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.

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.

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 retention 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.

References


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.