

Effective searching procedures for relevant references on quantitative topics

Cristian Homescu

May 2021

This document describes various approaches that were found to be effective within context of searching and identifying relevant references on quantitative topics.

Table of Contents

1	Search through Google Scholar	2
2	Use academic search engines.....	2
3	Search through academic research databases.....	2
4	Use Dimensions.ai online search engine	3
5	Use Semantic Scholar online repository.....	3
6	Use PapersWithCode online repository	3
7	Use ConnectedPapers.com online search engine (network-based)	4
8	Search arXiv online repository	4
9	Search SSRN online repository	5
10	Use JabRef to search within bibliography (.bib) files	6

1 Search through Google Scholar

Some relevant links: [link1](#), [link2](#), [link3](#), [link4](#), [link5](#)

One can also leverage the “Cited by” found under each article. See an example below:

The screenshot shows the Google Scholar interface with search results for 'Articles'. The top bar indicates '10 results (0.02 sec)'. On the left, there are filters for 'Any time', 'Since 2020', 'Since 2019', 'Since 2016', and 'Custom range...'. Below these are sorting options: 'Sort by relevance' and 'Sort by date' (selected). There are also checkboxes for 'include patents' and 'include citations', and a 'Create alert' button. The main content area displays three search results. The first result is 'A review of two decades of correlations, hierarchies, networks and clustering in financial markets' with a link to '[PDF] arxiv.org'. The second result is 'Learning Undirected Graphs in Financial Markets' by J.V.M. Cardoso and D.P. Palomar, with a link to '[PDF] ijeeee.org'. The third result is 'Understanding Machine Learning for Diversified Portfolio Construction by eXplainable AI' by M. Jaeger, S. Krügel, D. Marinelli, and J. Papenbrock, with a link to '[PDF] isfa.fr'. Each result includes a brief abstract and a 'Cited by' link.

Articles 10 results (0.02 sec)

Any time
Since 2020
Since 2019
Since 2016
Custom range...

Sort by relevance
Sort by date

☐ include patents
☐ include citations

Create alert

A review of two decades of correlations, hierarchies, networks and clustering in financial markets
☐ Search within citing articles
Articles added in the last year, sorted by date **Abstracts** Everything

Learning Undirected Graphs in Financial Markets [PDF] arxiv.org
J.V.M. Cardoso, D.P. Palomar - arXiv preprint arXiv:2005.09958, 2020 - arxiv.org
23 days ago - We investigate the problem of learning undirected graphical models under Laplacian structural constraints from the point of view of financial market data. We show that Laplacian constraints have meaningful physical interpretations related to the market index ...
☆ 99 All 4 versions

[PDF] Stock Movement Modeling Based on the Analysis of Negative Correlation [PDF] ijeeee.org
K Chansilp, K Kerdprasop, P Chuaybamroong... - ijeeee.org
69 days ago - This research presents the data-driven modeling method to derive a combined trading model from the analysis of negative correlations among the top-five active stocks from each sector of the Thailand stock market. The negative movements are computed from ...
☆ 99

Understanding Machine Learning for Diversified Portfolio Construction by eXplainable AI [PDF] isfa.fr
M Jaeger, S Krügel, D Marinelli, J Papenbrock... - Available at ..., 2020 - papers.ssrn.com
131 days ago - In this paper, we construct a pipeline to investigate heuristic diversification strategies in asset allocation. We use machine learning concepts ("explainable AI") to compare the robustness of different strategies and back out implicit rules for decision ...
☆ 99 All 3 versions

2 Use academic search engines

[Top list of academic search engines](#) includes:

- [Google Scholar](#)
- [Microsoft Academic](#)
- [Bielefeld Academic Search Engine BASE](#)
- [CORE](#)
- [Science.gov](#)
- [Semantic Scholar](#)
- [Baidu Scholar](#) (NOTE: as of now interface only in Chinese)

