

Tools for initially scoping the literature

Terena Solomons, Research Librarian – Western Australian Group for Evidence Informed Healthcare Practice, a Joanna Briggs Institute Centre for Excellence
t.solomons@curtin.edu.au

Identifying and selecting appropriate keywords and indexed terms is crucial for developing robust search strategies to locate studies for systematic and scoping reviews. At the HLA PD online event in November 2020 I gave a 5 minute Lightning Talk on some of the text-mining and other tools I use for initially scoping the literature to gather keywords and indexed terms prior to developing search strategies for systematic and scoping reviews. This article is an adaptation of that presentation, the recording of which is available:

<https://www.alia.org.au/groups/HLA/hla-professional-development-education-and-training>

For the past 7 years I have worked as a Research Librarian with the Western Australian Group for Evidence Informed Healthcare Practice, conducting literature searches for JBI Systematic and Scoping Reviews plus Evidence Summaries on a variety of topics.

Using the following Scoping Review protocol as a case study I will give an overview of text-mining and other tools used to assist in developing the final search strategy (which is in Appendix 1 of the protocol):

Capper, T., Brown, J., Donovan, H., Hegney, D., Williamson, M., Cusack, L., Solomons, T., & Wilson, S. (2020). Individual and environmental factors that influence longevity of newcomers to nursing and midwifery: a scoping review protocol. *JBI Evidence Synthesis*, 18(6), 1271-1277. <https://doi.org/10.11124/jbies-20-00003>

The objective of this scoping review was to identify and map the literature of why nurses and midwives leave or remain in the profession within the first three years of practice.

The mnemonic for Scoping Reviews is PCC – Population, Context, Concept.
Population: newcomers to nursing & midwifery within the first 3 years of graduating
Concept: Intention to stay or leave the profession
Context: Any clinical setting, hospitals, clinics etc

English is such a rich and varied language and often some words can have multiple meanings. For example the word 'retention'. It can mean retention of staff in a human resources context, but it is also used for urinary retention, fluid retention,

knowledge retention. From my initial scoping of the literature I also discovered that the term attrition also appears in the dentistry literature for tooth attrition.

It is really important to be aware of these multiple meanings and it is through the initial scoping of the literature that these are discovered. I searched the term 'career longevity' which unearthed numerous articles on the racing career longevity of racehorses. Type 'career longevity' into pubmed reminer and you'll see the second journal listed is Equine Vet Journal.

Mesh on Demand

For health-related topics, the first text-mining tool I use is MeSH on Demand <https://meshb.nlm.nih.gov/MeSHonDemand>. This is a software tool that analyzes text and matches it against Pubmed data to generate MeSH - Medical Subject Headings. Copy and paste the text from a systematic review protocol or research proposal into the MeSH on Demand tool and it suggests MeSH subject headings and lists 10 articles based on the text entered. It is machine generated so there is no human intervention or evaluation. When selecting the text to enter into MeSH on Demand it is important that it includes the key points and important words. Generally, I'll include the research question and key background information for the review topic. A similar tool to MeSH on Demand is HelioBLAST <https://helioblast.heliotext.com/>.

PubMed Reminer

PubMed Reminer <https://hgserver2.amc.nl/cgi-bin/miner/miner2.cgi> provides a different overview of PubMed data showing publication trends, prolific authors, journal titles, frequent words & MeSH Subject Headings. I use this tool to see who the key authors are, the trends in publishing over the years, MeSH subject headings and frequently occurring words.

Below is a screenshot of PubMed Reminer searching (new graduate nurse OR new graduate midwife) AND (turnover OR attrition OR retention)

The screenshot shows the PubMed PubReMiner interface. At the top, it says "Your query resulted in 361 references". Below this, there are buttons for "Goto PubMed with query" and "Create CV output". A search box contains the query: "(new graduate nurse OR new graduate midwife) AND (turnover OR attrition OR retention)". Below the search box, there is a dropdown for "AbstractLimit: 7500" and a "Search with Manual Adjustment" button.

Below the search box, there is a table with columns: # OR Year, # OR, Journal, # OR, Author, # Count OR, Word, # OR, Mesh. The table lists various authors and journals, along with their counts and the words they are associated with. The Mesh column lists various MeSH terms.

