Research White Paper

EPC Methods: An Exploration of the Use of Text-Mining Software in Systematic Reviews

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Research White Paper

EPC Methods: An Exploration of the Use of Text-Mining Software in Systematic Reviews

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No.:
290-2012-00004-C (SRC)
290-2012-00009-I (Vanderbilt)
290-2012-00011-I (ECRI)

Prepared by:
Scientific Resource Center
Portland, OR

Investigators:
Robin Paynter, M.L.I.S.
Lionel L. Bañez, M.D.
Elise Berliner, Ph.D.
Eileen Erinoff, M.S.L.I.S.
Jennifer Lege-Matsuura, M.S.L.I.S.
Shannon Potter, M.L.I.S.
Stacey Uhl, M.S.S.

AHRQ Publication No. 16-EHC023-EF
April 2016

This report is based on research conducted by the Scientific Resource Center, Vanderbilt, and ECRI Evidence-based Practice Centers (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract Nos. 290-2012-00004C [SRC], 290-2012-00009I [Vanderbilt], and 290-2012-00011I [ECRI]). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help health care decision makers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

Suggested citation: Paynter RA, Bañez LL, Berliner E, Erinoff E, Lege-Matsuura J, Potter S, Uhl S. EPC Methods: An Exploration of the Use of Text-Mining Software in Systematic Reviews. Research White Paper. (Prepared by the Scientific Resource Center and the Vanderbilt and ECRI Evidence-based Practice Centers under Contract Nos. 290-2012-00004-C [SRC], 290-2012-00009-I [Vanderbilt], and 290-2012-00011-I [ECRI].) AHRQ Publication 16-EHC023-EF. Rockville, MD: Agency for Healthcare Research and Quality; April 2016. www.effectivehealthcare.ahrq.gov/reports/final/cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although they may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

If you have comments on this Methods Research Project they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by e-mail to epc@ahrq.hhs.gov.

Sharon B. Arnold , Ph.D.
Acting Director
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Arlene S. Bierman, M.D., M.S.
Director
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Lionel L. Bañez, M.D.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Elise Berliner, Ph.D.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Health Research and Quality

Key Informants

In designing the study questions, the EPC consulted eight Key Informants (KIs), who were chosen to represent different perspectives on the use of text-mining tools within the systematic review process (i.e., from an organizational/programmatic perspective and from an end-user information specialist/research team member perspective). Informants are not involved in the analysis of the evidence or the writing of the report. Therefore, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual KIs.

KIs must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their role as text-mining software users, individuals with potential conflicts may be retained. The Task Order Officer and the workgroup endeavored to balance, manage, or mitigate any conflicts of interest.

The list of Key Informants who provided input to this report follows:

Ahmed Abou-Setta, M.D., Ph.D.
Assistant Professor
University of Manitoba
Winnipeg, Manitoba, Canada

Wichor M. Bramer, M.Sc.
Biomedical Information Specialist
Medical Library
Erasmus Universiteit
Rotterdam, the Netherlands

Julie Glanville, M.Sc.
Associate Director
York Health Economics Consortium Ltd.
University of York
York, United Kingdom

Elke Hausner, M.Sc.
Research Associate
Information Management
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Cologne, Germany

Mike Kelly, Ph.D., Hon. FRCP, FRCPE, FFPH
Senior Visiting Fellow
Department of Public Health and Primary Care
Institute of Public Health
University of Cambridge
Cambridge, United Kingdom

Claire Stansfield, M.Sc.
Information Scientist
Evidence for Policy and Practice Information and Coordinating-Centre
University College London
London, United Kingdom

James Thomas, Ph.D., M.A., M.Mus., GGSM
Associate Director
Evidence for Policy and Practice Information and Coordinating-Centre
University College London
London, United Kingdom

Thomas Trikalinos, M.D., Ph.D.
Associate Professor and Director
Center for Evidence-based Medicine
Brown University
Providence, RI, USA

Acknowledgments

The authors gratefully acknowledge the continuing support of our AHRQ Task Order Officers, Lionel L. Bañez, M.D., and Elise Berliner, Ph.D. In addition, we thankfully acknowledge Michael Phillips, Kitty Donahue, Helen Dunn, Tracey Monastero-Stem, and Jessica Gontarek for their document preparation support.

Peer Reviewers

Prior to publication of the white paper, we sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential nonfinancial conflicts may be retained. The Task Order Officer and the EPC work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified

The list of Peer Reviewers follows:

Karen A. Robinson, Ph.D., M.Sc.
Director
Johns Hopkins University Evidence-based Practice Center
Baltimore, MD

Nila Sathe, M.A., M.L.I.S.
Assistant Director
Vanderbilt University Evidence-based Practice Center
Nashvillle, TN

EPC Methods: An Exploration of the Use of Text-Mining Software in Systematic Reviews

Structured Abstract

Objective. This project's goal was to provide a preliminary sketch of the use of text-mining tools as an emerging methodology within a number of systematic review processes. We sought to provide information addressing pressing questions individuals and organizations face when considering utilizing text-mining tools.

Methods. We searched the literature to identify and summarize research on the use of text-mining tools within the systematic review context. We conducted telephone interviews with Key Informants (KIs; n=8) using a semi-structured instrument and subsequent qualitative analysis to explore issues surrounding the implementation and use of text-mining tools. Lastly, we compiled a list of text-mining tools to support systematic review methods and evaluated the tools using an informal descriptive appraisal tool.

Results. The literature review identified 122 articles that met inclusion criteria, including two recent systematic reviews on the use of text-mining tools in the screening and data abstraction steps of systematic reviews. In addition to these two steps, a preliminary exploration of the literature on searching and other less-studied steps are presented. Support for the use of text-mining was strong amongst the KIs overall, though most KIs noted some performance caveats and/or areas in which further research is necessary. We evaluated 111 text-mining tools identified from the literature review and KI interviews.

Conclusions. Text-mining tools are currently being used within several systematic review organizations for a variety of review processes (e.g., searching, screening abstracts), and the published evidence-base is growing fairly rapidly in breadth and levels of evidence. Several outstanding questions remain for future empirical research to address regarding the reliability and validity of using these emerging technologies across a variety of review processes and whether these generalize across the scope of review topics. Guidance on reporting the use of these tools would be useful.

Contents

Introduction

Background

Systematic reviews have been defined as “attempts to collate all empirical evidence that fits prespecified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn.”[1] While the production of systematic reviews is a cornerstone of evidence-based practice, the cost and time required to conduct many systematic reviews are concerns.[1,2] In recent years, the Agency for Healthcare Research and Quality (AHRQ) Effective Health Care Program through the Evidence-based Practice Center (EPC) Program has been engaged in methodologic research on how systematic reviews can be conducted more efficiently, including research on rapid review methods, the utilization of text mining/machine learning in the screening process and for updating reviews, and the development of Abstrackr.[3-5]

For the purposes of this white paper, we adopted Thomas et al.’s broad definition of text mining within the systematic review context as a method “to retrieve information from unstructured text and to present the distilled knowledge to users…. [which] comprises three major activities: information retrieval which retrieves texts relevant to the user’s query; information extraction which identifies and extracts snippets of textual fragments related to the query; and data mining, which finds direct and indirect associations among the pieces of information extracted from texts.”[6] For readers interested in more specific categories of tools and their functionality, Miner et al. have defined the following seven areas of practice within the text-mining field:[7]

- **Document classification** – grouping and categorizing snippets, paragraphs, or documents, using data mining classification methods, based on models trained on labeled examples
- **Document clustering** – grouping and categorizing terms, snippets, paragraphs, or documents, using data mining clustering methods
- **Information retrieval** – storage and retrieval of text documents, including search engines and keyword search
- **Concept extraction** – grouping of words and phrases into semantically similar groups
- **Information extraction** – identification and extraction of relevant facts and relationships from unstructured text; the process of making structured data from unstructured and semistructured data
- **Natural language processing** – low-level language processing and understanding tasks (e.g., tagging part of speech); often used synonymously with computational linguistics
- **Web mining** – data and text mining on the Internet, with a specific focus on the scale and interconnectedness of the web

Table 1 offers a brief overview of how text-mining tools have been used in various systematic review processes, followed by their relative advantages and disadvantages compared to traditional methods.

Table 1. Text-mining tool use(s) and advantages/disadvantages by SR process step

SR Process Step	How TM Used	Advantages / Disadvantages
Literature Search	Identification of: - keywords - synonyms - subject terms Filter creation	**Identification of keywords, etc.:** Advantage: Supports easy review of a far larger corpus of preliminary results for identification of keyword/subject search terms than would be feasible otherwise (potential to improve time efficiency, use of reproducible/objective method, and improved search strategies) Disadvantage: Most TM tools geared to search Medline/PubMed so will likely not be useful for topics in other disciplines **Filter creation:** Advantage: Creation of reusable tool to identify citations in a database (use of reproducible/objective method and improved search strategies) Disadvantage: 1) Filter development takes time that may/may not be warranted given search topic 2) Filters tend to be very sensitive and may return too many results to be useful to review team
Screening Citations	Prioritization of most relevant citations first Fulfilling second screener role	**Prioritization:** Advantage: Relevant citations are displayed first for screening review, so review team can begin work on these while completing review of all citations (potential to improve process) **Second Screener:** Advantage: TM tool determines relevant citations and compares against human screener's selections (potential for time efficiency) Disadvantage: Potential for missing relevant citations
Abstracting Data	Information extraction: - Population - Intervention - Condition - Outcome	Advantage: Validation of human extracted data (potential for time efficiency and increased accuracy) Disadvantage: Not currently ready to be used without human oversight; additional TM tool development and evaluation required
Quality Appraisal	Risk of bias	Advantage: Validation of human appraisal (potential for time efficiency and increased accuracy) Disadvantage: Not currently ready to be used without human oversight; additional TM tool development and evaluation required
Review Updating	Identification of new studies	Advantage: Potential for time efficiency Disadvantage: Not currently ready to be used without human oversight; additional TM tool development and evaluation required

TM = text mining; SR = systematic review

As of this writing, official methods guidance on the use of text-mining tools in systematic reviews has yet to be issued by any of the following prominent organizations in their methods publications: AHRQ Effective Health Care Program (*Methods Guide for Effectiveness and Comparative Effectiveness Reviews*), Campbell Collaboration (*Campbell Collaboration Systematic Reviews: Policies and Guidelines*), Cochrane Collaboration (*Cochrane Handbook for Systematic Reviews of Interventions*), Evidence for Policy and Practice Information and Coordinating Centre (*Methods for Conducting Systematic Reviews*), Institute of Medicine (*Finding What Works in Health Care: Standards for Systematic Reviews*), or the Joanna Briggs Institute (*Joanna Briggs Institute Reviewers' Manual*).[1,2,8-11] While the European Network for Health Technology Assessment's *HTA Core Model® version 2.1; 2015* does not contain guidance on text-mining, it recently published another guideline titled, *Process of Information Retrieval for Systematic Reviews and Health Technology Assessments on Clinical Effectiveness*, that advises using word frequency analysis to develop search strategies.[12,13] In the future, such guidance will hopefully be published to give greater clarity to when its use is advisable and standards for reporting its use in systematic reviews.

