

SWIFT: A Text-mining Workbench for Systematic Review

Ruchir Shah, Sciome

During the initial stages of a systematic review, it is often necessary to assess the relevance of tens or even hundreds of thousands of potentially applicable bibliographic citations. Although this is an inherently labor-intensive endeavor, utilization of text- and data- mining algorithms has the potential to dramatically increase the efficiency and speed with which a systematic review can be conducted. Here we introduce SWIFT (Sciome Workbench for Interactive computer Facilitated Text-mining), which is being integrated into the OHAT systematic review process.

Currently, SWIFT incorporates powerful, state-of-the-art methods into an interactive, visual workbench that allows users to search, organize, explore, prioritize, visualize and extract information from very large bibliographic databases. For example, SWIFT can be used to automatically cluster documents and uncover the primary themes discussed within a body of literature. In addition, the software can automatically tag documents according to health endpoint, exposure, evidence stream, or other relevant terminologies. Given a small seed set of known relevant and irrelevant documents, users can also utilize SWIFT to train powerful machine learning models, which can then successfully prioritize the remaining documents according to their estimated relevance. The success of SWIFT's machine learning is currently being benchmarked using historical datasets from previously conducted systematic reviews in which studies were manually screened for relevance by 2 independent screeners. The data suggests that by using SWIFT the required time to screen citations can be reduced by at least 50 percent.

Although SWIFT is under active development, the tool is currently being assessed by OHAT for use during problem formulation to help identify and screen studies for determining whether a systematic review is warranted and/or to prioritize and refine questions addressed in a systematic review. SWIFT is also being used to prepare "scoping reports" for topics to help identify current research trends for complex topics with large literature sets (e.g., environmental influences on epigenetics and the extent of the literature on endocrine disrupting chemicals). The current version of SWIFT focuses on using PubMed abstracts; however, we also have a set of related tools that are designed to work on full text PDF documents, and are currently working on ways to incorporate this functionality into SWIFT. We plan to make a free-of-charge version of SWIFT publically available in Fall 2015.