8-2022

Digital tools - Supporting systematic reviews & evidence synthesis. Where are we now and what might the future look like?

Aaron TAY
Singapore Management University, aarontay@smu.edu.sg

Digital tools - Supporting systematic reviews & evidence synthesis. Where are we now and what might the future look like? ALIA HLA Lunchtime Seminar

Aaron Tay (Lead, Data Services)* ***Not a Health Science Librarian**

Singapore Management University Libraries

12 Aug 2022

 Aarontay@gmail.com

 @aarontay

 https://musingsaboutlibrarianship.blogspot.com

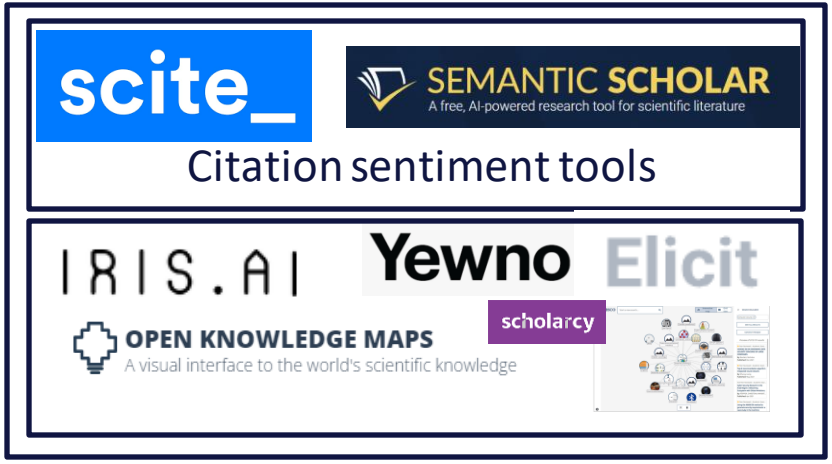
Myriad of new classes of discovery tools

New mega citation indexes

Citation based literature mapping services

Science mapping tools

Research Graphs



SMU
SINGAPORE MANAGEMENT UNIVERSITY

The world we used to live in*.....

* Bibliometric databases

The world we now live .. mega bibliographic databases

Estimated >200 million records*

Google Scholar — Link

BASE — Link

Microsoft Academic ✖ — Link(Successor)

LENS.ORG — Link

CORE — Link

Semantic Scholar — Link

Estimated >100 million records*

OpenAIRE — Link

Crossref — Link

Dimensions — Link

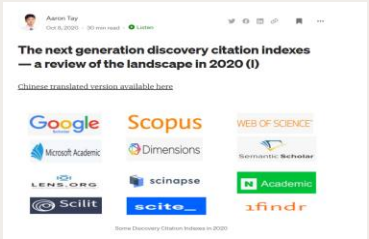
scite_ — Link

MENDELEY — Link

Open Access | Published: 06 May 2022

Search where you will find most: Comparing the disciplinary coverage of 56 bibliographic databases

Michael Gusenbauer *Gusenbauer (2022)



Blog post

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Is a single source database search a realistic dream?

Research article | Open Access | Published: 09 January 2013

Is the coverage of google scholar enough to be used alone for systematic reviews

Jean-François Gehanno, Laetitia Rollin & Stefan Darmoni

BMC Medical Informatics and Decision Making **13**, Article number: 7 (2013) | Cite this article

26k Accesses | **163** Citations | **163** Altmetric | Metrics

Research article | Open Access | Published: 26 October 2013

Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough

Martin Boeker, Werner Vach & Edith Motschall

BMC Medical Research Methodology **13**, Article number: 131 (2013) | Cite this article

16k Accesses | **89** Citations | **124** Altmetric | Metrics

Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources

Michael Gusenbauer [1], Neal R Haddaway [2 3]

Affiliations + expand

PMID: 31614060 PMCID: PMC7079055 DOI: 10.1002/jrsm.1378

Ideal academic search engine to support structured search

- Size of index – ideally has good retrospective coverage, and coverage of grey literature and non-English literature
- Functionality – support of
  - Controlled vocabulary
  - Exact Phase searching
  - Field searching
  - Boolean searching
    - AND/OR/NOT
    - Parentheses
    - Proximity operators
    - Truncation, Wildcard support
  - No or high maximum search query character limit
  - Post query filter support (See Jeroen Baas comparison of 51 search engine filters)

Adapted from Gusenbauer & Haddaway (2019)

Ideal academic search engine to support structured search

- Transparency and openness (reproducibility)
  - Transparency and openness of sources
  - **Open Data**
    - Size of batch downloads via web interface
    - Open Licensing of Data
  - Is code (relevancy algo etc) transparent or a black box?
    - Reproducibility of search results across different times and locations

Comparison of Google Scholar and Microsoft Academic Graph (MAG)

	Google Scholar	Microsoft Academic
Size of index	Largest index* , Excellent coverage of non-English and Grey literature	2nd largest in most subject areas*
Support of Gusenbauer & Haddaway (2019) search features	Bad	Very bad - "Semantic Search"
Are algo black boxes?	Yes	Yes
Restrictions in downloads in Web interface?	Yes	Yes
Licensing of data	Closed	Open Data – ODC-BY

*

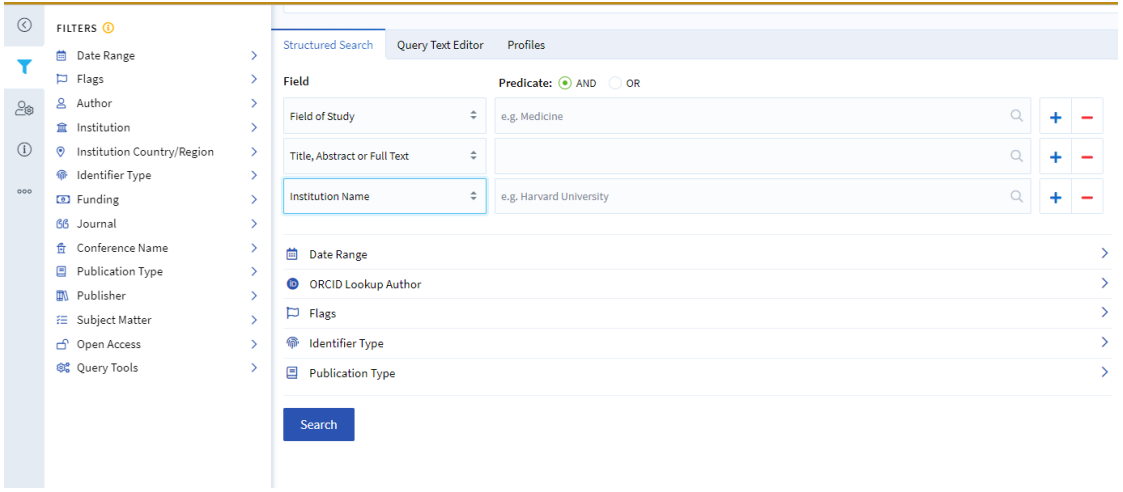
*Gusenbauer (2022)

