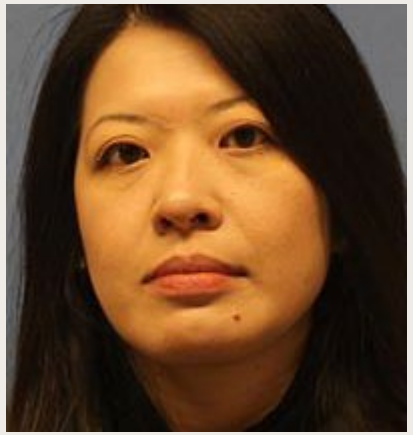
NIPH
Norwegian Institute of Public Health

How to speed up information retrieval

Elisabet Hafstad

May 25 2023



Lien Nguyen

Marit Johansen

Ingrid Harboe

Aims, findings, and suggested target areas for automation of information retrieval: final report 2022 (Nguyen 2023)

Themes covered

- Steps in information retrieval for evidence synthesis
- Tools and automation for searching
- Librarian time spent on systematic review tasks
- Automation and more to optimize the search process

«Glossary» for today

- Information retrieval = searching for and selecting reports of studies to include in evidence syntheses.

- Evidence syntheses <-> systematic reviews, mapping reviews and more
- Librarian <-> information specialist <-> anyone conducting systematic searches for evidence synthesis purposes
- Traditional search <-> keyword-based Boolean search
- Software <-> digital tools

Steps in information retrieval for evidence synthesis

- Check for similar reviews
- Design the systematic search
  - Select relevant databases/sources
  - Identify relevant review and search concepts (e.g., PICO)
  - Identify relevant search terms
  - Combine search terms and concepts
  - Adapt to each database/source
  - Peer review, and if necessary, revise search strategies (PRESS)
- Run the systematic search
- Deduplicate the search results
- Screen the titles and abstracts
- Obtain the full text
- Screen the full text
- Conduct a citation analysis
- Report the retrieval process
- (Update the systematic search)

Steps in information retrieval for evidence synthesis

- Check for similar reviews
- Design the systematic search
  - Select relevant databases/sources
  - Identify relevant review and search concepts (e.g., PICO)
  - Identify relevant search terms
  - Combine search terms and concepts
  - Adapt to each database/source
  - Peer review, and if necessary, revise search strategies (PRESS)
- Run the systematic search
- Deduplicate the search results
- Screen the titles and abstracts
- Obtain the full text
- Screen the full text
- Conduct a citation analysis
- Report the retrieval process
- (Update the systematic search)

Screening tools to speed up information retrieval

Mission accomplished?

[..] research suggests that adopting automation can reduce the need for manual screening by at least 30% and possibly more than 90%, although sometimes at the cost of up to a 5% reduction in sensitivity (O'Mara-Eves et al 2015 – in Cochrane Handbook 4.6.6.2 Automating the selection process)

Priority screening, a machine learning function in EPPI-reviewer, reduced screening time by 60 % [..] (Muller 2021)

Digital tools for searching

- Design the systematic search
  - Select relevant databases/sources
  - Identify relevant review and search concepts (e.g., PICO)
  - Identify relevant search terms
  - Combine search terms and concepts
  - Adapt to each database/source
  - Peer review, and if necessary, revise search strategies (PRESS)
- **Run the systematic search**
- **Deduplicate the search results**
- **Obtain the full text**
- **Conduct a citation analysis**

How to read:
Red: Intellectual process difficult to automate?
Bold: More hits / relevant studies -> more hours

Automation of information retrieval

Aims, findings, and suggested target areas for automation of information retrieval: final report 2022 (Nguyen 2023)

We looked for tools

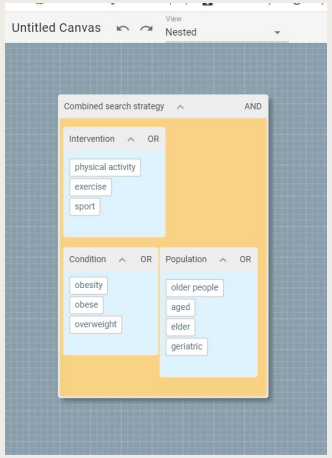
- that (semi-)automate > 2 steps of current search practice (keyword-based Boolean)
- use machine learning or text mining
- not requiring extensive upgrading of skills (i.e. coding/programming)

SR Toolbox

Digital tools for searching

[2Dsearch]

