Centre for Reviews and Dissemination

Finding needles in (giant) haystacks: use of text mining to support study selection by reducing screening workload

Ian Shemilt

Seminar

James Thomas[1] and Ian Shemilt[2]

Tuesday 11 September 2012

1:30 - 2:30pm

A19/20 Alcuin A/B

[1]*Associate Director, EPPI-Centre, Social Science Research Unit, Institute of Education*
[2]*Senior Research Associate, Behaviour and Health Research Unit, University of Cambridge*

Locating and selecting studies are important initial stages in the conduct of systematic reviews and compilation of specialist tertiary literature databases such as NHS EED and DARE. In order to ensure they retrieve all eligible study records, electronic search strategies sometimes need to be highly sensitive and are usually executed in multiple databases. Conventionally, the process of selecting eligible studies requires the manual screening of all records retrieved. However, when the number of records retrieved is large (or very large), this process becomes time consuming (or impractical). In these circumstances, the use of text mining technologies to prioritise records for manual screening can expedite study selection and reduce screening workload.

This seminar will introduce the use of text mining to support study selection by drawing on the examples of large-scale scoping reviews of public health evidence, in which electronic searches retrieved >800,000 and >1 million records. In these reviews, the use of text mining reduced manual screening workload by 90% and 88% compared with conventional methods. This equated to absolute reductions in screening workload of >430,000 and >378,000 records (i.e. if we had used conventional methods, we would have needed to screen >430,000 and >378,000 additional title and abstract records to identify the same numbers of potentially eligible records as we identified in practice by screening records prioritised by text mining). Discussion will cover strengths and limitations of this approach and potential applications in reviews and related activities undertaken by CRD and its affiliates.