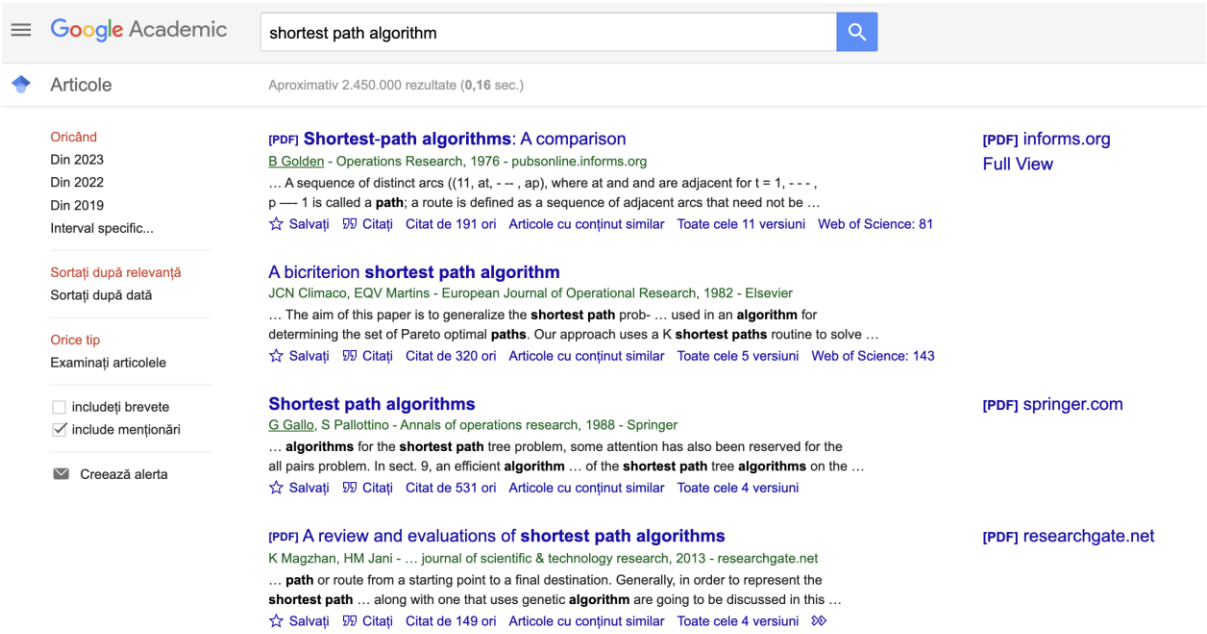
Lab 1 Algorithms and Data

When we want to solve a problem, it is useful if we first look at how other people have approached that problem. For this we can use scientific databases, such as Google Scholar, SCOPUS, or Web of Science. Algorithms also process data and it would be useful if we had a set of input data on which to test different algorithms in terms of performance (running time) and accuracy of the results obtained.

Google Scholar

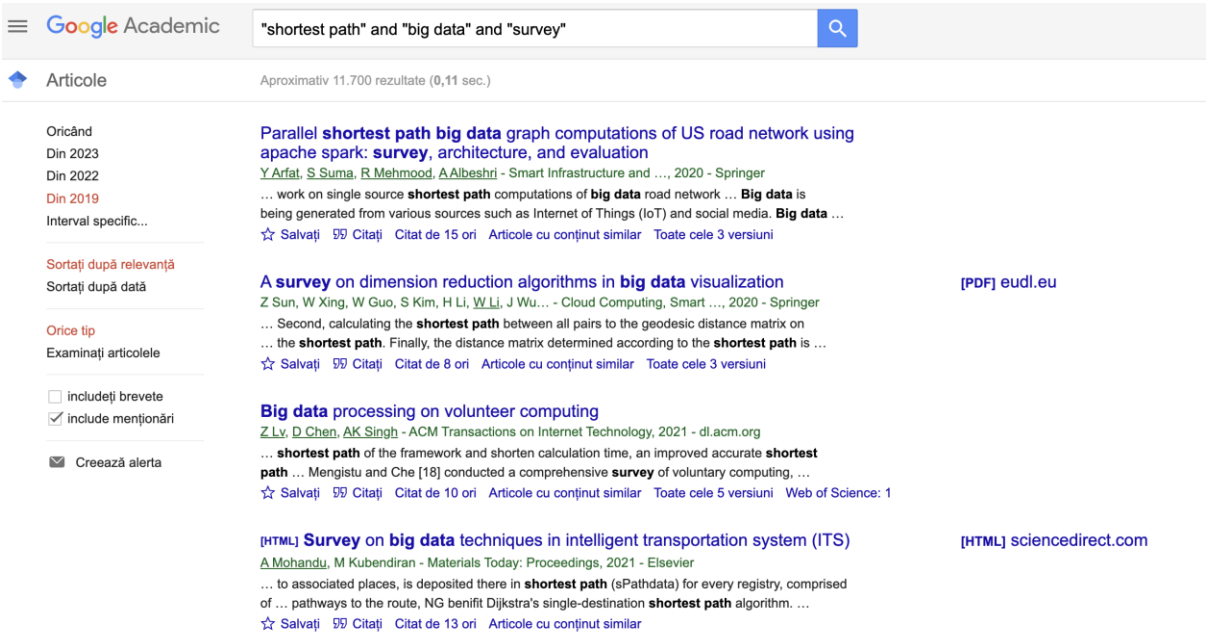
Google Scholar (https://scholar.google.com/) is a search engine for specialist (scientific) articles. In the search bar we can write keywords that can be joined by logical operators.