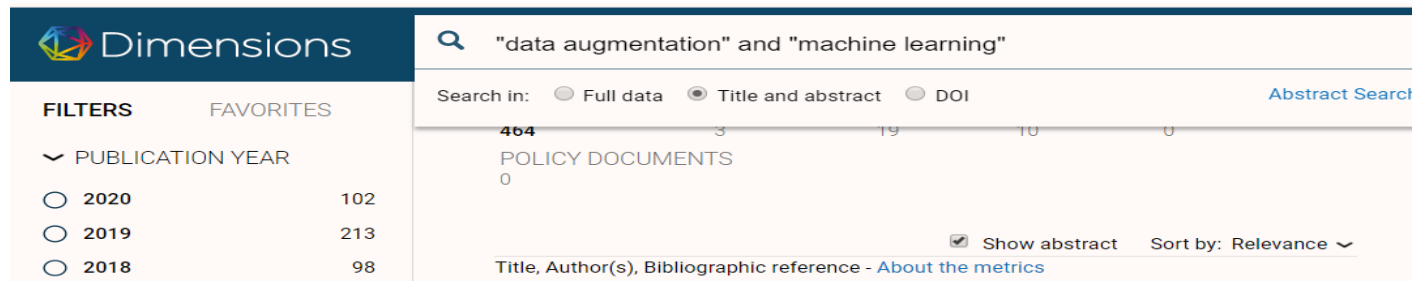
3 Search through academic research databases

[Top list of academic research databases](#) includes:

- [Scopus](#)
- [Web of science](#)
- [PubMed](#)
- [ERIC](#)
- [IEEE Xplore](#)
- [ScienceDirect](#)
- [Directory of Open Access Journals DOAJ](#)
- [JSTOR](#)

4 Use Dimensions.ai online search engine

A very recent addition to searching through publications is Dimensions.AI (see [link](#)). Searching can be based on keywords, years, title, abstract, etc.



5 Use Semantic Scholar online repository

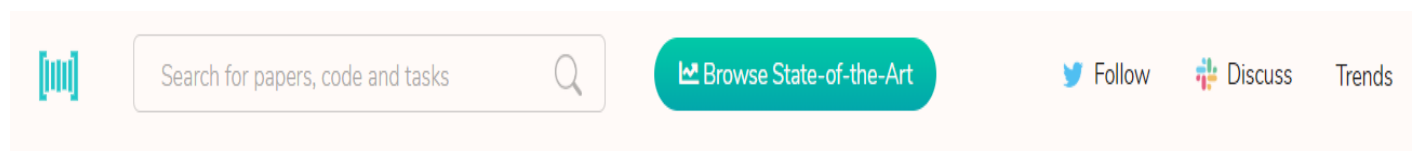
[Semantic Scholar](#), which is based at Allen Institute for AI, combines Natural Language Processing, Machine Learning, Human Computer Interaction, and Information Retrieval:

- Search through almost 190 million research works from all fields of science, with automatic extraction of abstracts, tables, figures, and citations
- Quantify impact of an article with statistics that highlight the volume and intent of the paper's citations, illuminating the influence of the research
- Provides links to related GitHub repositories, clinical trial data, presentations, videos, and other supplemental content to help reproduce the results of a paper and put it in context



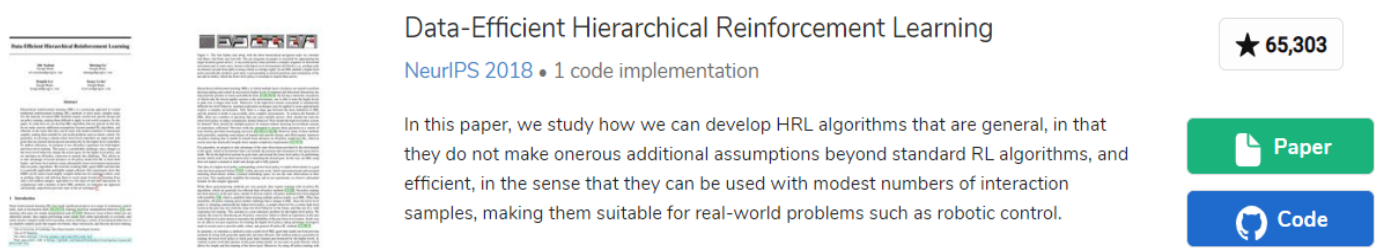
6 Use PapersWithCode online repository

[Papers with code](#) highlights trending Machine Learning research and corresponding code(s) to implement it.



