Resources to accompany DLF Workshop: Beyond Boolean Search

Anna Willoughby, Dr. Kaylee Arnold; 12 November 2023

Synthesizing literature is a primary task for scientists. Below are approaches and tools to aid in (I) locating and (II) organizing literature lists.

I. Locating literature

1. Systematic searches. Be aware each database has a unique corpus and the best strategy is to iterate across several. Retain your queries and outputs; *citation managers a must*
    a. Decide on target subscription databases through your library
        - Databases relevant for natural history and taxonomy:
            - Web of Science. You can explore different journals included through the *free* Master Journal List (https://mjl.clarivate.com/)
                - Zoological Record
                - Biological Abstracts (BIOSIS),
            - Centre for Agricultural Bioscience International (CABI)
            - Wildlife and Ecology Studies
            - Medical Literature Analysis and Retrieval System Online (MEDLINE),
            - Excerpta Medica Database (EMBASE)
            - Science Citation Index (SciSearch)
            - Google Scholar (https://scholar.google.com/)

        ii. Iteration software available (*beware of pagination limits!*)
            - Publish or Perish (https://harzing.com/resources/publish-or-perish)

    b. Collate relevant journals to query or manually parse

        i. Repositories of Journal titles:
        - Web Of Science Master Journal List (*Need a free account* (https://mjl.clarivate.com/collection-list-downloads).
        - Wikipedia/Wikidata (https://en.wikipedia.org/wiki/Lists_of_academic_journals)

        For relevant journals, systematically review table of contents

    c. Perform query through public repositories

        i. OpenAlex (https://openalex.org/):
        - Biodiversity Heritage Library
        - Internet archive (https://archive.org/):

2. Snowballing. This method builds from “perfect papers”, so you can either start with a target paper or utilize references identified in your systematic search above.

    i. *Citation chaining*
    - Citation trees (https://www.citationtree.org/) - this is a citation network which will show you the cited references in **a target paper** as well as references that cite that paper.
    - InCiteful (https://inciteful.xyz/) - this is a citation network which will show you the cited references in **a target paper** as well as references that cite that paper. + beyond

- Local Citation Network (https://localcitationnetwork.github.io/) - this website provides citation network and **co-authorship networks** of **full reference lists.** By looking at outliers, you may find an author or reference that you should perform further citation diving on.

ii. Person- an Place-based snowballing

- Query from authors (person-based)
  - Unique author identifiers: ORCID ID; Wikidata
  - Author-affiliate websites: ResearchGate.net or Academia.edu
  - Co-author networks: Local Citation Network
- Query from author-affiliations or local publishers
  - Check local university and college websites and libraries for thesis or dissertation repository

3. Text mining. Traditional bibliographic database queries are limited to search-term detection in titles, abstracts or keywords. Recent NLP approaches allows word detection in full-texts.

i. General Index (https://archive.org/details/GeneralIndex) - this repository breaks down millions of papers into n-grams for document identification. *Storage capacity needs (7 TB) makes working with this dataset difficult.*

II. Organizing literature lists

Scientists utilize literature lists in a variety of ways. This includes:

1) storing literature locally for reference during research, often in a citation manager.
2) documenting literature sources for data provenance within flat files
3) publishing literature lists in standard bibliographic form with the final product

These different uses require file storage in different file formats, including .txt, .doc, .bib, or .csv.

1. Transforming text into parsed bibliographic information (.bib)
   a. AnyStyle (https://anystyle.io/) - keep literature lists <100 references
2. Checking for digital object identifier (doi)
   a. CrossRef Simple Text Query (https://www.crossref.org/documentation/retrieve-metadata/simple-text-query/)
3. Transforming bibliographic information (.BibTex) into parsed text (.csv)
   a. https://paperpile.com/t/ris-to-csv-converter/
4. Scaling literature reviews abstract screening: metagear through R

Further reading:

Booth, A., Sutton, A., and Papaloannou, D. 2016. Successful approaches to a successful literature review. SAGE Publications. ISBN 9781473952812.

Dekkers, R., Carey, L. and Langhorne, P. (2022). Making Literature Reviews Work: A Multidisciplinary Guide to Systematic Approaches. Springer, London.

Pickering, C., & Byrne, J. (2014). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development,* 33(3), 534-548.

Wohlin, C., Kalinowski, M., Felizardo, K. R., and Mendes, E. (2022). Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. *Information and Software Technology*, *147*, 106908.