

Literature search, search engines

Wuhao Wang

2021/8/21

Abstract

ABSTRACT: In the context of higher education, searching for support material is always necessary in doing research and writing a paper. This paper will briefly introduce how search engines work. After that, some basic strategies and search tools will be mentioned. This paper will also illustrate why those strategies would work by correlating to how search engines work.

Introduce

The availability of scientific information is growing faster and faster^[1]. Although that means Researchers and students have way easier access to the material they used to dream about, they need extra effort to locate their needed information from the growing scale of database. Under such circumstances, advanced searching skills and strategies could come to rescue and make the searching process more efficient.

Literature search is about searching several key words and setting limitations in search engines and then finding related information. In this process, researchers usually need to narrow the results several times.

Search engines work like a mapping tool, they collect the information and classify them by various standards. Once people need information, they can search keywords in search engines and get feedback.^[2]

1 Web crawler

Before diving into the topic of how search engines work, what is a web crawler should be mentioned. The information on the Internet is stored on countless servers. Any search engine that wants to answer the user's search first needs to store the web page on its own local server, which depends on the web crawler^[3]. Crawlers send requests to various websites and store the returned pages. To find more websites, crawlers will visit available links in every found websites. In this way, web engines like Google will have access to all web pages on the internet.

2 How search engines work

Like using a dictionary, search engines also have its own dictionary: inverted list. Inverted list will store words and its corresponding location just like what hash map does^[4]. When a word is fed into a search engine, the search engine will check the list and return a score of

every possible website. After that, other factors like update date, population and even business relationship will be considered as a part of score^[5]. Finally, the search engine would have a rank list and then present the result with this list.

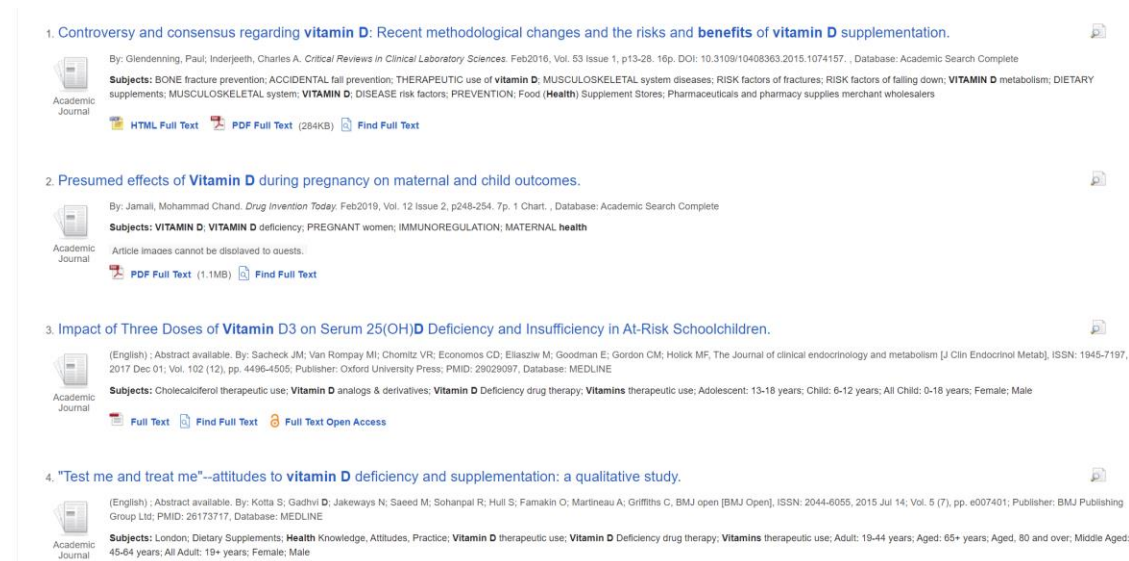
From the sorting process, a conclusion can be extended: the more keywords typed, the less accurate the result. Because the score for each link is the sum of all the keywords' scores (ignoring other factors), if a particular keyword scores particularly high, it is likely that the link will outweigh the rest. In picture1, the third link, which contains only vitamin D, scored higher than the fourth link, which seemed to be more relevant. In the following section, search strategies will be presented.

3 Search skills and strategies

There are two important aspects about literature search, one is to write down the keywords, the other is to choose the search engine^[6].