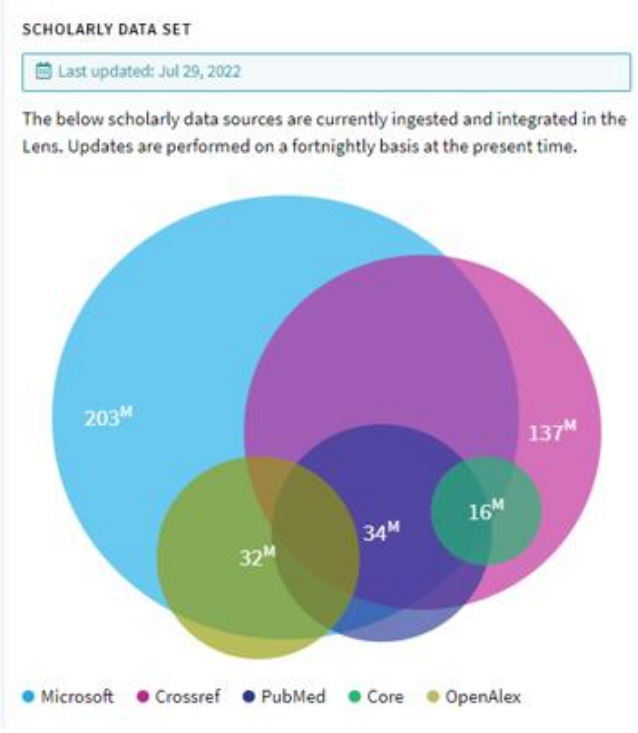
Comparison of Google Scholar and Microsoft Academic Graph (MAG)

"There has recently been an upsurge in using semantic search engines over traditional ones, as is evident in the birth of Semantic Scholar (2015), the relaunch of Microsoft Academic (2017), and the expected launch of *Meta*, a project of the Zuckerberg foundation. These semantic search engines tend to be designed to reward exploratory rather than systematic search behavior. … Our findings indicating that these systems are inadequate to be used as principal systems in systematic searches support this notion." (Gusenbauer & Haddaway, 2019, p. 211)

*

What is Lens.org?





Lens.org

	Lens.org
Size of index	4th largest index (Combines MAG+Crossref+Pubmed etc)*
Support of Gusenbauer & Haddaway (2019) search features	Excellent
Are algo black boxes?	Yes
Restrictions in downloads in Web interface?	Minimal, 50k batch limit
Other methods of data access	API (paid service)

*

The power of Open Scholarly data – Single source database is possible?

RESEARCH ARTICLE

Cost-effectiveness of Microsoft Academic Graph with machine learning for automated study identification in a living map of coronavirus disease 2019 (COVID-19) research [version 1; peer review: 2 approved with reservations]

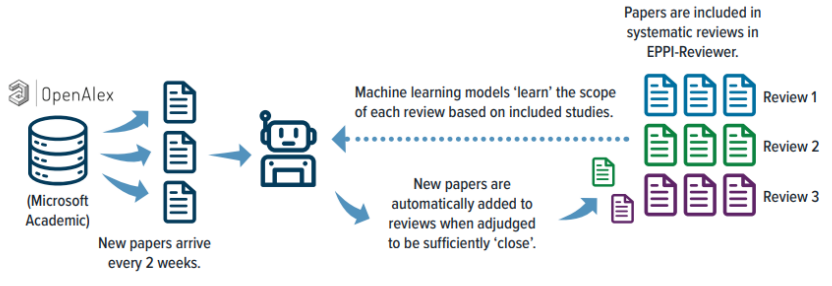
Ian Shemilt[1*], Anneliese Arno[1*], James Thomas[1*], Theo Lorenc[2], Claire Khouja[2], Gary Raine[2], Katy Sutcliffe[1], D'Souza Preethy[1], Irene Kwan[1], Kath Wright[2], Amanda Sowden[2]

RESEARCH AND DEVELOPMENT:
Can MAG Be Used as a Single Source for Locating Studies for Systematic Reviews?

The MAG dataset (now updated by OpenAlex)[4] might have the potential to serve as a single source of references for systematic reviews, though research is needed to determine whether it *can* function as a single-source database. In addition, because the dataset contains more than 200 million records, it would be necessary to establish whether records could be identified with sufficient accuracy. We used a case study of a collaborative project between the EPPI-Centre and the University of York, to answer two parallel research questions:

1. Does the MAG dataset contain all the studies that we need? *(sensitivity)*
2. Can relevant studies be identified efficiently? *(specificity)*

Figure 4. EPPI-Reviewer auto-update workflow



Source: Diagram created by author.

https://ktdrr.org/products/info-briefs/KTDRR-Brief-Keeping-SR-UpToDate-508.pdf

The power of Open Scholarly data

- "For maintaining our living systematic map of COVID-19, we found that using Microsoft Academic as a single-source database is more cost-effective than searching MEDLINE and Embase and then following conventional methods for deduplicating and screening results. The literature search tool yields higher numbers of relevant records at a lower cost"

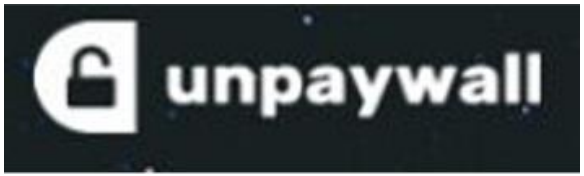
- "We found that nearly all the records in MEDLINE/Embase (99%) also were available in MAG (99%). In addition, we found that MAG contained 743 records that we did not find in our MEDLINE/ Embase searches. This discovery was concerning. Although we searched what are widely considered to be the main sources of research in health, we achieved a recall of only 83% of the records that we sought"