Purpose of This Report

This project's overall aim was to investigate the use of text-mining tools as an emerging methodology within the context of systematic reviews. This preliminary sketch covers three core areas:

1. To describe the state of published evidence on the use of text mining within systematic review processes (i.e., searching, abstract screening);
2. To better understand issues arising from the use of text-mining technologies from an organizational perspective (i.e., senior investigators) and from a systematic review team member (i.e., information specialists) perspective;
3. To identify and create a core list of text-mining tools that have been used within a systematic review, develop a descriptive tool to broadly characterize them (e.g., "trialability," free vs. fee, algorithm type), and apply that to the list of tools.

Methods

General Approach

This project's overall aim was to provide a snapshot of the state of knowledge on the use of text mining in systematic reviews, providing groundwork for future methodologic work in this area. Given the project's exploratory nature, we adopted a multipronged approach to illustrate how text-mining tools have supported various steps in systematic review processes and, secondarily, different types of reviews. We conducted a literature review to identify existing research on text-mining use. We augmented this information with insights from Key Informants to capture senior investigator/organizational perspectives and information specialist/research team member perspectives on text mining. Lastly, we provided a descriptive evaluation of specific text-mining tools/software used to support systematic review processes.

A workgroup composed of members from the EPCs, the Scientific Resource Center (SRC), and AHRQ participated in weekly workgroup teleconference calls over a three-month period to discuss the direction and scope of the project, assign and coordinate tasks, collect and analyze data, and discuss and edit draft documents. The workgroup consisted of three professional librarians (EPC and SRC members), an EPC Project Manager, an EPC Senior Analyst, and two AHRQ Task Order Officers.

Initially, this exploratory research project intended to cover all steps within the systematic review process equally across the literature review and interviews; however, our emphasis changed early on because we found several recent existing systematic reviews that covered screening and data abstraction.[14-16] Thus, this preliminary sketch of the use of text-mining tools within systematic review processes will attempt to more comprehensively cover searching and other less well-studied steps while summarizing the existing systematic reviews.

Text mining covers various techniques and tools used to detect patterns and extract knowledge from unstructured natural language text. Text mining uses statistical approaches to explore (e.g., co-occurrence, frequencies of words) and categorize (e.g., clustering, classification) text-based information to support knowledge discovery while minimizing human effort. We considered a text-mining tool to be any software or application to aid the process of text mining. We included resources that our Key Informants identified as text-mining tools although they are traditionally used for other purposes (e.g., EPPI-reviewer, EndNote, Microsoft Excel).

Literature Review

We searched a range of bibliographic databases and gray literature sources to identify candidate publications. We limited bibliographic searches by publication date (2005 – 2015) and to English-language publications due to time constraints to complete the research project. In addition to major biomedical databases, we searched the computer/information science literature to improve recall of relevant content. Time constraints precluded a full systematic review; however, we used the following inclusion criteria:

- Does this article address text mining within the context of the systematic review process?
- Does this article address an area of text mining that is of interest to this report?

Publications focused on text mining of electronic health records and administrative datasets (although of interest), were outside the scope of this white paper. Publications that focused solely on technical aspects of text-mining algorithms were excluded.

Our searches identified 1,473 candidate citations. After duplicate removal, 670 unique citations were uploaded to DistillerSR for review. The full text of 122 articles was retrieved for data abstraction. We noted whether text mining was used in the searching, screening, data-extraction, updating, or other parts of the systematic review process.

Given this review's rapid nature, we decided to rely on two 2015 systematic reviews that covered the use of text mining in the screening and data abstraction part of the systematic review process rather than conducting a de novo review for those areas.[14,15] We discuss these reviews and studies focusing on searching and updating processes later in this report. Full details of strategy development, databases searched, and search strings are available in Appendix A.

Key Informant Interviews

We conducted Key Informant (KI) interviews with senior investigators from systematic review research organizations (n=4) and information specialists who have used text-mining tools to develop systematic review search strategies (n=4) to form a preliminary understanding of the experiences and insights of researchers who have used these tools. We decided to split the KIs into two groups for two reasons: 1) We thought the senior representatives were likely to have more experience with the use of text-mining tools in the screening phase of systematic reviews (reflecting the more extensive published literature that exists on this phase overall) and 2) the desire of the workgroup to focus on the searching phase to begin fleshing out the use of text mining in this step in greater detail. In compliance with the Paper Work Reduction Act Office of Management and Budget regulation (5 C.F.R. § 1320), the sample of KIs was limited to nine or fewer nonfederal employees. One of this report's team members conducted the interviews during July and August 2015 using semistructured interview instruments. At least two additional team members also attended each interview.

We identified potential KIs in the following ways: 1) by reviewing authors of relevant published literature, 2) by sending emails to librarian discussion lists to recruit potential participants (i.e., Cochrane IRMG, MLA Expert Searching, HTAi ISG-Info Resources), and 3) via contacts within the systematic review community. We invited 14 individuals to participate as KIs in an (approximately) 60-minute individual telephone interview; eight agreed and were interviewed, and six declined. KIs are listed in the *Key Informants* section of this report and are quoted anonymously in the text. All KIs had experience using text-mining tools in multiple reviews. All information specialist KIs are masters-level medical librarians, so information specialist and librarian are used synonymously hereafter.

All interviews were intended to be audio-recorded and transcribed; however, due to technical issues, two of the interviews were not recorded; for these interviews notes taken by three workgroup members were analyzed instead (these comments appear inside square brackets in Table C-2 to distinguish them from actual quotes). Scientific Resource Center methods research projects fall under the Portland VA Research Foundation Institutional Review Board's blanket ethics nonapplicable exemption; thus, no approval was sought for this project. At the beginning of each call, we asked KIs for permission to record the call for later analysis and to be quoted anonymously; all KIs verbally agreed to these conditions. Each KI completed an "EPC Conflict of Interest Disclosure Form" before being interviewed, and no disclosed conflicts precluded participation by any of the informants. All participants received a copy of the questions ahead of the scheduled conference call.

Interview Guide

The workgroup developed the interview guide through review and discussion over multiple iterations. We developed two separate sets of questions, one for senior investigators/ organizational representatives and one for librarians/research team members. Please see Appendix B for a copy of the interview guide.

Data Analysis

Transcripts were analyzed using a constructivist grounded theory approach in NVivo™ 10 software by one investigator with qualitative analysis experience who developed the descriptive coding structure and themes.[17-21] The larger workgroup reviewed the original transcripts and critiqued the analysis. Please see Appendix C for a table outlining specific text-mining tools with more extended comments by librarians.

Tools Catalog

We compiled a list of text-mining tools identified in the literature and from broad web-based searches. We created a table to summarize features and describe characteristics, accessibility, and potential applications to the systematic review process.

Two team members examined prespecified characteristics and cross-referenced features with those mentioned by Key Informants and identified in the literature search. We elected to focus on components and features likely relevant to topic refinement, literature searching, study selection, and data extraction for systematic reviews. We informally evaluated the potential for a tool feature to support one or more of these key steps of the systematic review. Our subjective assessment of tool utility and relevance was informed by the team's collective experience developing and executing comprehensive literature searches, as well as from the requisite knowledge of the selection, extraction, and appraisal process derived from guidance and standards issued by the EPC Program for conducting comparative-effectiveness reviews and international reporting standards of various stages in a comparative-effectiveness review.[22]

We did not include information-processing products or services (e.g., Doctor Evidence) unless they were mentioned specifically by the Key Informants (e.g., EndNote, DistillerSR, EPPI-Reviewer). We did not examine machine learning or tools designed to extract or describe name relationships exclusively (e.g., genetic and biologic entity recognition).The term "text mining" frequently captures tools designed to extract and classify granular information from the molecular biology literature. Although similar in concept and underlying mechanism, we did not include those in our catalog. Readers who are interested in detailed explanations and comparisons of the component tasks and methods (e.g., preprocessing, context representation, content selection) will find ample information elsewhere, particularly within the bioinformatics, computer and engineering sciences, and biostatics literature.[23-29]

We rated a tool as applicable to systematic reviews if the tool was designed to support systematic review conduct or could be adapted to improve or augment existing systematic review tools or methods. We assessed each tool for functionality to enhance a) topic refinement, scope, or question development; b) searching or retrieval of literature or candidate data; c) screening or eligibility assessment; or d) data extraction or synthesis. We included text-mining tools with features to support overall quality or efficiency of one or more steps in the review process.

Table 2 lists the labels and definitions for the variables that we prespecified for the characterization of text-mining tool features. Given the varying levels of sophistication of tools and the technical support required for installation and/or setup, we did not test the tools or applications for relative performance or precision. As a preliminary assessment, our group focused on availability, capability, stability, and usability. Where possible, we listed key features. We established definitions for the categories and choices to ensure a degree of comparability and enable meaningful classifications. We defined “tool” as any application, resource, software program, software feature, open-source code, or web-based resource intended to automate or facilitate information analysis.

Table 2. Prespecified items to characterize text-mining tools

Item	Description
Name	Name of the tool
Acronym	Acronym or alternate name of the tool
URL	Tool (or file download) URL
Availability	Freely available to any user or proprietary/commercial product
Cost	Cost in U.S. dollars
Web Based	Accessible for use via the internet using a URL
Platform	Supported operating system for tools that require download or installation
Developer	Company, institution, or individual credited as the developer or maintainer of the tool
Multiuser	Capable of supporting multiple concurrent users for collaboration
Literature	Publication describing tool development, validation, or evaluation
Applicability	Tool designed or capable of supporting systematic review conduct/methods
Prior Use in a Systematic Review	Tool used previously to support a systematic reviews
Systematic Review Support	How the tool has been or could be used to support systematic review conduct/methods
Deployment Status	Tool functionality and availability for use (i.e., not in development or pilot status)
Tech Level	Technical expertise or support expected for an average user to install, customize and/or use the tool
Features	Functions and features of text mining
Help	Availability of instructions or documentation to support tool use and/or installation
Comments	Comments
Registration	Tool requires registration or account

Prioritization of Tools Assessment

We prioritized our assessment of tools based on the following: tools that were out of scope or unlikely applicable to systematic reviews were rated as “low” or “no” applicability to the systematic review process (e.g., gene or protein data-mining tools); we did not download or install software to evaluate, focusing for this assessment on those tools that are available via the web. We did not evaluate proprietary products and ceased assessment when testing a strategy or text document did not work properly.