3.1 narrow keywords

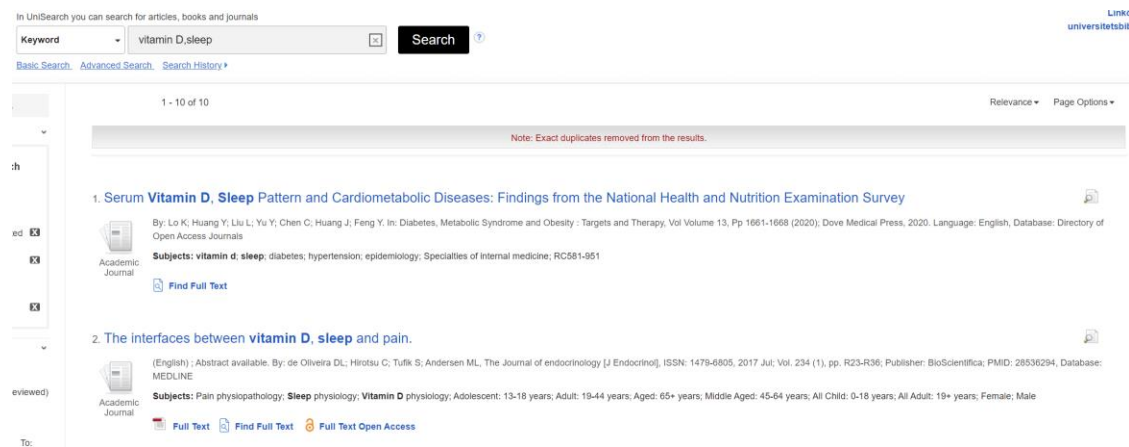
When choosing keywords, we usually choose words that can reflect the characteristics of the topic itself. However, in many cases, a topic will have many branch topics, which leads to the keyword search results are more likely to be relevant rather than helpful. In this case, the method of narrowing the scope of keywords should be adopted. For example, when searching for the topic "vitamin D benefits human health", researchers may choose the keywords "vitamins" and "benefit health". The picture below shows the result from Liu library search system.



picture1 "vitamins" and "benefit health"

As the picture1 shows, the previous four literatures are only relevant, but not hitting the point. Therefore, refining the keywords is the strategy that should be adopted, and it is also helpful regarding sorting collected material. Divide the topic into several parts. Health consists of physical and mental health, and one of the standards of physical health is sleeping

quality. So new keywords could be “vitaminD ” and “sleep”. As shown in Figure 2, the second article hits the right keywords. So researchers can combine all the results together and form their conclusion about “how vitaminD benefits health”.



picture2 “vitamins” and “sleep”

3.2 find appropriate search engine

Many search engines have their own unique databases, and sometimes researchers need to search specific databases to get the information they need. In China, academic papers can be searched on “CnKI” or “Wanfang”, daily life knowledge and daily news can be searched on “Baidu”, and computer industry related articles can be searched on “CSDN”^[7]. For example, when researchers want some information about M1 chip, if they search “M1 chip” in “Baidu”, they may get advertisement from online shop selling Apple computers in the first link whereas in “CSDN”, they are more likely to get information given by experts.

Reference

- [1] DONATO H M. How to do an effective literature search? The use of push and pull strategies.[J]. Acta Medica Portuguesa, 2013, 26(4): 471–475.
- [2] MORAN M, HUNT B. Search Engine Marketing, Inc. : driving search traffic to your company’s website, third edition.[M]. IBM Press, 2015.
- [3] SHARMA A K, SHRIVASTAVA V, SINGH H. Experimental performance analysis of web crawlers using single and Multi-Threaded web crawling and indexing algorithm for the application of smart web contents.[J]. Materials Today: Proceedings, 2021, 37(Part 2): 1403–1408.
- [4] WIKIPEDIA. Inverted index[J]. Wikipedia. Access date: 2021-08-21.URL: https://en.wikipedia.org/wiki/Inverted_index
- [5] PATIL S P, PAWAR B V. REMOVING NON-RELEVANT LINKS FROM TOP SEARCH RESULTS USING FEATURE SCORE COMPUTATION. [J]. Bulletin of Pure & Applied Sciences-Mathematics, 2018, 37E(2): 311–320.

[6] LOO J L. Building Users' Search Skills for Systematic Reviews: Development of Self-Directed Learning Through Qualitative Synthesis of Guidelines.[J]. 2020.

[7] BAIDU. how to search material[J]. Baidu. Access date: 2021-08-21.URL:
<https://jingyan.baidu.com/article/fdbd4277a0c8efb89f3f4847.html>