https://ktdrr.org/products/info-briefs/KTDRR-Brief-Keeping-SR-UpToDate-508.pdf

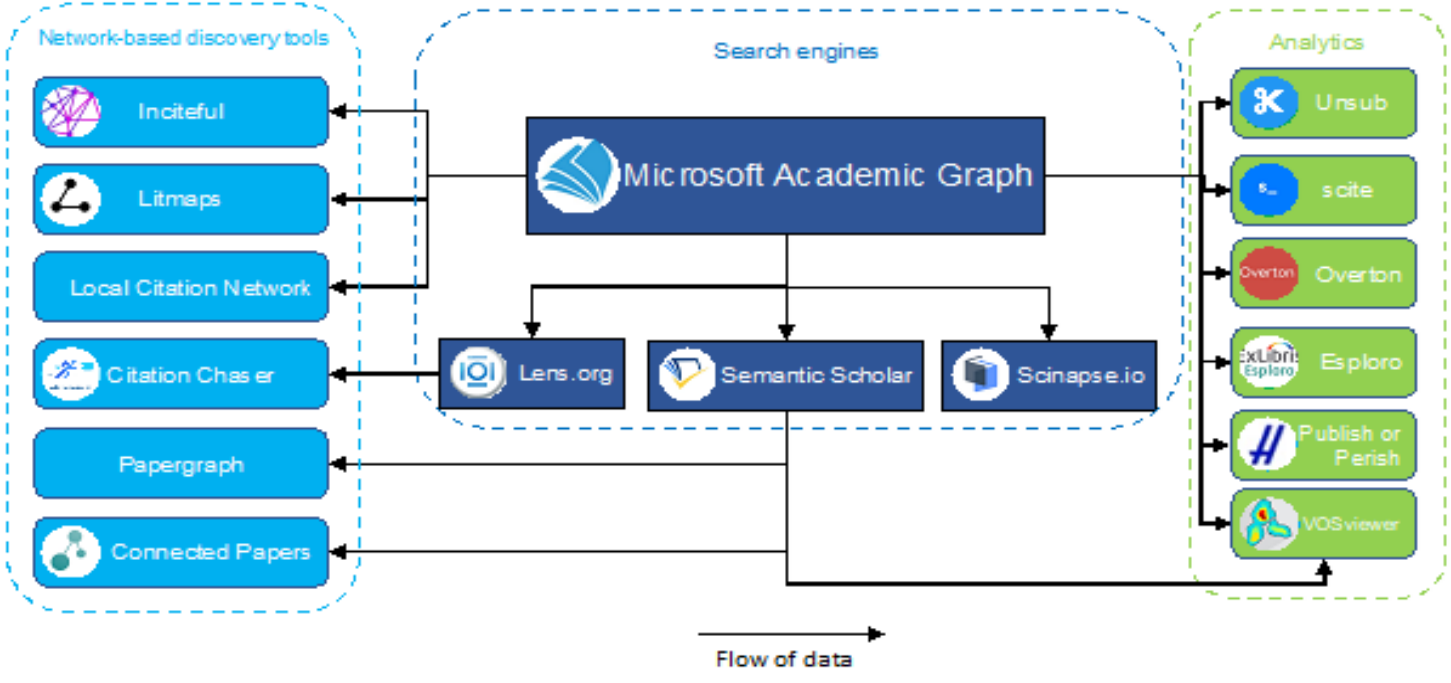
What is different this time- flood of Open Scholarly Metadata

Crossref, Datacite, ORCID, ROR, NIH

What is different this time- flood of Open Scholarly Metadata



Openness is not enough – The closure of Microsoft Academic Dec 31, 2021



Goodbye, Microsoft Academic – Hello, open research infrastructure?

The importance of Open, sustainable infrastructure for Open search tools

- Governance, Sustainability and Insurance (Principles of Open Scholarly Infrastructure)
  - Who determines the future of the search engine ? Is it Stakeholder governed (**Governance**)
  - Are there sustain contingency fund to support operations? Is there adequate revenue generated based on services provided to sustain the search engine? (**Sustainabilty)**
  - Are the data, code made as open as possible (**Insurance**)

Crossref CORE OpenAIRE our research OpenCitations DataCite FIND, ACCESS, AND REUSE DATA ROR DRYAD Europe PMC

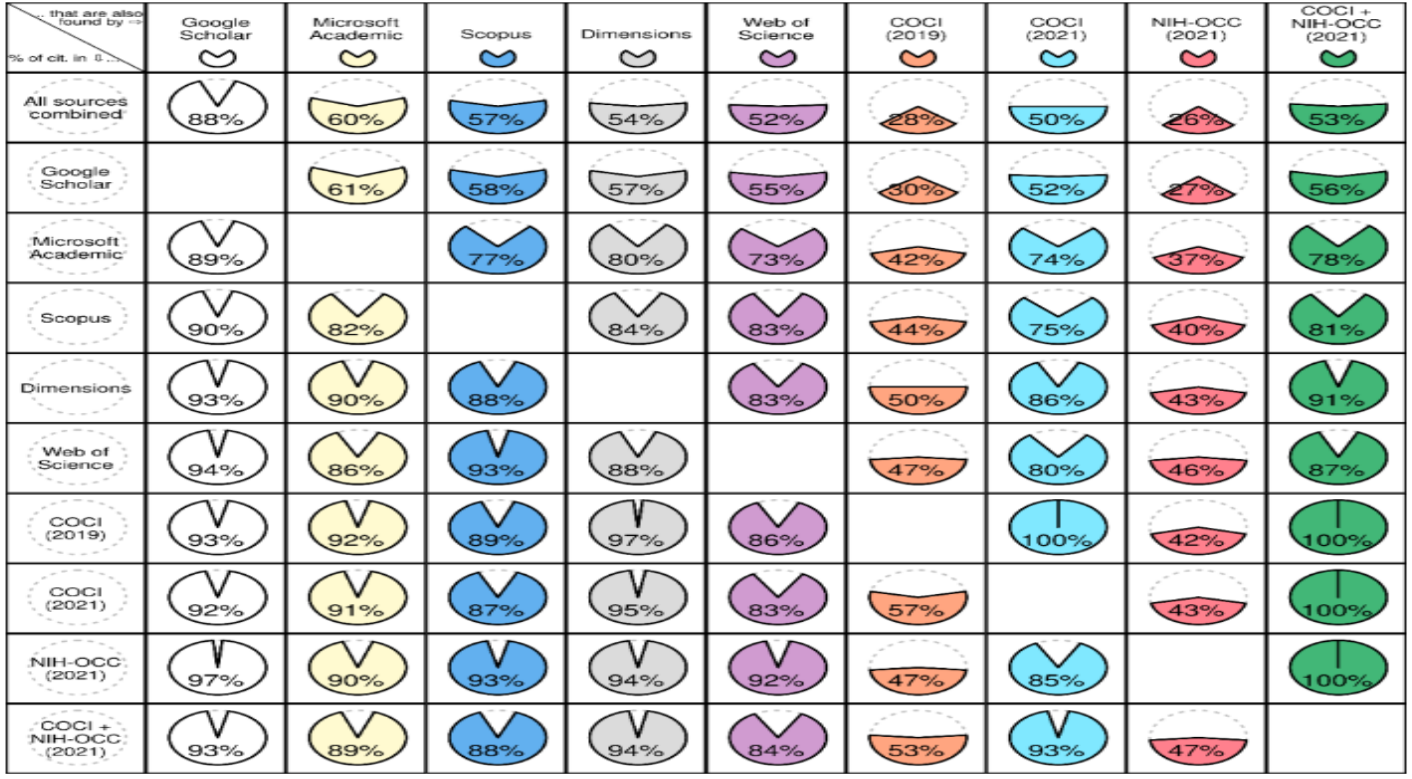
Organizations committed to Principles of Open Scholarly Infrastructure

WHAT CAN YOU DO IN A WORLD WHERE ALL THIS SCHOLARLY INFO AND FULL TEXT IS OPEN AND AVAILABLE?