Results

Overview

We searched 18 databases (please see Appendix A for the search strategies) and retrieved 1,473 records. We identified 122 relevant publications. We provide a narrative synthesis of the literature below. Most of the literature we reviewed studied text mining within a single context in the systematic review process. Given the lack of overlap, we present our findings in their typical order within the systematic review process: searching, screening, data extraction, critical appraisal, and updating. Following the literature review section, we summarize the Key Informant interviews and summarize the information from our review of individual text-mining tools.

Literature Review

Searching

Information retrieval is one of the earliest tasks within the systematic review process and has a profound impact on the review's comprehensiveness. A challenge librarians face is identifying the universe of concepts, text words, and controlled vocabulary terms relevant to the review topic. The search strategy's quality depends on the librarian's experience and skill. As Hausner et al. note, concept-based approaches are subjective and depend on the information specialist's knowledge of the topic under investigation. Given the complex nature of many topics, it is difficult to know when a strategy is complete.[30]

One way text mining is applied within the search stage of a systematic review is identification of keywords and controlled vocabulary terms for the search strategy. Typical strategy development involves exploratory searches followed by scrutiny of keywords and indexing by information specialists. Although effective, this process is time-consuming and limited by the librarian's understanding of the topics and controlled vocabularies. It is also difficult to capture this iterative process in the review documentation and thereby affects the transparency of the review process.

In our results set, the most common use of text mining in the systematic review search process was objective topical filter development.[30-41] The specific topics are noted in Table 3. Although the topics studied are not directly related, they have a couple of features in common: They share a complex and diffuse nature that is not covered well by current controlled vocabularies used to index bibliographic databases, and they are also multidisciplinary and require searches of diverse resources to ensure comprehensive retrieval.

Table 3. Topical filters

Publication	Topic
Balan et al.[34]	Cognitive rehabilitation
Damarell et al.[37]	Heart failure
Hausner et al.[30]	Brachytherapy for treating prostate cancer
Iansavichus et al.[39]	Chronic kidney disease
Kok et al.[35]	Prognosis of work disability
Li et al.[38]	Nephrology
O'Mara-Eves et al.[31]	Community engagement
Petrova et al.[36]	Health-related values

Table 3. Topical filters (continued)

Publication	Topic
Shemilt et al.[40]	Choice architecture; Economic environment
Simon et al.[32]	Nurse staffing research
Tanon et al.[33]	Patient safety
Thompson et al.[41]	Transborder drug interventions

There were several general approaches to developing strategies in the literature we reviewed. The first approach assessed word frequency in citations as presented by a stand-alone application. Tools such as PubReMiner provide a user interface to analyze PubMed output.[42] The program generates frequency tables from the results set outlining the number of records by text word, controlled vocabulary heading, year, substances, county, etc. Balan,[34] Kok,[35] and Hausner[30] used this approach in their studies. Tanon,[33] Petrova,[36] Hausner,[30]and Poulter[43] used EndNote, a citation management application, to generate word frequency lists. This technique is limited to use with words appearing in citation titles, abstracts, and controlled vocabulary terms.

The second approach is automated term extraction.[6] This approach can also be used with citations, abstracts, and controlled vocabulary terms but is extensible to the full text of documents. These tools also generate word frequency tables, but many are limited to single word occurrences. This limits the utility since many controlled vocabulary terms are phrases and the relevance of single text words is best assessed in context. Tools such as Antconc, Concordance, and TerMine extract phrases and combination terms. Programs such as MetaMap and Leximancer add a semantic layer to the process by using tools provided through the National Library of Medicine’s Unified Medical Language System. Words and phrases identified in the corpus are expanded through mapping to Metathesaurus concepts and may be clustered according to semantic relationships associated with the concepts.[38]

The tools described in the literature we reviewed (see Table 3, Table 4) employ different algorithms. However, the overall approaches were similar. The first step is creating a developmental set to train the text-mining application. Several methods were used to generate these sets. The most common was creating a corpus of included references from completed systematic reviews on the topic of interest.[31,32] Variations of this approach included manually created sets based on author knowledge and curated bibliographies, reference sets from clinical practice guidelines, and PubMed click-through data.[33,35-38,44] In addition to the training set, another corpus representing the general literature, usually created by randomly sampling citations from PubMed, is also presented to the algorithm. Only words and phrases that are “overrepresented” in the training set are considered for inclusion in the search strategy. For example, Simon et al. included terms from the development set that were prevalent in two percent or less of the population set.[32]

This approach has inherent problems. Petrova et al. note that “the reported frequencies for text words did not necessarily reflect the number of abstracts in which a word appears. It is the latter that would be a true indicator of sensitivity.”[36] Also, “the term extraction algorithm depends on the content of the documents supplied to it by the reviewer.”[31]

Most study groups took a diagnostic framework approach and reported the recall, precision, and number needed to read for their objectively derived strategies. Gold standard comparator groups were generated using PubMed HSR Queries, existing curated subject bibliographies, and strategies used to create existing systematic reviews.[32,33]

Study results, text-mining tools reported in the studies, and other data are presented in Table 4.

Table 4. Searching

Study	Topic	Tools	Metric	Precision†	Sensitivity†	Other†	Complementary to manual development?‡	Comments
O'Mara-Eves et al.[31]	Community engagement	TerMine	C-value >5.0*				Yes	Identified 28.5% of included studies through text-mining terms.
Simon et al.[32]	Nurse staffing levels and patient outcome	TM in R	Word frequency – terms that appear in at least 5% of references	Sensitive: 0.3% Precise: 14.7% Balanced: 1.8%	Sensitive: 100% Precise: 83.3% Balanced: 83.3%	NNR Sensitive: 297 Precise: 7 Balanced: 56	Yes	Word frequency from term document matrix. Worked well at identifying hard to detect concepts (nurse staffing/patient outcome studies). Only worked reliably for single word terms.
Tanon et al.[33]	Patient safety	EndNote Excel	Word frequency	Sensitive: 8.28% Precise: 51.35% Balanced: 44.25%	Sensitive: 100% Precise: 45.78% Balanced: 92.77%	NNR Sensitive: 12 Precise: 2 Balanced: 2 Sensitivity* Precision Sensitive: 8.28% Precise: 23.51% Balanced: 41.05%	Yes	Strategies presented for Medline, EMBASE, and CINAHL separately. Reporting Medline results for most sensitive, precise and balanced approaches in this table. Refer to article for EMBASE and CINAHL statistics. Text-mining used and described but not the focus of the research.
Balan et al.[34]	Cognitive rehabilitation	Anote2 askMEDLINE BioRAT Carrot KH Coder LingerCat Medline Ranker MEDSUM PubReMiner Quertle Text to matrix generator Textpresso VisualText		Not reported	Not reported		Not stated	"Methodologically speaking, we conclude that TM was helpful in getting an overall perspective on a huge corpus of literature with some level of detail, intentionally limited to handle complexity. Richer information can be extracted using more complex TM methods focused on narrower topics, but this requires extensive training and knowledge."

Table 4. Searching (continued)

Study	Topic	Tools	Metric	Precision†	Sensitivity†	Other†	Complementary to manual development?‡	Comments
Kok et al.[35]	Prognosis of work disability	PubReMiner TerMine			Sensitive: 90% Precise: 68%	NNR Sensitive: 20 Precise: 10	Not stated	Used in combination with the Yale methodological filter for prognosis and natural history. Text-mining used and described but not the focus of the research.
Petrova et al.[36]	Health-related values	Concordance EndNote SPSS 14.0	Word frequency	Internal validation:86.6% External validation – food: 63.6% External validation – dentistry: 82.6%	Internal validation: 76.8% External validation – food: 70.1% External validation – dentistry: 47.1%		Yes	Authors analyzed frequency of full MeSH headings and phrases.
Li et al.[38]	Nephrology	MetaMap	Parsed PubMed clinical query click through	Dev: 94.2% Validation: 94.6%	Dev: 87.8% Validation: 91.3%		No	Method for generating topic-specific search filters. Conclude that the automated method is comparable to manually created filters.
Thompson et al.[41]	Transborder interventions for drug control	Leximancer	Word frequency Co-occurrence				Yes	Links concepts and themes graphically.
Iansavichus et al.[39]	Chronic kidney disease	Not specified					Yes	Automation used to create strategies from manually selected terms. No details of automation included in study.
Choong et al.[45]	Not topic specific	ParsCit Microsoft Academic Search		Citations: 97.7% Abstracts: 92.1% Full text: 91.9%	Citations: 66.7% Abstracts: 54% Full text: 53.3%	F1 Score Citations: 0.793** Abstracts: 0.681 Full text: 0.674	Yes	Automated snowballing.
Hausner et al. 2012[30]	Brach-therapy for treatment of prostate cancer	TM in R EndNote PubReMiner	Word frequency				Yes	

Table 4. Searching (continued)

Study	Topic	Tools	Metric	Precision†	Sensitivity†	Other†	Complementary to manual development?‡	Comments
Hausner et al. 2015[46]	Not topic specific	Wordstat EndNote AntConc	Word frequency, z-score >20		Objective: 96% Comparator: 86% (94% when 1 SR was excluded from the analysis)	Noninferiority test	Yes	Goal was establishing noninferiority. Prospective study in progress.
Damarell et al.[37]	Heart failure	Concordance		Post-hoc precision estimate: 75%	Test: 98.2% Validation: 97.8%		Yes	Used clinical practice guidelines in place of systematic reviews to create training set.
Poulter et al.[43]	Not topic specific	MScanner				ROC areas between 0.97 and 0.99	Yes	Classifier to create training sets for text mining. Use case is automated updating of large topical bibliographies.

*See Frantzi et al. for a detailed description of this metric.[47]
**See Powers for a detailed description of this metric.[48]

CINAHL = Cumulative Index to Nursing and Allied Health Literature; MeSH = medical subject heading; NNR = number needed to read (number of papers needed to read to identify a relevant paper); ROC = receiver operating characteristic; SR = systematic review; TM = text mining

† Sensitive strategies are maximized for comprehensive recall. Precise (specific) searches are maximized for relevant recall. Balanced searches are intended to maximize recall without sacrificing relevance. They are not as comprehensive as a sensitive search, but are more comprehensive than precise searches.

‡Complementary to manual development means that text mining methods were used in addition to traditional strategy development by information professionals.

All the studies represented in Table 3 found benefit in automating term selection for systematic reviews, especially those comprising large unfocused topics. Balan et al. found that "the benefits of text-mining are increased speed, quality and reproducibility of text process boosted by rapid updates of the results."[34] They also found that text-mining "revealed trends in big corpora of publications by extracting occurrence frequency and relationships of particular subtopics."[34] Petrova et al. similarly noted, "Word frequency analysis has shown promising results and huge potential in the development of search strategies for identifying publications on health-related values. Other "diffuse topics, such as change (both in healthcare organizations and of health behaviors), communication, social support, learning, and teaching may also lend themselves to effective exploration for the purposes of search strategy design through these or similar techniques for the field of the health information sciences."[36] In their studies, Hausner et al. proposed and validated that the objective approach to creating search strategies was noninferior to the manual conceptual approach.[30,46] Search documentation for reviews using the automated approach include the word frequency tables and seed references included in the test set. The group is currently running a prospective head-to-head study comparing these methods.