# OR Year	# OR	Journal	# OR	Author	# Count OR	Word	# OR	Mesh
15	2020	J Nurs Adm	8	LASCHINGER HK	325	NURSE*	311	organization & administration
34	2019	J Nurs Staff Dev	6	NURSE P	314	NEW*	201	Humans
30	2018	J Nurs Manag	6	BEHZADIFAR M	291	GRADUATE*	183	psychology
22	2017	Nurse Educ Today	5	LALONDE M	275	NURS*	117	education
26	2016	J Contin Educ Nurs	4	AREMILO	224	RETENT*	100	Personnel Turnover
34	2015	J Adv Nurs	4	ASSADI R	213	HUMAN*	182	Nursing Staff, Hospital
22	2014	Nurse Educ Pract	4	DHARMARATNE SQ	203	EDUCATE*	76	Female
15	2013	Int J Nurs Stud	4	EL-KHATIB Z	194	METHOD*	74	statistics & numerical data
26	2012	J Clin Nurs	4	FISCHER E	177	HEALTH	73	Adult
19	2011	Nurs Econ	4	HAJ-MIRZAJANI	169	ORGANISE*	72	Male
8	2010	J Prof Nurs	4	HALFER D	165	PROGRAM*	71	Job Satisfaction
13	2009	Nurs Leadersh (Tor Ont)	4	HALL BJ	157	CONCLUS*	70	methods
17	2008	J Nurses Prof Dev	4	HASSEN HY	156	DEVELOP*	69	standards
12	2007	AORN J	4	JONAS JB	156	TURNOVER	60	Attitude of Health Personnel

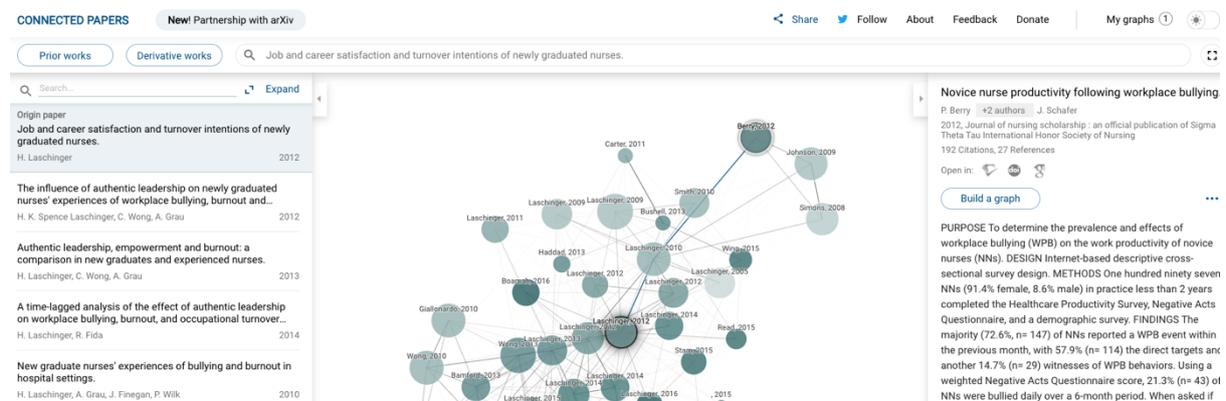
From the PubMed Reminer tool I discovered the author Laschinger (2012) had written about the experiences of newly graduated nurses and had written this highly cited paper: "Job and career satisfaction and turnover intentions of newly graduated nurses".

Connected Papers

This leads me to the next tool which is Connected Papers

<https://www.connectedpapers.com/>. This is a visual tool to help researchers find and explore papers relevant to their field. The papers are arranged according to their similarity. According to the creators of Connected Papers, even papers that do not directly cite each other can be strongly connected based on a "similarity metric based on the concepts of Co-citation and Bibliographic Coupling" (Etian et al, 2020).

Below is a screenshot of the Connected Papers graph where I've entered the Laschinger article "Job and career satisfaction and turnover intentions of newly graduated nurses". In real time when you hover over an author it displays the title. I could see that the group of authors in the top right corner (Johnson, Berry & Simons) were looking at bullying in the workplace. The group of authors in the bottom left of the display were looking at authentic leadership and newly graduated nurses.