- Title/author/abstract/subject
- References (relationships!)
- Affiliations
- Funding info
- Altmetrics
- Open access status
- Others? - Open peer review data etc

Full Text

Coverage comparisons for citations – WOS vs Open sources as of Sept 2021



% of cit. in ⇩ ... that are also found by ⇨	Google Scholar	Microsoft Academic	Scopus	Dimensions	Web of Science	COCI (2019)	COCI (2021)	NIH-OCC (2021)	COCI + NIH-OCC (2021)
All sources combined	88%	60%	57%	54%	52%	28%	50%	26%	53%
Google Scholar		61%	58%	57%	55%	30%	52%	27%	56%
Microsoft Academic	89%		77%	80%	73%	42%	74%	37%	78%
Scopus	90%	82%		84%	83%	44%	75%	40%	81%
Dimensions	93%	90%	88%		83%	50%	86%	43%	91%
Web of Science	94%	86%	93%	88%		47%	80%	46%	87%
COCI (2019)	93%	92%	89%	97%	86%		100%	42%	100%
COCI (2021)	92%	91%	87%	95%	83%	57%		43%	100%
NIH-OCC (2021)	97%	90%	93%	94%	92%	47%	85%		100%
COCI + NIH-OCC (2021)	93%	89%	88%	94%	84%	53%	93%	47%	

Fig. 1. Percentage of citations found by each database, relative to all citations (first row), and relative to the number of citations found by the other databases (subsequent rows).

[Coverage of open citation data approaches parity with Web of Science and Scopus]

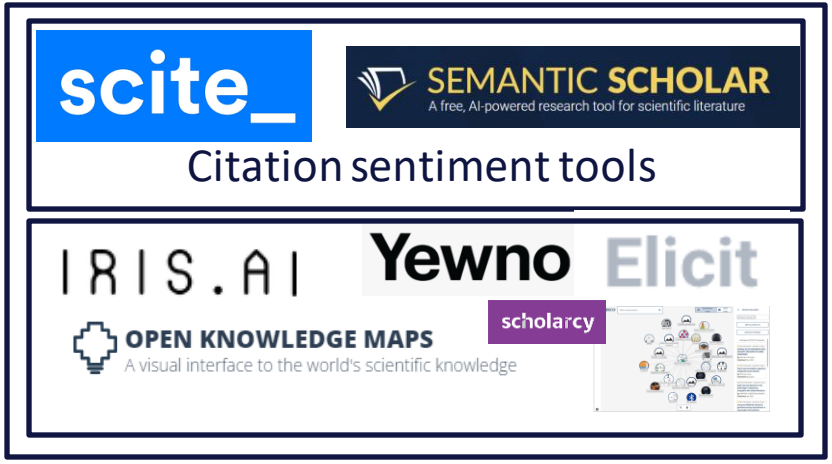
That is why the world of discovery + search is now complicated

New mega citation indexes

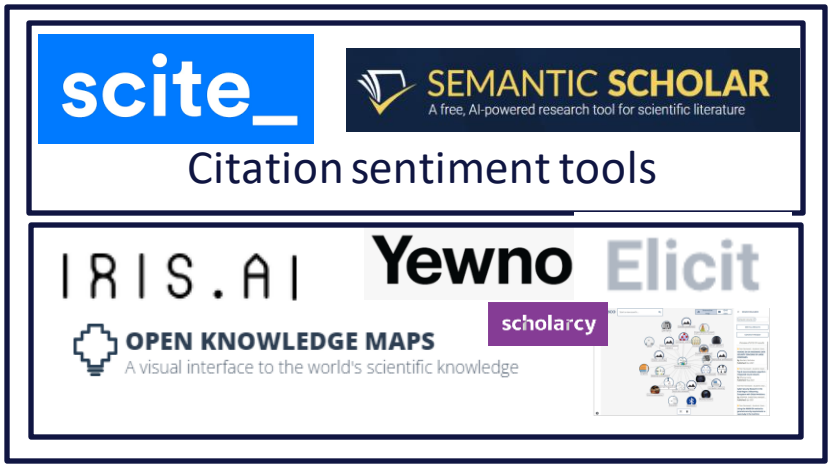
Citation based literature mapping services

Science mapping tools

Research Graphs



Full-text extraction



- Tools based on full-text (NLP, Large language models) tend to be less common due to lack of availability of full text (Open Access is still rising) + greater computation power needed for computation – so mostly expensive commercial tools'
- Some examples of interesting tools that rely on processing full-text
  - Scite.ai - Classifies citation by citation sentiment/context/type - Mentioning cite vs supporting cite vs contradicting cites
  - Semantic Scholar – Classifies citations by citing background/citing methods/citing results and identifying "influential cites"

Elict.org example – magic of NLP – Large Language Models for data extraction

Elicit | what is the sensitivity rate of covid-19 ART tests | FAQ | Tasks | Starred

Paper title	Takeaway from abstract	PDF	Year	Citations	Region	Question-relevant summary
Covid-19: Point of care test reports 94% sensitivity and 100% specificity compared with laboratory test	The CovidNudge real time PCR platform can test results in 90 minutes with 94% sensitivity and 100% specificity.	PDF	2020	2	UK	The sensitivity rate of covid-19 art tests is 94%.
Scaling up COVID-19 rapid antigen tests: promises and challenges	Antigen-detection rapid diagnostic tests for COVID-19 should be replaced by a combination of molecular and serological tests.	PDF	2021	108	-	-
Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19	RT-PCR remains the gold standard for the diagnosis of COVID-19 in sputum samples.	PDF	2020	195	-	The sensitivity rate of covid-19 art tests is 84.5%.
Diagnostic Accuracy of Rapid Antigen Tests for COVID-19 Detection: A Systematic Review With Meta-analysis	The pooled sensitivity and specificity of the rapid antigen tests against the reference test (the real-time PCR) were 69% and 99% respectively.	PDF	2022	1	-	The sensitivity rate of covid-19 art tests is 69%.
Covid-19: Lateral flow tests are better at identifying people with symptoms, finds Cochrane review	The Covid-19 test had a sensitivity of 58% in identifying people with symptoms.	PDF	2021	6	-	The sensitivity rate of covid-19 art tests is 72%.
Sensitivity, Specificity and Predictive Values of Molecular and Serological Tests for COVID-19: A Longitudinal Study in Emergency Room	Molecular tests for SARS-CoV-2 infection showed excellent specificity, but significant differences in sensitivity.	PDF	2020	16	Italy	The sensitivity rate of covid-19 art tests is >90%.
Sensitivity, specificity and predictive values of molecular and serological tests for COVID-19. A longitudinal study in emergency room.	Molecular tests for SARS-CoV-2 infection showed excellent specificity, but significant differences in sensitivity.	PDF	2020	12	Italy	The sensitivity rate of covid-19 art tests is 91.8%.
Diagnostic Accuracy of COVID-19 Antibody Tests Authorized by FDA Philippines: A Systematic Review and Meta-Analysis	Clinical accuracy of COVID-19 antibody tests was excellent with high specificities.	PDF	2021	1	Philippines	The sensitivity rate of covid-19 art tests is 68.5%.