Text mining can be incorporated at various points in search strategy development. Although most of the literature describes identification of keywords for the strategy, Choong et al. suggest an automated text-mining approach to "snowballing." "Snowballing" is the process in which relevant references cited in retrieved literature are added to the search results and usually is performed after the main literature search is completed. Choong et al. found that "Snowballing is automatable and can reduce the time and effort of evidence retrieval. It is possible to reliably extract reference lists from the text of scientific papers, find these citations in scientific search engines, and fetch the full text and/or abstract."[45]

Although it seems promising, text mining has not become a standard tool for creating systematic review search strategies. Simon et al. note that "the described development process for an empirical search strategy is a useful – though technically demanding – approach to building performance-oriented strategies."[32] Balan et al. concluded, "Methodologically speaking, we conclude that text-mining was helpful in getting an overall perspective on a huge corpus of literature with some level of detail, intentionally limited to handle complexity. Richer information can be extracted using more complex text-mining methods focused on narrower topics, but this requires extensive training and knowledge." They also commented that "A decision factor to use text-mining relates to how profitable and how difficult the tools may be."[34]

One common limitation we observed in the literature was that many of the tools depend on output from PubMed/MEDLINE. Citations retrieved from this resource are important for systematic reviews but do not represent the entire population of literature relevant for health-care-related systematic reviews. Other limitations are related to the nature of the literature base itself. For example, extraction tools that do not use semantic expansion may miss relevant studies. Damarell et al. found that although their filter improved recall for heart failure–related topics, some studies were missed because they mentioned specific symptoms/syndromes rather than the underlying condition.[37]

Most authors recommend incorporating text-mining processes as an adjunct to employing experienced information professionals. O'Mara-Eves et al. conclude that text mining "should never be used on its own but rather in conjunction with the expertise and usual processes that are followed when developing a search strategy."[31] Interestingly, some authors argue that when an objective approach to text-mining is applied, further approaches such as obtaining expert knowledge or reading background literature may no longer be necessary to develop reliable search strategies.[30,46]

Screening

After searching, the next step in the systematic review process is screening the retrieved citations for relevancy to the research questions. This requires analysts to review each retrieved item and compare it to a predetermined list of inclusion and exclusion criteria. The full text of included citations is obtained for further review, data abstraction, and analysis.

O'Mara-Eves et al. published a systematic review on this topic in January 2015.[14] Because of this review's currency and comprehensiveness, we are using it as the basis for our review of text mining in the systematic review screening process.

The O'Mara-Eves et al. review comprises 44 studies (27 retrospective studies, 17 prospective studies). Across these studies, text mining was incorporated into the screening process for multiple purposes. One major use was prioritizing citations for manual screening. This had the advantage of human review of all the citations, but it provided efficiencies by presenting the most relevant citations first. This concentrated the document-retrieval activity earlier in the process so data abstraction could proceed in tandem with review of the machine-designated "less relevant" citations. Some programs used visualization methods to group "like" citations. This allowed researchers to more rapidly assess the citation groups and make inclusion/exclusion decisions. Another variation on this method was rating the difficulty of screening individual citations. More challenging citations would be assigned to more experienced researchers, again speeding the overall process.

Some studies reported using text-mining techniques for automated citation inclusion/exclusion decisions. Most commonly, the automated screening would fulfill the role of second screener to meet recommendations for dual screening of citations.

As mentioned in the searching section of this report, text mining is highly dependent on the set of citations used to train the algorithm. O'Mara-Eves et al. define active learning as "an iterative process whereby the accuracy of the predictions made by the machine is improved through interaction with reviewers."[14]

Creating training sets for systematic review screening provides challenges not present in other text-mining use cases. Because of the comprehensive nature of systematic reviews, search retrieval tends to include many more irrelevant than relevant citations, leading to "imbalanced datasets." This problem has been addressed several ways. One approach is assigning greater weight to included citations than excluded citations in the training algorithm. Another approach is using under-sampling techniques, which can be done randomly or aggressively. Aggressive under-sampling ranks excluded citations in terms of similarity to included citations. Those most similar are thrown out of the set, skewing the remaining set. This ensures that equivocal citations will be included and helps prevent false-negative exclusions.

False negatives (deeming a citation irrelevant when it should have been included in the review) are more problematic than false positives since these publications can be excluded at the full article review stage. One method of managing this problem is implementing "voting or committee approaches for ensuring high recall."[14] This can be implemented by running multiple classifiers simultaneously and counting the "votes" for inclusion or exclusion. Disputed items can be forwarded for manual review. Another approach is including the citation if any classifier recommends inclusion. O'Mara-Eves et al. note that implementers of text-mining algorithms should "consider whether the amount and/or quality of the training data make a difference to the ability these modifications to adequately penalize false negatives. The reason for this is that, if used in a 'live' review, there might be only a small number of human-labelled items in the training set to be able to determine whether the classifier has incorrectly rejected a relevant study."[14]

Another training problem is that a set of citations may not be representative of the entire population of relevant documents. This imposes a risk of "hasty generalization." Processes recommended to avoid this problem are incorporating reviewer domain knowledge and employing patient active learning. In this approach, classifiers are targeted on different "views" of the citations such as titles, abstracts, and controlled vocabulary terms. O'Mara-Eves et al. noted that human input resulted in a decline in recall when active learning was added to a support vector machine or decision tree classifier but made no difference to the recall of a naïve Bayes classifier. They found this intriguing and recommend further research in this area.[14]

Creating training sets for systematic review updates presents unique problems. Although it may seem an easier task because there is already a set of included citations for training the algorithm, concept drift may have occurred. Concept drift is a phenomenon in which "data from the original review may cease to be a reliable indicator of what should be included in the new one."[14] Training sets might not be representative of those available when conducting a "new" review. Also, biases may have been introduced by overly inclusive reviewers for the report's previous iteration.

Where possible, the 44 studies in the O'Mara-Eves et al. systematic review were evaluated for workload reduction. The authors evaluated the algorithms or text-mining methods employed in the included studies. Within this umbrella are classifiers and the options for using them (kernels) and feature selection for the algorithms (titles, abstracts, MeSH headings), including the effect of different combinations on performance. They also evaluated the effectiveness of methods for implementing text mining. These metrics include the F measure (harmonic mean of precision [positive predictive value] and recall [sensitivity]), work saved over sampling (WSS), and utility. Reported evaluation metrics had subjective elements, which made it difficult to compare across studies. Individual study results are available in the O'Mara-Eves systematic review.[14] Almost all papers considered text mining a promising method to reduce workload during screening.

O'Mara-Eves et al. suggested elements for consideration before broadly implementing text mining. First, the program should be available to systematic reviewers without the need for a computer scientist to write code or process text for individual reviews. At the time of fact checking, the authors identified only six such systems:

- Off-the-shelf for systematic review:
    - Abstrackr
    - EPPI-Reviewer
    - GAPScreener
    - Revis
- Generic – require some training
    - Pimiento
    - RapidMiner

Replicability, scalability, and suitability should also be considered. Only one study reported in the review was a replication study**.** Although some studies used the same dataset, it was impossible to directly compare the studies. Scalability is still questionable. The evaluation datasets were relatively small compared with typical systematic review retrieval sets. With few exceptions, most datasets included fewer than 5,000 citations. Suitability also requires additional study. Only a few types of evidence bases have been evaluated to date, mostly in the domains of biomedicine and software engineering.

O'Mara-Eves et al. conclude, "On the whole, most [studies] suggested that a saving in workload of between 30% and 70% might be possible (with some a little higher or a little lower than this), though sometimes the saving in workload is accompanied by the loss of 5% of relevant studies (i.e., a 95% recall)."[14] They noted that the approaches so far have been based on citations, abstracts, and metadata rather than full text. They recommend:

- Systematic reviewers should work together across disciplines to test these approaches.
- Text mining for prioritization is ready for implementation.
- Text mining as a second screener may be used cautiously.
- Text mining as the only means of excluding articles is not yet ready for use.

One of the tools the O'Mara-Eves review mentions is worthy of additional discussion since it was developed by members of an EPC. The Abstrackr development team has multiple publications tracking the evolution of Abstrackr.[49-51] Abstrackr uses a semi-automated screening algorithm that incorporates labeled terms and timing data into an active learning framework. The algorithm was developed for imbalanced datasets and is intended as an add-on to manual processes. It uses a pool-based active learning approach using the LibSVM support vector machine. The SIMPLE active learning strategy trains the algorithm by presenting the most ambiguous citations for labeling first. It continues presenting citations until a predefined stopping criterion is met. After experimentation, the developers selected 50% as the cut-off point. As of 2012, the developers had used Abstrackr in more than 50 systematic reviews.

Rathbone et al., at the Centre for Research in Evidence-based Practice in Australia, have also studied Abstrackr. Their study included four systematic reviews representative of different types of evidence bases (diagnostics, multiple-intervention, small homogenous set, multiple study types), and their metric was workload savings. The authors chose Abstrackr for evaluation over other text-mining tools because "existing literature indicates that the recall accuracy of Abstrackr is very high… and therefore, a promising predictive text-mining tool for systematic reviews where the primary goal is to identify all relevant studies."[52] The authors conclude that "Semi-automated screening with Abstrackr can potentially expedite the title and abstract screening phase of a systematic review. Although the accuracy is very high, relying solely on its predictions when used as a stand-alone tool is not yet possible. Nevertheless, efficiencies could still be attained by using Abstrackr as the second reviewer thereby saving time and resources."[52]

Data Extraction

After the full-text articles have been retrieved and the inclusion decision verified, members of the systematic review team begin extracting data elements relevant for their review topic. Since data abstraction is a form of information extraction, this process has also been studied in the context of text mining.

Information extraction can include name entity recognition (concept extraction) and association (relationship) extraction. Jonnalagadda et al. published a systematic review focused on automating data extraction in systematic reviews in June 2015.[15] This section will focus mainly on this work, with the addition of several studies that may be of specific interest to the AHRQ EPCs.