Example for "shortest path algorithm"



We can see that the first article was published in 1976. From the title we can deduce that the article compares several algorithms for calculating the shortest path.
On the right side of the article we have the link to the open-access PDF version of the article (if there is an open-access version). The open-access variant exists for the first, third and fourth articles, but not for the second.
On the left side we can refine the results by choosing a specific range of publication year, type, and whether to order by date or relevance.

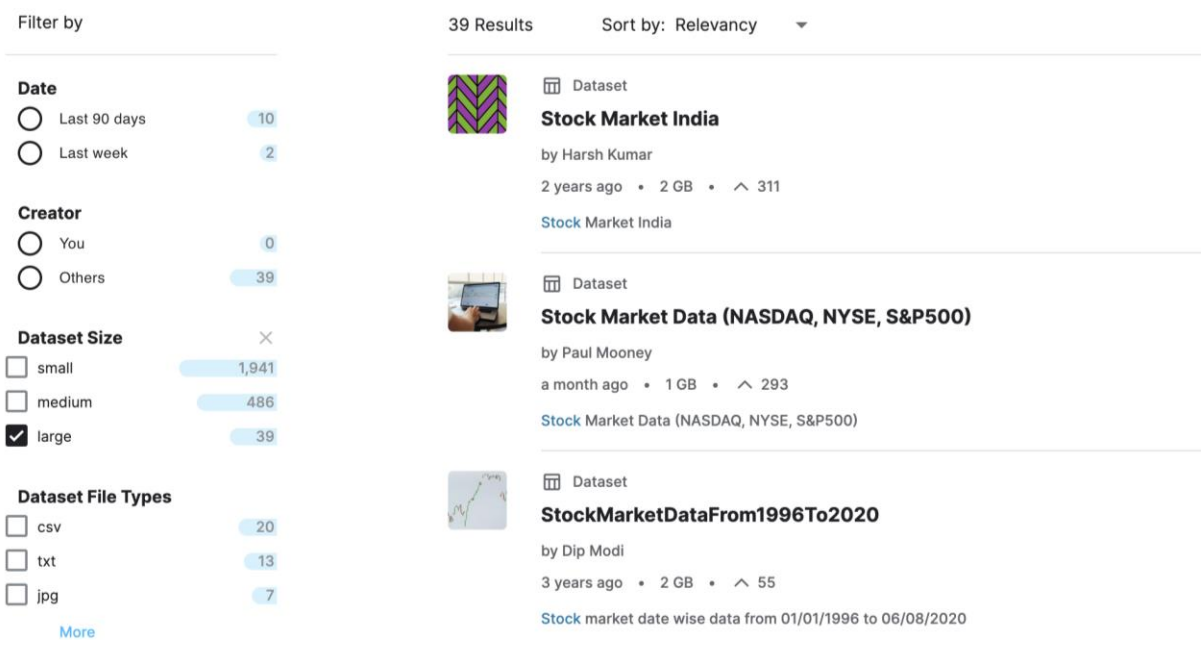
Example for “ “shortest path” and “big data” and “survey” “



Notice that “**and**“ is a logical operator which makes sure that all complex keywords will be found in the same article. You can also use “**or”, “not”**.

Kaggle

Kaggle (https://www.kaggle.com/) is a platform that hosts datasets. You can use the search bar on the top right to enter keywords. After the search, on the right side we have different filters. A useful filter for our matter is "Dataset Size" where we should choose the "large" option. Below we have a print-screen for the "stock" search and the "large" filter.



HuggingFace

https://huggingface.co/

Lab/Homework exercises

1. Choose a problem you are interested in.
2. Look for a survey that cites articles that address the same issue.
3. Search for a data set that is related to the proposed problem.