Show more

Add info

Paper title
Takeaway from abstract ×
PDF × Year ×
Citations × Region ×
Question-relevant sum... ×
+ Add Column

Metadata
Source
Authors
Journal
Influential citations
DOI
Funding source
Population studied
Number of participants
Number of studies
Population characteristics

Download csv

CITATION BASED LITERATURE MAPPING SERVICE*

1. Typically web-based service with built-in citation network/academic knowledge graph

2. Accepts input of one or more "seed" papers

3. From 1+2 , "recommend" papers via bibliometric methods/algo - a visualization is also usually provided

4. User friendly interface, avoids bibliometric jargon

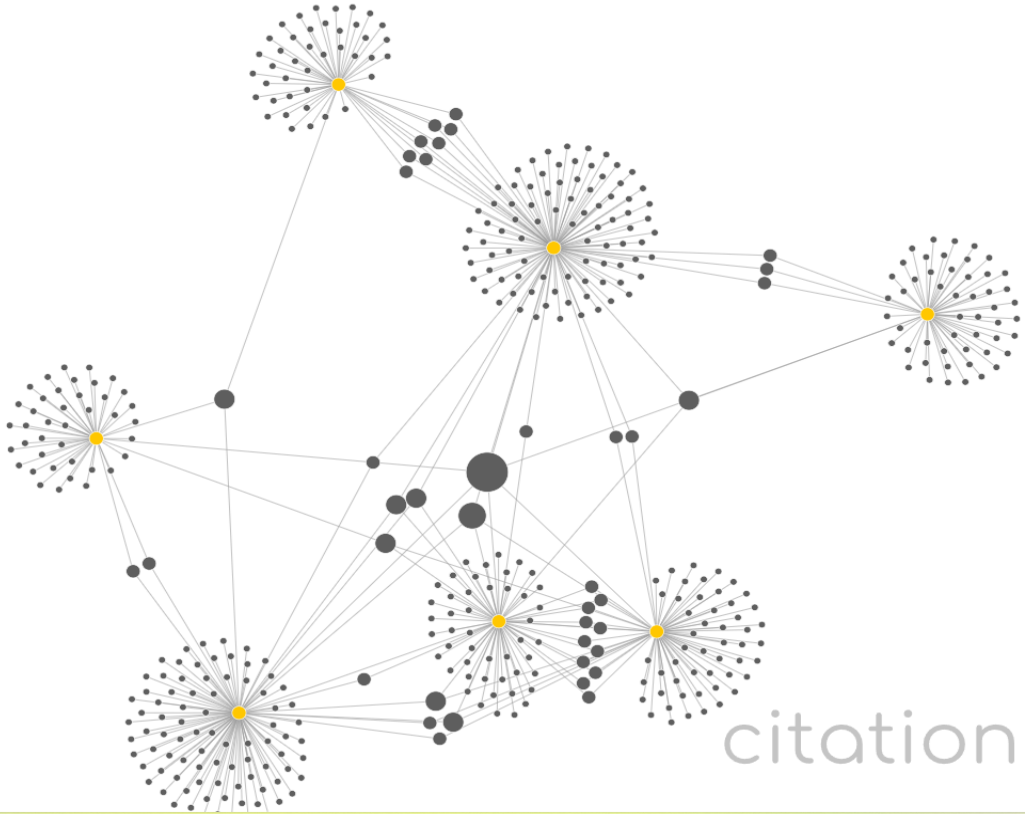
Examples includes ConnectedPapers, Citation Gecko, Research Rabbit, LitMaps, Inciteful, Cocites - See List actively maintained by me!

Recommended Papers

- The Pricing of Audit Services: Theory and Evidence
  Simunic
  Journal of Accounting Research
- The Role of Risk Management and Governance in Determining Audit Demand
  Knechel
  Journal of Business Finance & Accounting
- Auditor Selection and Audit Committee Characteristics
  Abbott
  Auditing: A Journal of Practice and Theory
- Board Characteristics and Audit Fees
  Carcello
  Contemporary Accounting Research
- Relation between External Audit Fees, Audit Committee Characteristics and Internal Audit
  Goodwin-Stewart
  Accounting & Finance
- Audit Fees: A Meta-Analysis of the Effects of Supply and Demand Attributes
  Hay
  Contemporary Accounting Research
- Evidence on the Impact of Internal Control and Corporate Governance on Audit Fees
  Hay