The Jonnalagadda et al. review comprises 26 studies. The authors created a table of extracted elements as identified in several systematic review standards and determined which elements had been extracted in the studies they reviewed. The “standards” include:

- Cochrane Handbook for Systematic Reviews
- PICO (Population, Intervention, Comparison, Outcomes Framework)
- PECODR (Patient-Population-Problem, Exposure-Intervention, Comparison, Outcome Duration and Results Framework)
- PIBOSO (Population, Intervention, Background, Outcome, Study Design, Other Framework)
- STARD (Standards for Reporting of Diagnostic Accuracy initiative)
- CONSORT (The Consolidated Standards of Reporting Trials)

Various studies had extracted population-related elements, including the total number of participants, demographic information (age, ethnicity, nationality, sex), and condition-related elements such as comorbidity and spectrum of presenting symptoms. Intervention-related elements included specific interventions, intervention details, total number of intervention groups, and current treatments for the condition. Outcomes-related information included both collected and reported outcomes and time points. Additional elements included:

- Comparators
- Sample size
- Overall evidence
- Generalizability
- External validity
- Research questions and hypotheses
- Study design
- Total study duration
- Sequence generation
- Allocation sequence concealment
- Blinding
- Methods for generating allocation sequence and implementation
- Key conclusions of study authors.

The Jonnalagadda et al. review lists an additional 28 elements, which have not yet been the subject of data extraction studies.

The accuracy of results was measured with the F metric. Studies reported data abstraction at the sentence, abstract, and full-text levels using a variety of approaches, including:

- Conditional random fields (lexical, syntactic, structural, sequential data)
- Multiple supervised classification techniques (MeSH semantic type, word overlap with title, punctuation marks on random forests, naïve Bayes, support vector machines, multi-layer perceptron classifiers)
- Naïve Bayes classifier and structured abstracts
- Statistical relational learning-based approach (kLog)
- Multistep processes:
    - Infer latent topics from documents, use logistical regression to determine probability that a criterion belongs to a topic
    - SVM classifier identification of sentences followed by manually crafted extraction rules

In many studies, the elements were identified and highlighted but not extracted from their context.

The F-score varied greatly between element types and studies. The review authors were unable to compare between studies because of the heterogeneity in data sets and methods. They conclude: “Most of the data elements that would need to be considered for systematic reviews have been insufficiently explored to date, which identifies a major scope for future work.” They suggest that automated data extraction might initially be useful to validate single reviewer manual data extraction followed by automated extraction as technology evolves.[15]

Specific Examples

Multiple studies addressed the use of text mining for detecting bias. Marshall et al. discuss use of the RobotReviewer tool to assess risk of bias using domains defined by the Cochrane Risk of Bias Tool.[53] They used a multitasking systems vector machine approach that purportedly exploits correlations between bias types. Their tool gauges whether a report is at a low risk of bias and extracts supporting statements. They used a nontypical approach to create their training set. Rather than using a curated set, they used structured data in existing databases within the Cochrane Library. They selected the Cochrane databases because they are rich in terms for bias assessment. The authors concluded that the tool was not ready to replace human review but could help prioritize assessment and could potentially allow review by one analyst rather than two. A major limitation is that the tool provides one risk-of-bias assessment for the paper rather than by outcome.

Hsu et al. focused on extracting statistical analyses using their automated sequence annotation pipeline (ASAP).[54] Their approach incorporates three annotators: concept, statistics, and clinical trials and requires full-text articles. ASAP runs all three annotators concurrently. The authors found that it was inconsistent in capturing variability in independent and dependent variables and hypothesis testing. They conclude: “Our system is a step towards automating the identification of key reported statistical findings that would contribute to the development of a Bayesian model of a complex disease.”[54] They note that while tools exist to extract data from the abstract only a small portion of the relevant information is reported there. They also noted that these tools do not provide the context necessary to interpret the extracted information. “We attempt to not only classify sentences related to the statistical analyses, but also characterize the values reported in these sentences to populate the data mode. This allows the computer to assist in assessing the validity of reported information and enables this information to be used for meta-analysis and probabilistic disease modeling.”[54] ASAP coordinates published study information with information from the protocol in the Clinicaltrials.gov record.

Shao et al. also used ClinicalTrials.gov data to address bias.[55] Their Aggregator clustering tool is designed to detect multiple publications derived from the same trial. Their study was based on a set of Medline articles containing one or more National Clinical Trial (NCT) registry number. There were two training sets. The positive set comprised articles with the same NCT numbers, while the balanced negative set comprised articles with the same conditions and interventions but different NCT numbers. The classifier was focused on multiple features, including:

- Rank in related articles
- Number of shared author names
- Affiliation similarity
- Shared email

- Publication type similarity
- Support type similarity (grants)
- Email domain
- Shared country
- Shared substance names
- All-capitalized words in title
- All-capitalized words in abstract or CN field

The training set reached 0.881 precision, 0.813 recall with F1=0.843. The validation set (composed of citations from 5 Drug Effectiveness Review Project systematic reviews) was calculated to have 0.877 precision, 0.833 recall with F1=0.854. They encountered two types of errors: Splitting errors occurred when the model missed articles from the same trial discussing different aspects of the study. Lumping errors occurred when the model incorrectly identified publications with shared authors and topics but different trials. While still a model, the authors plan to incorporate it into their pipeline tool.[55]

Cohen et al. describe the randomized controlled trial (RCT) tagger.[56] This tool predicts whether a study is an RCT based on the citation, abstract, and MeSH headings. The model can be used with or without the MeSH headings. RCT tagger is a web-based tool (http://arrowsmith.psych.uic.edu/cgi-bin/arrowsmith_uic/RCT_Tagger.cgi) and returns a list of abstracts from PubMed with an RCT confidence rating. It shares the pipeline with Aggregator so the RCTs can be further analyzed for trial source within the same search session. Cohen et al. found that many RCT citations were not classified with the RCT tag and vice versa.

Updating

After a systematic review has been completed and published, one major challenge is determining whether changes in the evidence base necessitate updating the report. Cohen et al. have addressed this problem in 2008 and 2012 publications.[57,58] The authors envision an automated alerting system that notifies a team that a study likely to meet inclusion criteria has been published as soon as that publication has been indexed in Medline. They found that "review experts are more willing to trade off recall for precision for the New Update Alert task, as compared to the work prioritization task that we have previously studied. In particular, the principal investigator of the Drug Effectiveness Review Project (one of the senior authors of this paper) consistently preferred a recall of 0.55 and the achievable precision corresponding to that level of recall over all other available levels of recall between 0.99 and 0.55." Although recall consistently reached 0.55 in the training set, it varied from 0.134 to 1.0 in the test topics. The authors attributed this to small sample sizes in the test sets; the largest topic set achieved recall of 0.50. They believe that a systematic review expert using a live alert system could use this approach effectively. They also note that it could be useful for prioritization of review updates between topics since it would facilitate comparing the number of citations that meet alert criteria as opposed to the gross number of citations captured in the search alerts.

Dalal et al. also considered text mining for report updating.[59] Their training set retrieved only PubMed citations that had been indexed with MeSH headings for simulated comparative-effectiveness research reports. The authors "evaluated statistical classifiers that used previous classification decisions and explanatory variables derived from MEDLINE indexing terms to predict inclusion decisions. This pilot system reduced workload associated with screening two simulated comparative effectiveness review updates by more than fifty percent with minimal loss of relevant articles."[59]

Key Informant Interviews

We interviewed eight Key Informants (KIs). To get a range of views on the implementation and use of text mining, we set out to interview two different groups: four senior investigators representing an organizational perspective and four librarians for a research team member perspective. Because the preponderance of published literature to date has focused on the use of text mining in the screening phase, our workgroup decided to focus on the searching phase and interviewing librarians to gain a fresh perspective. Below, we provide a narrative summary of integrated findings for each group. Please see Appendix C for question themes and quotes from each group and Appendix D for KI comments on specific text-mining tools.

Summary of Integrated Findings

Senior Investigator/Organizational Perspective

Motivation to use text-mining tools: Most of the KIs were interested in process improvement but came to it from different angles (e.g., medical versus social science topics, systematic review versus scoping review), with most of the comments relating to the use of text mining in the screening process to overcome problems associated with large result sets.

Cost and time efficiencies: Software and staffing costs to create a text-mining tool seemed difficult to calculate for the interviewed KIs because many of the tools they use have been developed over a long time, working alone or in conjunction with colleagues and existing staff. Two of the KIs use text-mining tools to prioritize records for screening, so records with the highest probability of being on-topic are shown at the beginning of the result set; thus, research team members can begin abstracting and analyzing included records earlier in the systematic review process than would be typical if research teams had to wait for the screening process to be completed first. One KI was involved in a large-scale scoping review in which text mining allowed their team to complete the project; it would otherwise have proved impossible with a traditional screening approach.

"Integratability" into existing workflows was mentioned by two KIs, both in terms of making it easier for staff to work without moving between multiple systems to complete a task and creating a user-friendly, front-end interface because many text-mining tools otherwise require some technological expertise to run.

Organizational and technological facilitators: Perhaps not surprisingly, organizational leadership seemed to be the critical factor in the decision to move forward with implementation. Information technology (IT) infrastructure and IT staff support varied across organizations. Three KIs noted its importance to the success of their projects, while the fourth KI lamented that no specific organizational budget line existed to aid its development.

Organizational and technological barriers: KIs mostly reported high staff acceptance to adopting text mining; however, staff were also mentioned as an organizational barrier, specifically librarians/information specialists who may feel their work has been deskilled. Two KIs expressed concerns regarding the systematic review community's reception of text-mining/machine learning use to support reviews because of human decision-making over computer bias. While KIs were generally optimistic about the future integration of text-mining tools into systematic reviews, two expressed some hesitation—that at present they should not be used blindly, but rather with a knowledge of their strengths and limitations because they are in their infancy developmentally.

Areas mentioned as needing more research and guidance included:

- Developing time and accuracy metrics to allow formal evaluation
- Developing metrics to evaluate the value of text mining

Reporting of text-mining tool use in the final review varied, with most KIs citing a lack of standards as problematic to transparently conveying what was done.

Librarian/Research Team Member Perspective

Motivation to use text-mining tools: All the librarians cited objectivity as one of their prime motivations to use text-mining tools to develop systematic review search strategies in a more rigorous manner. Generally, they expressed confidence in the resulting keywords and synonyms found as being more comprehensive and faster/easier to identify than the typical iterative approach of reviewing titles and abstracts for these. One KI also mentioned that text-mining brings out more subtle aspects of how a topic is (non-obviously) connected to other topics (e.g., how diabetes is implicated in several other childhood conditions that could/should also be searched for a review). Overall, the KIs noted that the time required to create the search strategy was decreased, and confidence in the resulting list of keyword/synonyms was high. While all of them recommended using text-mining tools, they also expressed a variety of qualifications:

- Some tools, like VOSviewer, while presenting intriguing visualization results, need to be more carefully assessed to determine how they can best be used to improve a search.
- Find the tools useful to create a search strategy but that they are not the end all and be all.
- Difficulty surrounds evaluating whether a corpus used to train a filter was indeed representative of the material it was developed to find.
- Complex topics (e.g., health services research) may be better suited to using text-mining approaches, whereas using "traditional" keyword/subject searches might be more suitable for straightforward one-drug/one-indication topics.