[Litsearchr]



[Nested Knowledge]

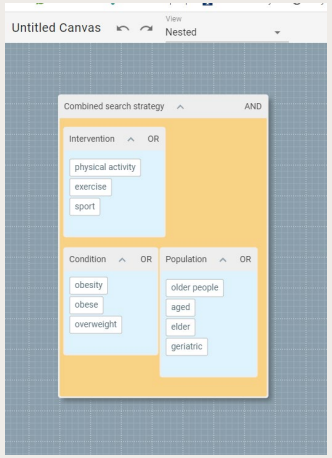
Qinsight (Quertle)

Digital tools for searching

2Dsearch

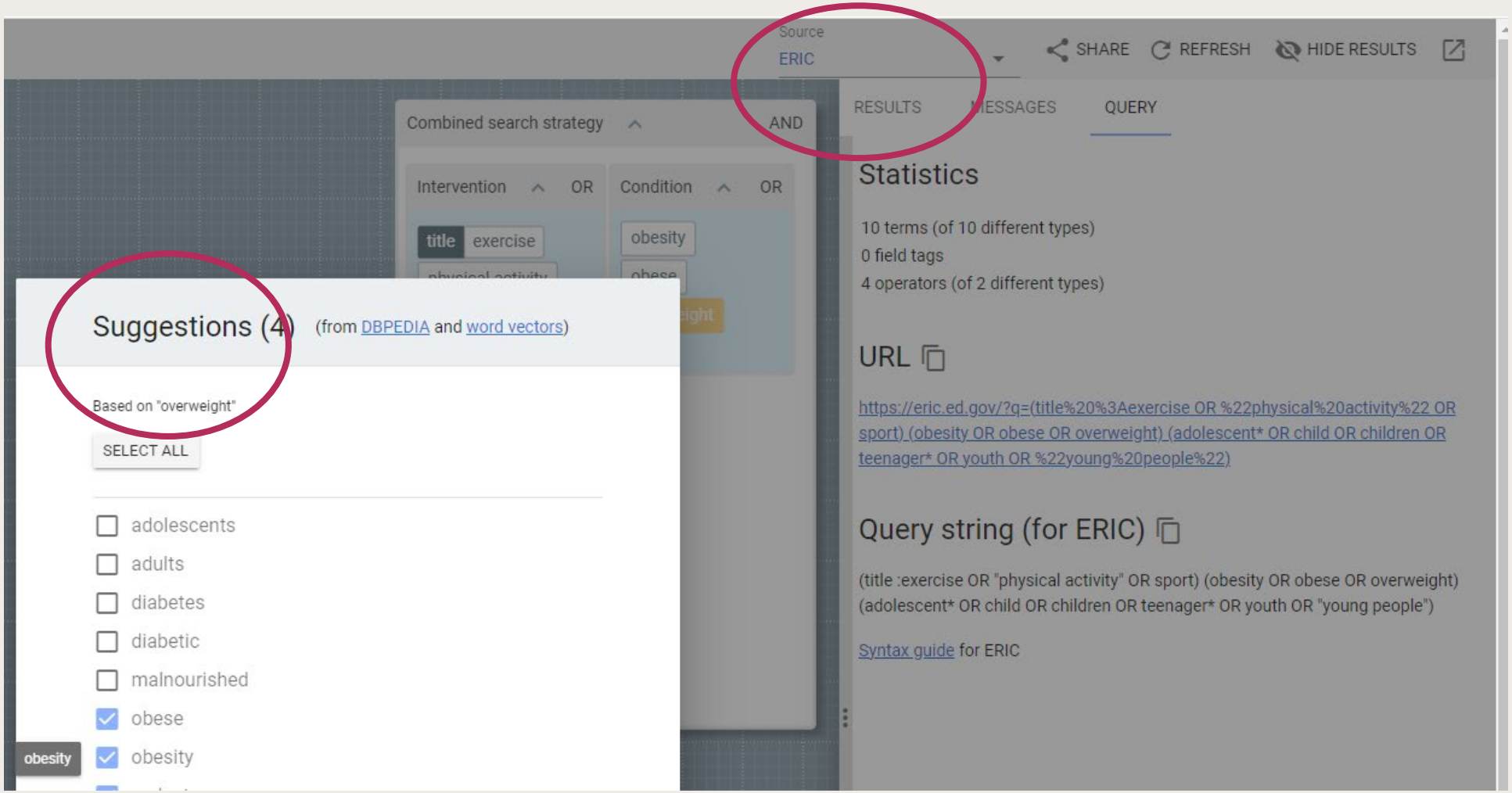
Litsearchr

Nested Knowledge



Qinsight (Quertle) - discontinued

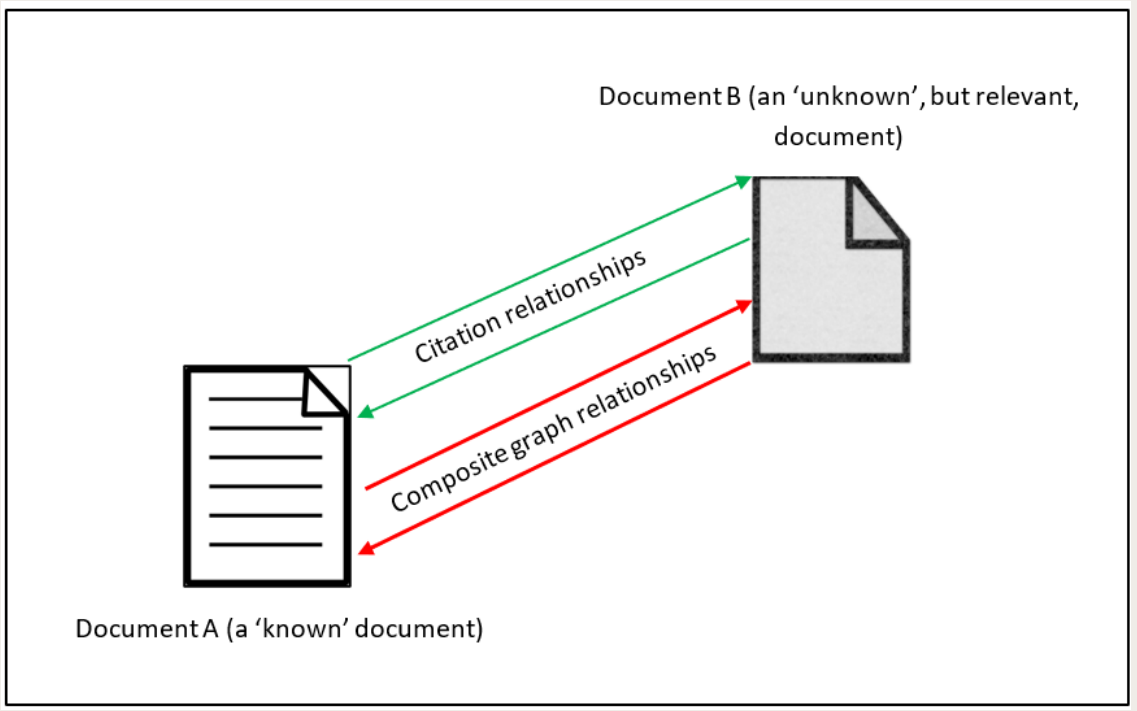
2Dsearch


Source
ERIC
SHARE
REFRESH
HIDE RESULTS
RESULTS
MESSAGES
QUERY
Combined search strategy
AND
Intervention
OR
Condition
OR
title
exercise
obesity
Suggestions (4)
(from DBPEDIA and word vectors)
Based on "overweight"
SELECT ALL
adolescents
adults
diabetes
diabetic
malnourished
obese
obesity
obesity
Statistics
10 terms (of 10 different types)
0 field tags
4 operators (of 2 different types)
URL
https://eric.ed.gov/?q=(title%20%3Aexercise OR %22physical%20activity%22 OR sport) (obesity OR obese OR overweight) (adolescent* OR child OR children OR teenager* OR youth OR %22young%20people%22)
Query string (for ERIC)
(title :exercise OR "physical activity" OR sport) (obesity OR obese OR overweight) (adolescent* OR child OR children OR teenager* OR youth OR "young people")
Syntax guide for ERIC

OpenAlex via EPPI-Reviewer

EPPI Reviewer

OpenAlex



https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3818

OpenAlex via EPPI-Reviewer

«Seed articles» from screened results as input

These tests suggest that studies retrieved from OpenAlex were more than three times as likely to be relevant than those identified from traditional searches.
(Borge 2023)

EPPI Reviewer

OpenAlex

Automation of searches for evidence syntheses

Some challenges

Automation of searches for evidence syntheses

Some challenges

- Trust and control
- Transparency and explainability
- Reproducibility and replicability

Time spent searching?