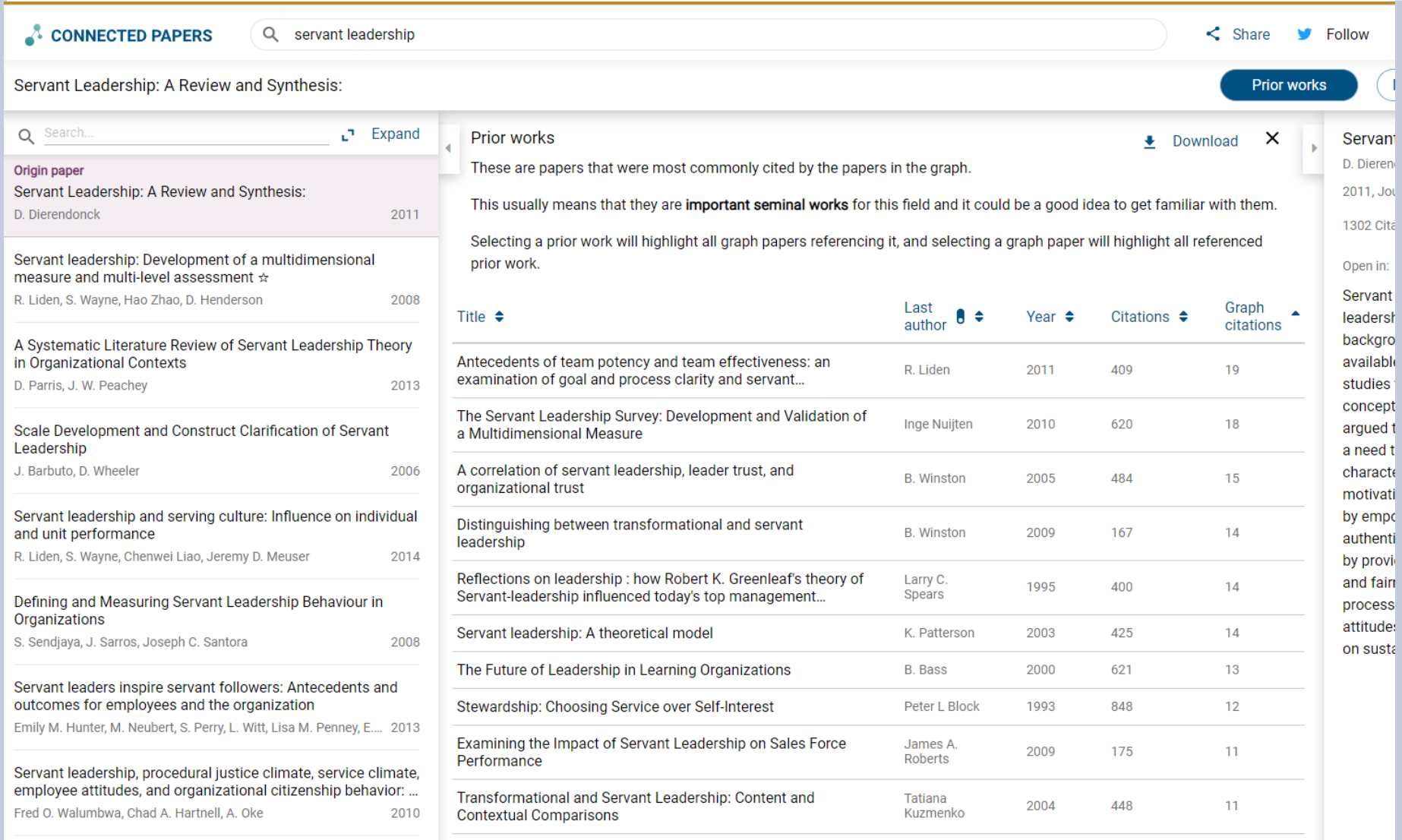
Add as seed | Export

Seed Papers
Papers Cited-By Seed Papers
Papers Citing Seed Papers



Save Image

citation gecko


CONNECTED PAPERS
servant leadership
Share
Follow
Servant Leadership: A Review and Synthesis:
Prior works
Search...
Expand
Origin paper
Servant Leadership: A Review and Synthesis:
D. Dierendonck 2011
Servant leadership: Development of a multidimensional measure and multi-level assessment ☆
R. Liden, S. Wayne, Hao Zhao, D. Henderson 2008
A Systematic Literature Review of Servant Leadership Theory in Organizational Contexts
D. Parris, J. W. Peachey 2013
Scale Development and Construct Clarification of Servant Leadership
J. Barbuto, D. Wheeler 2006
Servant leadership and serving culture: Influence on individual and unit performance
R. Liden, S. Wayne, Chenwei Liao, Jeremy D. Meuser 2014
Defining and Measuring Servant Leadership Behaviour in Organizations
S. Sendjaya, J. Sarros, Joseph C. Santora 2008
Servant leaders inspire servant followers: Antecedents and outcomes for employees and the organization
Emily M. Hunter, M. Neubert, S. Perry, L. Witt, Lisa M. Penney, E.... 2013
Servant leadership, procedural justice climate, service climate, employee attitudes, and organizational citizenship behavior: ...
Fred O. Walumbwa, Chad A. Hartnell, A. Oke 2010
Prior works
Download
These are papers that were most commonly cited by the papers in the graph.
This usually means that they are important seminal works for this field and it could be a good idea to get familiar with them.
Selecting a prior work will highlight all graph papers referencing it, and selecting a graph paper will highlight all referenced prior work.
Title
Last author
Year
Citations
Graph citations
Antecedents of team potency and team effectiveness: an examination of goal and process clarity and servant... R. Liden 2011 409 19
The Servant Leadership Survey: Development and Validation of a Multidimensional Measure Inge Nuijten 2010 620 18
A correlation of servant leadership, leader trust, and organizational trust B. Winston 2005 484 15
Distinguishing between transformational and servant leadership B. Winston 2009 167 14
Reflections on leadership : how Robert K. Greenleaf's theory of Servant-leadership influenced today's top management... Larry C. Spears 1995 400 14
Servant leadership: A theoretical model K. Patterson 2003 425 14
The Future of Leadership in Learning Organizations B. Bass 2000 621 13
Stewardship: Choosing Service over Self-Interest Peter L Block 1993 848 12
Examining the Impact of Servant Leadership on Sales Force Performance James A. Roberts 2009 175 11
Transformational and Servant Leadership: Content and Contextual Comparisons Tatiana Kuzmenko 2004 448 11
Servant
D. Dieren
2011, Jou
1302 Cita
Open in:
Servant
leadersh
backgro
available
studies
concept
argued t
a need t
characte
motivati
by empo
authenti
by provi
and fairn
process
attitudes
on susta

KEY TO UNDERSTANDING NEW TOOLS

Underlying data
e.g. NIH-OCC

This affects coverage of the tool. Get a sense of what the tool can never find, recency etc

Technique or method used
e.g. cocitation, documented in this paper

Might be totally black box, partly documented or totally open with formulas or codes specified

Use case + Business model

A lot of tools are now going "freemium", e.g. Connected Papers, LitMaps.

Other factors to consider

- Recency of coverage
  - Two tools might use the same dataset e.g. OpenAlex, but one tool might be pulling the info real time via API (update by source immediately), another might be downloading the data in full-text every x weeks , cleaning before being loaded into tool (might have time lag).
- Price?
- Additional features – collaborative functions, Sync with Zotero collections, visualization capabilities

Tool	Index used	Transparency of method	Main method
Citation Gecko	OpenCitations Index of Crossref open DOI-to-DOI Citations(COCI) & OpenCitations Corpus (OCC)	High, documented, **Open Source**	Identify papers which are highly cited or citing seed papers
Connected Papers	Semantic Scholar Academic Knowledge Graph	Low, loosely explained	Similarity metric based on cocitation & bibliometric coupling
ResearchRabbit	MAG, OpenAlex	Depends on method used. Mostly not explained	Unknown for "similar work", "earlier work", "later work"
Litmaps	Crossref, Semantic Scholar	Low,	Select papers within 1-2 citation

ARE CITATION BASED LITERATURE MAPPING SERVICE USED IN SR?