Keyword/synonym tools: Some of the tools are easy to use and can be learned to use quickly, especially the keyword/synonym type tools like PubReminer and GoPubMed. Integration with other databases or software is often not as seamless as desired because files may need to be reformatted to get data into/out of the tools. Some tools, like PubReminer, are easy to work with while others are not. In addition to generating lists of terms to use in a search, some KIs noted these tools can also be helpful in identifying terms that can be excluded from the result set. Please see Appendix C for more comments on specific text-mining tools.

Filter tools: Most of the librarian comments focused on the keyword/synonym-type tools rather than filter-type tools. One KI mentioned that due to the sensitivity of filters and greater retrieval of records that it is sometimes more efficient to approach the search via the "traditional" keyword/subject approach because it does not take as long to develop the search or screen the results. The published literature has more articles on filter development, so this focus on keyword/synonym tools was unanticipated and bears further scrutiny in future research to better understand which types of tools are more useful and under what conditions (e.g., straightforward questions versus complex questions, systematic review organization versus one-off research project).

IT environment: KIs generally had few problems using text-mining software online or installing it locally, if necessary; however, some issues did arise with an organizational server's security settings for one KI. Given local institutional IT risk tolerance, access to and/or downloading programs seems likely to be an issue for some research teams wanting to use these tools. One KI commented that using more complicated tools like General Architecture for Text

Engineering (GATE) and VOSviewer required the help of IT staff to run correctly; though most KIs found they could run most of the keyword/synonym-type tools with no assistance.

Reporting the use of tools to develop the search strategy was variable, from not reporting using them to reporting the use of filter-type tools but not keyword/synonym tools. KIs who do not report keyword/synonym tools noted that other "background" methods used to develop keywords and synonyms are not typically reported; thus, these tools should not be a similar case. Performance evaluation, much less comparative performance evaluation, of these tools has not yet been researched, so it is not yet known whether using one tool or another may bias strategy development.

Identification of Individual Text-Mining Tools

We cast a broad net to identify text-mining tools and applications and retrieved many. We assessed 111 text-mining tools. To provide a meaningful summary, we narrowed the retrieval to a subset of tools that met one or more of the following criteria: (1) described in the literature and deemed as relevant or useful to a systematic review; (2) used and reported as a methodologic resource in a systematic review publication; or (3) mentioned by a KI during an interview (see Appendix D). In addition, we expanded the list to include those tools that met all of the following: (1) free and Web-based (i.e., not requiring download or license); (2) high likelihood of relevance to one or more steps in the systematic review process; and (3) high degree of confidence in the tool's stability and usability (i.e., a reliable connection, existence of help documentation, and/or literature references).

The findings from our preliminary assessment of tools (Appendix E) suggests that 73 (66%) were referenced in the literature captured by the literature review and 19 (17%) were identified by KIs (Table 5). Most of the tools (79%) we examined were available without cost via the web or through download of open source code. Some tools mentioned in papers published just a few years ago were no longer supported or functional. Fourteen resources were unavailable, retired, or nonfunctional at the time of our assessment.

We designated 89 of the 111 as potentially applicable or useful to the conduct of systematic review (i.e., designed or could easily be modified to support 1 of the core steps for systematic literature review). Of those we were able to test (i.e., tools that did not require download or installation), 64 (57%) included a feature to support one or more of the key steps in the systematic literature review process. Most tools (n=52) supported searching, 44 supported scoping, 15 supported the screening process, and 14 aided information extraction.

Table 5. Summary of tools identified by key informants

Name	Availability	Tech Level	ATM DTM	DCT DCL	TCA TCL TCT	VIZ	WFA	Comments
Abstrackr	Free	Low		●				Systematic review support
AntCont	Free	Medium			●		●	Requires installation
Carrott2	Free	High		●				Requires installation
EndNote	Commercial Product	Low					●	Requires license and installation
EPPI-Reviewer 4	Free Trial	Medium	●	●				Requires registration
Excel	Commercial Product	Low					●	Requires customized coding
GAPScreener	Free	Medium			●			Requires installation
GATE	Free	High			●			Requires installation

Table 5. Summary of tools identified by key informants (continued)

Name	Availability	Tech Level	ATM DTM	DCT DCL	TCA TCL TCT	VIZ	WFA	Comments
GoPubMed	Free	Low	●				●	Requires registration; PubMed interface
KNALIJ	Free	Medium						Retired; not working
Lingo3G	Free Trial	High		●		●		Requires installation
Mimir	Free	Medium			●			Requires installation
PubReMiner	Free	Low			●		●	PubMed interface
TerMine	Free	Medium	●					Web version for demonstration; register for batch service
tm for R	Free	High	●					Requires coding / syntax
VOSviewer	Free	Medium			●	●		Requires installation
Voyant Tools	Free	Medium			●	●	●	Web-based
WordStat / SimStat	Commercial Product	High		●		●	●	Requires license and installation

ATE = general architecture for text engineering; ATM = automatic term recognition; DCL = document classification; DCT = document clustering; DTM = document term matrix; IEX = information extraction; IR = information retrieval; NLP: natural language processing; TCA = text categorization; TCL = text cluster; TSM = text summarization; VIZ: visualization; WFA: word frequency analysis

Discussion

What Was Known and What This Paper Adds

Text-mining techniques have been explored at every phase of the systematic review production process. Screening and data extraction have been studied more extensively than information retrieval and specialized tasks such as quality assessment and updating. Common limitations were a dependence on PubMed/Medline content, nonstandardized tools that require a familiarity with programming, and heterogeneity between studies that precludes analysis. Most authors and KIs considered text mining and machine learning to be very promising to reduce workload for systematic review teams, although they suggest adding these tools as adjuncts rather than as a means of replacing existing tools or displacing team members. The interviews conducted with senior investigators from systematic review organizations and librarians are the first to investigate the experiences and perspectives of these groups with text-mining tools. While the interviews revealed an enthusiasm for the continuing development in the use of these technologies, they also suggest some presumptive benefits have not been adequately studied to date, nor has the scope of their utility been fully described. Our literature review identified no descriptive or performance appraisal instruments of text-mining tools within the context of systematic reviews in medicine and health sciences, indicating a need to develop additional decision support and analysis tools.

Based on this preliminary exploration, we believe that text-mining tools will be increasingly featured within the repository of resources to support the conduct of systematic reviews. Text mining and its related forms of automated knowledge discovery tasks are unlikely to displace current literature retrieval and management tools. Rather, we expect that text mining will augment existing literature-retrieval techniques and information extraction tools. Text mining may be a promising resource to detect unknown relationships and patterns from an expansive archive on unstructured data, but it offers a limited "return on investment" for well-developed processes such as systematic literature searching using controlled vocabulary. Furthermore, the benefits of using a text-mining tool may be limited, in part, by the complexity and size of the relevant literature for a given topic.

Challenges and Barriers

One key barrier to consider when investigating text-mining tools to support the systematic review methods is lack of information about underlying rules, ontologies, and algorithms. Individuals without experience or training are unlikely to understand how the system or tool works and may be deterred from adopting a tool that does not provide installation instructions or a user guide. Therefore, in our assessment, we included a category for "help" to distinguish the sources that may be a better starting place for a potential user with limited knowledge of text-mining technology. For the tools that provided documentation, the usefulness varied; however, those tools with documentation appeared more accessible among the expansive and diverse set of options. The GATE open-source software is one project that explicitly addresses this challenge, noting that, over 20 years, developers have "spent a great deal of effort deconstructing the black box semantic platform" and, that in addition to being open source, GATE-based solutions include "numerous plug and play components and users can see where the rule sets or the ontology or the algorithms fit" to offer users greater insight into how their systems work.[60]

Pragmatic Considerations

Based on our literature review, interviews with KIs, and descriptive assessment of text-mining tools, we offer the following pragmatic considerations:

- Socio-technological barriers to adoption (e.g., staff integration issues, methodologic criticism by others in the systematic review community.
- Organizational IT security systems may forestall or hamper the functionality of some text-mining tools.
- “Integratability” of text-mining tools into existing workflows and with other organizational IT systems.
- Reporting use of text-mining tools is undeveloped by well-known systematic review methodology organizations, resulting in their use often going unreported or heterogeneously reported.

For current non-users, the decision to adopt text-mining into the systematic review process will likely be influenced by the first three considerations. With regard to the fourth consideration, guidelines for reporting the use of text-mining tools in prospective systematic reviews would be optimal and are likely to aid in assessment of reproducibility and validation as well as comparative evaluation of these tools.

Additionally, it is imperative that one or more of the methodologists on the systematic review team understand the parameters, limitations, and implications of the text-mining features and functions. Many tools rely on algorithms or statistical analyses that may be unfamiliar to the individuals who typically manage literature searching and retrieval, conduct screening, and/or data extractions. Thus, teams should consider consulting with a biostatistician and/or a computer programmer before introducing a text-mining application into the systematic review to ensure that the tool generates meaningful and reliable results that can be replicated.

Strengths and Limitations

To our knowledge, this is the first narrative review of published research on text-mining tools use across all types of reviews and all steps in the systematic review process from a wide variety of disciplines (e.g., computer science, library and information science, medicine). It should be noted that due to time constraints, this project is not a comprehensive review, so additional unidentified materials of interest may exist. Results were limited to English-language materials published since 2005.

We solicited a range of perspectives by interviewing KIs from two groups of systematic review researchers a novel source of information on this topic; however, our sample size was small so these preliminary findings bear further investigation. Other investigators, librarians, or end-user groups might have different perspectives than those we interviewed for this project. Due to unsuccessful recruitment attempts, the workgroup did not include a computer scientist, which may have influenced this report’s conduct or conclusions.

Lastly, this study includes the first compilation of text-mining tools used in published systematic reviews, described in the research literature, or mentioned by one of the KIs. We created a descriptive tool to characterize and compare broad attributes (e.g., cost, desktop- or server-based) of available tools. Creating a comparative performance evaluation test and conducting the requisite evaluation of the tools was out of the scope of this project.

This is not necessarily an exhaustive index or a validated method of feature analysis or tool evaluation. Several assessments (e.g., whether a tool is applicable or could be used to support a systematic review) were based on the subjective judgment of team members. However, we relied on informal thresholds to apply consistent judgments. The term “text mining” frequently captures tools designed to extract and classify granular information from the molecular biology literature. Although similar in concept and underlying mechanism, we did not include those in our catalog. Assessments were conducted by information specialists who primarily conduct literature searches for systematic reviews of health care interventions and is therefore biased to reflect the experience, skills, and vocabulary reflective of this discipline. An evaluation of the same tools by experts from other domains (e.g., computer science, genetics, engineering, bioinformatics) is likely to generate different conclusions about the features and usefulness.

Future Directions

Tools

Next steps for examining the features and usefulness of text-mining tools to support systematic reviews should include simulation and replication studies using actual systematic reviews and/or case studies to assess performance.[34] Our assessment was preliminary and limited to mapping characteristics and judging usability among a subset of text-mining tools most likely to support systematic review methods. With numerous sources of tools and a variety of ways to leverage text-mining technology, it is critical to consider how to cull a subset of those most likely to improve quality or boost efficiency of existing processes.