Time spent searching

	# Participants (# Reviews)	Average
Bullers 2018	105 (1-500)	30.7 h (SD=30,0)
Saleh 2014	17 (1 ≥ 30)	24 h (incl. 7 h on grey lit.)
Gann 2013	9 (17)	23 h
Clark 2020	1 (1)	5.4 h
Bramer 2018	1 (37)	1 h 13 min

Time spent searching

	# Participants (# Reviews)	Average	Median
Bullers 2018	105 (1-500)	30.7 h (SD=30,0)	22 h (2 –219 h)
Saleh 2014	17	24 h (incl. 7 h grey lit.)	7.9 h (2 – 113 h) (grey: 20 min – 58 h)
Gann 2013	9 (17)	23 h	
Clark 2020	1 (1)	5.4 h	
Bramer 2018	1 (37)	1 h 13 min	1 h (20 min – 3 h 10 min)

Variability

Factors that impact time spent searching

- Characteristics of the studies
  - mode of data collection
  - IR tasks included
- Level of librarian experience
- Search methods, including tools & automation
- Number of search hits
- Topic
  - terminology
  - number of databases/sources needed
  - portion of non-journal research output
  - complexity
  - breadth of scope

Next steps at NIPH?

Next steps at NIPH

- Digital tools and automation – to be continued
- Review, standardize, and improve our current workflow

Summary

- Large variability in time spent searching – 1-200 hours
- Many human-in-the-loop information retrieval tools
- Artificial intelligence challenges important values in evidence synthesis
- “Make haste slowly”

- Thanks for having me!

References

- Borge, TC, Ames H, Jardim PJ, Meneses-Echavez, JF, Himmels J, Rose C, Hestevik, C, Muller AE. Implementation of machine learning in evidence syntheses in the Cluster for Reviews and Health Technology Assessments: Final report 2021-2022 [Implementering av maskinlæring i kunnskapsoppsummeringer i klynge for vurdering av tiltak: Sluttrapport 2021-2022] Oslo: Folkehelseinstituttet, 2023
- Bramer WM, Rethlefsen ML, Mast F, Kleijnen J. Evaluation of a new method for librarian-mediated literature searches for systematic reviews. Res Synth Methods. 2018;9(4):510-520. doi: 10.1002/jrsm.1279
- Bullers K, Howard AM, Hanson A, Kearns WD, Orriola JJ, Polo RL, Sakmar KA. It takes longer than you think: librarian time spent on systematic review tasks. J Med Libr Assoc. 2018;106(2):198-207. doi: 10.5195/jmla.2018.323
- Clark J, Glasziou P, Del Mar C, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. J Clin Epidemiol. 2020;121:81-90. doi: 10.1016/j.jclinepi.2020.01.008
- Gann LB, Pratt GF. Using library search service metrics to demonstrate library value and manage workload. J Med Libr Assoc. 2013;101(3):227-9. doi: 10.3163/1536-5050.101.3.015

References (2)

- MacFarlane A, Russell-Rose T, Shokraneh F. Search strategy formulation for systematic reviews: Issues, challenges and opportunities. Intelligent Systems with Applications. 2022:200091. doi: 10.1016/j.iswa.2022.200091
- Muller AE, Ames H, Himmels J, Jardim PJ, Nguyen L, Rose C, Van de Velde S. Implementation of machine learning in evidence syntheses in the Cluster for Reviews and Health Technology Assessments: Final report 2020-2021. Oslo: Norwegian Institute of Publich Health, 2021. Available from: https://www.fhi.no/en/publ/2021/maskinlaringslagets-sluttrapport-2020-2021/
- Nguyen L, Hafstad E, Johansen M, Harboe I. Mål, funn og forslag til satsningsområder for automatisering av informasjonsgjenfinning: sluttrapport 2022. [Aims, findings, and suggested target areas for automation of information retrieval: final report 2022]. Oslo: Folkehelseinstituttet, 2023. Available from: https://www.fhi.no/en/publ/2023/Aims-findings-suggested-target-areas-for-automation-of-information-retrieval/
- Saleh AA, Ratajeski MA, Bertolet M. Grey Literature Searching for Health Sciences Systematic Reviews: A Prospective Study of Time Spent and Resources Utilized. Evid Based Libr Inf Pract. 2014;9(3):28-50. doi: 10.18438/b8dw3k

elisabetvivianne.hafstad@fhi.no