Yes, somewhat

Straightforward Citation searching/citation chasing

C30: Searching reference lists (**Mandatory**)	
Check reference lists in included studies and any relevant systematic reviews identified.	Searches for studies should be as extensive as possible in order to reduce the risk of publication bias and to identify as much relevant evidence as possible.

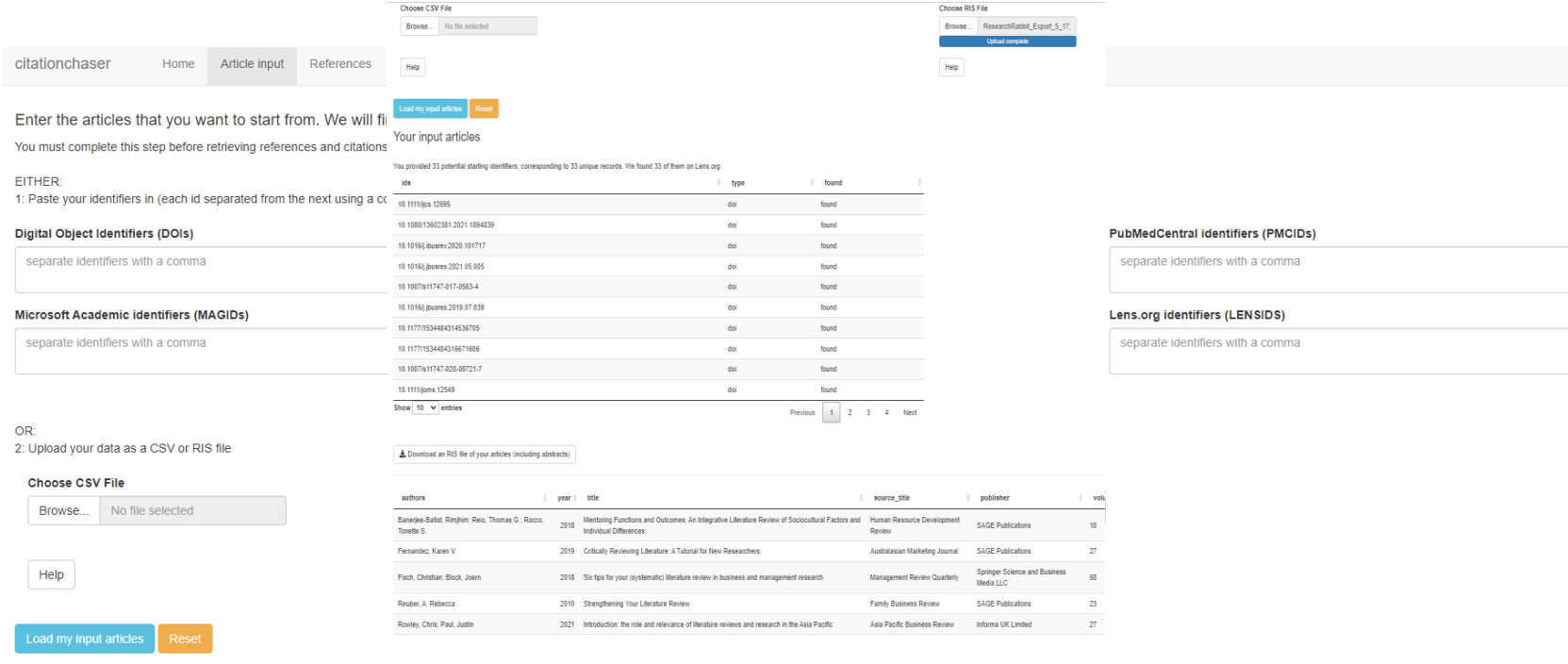
MECIR Box 4.3.e.

Straightforward Citation searching/citation chasing

"It is good practice to carry out forwards citation searching on reports of studies that meet the eligibility criteria of a systematic review. Thus forwards citation searching usually takes place after the results of the bibliographic database searches have been screened and a set of potentially includable studies has been identified (Briscoe et al 2020a)… Since researchers may selectively cite studies with positive results, forwards citation searching should be used with caution as an adjunct to other search methods in Cochrane Reviews"

4.S1 Technical Supplement to Chapter 4:
Searching for and selecting studies , P.5

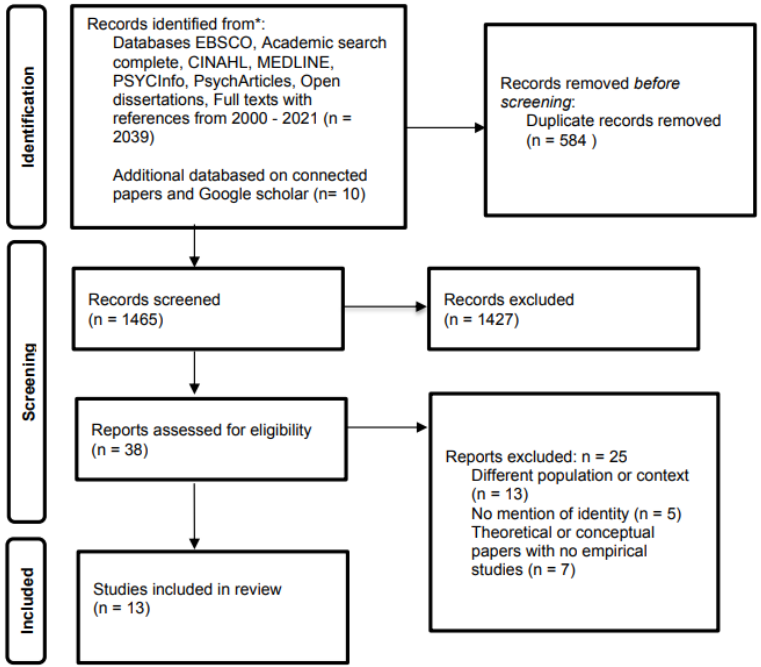
Citation Chaser, straight forward – forward and backward citation searcher



NR Haddaway, MJ Grainger, CT Gray (2021) citationchaser: An R package and Shiny app for forward and backward citations chasing in academic searching. Zenodo, doi:10.5281/zenodo.4533747

"Connected Papers was used to search for related papers"

Figure 1: PRISMA Flow diagram of selection process



The search strategy was co-created as a team and with a university librarian, detailed in Table 1, which was inputted into the University of Liverpool database, on 21st November 2021 and rechecked on 17th May 2022 using EBSCO to access journals from Academic search complete, CINAHL, MEDLINE, PsycINFO, Psych Articles, and University of Liverpool Catalogue was used for this. Hand searching of references from papers was conducted. Connected papers website was used to search for related papers. Figure 1 reports the selection procedure.

Gupta, V., Golding, L., Eames, C., Greenhill, B., Qi, R., Allan, S., ... & Fisher, P. (2022). Understanding the identity of lived experience researchers and providers: A conceptual framework and systematic narrative review.