See in particular the [State of the art](#) and the [Methodology](#) subsections.

Below is a snapshot example of an article with both PDF and code:



7 Use ConnectedPapers.com online search engine (network-based)

The [website](#) (released in June 2020) is a visual tool to analyze more than 50,000 research works (articles, working papers, etc.) and selects the few dozen with the strongest connections to the original research work.

In the graph, papers are arranged according to their similarity, based on the concepts of Co-citation and Bibliographic Coupling. That means that even papers that do not directly cite each other can be strongly connected and very closely positioned.


Their algorithm then builds a Force Directed Graph to distribute the papers in a way that visually clusters similar papers together and pushes less similar papers away from each other. Their database is connected to the Semantic Scholar Paper Corpus


Explore connected papers in a visual graph


To start, enter a paper identifier

Build a graph


You can try:


Paper DOI


arXiv
Paper URL


Paper Title


Semantic Scholar
Paper URL


PubMed
Paper URL

8 Search arXiv online repository

Search on arXiv can be efficiently performed through “Advanced Search” (see [link](#))

Title

AND

Abstract

Add another term

+

Subject

All classifications will be included by default.

☐ Computer Science (cs)

☐ Physics

all

☐ Economics (econ)

☐ Quantitative Biology (q-bio)

☐ Electrical Engineering and Systems Science (eess)

☐ Quantitative Finance (q-fin)

☐ Mathematics (math)

☐ Statistics (stat)

☒ Include cross-listed papers

☐ Exclude cross-listed papers

Date

☐ All dates

☐ Past 12 months

☐ Specific year

☒ Date range

From

to

When limiting by date range, the lower bound of the "from" date and the upper bound of the "to" date are used. For example, searching with **From:** 2012-02 and **To:** 2013 will search for papers submitted from 2012-02-01 to 2013-12-31.

Searching by Author Name

- Using the **Author(s)** field produces best results for author name searches.
- For the most precise name search, follow **surname(s), forename(s) or surname(s), initial(s)** pattern: example Hawking, S or Hawking, Stephen
- For best results on multiple author names, **separate individuals with a ;** (semicolon). Example: Jin, D S; Ye, J
- Author names enclosed in quotes will return only **exact matches**. For example, "Stephen Hawking" will not return matches for Stephen W. Hawking.
- Diacritic character variants are automatically searched in the Author(s) field.
- Queries with no punctuation will treat each term independently.

Tips

Wildcards:

- Use ? to replace a single character or * to replace any number of characters.
- Can be used in any field, but not in the first character position. See Journal References tips for exceptions.

Expressions:

Search using Quant-Arxiv Sanity Preserver and, respectively, Arxiv Sanity Preserver

Link: <http://quant-arxiv-sanity.com> and <http://arxiv-sanity.com>

Quant-Arxiv Sanity Preserver

Adapted in spare time by @pepicello from ASP to accelerate research.
Serving last 2845 papers from q-fin.[PR|PM|TR|MF|CP|ST|GN]

[most recent](#)[top recent](#)[top hype](#)[friends](#)[discussions](#)[recommended](#)[library](#)[donate](#)[Only show v1](#)[Last day](#)[Last 3 days](#)[Last week](#)[Last month](#)[Last year](#)[All time](#)

Top papers based on people's libraries:

Arxiv Sanity Preserver

Built in spare time by @karpathy to accelerate research.
Serving last 113686 papers from cs.[CV|CL|LG|AI|NE]/stat.ML

[most recent](#)[top recent](#)[top hype](#)[friends](#)[discussions](#)[recommended](#)[library](#)[Only show v1](#)[Last day](#)[Last 3 days](#)[Last week](#)[Last month](#)[Last year](#)[All time](#)

Top papers based on people's libraries:

9 Search SSRN online repository

Most efficient search through SSRN website is done through Browse (see snapshot below), which includes:

- Advanced search
- Browse by Network and Subject Area
- Browse by JEL code



In Advanced Search one can search in title and abstract, as well as title, abstract, keywords and full text. Searching can be done for a specific period (last week, last month, etc.) or for all dates

When Browse by JEL codes, some JEL classification codes are more relevant than others. In particular

G11 - Portfolio Choice; Investment Decisions

G12 - Asset Pricing; Trading Volume; Bond Interest Rates

G01 - Financial Crises

G02 - Behavioral Finance: Underlying Principles

G13 - Contingent Pricing; Futures Pricing

G14 - Information and Market Efficiency; Event Studies; Insider Trading

G15 - International Financial Markets

C11 - Bayesian Analysis: General

C12 - Hypothesis Testing: General

C51 - Model Construction and Estimation

C52 - Model Evaluation, Validation, and Selection

C53 - Forecasting and Prediction Methods; Simulation Methods

C31 - Cross-Sectional Models; Spatial Models; Treatment Effect Models; Quantile Regressions; Social Interaction Models
C32 - Time-Series Models; Dynamic Quantile Regressions; Dynamic Treatment Effect Models; Processes; State Space Models
C33 - Panel Data Models; Spatio-temporal Models
C34 - Truncated and Censored Models; Switching Regression Models

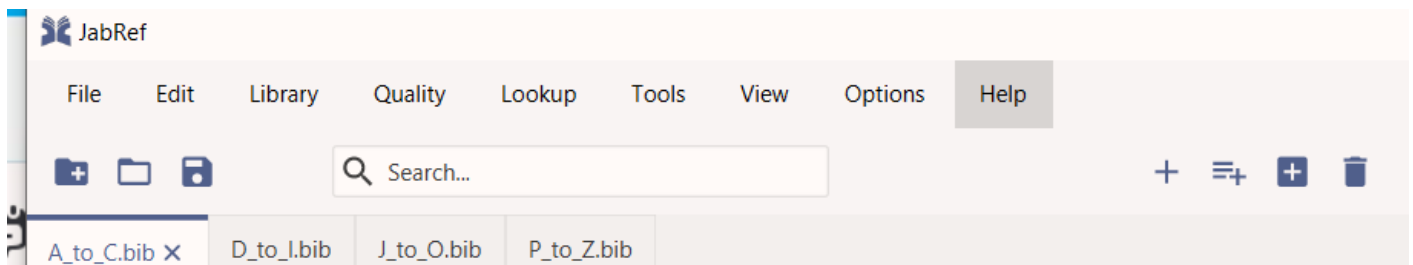
10 Use JabRef to search within bibliography (.bib) files

A bibliography (.bib file) was provided for each QWIM Project (see Bibliography folder)

Most of the entries include the abstract, and thus the search can be done across the abstracts as well.

The latest version of JabRef is 5.1 The documentation (including for search) can be found at link <https://docs.jabref.org>, with info on search at this [section](#)

Search can be done through keywords placed in the search window. (see snapshot below)



One can specify whether the keywords are located everywhere (default), or in title, or in abstract, or in optional fields, or in title and/or abstract, etc.

There are two search modes in JabRef: Normal Search and Advanced Search

Normal search

In a normal search, the program searches your database for all occurrences of the words in your search string, once you entered it. Only entries containing all words will be considered matches. To search for sequences of words, enclose the sequences in double quotes. For instance, the query **portfolio "mean variance"** will match entries containing both the word "portfolio" and the phrase "mean variance".

