, systematic review, and according to publication date, one year, five year, ten year, you can add custom range, and in addition to these filters, we can have filters according to species, article language, sex, age, others. For example, here we have got the first article as heart

disease in children for searching keyword heart disease. So, in the title, you have citation information, then PMID number.

PMID means PubMed-ID and you can cite and share. Articles in PubMed have permanent links which you can share on Twitter, Facebook and for citing, it has citation Style AMA, APA, MLA and LM. You can also email the records. So, in the document result page, you can see the first document. On the top of the document, it has the citation information. It will tell you what kind of article it is, review or article, research type, author's name, DOI, abstract, keywords, similar articles at the bottom and on the right hand side, you can share the article using permalink, abstract, MeSH terms, all are there. How to cite the article, and where the article is available. After this document result, we are moving to how we can save the results, or how we can export the results in PubMed. So, you can select for saving the result, you can have all the results on this page, all results, or the selected results. For all results, you can have citations exported or saved for the first 10,000 citations can be saved in your file and you can create summary, PubMed ID, abstract text or CSV file. You can export the CSV file and how the CSV file looks like. So, it has PMID, PubMed ID, title, author, citation information, the name of the book or journal, year, date, PMC ID, DOI, etc.

So, there are various search tags used in PubMed for building a query. So, these are listed in this slide and how to use these tags in basic and advanced search or query building, we will see in our next slide using the search builder. So, in PubMed advanced search builder, you can search according to different fields. For example, I want to search according to the date of publication. So, in this example, we are taking 2020/01/01 to 2023/01/22, with the key term and we are adding the key term with AND.

So, now our query is COVID-19 AND the date of publication is 2020/01/01 to 2023/01/22. And you can see that in history search details are coming and we have retrieved 3,34,294 articles. So, you can use the date of publication or publication date as Tag [DP], and always use square brackets [] while using search tags. So, we can use Boolean operators for this COVID-19 AND the date of publication. So, if you see the results, all the results pertaining to March 2020 are there, and you can even combine date ranges by putting colon first date to the present. See for example, 2020:2021, we can combine both the dates of publication using colon.

See in first result, we have August 2020 and in another March 2021. We can build the query using the date of publication as a search tag and we can also truncate the search term. Truncation means using wildcard asterisk (*). So, breastfeed* can be breastfeeding, breastfeed. So, in this example, we are using title and abstract search tag with date of publication. So, either in title or in abstract, you will find breastfeeding and that the date of publication is 2020, 1st January 2020. So, here you see in e-publication, it is 2020, but

12th June, see this is e-publication. In another, it's 2020, the 1st January. So, you can see in both, breastfeeding is in the abstract.

Now we come to use Boolean operators. Boolean operators AND, OR, NOT can be used using parentheses to nest the concept and can be processed as a unit. So, you can put in (lead poison AND humans), or (lead poison OR humans), you can put in the bracket and humans even. Okay. So, that lead poison will be processed as only a concept. So, in PubMed, search is always done from left to right when we are using Boolean operators. So, see in the first result, lead poison AND humans are there. So, in this, we have got lead poisoning and in another lead poisoning in children and adults, lead poisoning in US children. And when we are writing lead poison, NOT humans, so we are getting just lead poison. Okay. And if we are writing lead poison OR humans, then we are getting either lead poison or humans. So, we are here getting the lead.

And in PubMed, Proximity search can be used by writing search terms in double quotes, then search tag field, then the number of terms and N stands for number of terms. So, for example as I write here, I have given “COVID-19 virus”[Title:~2]. Okay. And in another slide, I have given “COVID-19 virus”[Title/Abstract:~2]. So, see in both COVID-19 virus in Title, or in Title/Abstract in both. COVID-19 virus has only two letters gap, two words gap, two letters gap. Okay. Here in Abstract, in this Title here in another slide, Title or Abstract. Okay. So, the quote is not found. If you want to search all terms together without having any distance, then put a 0. Okay. If you don't want to have any separation in terms, then put 0. Like “Cognitive impairment in multiple sclerosis”. So, here I have put TIAB means Title/Abstract search ~0. So that this keyword comes together without having any words.

So, now we move to a demo session of PubMed so that you will understand how we can conduct the search in PubMed. First, let me see how we can log in. I have already told you that you can create an account in PubMed using these, any of these. And I am using ORCID. So, now we can do our search.

So, I have already told you that if you just start writing small keywords, then PubMed gives you the relative keywords and I'm searching for heart disease. So, here are your results. You can filter these results and can even email the results, or send to my clipboard, my bibliography collection, and citation managers. So, I'm telling you how you can export the information. So, all the results.

I have already told you that only 10,000 results can be exported at a time. I'm selecting a CSV file, or you can even select. Few results on this page I am selecting. I am selecting five results, the first five results. And I want to export these. Select file selection, CSV file. So, see this is the data, PMID, title, authors, citation information, first author, then journal or book, publication year, create date, PMC ID, NIHMS ID, and DOI. This is all

we have got in the CSV file, and you can even build an advanced query by clicking in add, date-publication 2020/01/01 to 2023/01/23. Now you can conduct the search. So, all articles which are published from 1st January 2020 to 23rd January 2023 are there on COVID-19.

Okay. This is for query building. We have built a query using advanced search features. But if we don't want to build a query using advanced search, we can do this and conduct the simple basic search by just writing the date of publication and connecting them using Boolean operators. AND for the search tag, you have to put that tag in square brackets. So, I am writing the date of publication or publication date as [BP], and now conducting the search. So, all articles published during the 1st January 2020 are there. You can see on COVID-19 and you can even combine this using double colon. This kind of query we have earlier built using advanced search for date range, you can combine the query date of publication or date range from past to present using colons.

And you can do it. You can conduct Boolean search by nesting them. So, (influenza OR birds). So, you can see either the influenza term is coming or birds are coming. You can. With this influenza coming, no bird is there. And we are putting (influenza AND birds). For and also you can do. So, you can nest the terms.

And for proximity operators, you can also. But putting all this. Now we are putting it on. So, avian influenza. So, for proximity operators, you can see in the title “avian influenza virus”[Title:~2]. Okay. Can have a gap, or two words. And you can also write [Abstract:~2], if you want to search. Avian influenza virus is in 2 proximity. In this proximity avian influenza viruses. You can see and if you do influenza a virus, some lateral gap is there in avian influenza and in between the name of that virus is there and then virus. Okay. If you don't want any gap, you can put zero, ~0. So, no gap is there in avian influenza virus in all the Search. In all Titles there is no gap.

So, thank you. This is from my side. You can write to us on the Discussion Forum for your queries. Thank you.