Before designing an evaluation or validation study, it is essential to consider the specific needs and define the expected value of a text-mining tool or technology within the systematic review process. The approach to selecting and benchmarking a tool varies by the intended application within the systematic literature review process. Automated text and document categorization, retrieval, and classification features may help reduce or redistribute workload, improve search sensitivity, prioritize a set of interventions or outcomes of interest, or refine the scope for a topic with a large literature base. Selection and analysis of features will differ based on the context and purpose. Existing case studies using text-mining tools or technologies described in the literature may be helpful in planning a detailed evaluation or performance assessment. Examples span disciplines and investigate various applications, including automated annotation of functional imaging experiments,[61] text mining to detect drug discontinuation rates,[62] word frequency analysis to validate a patient safety search strategy,[33] narrative text interrogation of administrative health data for injury surveillance,[63] and automated document classification to prioritize workload for systematic reviews.[58]

The underlying computational methods that power a tool are not always described for or apparent to the user. Thus, a robust evaluation requires either knowledge of the underlying technologies or access to descriptions of the tool features and development and technical resources for installation and application support. Key characteristics of text-mining tools selected for systematic review support should include reliability and transparency of the computational methods; ease of use, including support for different formats of text and data; and results that are replicable and compatible with the tools and workflow currently in use.

Conclusions

This research illustrates the state of the published evidence on the use of text-mining tools to support different systematic review processes and was conducted to provide the initial groundwork for future research into its utility to the systematic review community. To date, text-mining tools appear promising, but further research is warranted on: (1) actual costs, (2) actual workload reduction/time efficiencies, (3) whether its use is better suited to some types of review topics (e.g., well-defined clinical topics versus more diffuse public health topics), (4) when its use is most beneficial (e.g., is there enough added benefit for small result sets of <1,000 citations versus large result sets of >10,000 citations?), (5) development of evaluation metrics, and (6) head-to-head comparative performance evaluation of the tools themselves. Lastly, as these tools become more widely used in the review community, the need for clear reporting standards increases.

References

1. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.The Cochrane Collaboration; 2011. http://handbook.cochrane.org/. Accessed: 2016 Mar 24.
2. Committee on Standards for Systematic Reviews. Institute of Medicine. Eden J, Levit L, Berg A, et al, editors. Finding what works in health care: standards for systematic reviews. Washington (DC): National Academies Press; 2011 Mar 23.
3. Hartling L, Guise JM, Kato E, et al. EPC methods: An exploration of methods and context for the production of rapid reviews. Research White Paper. (Prepared by the Scientific Resource Center under Contract No. 290-2012-00004-C.) AHRQ Publication No. 15-EHC008-EF. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2015 Feb.
4. Hempel S, Shetty KD, Shekelle PG, et al. Machine learning methods in systematic reviews: identifying quality improvement intervention evaluations (Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2007-10062-I). Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2012 Sep. 55 p.
5. Dalal SR, Shekelle PG, Hempel S, et al. A pilot study using machine learning and domain knowledge to facilitate comparative effectiveness review updating (Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2007-10062-I). Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2012 Sep. 50 p.
6. Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. Res Synth Meth. 2011 Mar;2(1):1-14. PMID: 26061596.
7. Miner G, Elder J IV, Hill T, et al. Practical text mining and statistical analysis for non-structured text data applications. Waltham (MA): Elsevier, Inc.; 2012 Jan 25. 1000 p.
8. Methods guide for effectiveness and comparative effectiveness reviews, AHRQ Publication No. 10(14)-EHC063-EF. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2014 Jan.
9. Campbell Collaboration. Campbell Collaboration Systematic Reviews: policies and guidelines. Version 1.1. Norway: Campbell Collaboration; 2015 Jan 14. 47 p. (Campbell Systematic Reviews, Supplement 1).
10. Evidence for Policy and Practice Information and Co-ordinating Centre. EPPI-Centre Methods for Conducting Systematic Reviews. London (UK): Social Science Research Unit, Institute of Education, University of London; 2010.
11. Joanna Briggs Institute. Joanna Briggs Institute reviewer's manual: 2014 edition. Adelaide, South Australia: The Joanna Briggs Institute; 2014. 197 p.
12. European Network for Health Technology Assessment. EUnetHTA Joint Action 2, Work Package 8. HTA Core Model® version 2.1. European Network Health Technology Assessment; 2015. [453 p]. http://www.corehta.info/BrowseModel.aspx. Accessed 2015 Aug 04.
13. European Network for Health Technology Assessment. Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness. Copenhagen (Denmark): European Network for Health Technology Assessment; 2015 Jul. 70 p.
14. O'Mara-Eves A, Thomas J, McNaught J, et al. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev. 2015;4:5. PMID: 25588314.
15. Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. Syst Rev. 2015 Jun 15;4(1):78. PMID: 26073888.
16. Fleuren WW, Alkema W. Application of text mining in the biomedical domain. Methods. 2015 Mar;74:97-106. PMID: 25641519.

17. Charmaz K. Reconstructing grounded theory. In: Alasuutari P, Bickman L, Brannen J, editors. SAGE handbook of social research methods. London: Sage; 2008. p. 461-79.

18. OBrien BC, Harris IB, Beckman TJ, et al. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med. 2014 Sep;89(9):1245-51. PMID: 24979285.

19. Saldana J. Descriptive coding. In: Saldana J. The coding manual for qualitative researchers. Thousand Oaks (CA): Sage; 2013. p. 87-91.

20. Taylor B. Grounded theory. In: Taylor BJ, Francis K, editors. Qualitative research in the health sciences: methodologies, methods, and processes. New York: Routledge; 2013. p. 30-55.

21. Qualitative Solutions and Research (QSR). NVivo 10 for Windows: qualitative data analysis software. Melbourne (Australia): QSR International; 2015. http://www.qsrinternational.com/products_nvivo.aspx. Accessed 2015 Aug 12.

22. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700. PMID: 19622552.

23. Calders T, Esposito F, Hullermeier E, et al. Machine learning and knowledge discovery in databases. In: Proceedings, Part III. Springer Heidelberg, New York, Dordrecht, London. European Conference, ECML PKDD 2014; September 15-19, 2014; Nancy, France. 2014.

24. Felizardo KR, Nakagawa EY, Feitosa D, et al. An approach based on visual text mining to support categorization and classification in the systematic mapping. Proceedings of the 14th international conference on Evaluation and Assessment in Software Engineering (EASE'10); British Computer Society, Swinton (UK); 2010. 34-43 p.

25. Felizardo KR, Macdonell SG, Mendes E, et al. A systematic mapping on the use of visual data mining to support the conduct of systematic literature reviews. J Software. 2012;7(2):450-61.

26. Hotho A, Nurnberger A, Paaß G. A brief survey of text mining. 2005 May 13.

27. Kao A, Poteet SR, editors. Natural language processing and text mining. London (UK): Springer-Verlag; 2007.

28. Sizov G. Extraction-based automatic summarization. Theoretical and empirical investigation of summarization techniques. Norwegian University of Science and Technology; 2010 Jan.

29. Zhang Q, Segall R. Review of data, text and web mining software. Kybernetes. 2010;39(4):625-55.

30. Hausner E, Waffenschmidt S, Kaiser T, et al. Routine development of objectively derived search strategies. Syst Rev. 2012;1:19. PMID: 22587829.

31. O'Mara-Eves A, Brunton G, McDaid D, et al. Techniques for identifying cross-disciplinary and 'hard-to-detect' evidence for systematic review. Res Synth Meth. 2014 Mar;5(1):50-9. PMID: 26054025.

32. Simon M, Hausner E, Klaus SF, et al. Identifying nurse staffing research in Medline: development and testing of empirically derived search strategies with the PubMed interface. BMC Med Res Methodol. 2010;10:76. PMID: 20731858.

33. Tanon AA, Champagne F, Contandriopoulos AP, et al. Patient safety and systematic reviews: finding papers indexed in MEDLINE, EMBASE and CINAHL. Qual Saf Health Care. 2010 Oct;19(5):452-61. PMID: 20457733.

34. Balan PF, Gerits A, Vanduffel W. A practical application of text mining to literature on cognitive rehabilitation and enhancement through neurostimulation. Front Syst Neurosci. 2014;8:182. PMID: 25309356.

35. Kok R, Verbeek JA, Faber B, et al. A search strategy to identify studies on the prognosis of work disability: a diagnostic test framework. BMJ Open. 2015;5(5):e006315. PMID: 25991444.

36. Petrova M, Sutcliffe P, Fulford KW, et al. Search terms and a validated brief search filter to retrieve publications on health-related values in Medline: a word frequency analysis study. J Am Med Inform Assoc. 2012;19(3):479-88.

37. Damarell RA, Tieman J, Sladek RM, et al. Development of a heart failure filter for Medline: an objective approach using evidence-based clinical practice guidelines as an alternative to hand searching. BMC Med Res Methodol. 2011;11:12. PMID: 21272371.

38. Li J, Lu Z. Developing topic-specific search filters for PubMed with click-through data. Methods Inf Med. 2013;52(5):395-402.

39. Iansavichus AV, Hilderbrand AM, Haynes RB, et al. High performance information search filters for CKD content in PubMed, Ovid MEDLINE, and EMBASE. Am J Kidney Dis. 2015;65(1):26-32.

40. Shemilt I, Simon A, Hollands GJ, et al. Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. Res Synth Meth. 2014 Mar;5(1):31-49. PMID: 26054024.

41. Thompson J, Davis J, Mazerolle L. A systematic method for search term selection in systematic reviews. Res Synth Meth. 2014;5(2):87-97.

42. Slater L. Product review: PubMed PubReMiner. J Canadian Health Libr Assoc. 2012;33(2):106-7.

43. Poultor GL, Rubin DL, Altman RB, et al. MScanner: a classifier for retrieving Medline citations. BMC Bioinformatics. 2008;9:108.

44. Macdonald EM, Carroll A, Albiston D, et al. Social relationships in early psychosis: clinical data-mining for practice-based evidence. J Soc Work Res Eval. 2005 Fall-Winter;6(2):155-66.

45. Choong MK, Galgani F, Dunn AG, et al. Automatic evidence retrieval for systematic reviews. J Med Internet Res. 2014;16(10):e223.

46. Hausner E, Guddat C, Hermanns T, et al. Development of search strategies for systematic reviews: validation showed the noninferiority of the objective approach. J Clin Epidemiol. 2015 Feb;68(2):191-9. PMID: 25464826.

47. Franzi K, Ananiadou S, Mima H. Automatic recognition of multi-word terms: the C-value/NC-value method. Int J Digit Libr. 2000;3(2):115-30.