"explored related publications through a graphical tool (www.connectedpapers.com).

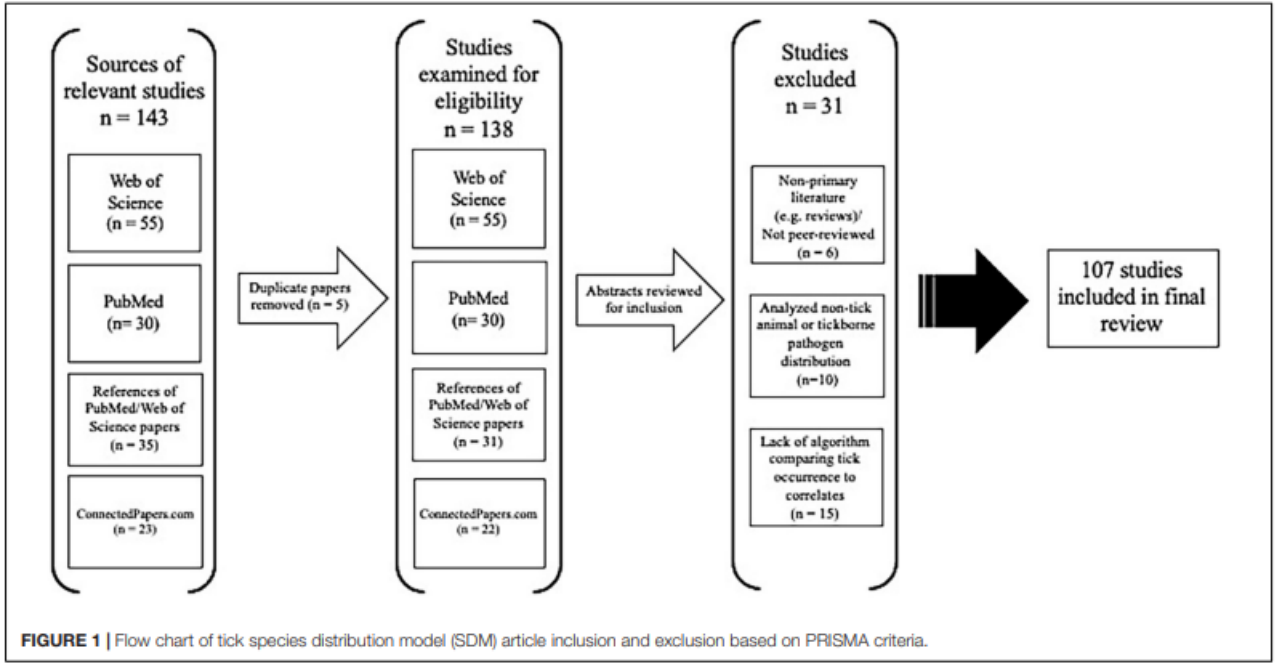


FIGURE 1 | Flow chart of tick species distribution model (SDM) article inclusion and exclusion based on PRISMA criteria.

Following the PRISMA-ScR checklist, we searched scientific databases for eligible articles, their references, and **explored related publications through a graphical tool** (www.connectedpapers.com).

Kopsco, H. L., Smith, R. L., & Halsey, S. J. (2022). A Scoping Review of Species Distribution Modeling Methods for Tick Vectors. *Frontiers in Ecology and Evolution*, *10*, 893016.

"**The seed papers of the high-appropriate papers are used as seed papers in the CitationGecko tool** (**www.citationgecko.com**, accessed on 9 November 2021) to get a deeper understanding of the interconnections between them and ensure that no relevant literature is ignored in this review"

Kaiblinger, A., & Woschank, M. (2022). State of the art and future directions of digital twins for production logistics: a systematic literature review. *Applied Sciences*, *12*(2), 669.

"**Additional sources were identified through the reference list of the eligible articles from the initial search and a co-citation method using the bibliographic coupling concept (www.connectedpapers.com).**"

"Further, we only used PubMed as our primary search database. Nonetheless, we felt that using PubMed with a relatively broad search strategy was most relevant for the review, and we performe"

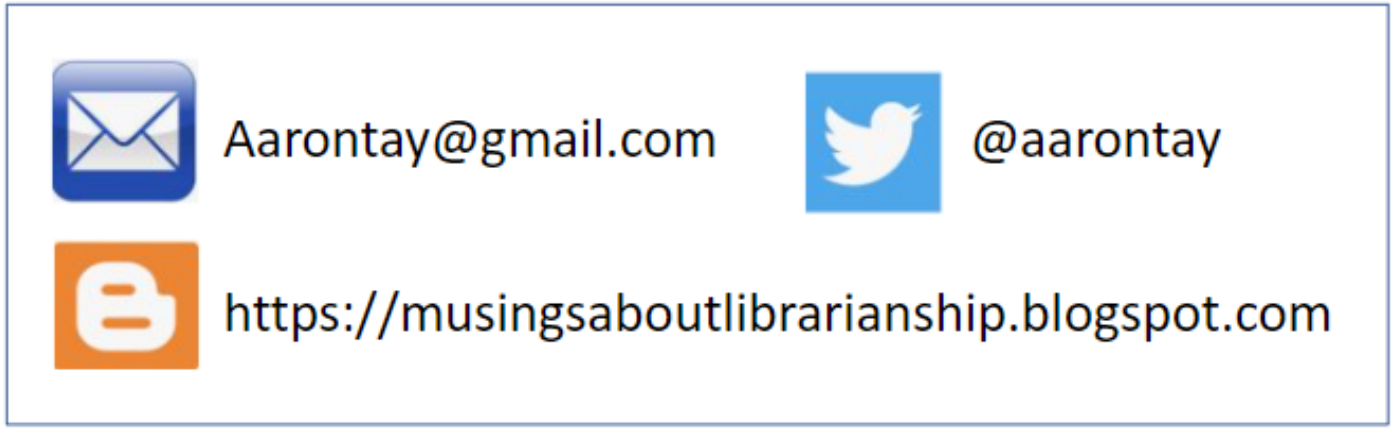
Willwacher, S., Kurz, M., Robbin, J., Thelen, M., Hamill, J., Kelly, L., & Mai, P. (2022). Running-Related Biomechanical Risk Factors for Overuse Injuries in Distance Runners: A Systematic Review Considering Injury Specificity and the Potentials for Future Research. *Sports Medicine*, 1-15.

ELICT.ORG -DEMO IF HAVE TIME

INTERESTED TO KNOW MORE?

1.List of Innovative Literature mapping tools
2.Citation based literature mapping tools - an update - tools offering premium accounts, the effects of the loss of MAG and use for evidence synthesis? (latest)
3. **Academia's missing references**

Thank You!



Acknowledgement: Phil Gooch (Scholarcy) for advice on text mining applications