I use Connected Papers as a way to initially find more similar articles at the beginning stages of a search.

EndNote Subject Bibliography

In my initial exploration of the literature I add relevant papers to an EndNote Library. EndNote has a feature that allows you to see what the most common keywords are in a group of references. Go to Tools – Subject – Bibliography and select Keywords, it will then generate a list of the keywords and number of times it comes up in that set of papers.

Interpret this list with caution. In this case study the word 'Transition' came up as the most popular term in the set of papers I discovered from initial searches. The

transition of nursing students to the work environment is an important environmental factor. After discussing this with the JBI research team it was decided not to actively search for the term Transition in the search strategy because it gathered too much 'noise' in the search results. It also meant that the search was biased towards finding articles on this factor.

Rayyan

Part of the methodology for systematic and scoping reviews is for 2 researchers to screen the Title & Abstracts of references, then the full text articles for inclusion in the review. This methodology is to mitigate bias. Rayyan (rayyan.ai) is a free software tool used by research teams for this process. I first heard of Rayyan from chatting with a Canadian librarian at the EAHIL conference in 2019. One of her colleagues Martin Morris gives a demonstration of Rayyan in this Webinar: [Management/Analysis Tools for Reviews: Rayyan with Martin Morris](#)

I have used Rayyan for managing the screening process for a number of JBI research teams and find it easy to use. Rayyan has an element of text mining to assist researchers in screening Title & Abstracts. The study by Olofsson (2017, p.280) discovered that the "text-based ranking function of Rayyan effectively assisted reviewers to identify records early in the review process".

I use Rayyan to confirm with the JBI team lead that we are both on the same page and in agreement with which references are "gold standard" or "seed references" from which the search strategy is developed. I export an initial corpus of references from EndNote to Rayyan and invite the team lead to go through the list and tag each reference with Include, Exclude, Maybe. Rayyan has the function to make this process blinded so we are not influenced by each other's decision.

From the included "seed references", the search strategy is grown/developed using the keywords used by authors in the titles and abstracts. I run the titles from these "seed references" through the databases I plan to search to see how these "seed references" are then indexed and note what subject headings (MeSH for Medline, Emtree for Embase etc) are used to describe them.

Acknowledgements

I first learnt about MeSH on Demand and PubMed Reminer text-mining tools from Julie Glanville when attending the HLA's Advanced Searching workshop in Melbourne in 2015.

One of silver linings of COVID has been the ability to attend online seminars offered around the world in real time. When it's daytime in the UK, it's late afternoon or evening in Perth. Last September I attended the "Advanced Search Strategy Design for Complex Topics: Text Analytics and Text Mining" facilitated by expert

searchers Julie Glanville and Carol Lefevbre <https://yhec.co.uk/training/> . This was a fantastic professional development opportunity and I encourage anyone interested in learning more about text-mining tools to consider attending their upcoming webinars in May & June 2021.

References

- Capper, T., Brown, J., Donovan, H., Hegney, D., Williamson, M., Cusack, L., Solomons, T., & Wilson, S. (2020). Individual and environmental factors that influence longevity of newcomers to nursing and midwifery: a scoping review protocol. *JBIE Evidence Synthesis*, 18(6), 1271-1277. <https://doi.org/10.11124/jbies-20-00003>
- Eitan, A., Smolyansky, E., & Harpaz, I. (2020). *Connected Papers*. Retrieved 6 Apr 2021 from <https://www.connectedpapers.com/>
- Laschinger, H. K. S. (2012). Job and career satisfaction and turnover intentions of newly graduated nurses. *Journal of Nursing Management*, 20(4), 472-484. <https://doi.org/10.1111/j.1365-2834.2011.01293.x>
- Olofsson, H., Brolund, A., Hellberg, C., Silverstein, R., Stenstrom, K., Osterberg, M., & Dagerhamn, J. (2017, Sep). Can abstract screening workload be reduced using text mining? User experiences of the tool Rayyan. *Research Synthesis Methods*, 8(3), 275-280. <https://doi.org/10.1002/jrsm.1237>