Advanced search

In order to search specific fields only and/or include logical operators in the search expression, a special syntax is available in which these can be specified. E.g. to search for entries whose author contains **miller**, enter:

```
author = miller
```

Both the field specification and the search term support regular expressions (see below). If the search term contains spaces, enclose it in quotes. Do *not* use spaces in the field specification!

As an example, to search for entries about mean variance, type:

```
title|keywords = "mean variance"
```

You can use **and**, **or**, **not**, and parentheses as intuitively expected:

```
(author = miller or title|keywords = "machine learning") and not author = brown
```

The `=` sign is actually a shorthand for `contains`.

Searching for an exact match is possible using `matches` or `==`.

Using `!=` tests if the search term is *not* contained in the field (equivalent to `not ... contains ...`).

The selection of field types to search (required, optional, all) is always overruled by the field specification in the search expression. If a field is not given, all fields are searched. For example, `video` and `year == 1932` will search for entries with any field containing `video` and the field `year` being exactly 1932.

Pseudo fields

JabRef defines the following pseudo fields:

Pseudo field	Purpose	Example
<code>anyfield</code>	Search in any field	<code>anyfield contains fruit</code> : search for entries having one of its fields containing the word fruit . This is identical to just writing <code>apple</code> . It may be more useful as <code>anyfield matches apple</code> , where one field must be exactly <code>apple</code> for a match.
<code>anykeyword</code>	Search among the keywords	<code>anykeyword matches apple</code> : search for entries which has the word apple among its keywords. However, as this also matches <code>pineapple</code> , it may be more useful in searches of the type <code>anykeyword matches apple</code> , which will not match <code>apples</code> or <code>pineapple</code>
<code>bibtexkey</code>	Search for citation keys	<code>bibtexkey == miller2005</code> : search for an entry whose BibTeX key is miller2005
<code>entrytype</code>	Search for entries of a certain type	

Regular expressions

JabRef uses Regular expressions (regex for short) as defined in Java (see [documentation](#) and [tutorial](#)). Regex can be used in the normal search mode and the advanced search mode

Regular expressions and casing

By default, regular expressions do not account for upper/lower casing. Hence, while the examples below are all in lower case, they match also upper- and mixed case strings.

Searching for entries with an empty or missing field

- `.` means any character
- `+` means one or more times

```
author != .+
```

Searching for a given word

- `\b` means word boundary
- `\B` means not a word boundary

`keywords = \buv\b` matches *uv* but not *lluvia* (it does match *uv-b* however)

`author = \bblack\b` matches *black* but neither *blackwell* nor *blacker*

`author == black` does not match *john black*, but `author = \bblack\b` does.

`author = \bblack\B` matches *blackwell* and *blacker*, but not *black*.

Searching with optional spelling

- `?` means none or one copy of the preceding character.
- `{n,m}` means at least *n*, but not more than *m* copies of the preceding character.
- `[]` defines a character class

`title =neighbou?r` matches *neighbour* and *neighbor*, and also *neighbours* and *neighbors*, and *neighbouring* and *neighboring*, etc.

`title = neighbou?rs?\b` matches *neighbour* and *neighbor*, and also *neighbours* and *neighbors*, but neither *neighbouring* nor *neighboring*.

`author = s[aá]nchez` matches *sanchez* and *sánchez*.

`abstract = model{1,2}ing` matches *modeling* and *modelling*.

`abstract = modell?ing` also matches *modeling* and *modelling*.

Searching for strings with a special character (() [] { } \ ^ - = \$! | ? * + .)

If a special character (i.e. () [] { } \ ^ - = \$! | ? * + .) is included in your search string, it has to be escaped with a backslash, such as `\}` for `}`.

It means that to search for a string including a backslash, two consecutive backslashes (`\\`) have to be used:

`abstract = xori{\\c{c}}o` matches *xoriço*.