SNOWBALLER

Computer-assisted snowball search for meta-analysis research

June Choe

Language, Education, and Reading Neuroscience Lab, School of Communication | Linguistics, Weinberg College of Arts and Sciences

The Challenge

The **literature search** process for **meta-analysis** is resource intensive. Much of it involves repetitive tasks done by hand, being prone to human error and requiring a lot of person-hours.

Goal

Develop a tool for ***automated literature search*** that can **assist** in meta-analysis research.

Research Question

How **reliable** and **helpful** are automated approaches compared to traditional search-by-hand methods?

Design Philosophy of *Snowballer*

A software for **snowball search** that runs on open-source scholarly databases, developed in consultation with meta-analysis researchers.

Snowball Search:

A literature search method where the citations and references from a starting set of core papers are followed and repeated for the newly found papers to accumulate relevant primary data sources. This runs until no new relevant papers are found, culminating in a comprehensive coverage of the target literature.

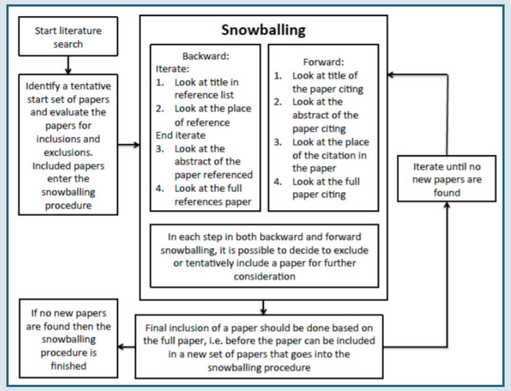


Diagram of the snowball search process from Wohlin (2014)

Validation Experiment

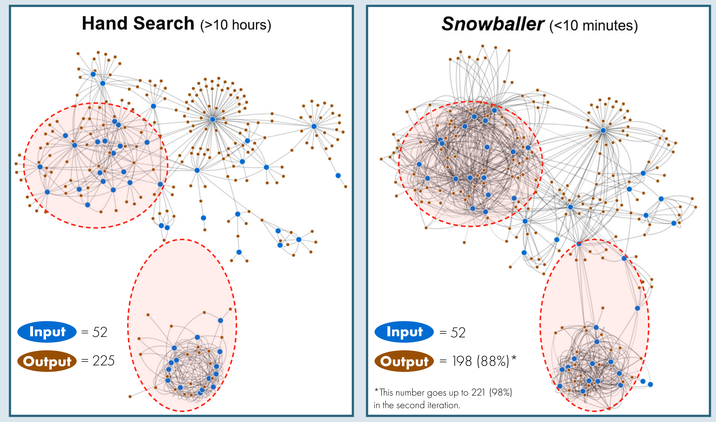
Experiment

The starting set of core papers from a separate meta-analysis project were passed into *Snowballer* and ran for one iteration.

A network of inputs and outputs was generated and compared between the results from the original search and the automated search.

Findings

In addition to having an advantage in **speed**, the comprehensive information about the citations and references between papers returned by *Snowballer* allows a clearer identification of **clusters**, **islands**, and **core papers** from the target literature.



Citation Networks

Pilot Study (Dec 2019 ~ ongoing)

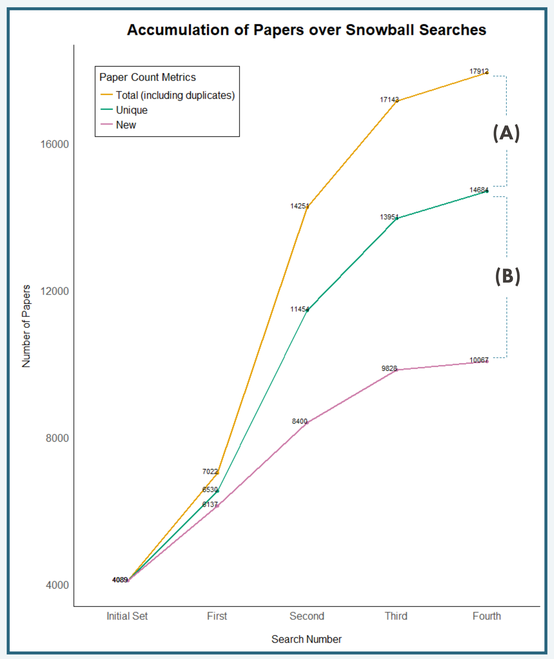
Implementation

Snowballer was used in a meta-analysis project in language development.

Step #1 - Pass a set of papers selected for inclusion as inputs to *Snowballer*.
Step #2 - Record newly found papers returned from *Snowballer*.
Step #3 - Screen papers for relevance based on title, abstract, and full text.
Step #4 - Mark the final set of papers for inclusion in analysis and repeat.

Findings

1. *Snowballer* performs well as a tool for snowball search.
   - Analysis of search data indicates that snowballer can **(A)** successfully narrow in on and **(B)** exhaust the target literature.
2. *Snowballer* assists in other areas of the meta-analysis process.
   - A comprehensive and automatically standardized output improves recordkeeping compared to traditional screen-as-you-search methods.
   - Additional data about the searched papers such as abstract and the number of connections to the input can inform screening decisions.



Accumulation of papers over snowball searches by different metrics

ID	Title	Year	Authors	Journal	Pub_type	DOI	Citations	References	Abstract	Density	Connections
2055886999	Early Identification of Dyslexia: Evidence from a Follow-Up Study of Speech-Language Impaired Children	1991	hugh w catts	annals of dyslexia	Journal Article	10.1007/BF02648084	102	41	A group of speech-language impaired children was administered a battery of standardized language tests and measures of phonological processing in kindergarten. Performance on these language measures was then compared to reading ability in first grade. Results indicated that children with semantic-syntactic language deficits had more difficulties in reading than did children with primarily speech articulation impairments. In addition, phonological processing measures were found to be good predictors of reading achievement. Results are discussed in terms of their implications for the early identification of developmental dyslexia.	3	1994068606, 1977349107, 2062064131

Example of an output returned from Snowballer

Additional data automatically fetched by Snowballer compared to hand search methods.

Conclusions

- *Snowballer* demonstrates that automated approaches can significantly cut down on time and resource without compromising a representative coverage of the target literature.
- The collection of auxiliary data at no additional cost can inform the screening process and improve recordkeeping practices for transparency and reproducibility.
- The retainment of all information used in the search and screening process can pave way for the development of other automated tools, such as text classification models.

Contact
yongchoe2020@u.northwestern.edu

Snowballer Demo

SCAN ME

youtu.be/HCuQLodZSt0

Snowballer Code

github.com/yjunechoe/Snowballer

References

[1] Arnab, S., Zhihong, S., Yang, S., Hao, M., Darrin, E., Bo-June, H., and Kuansan, W. (2015). An Overview of Microsoft Academic Service (MA) and Applications.
[2] Wohlin, C. (2014). Guidelines for Snowballing in Systematic Literature: Studies and a Replication in Software Engineering.

Acknowledgements

Many thanks are due to Professor Elizabeth Norton, Sean McWeeny, Jinnie Choi, and members of the LEARN lab for the intellectual and financial support throughout the duration of this project. This research has benefited from the feedback on the software design and the identification of need from Professor Meg Roberts and researchers in the Early Intervention Research Group. The validation experiment was made possible by an anonymous researcher who graciously offered to share their search data. Snowballer depends on many free and open-access packages published by the rOpensci community.