48. Powers DM. Evaluation: from precision, recall and F-factor to ROC, informedness, markedness & correlation. Adelaide (South Australia): Flinders University of South Australia; 2007 Dec. 24 p. (Technical report; no.SIE-07-001).

49. Wallace BC, Trikalinos TA, Lau J, et al. Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics. 2010;11:55. PMID: 20102628.

50. Wallace BC, Small K, Brodley CE, et al. Modeling annotation time to reduce workload in comparative effectiveness reviews. In: Conference Proceedings. 1st ACM International Health Informatics Symposium, IHI '10; November 11-12, 2010. p. 28-35.

51. Wallace BC, Small K, Brodley CE, et al. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In: Proceedings of the ACM International Health Informatics Symposium (IHI); 2012. p. 819-24.

52. Rathbone J, Hoffmann T, Glasziou P. Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. Syst Rev. 2015;4:80. Also available: http://dx.doi.org/10.1186/s13643-015-0067-6. PMID: 26073974.

53. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc. 2015 Jun 22;Epub ahead of print. PMID: 26104742.

54. Hsu W, Speier W, Taira RK. Automated extraction of reported statistical analyses: towards a logical representation of clinical trial literature. AMIA Annu Symp Proc. 2012;350-9. PMID: 23304305.

55. Shao W, Adams CE, Cohen AM, et al. Aggregator: a machine learning approach to identifying MEDLINE articles that derive from the same underlying clinical trial. Methods. 2015 Mar 1;74:65-70. PMID: 25461812.

56. Cohen AM, Smalheiser NR, McDonagh MS, et al. Automated confidence ranked classification of randomized controlled trial articles: an aid to evidence-based medicine. J Am Med Inform Assoc. 2015 May;22(3):707-17. PMID: 25656516.

57. Cohen AM, Ambert K, McDonagh M. Studying the potential impact of automated document classification on scheduling a systematic review update. BMC Med Inform Decis Mak. 2012 Apr 19;12(1):33. PMID: 22515596.

58. Cohen AM. Optimizing feature representation for automated systematic review work prioritization. AMIA Annu Symp Proc. 2008;121-5. PMID: 18998798.

59. Dalal SR, Shekelle PG, Hempel S, et al. A pilot study using machine learning and domain knowledge to facilitate comparative effectiveness review updating. Med Decis Making. 2013;33(3):343-55.

60. Cunningham H, Tablan V, Roberts A, et al. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. PLoS Comput Biol. 2013;9(2):e1002854. PMID: 23408875.

61. Turner MD, Chakrabarti C, Jones TB, et al. Automated annotation of functional imaging experiments via multi-label classification. Front Neurosci. 2013;7:240. PMID: 24409112.

62. Williams SA, Porter V, Zarotsky V, et al. An evaluation of literature on discontinuation rates of biologics in rheumatoid arthritis [abstract]. Arthritis Rheum. 2012 Oct;64(S10):483.

63. McKenzie K, Scott DA, Campbell MA, et al. The use of narrative text for injury surveillance research: a systematic review. Accid Anal Prev. 2010 Mar;42(2):354-63. PMID: 20159054.

64. Smalheiser NR, Zhou W, Torvik VI. Anne O'Tate: A tool to support user-driven summarization, drill-down and browsing of PubMed search results. J Biomed Discov Collab. 2008;3:2. PMID: 18279519.

65. Wojtusiak J, Michalski RS, Simanivanh T, et al. Towards application of rule learning to the meta-analysis of clinical data: an example of the metabolic syndrome. Int J Med Inform. 2009;78(12):e104-e111.

66. Fontelo P, Liu F, Ackerman M. askMEDLINE: a free-text, natural language query tool for MEDLINE/PubMed. BMC Med Inform Decis Mak. 2005;5:5. PMID: 15760470.

67. Kossenkov A, Manion FJ, Korotkov E, et al. ASAP: automated sequence annotation pipeline for web-based updating of sequence information with a local dynamic database. Bioinformatics. 2003 Mar 22;19(5):675-6. PMID: 12651736.

68. Divoli A, Attwood TK. BioIE: extracting informative sentences from the biomedical literature. Bioinformatics. 2005 May 1;21(9):2138-9. PMID: 15691860.

69. Papanikolaou N, Pavlopoulos GA, Pafilis E, et al. BioTextQuest(+): a knowledge integration platform for literature mining and concept discovery. Bioinformatics. 2014 Nov 15;30(22):3249-56. PMID: 25100685.

70. Ananiadou S, Rea B, Okazaki N, et al. Supporting systematic reviews using text mining. Soc Sci Comput Rev. 2009 Nov;27(4):509-23.

71. Chen H, Sharp BM. Content-rich biological network constructed by mining PubMed abstracts. BMC Bioinformatics. 2004 Oct 8;5:147. PMID: 15473905.

72. Fleuren WW, Verhoeven S, Frijters R, et al. CoPub update: CoPub 5.0 a text mining system to answer biological questions. Nucleic Acids Res. 2011 Jul;39(Web Server issue):W450-4. PMID: 21622961.

73. Yamamoto Y, Yamaguchi A, Yonezawa A. Building linked open data towards integration of biomedical scientific literature with DBpedia. J Biomed Semantics. 2013;4(1):8. PMID: 23497538.

74. Errami M, Wren JD, Hicks JM, et al. eTBLAST: a web server to identify expert reviewers, appropriate journals and similar publications. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W12-5. PMID: 17452348.

75. Kiritchenko S, de Bruijn B, Carini S, et al. ExaCT: automatic extraction of clinical trial characteristics from journal publications. BMC Med Inform Decis Mak. 2010;10:56. PMID: 20920176.

76. Tsuruoka Y, Tsujii J, Ananiadou S. FACTA: a text search engine for finding associated biomedical concept. Bioinformatics. 2008 Nov 1;24(21):2559-60. PMID: 18772154.

77. Yu W, Clyne M, Dolan SM, et al. GAPscreener: an automatic tool for screening human genetic association literature in PubMed using the support vector machine technique. BMC Bioinformatics. 2008;9:205. PMID: 18430222.

78. Marshall C, Brereton P. Tools to support systematic literature reviews in software engineering: a mapping study. In: Conference Proceedings. 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement; October 10, 2013. p. 296-9.

79. Doms A, Schroeder M. Go exploring PubMed with the Gene Ontology. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W783-6. PMID: 15980585.

80. Seo J, Shneiderman B. Knowledge discovery in high-dimensional data: case studies and a user survey for the rank-by-feature framework. IEEE Trans Vis Comput Graph. 2006 May-Jun;12(3):311-22. PMID: 16640245.

81. Eaton AD. HubMed: a web-based biomedical literature search interface. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W745-7. PMID: 16845111.

82. Giglia E. Quertle and KNALIJ: searching PubMed has never been so easy and effective. Eur J Phys Rehabil Med. 2011 Dec;47(4):687-90. PMID: 22222966.

83. Stansfield C, Thomas J, Kavanagh J. 'Clustering' documents automatically to support scoping reviews of research: a case study. Res Synth Meth. 2013 Sep;4(3):230-41. PMID: 26053843.

84. Lin JW, Chang CH, Lin MW, et al. Automating the process of critical appraisal and assessing the strength of evidence with information extraction technology. J Eval Clin Pract. 2011 Aug;17(4):832-8.

85. Sarkar IN, Schenk R, Miller H, et al. LigerCat: using "MeSH Clouds" from journal, article, or gene citations to facilitate the identification of relevant biomedical literature. AMIA Annu Symp Proc. 2009;2009:563-7. PMID: 20351918.

86. Fontaine JF, Barbosa-Silva A, Schaefer M, et al. MedlineRanker: flexible ranking of biomedical literature. Nucleic Acids Res. 2009 Jul;37(Web Server issue):W141-6. PMID: 19429696.

87. Theodosiou T, Vizirianakis IS, Angelis L, et al. MeSHy: mining unanticipated PubMed information using frequencies of occurrences and concurrences of MeSH terms. J Biomed Inform. 2011;44(6):919-26.

88. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc. 2010 May-Jun;17(3):229-36. PMID: 20442139.

89. Tenner H, Thurmayr G, Thurmayr R. Data mining with Meva in MEDLINE. In: Perner P, Brause R, Holzhutter HG, editors. Medical data analysis. Vol. 2868. Springer Berlin Heidelberg; 2003. p. 39-46.

90. States DJ, Ade AS, Wright ZC, et al. MiSearch adaptive pubMed search tool. Bioinformatics. 2009 Apr 1;25(7):974-6. PMID: 18326507.

91. Salgado D, Krallinger M, Depaule M, et al. MyMiner: a web application for computer-assisted biocuration and text annotation. Bioinformatics. 2012 Sep 1;28(17):2285-7. PMID: 22789588.

92. Aronson AR, Mork JG, Gay CW, et al. The NLM Indexing Initiative's Medical Text Indexer. Stud Health Technol Inform. 2004;107(Pt 1):268-72. PMID: 15360816.

93. Castleberry A. NVivo 10 [software program]. Version 10. QSR International; 2012. Am J Pharm Educ. 2014 Feb 12;78(1):25.

94. Cohen AM, Hersh WR, Peterson K, et al. Reducing workload in systematic review preparation using automated citation classification. J Am Med Inform Assoc. 2006 Mar-Apr;13(2):206-19. PMID: 16357352.

95. Raja K, Subramani S, Natarajan J. PPInterFinder--a mining tool for extracting causal relations on human proteins from literature. Database (Oxford). 2013;2013:bas052. PMID: 23325628.

96. Malheiros V, Hohn E, Pinho R, et al. A visual text mining approach for systematic reviews. In: Conference Proceedings. 2007 First International Symposium on Empirical Software Engineering and Measurement; September 20, 2007. p. 245-54.

97. Plikus MV, Zhang Z, Chuong CM. PubFocus: semantic MEDLINE/PubMed citations analytics through integration of controlled biomedical dictionaries and ranking algorithm. BMC Bioinformatics. 2006;7:424. PMID: 17014720.

98. Douglas SM, Montelione GT, Gerstein M. PubNet: a flexible system for visualizing literature derived networks. Genome Biol. 2005;6(9):R80. PMID: 16168087.

99. Zhang Y, Sarkar IN, Chen ES. PubMedMiner: mining and visualizing MeSH-based associations in PubMed. AMIA Annu Symp Proc. 2014;2014:1990-9. PMID: 25954472.

100. Cohen WW, Singer Y. A simple, fast, and effective rule learner. In: Proceedings of the Annual Conference of the American Association for Artificial Intelligence (AAAI); 1999. p. 335-42.

101. Mierswa I, Wurst M, Klinkenberg R, et al. YALE (now RapidMiner): rapid prototyping for complex data mining tasks. In: Proc ACM SIGKDD Int Conf on Knowl Discov Data Mining; 2006.

102. Pafilis E, O'Donoghue SI, Jensen LJ, et al. Reflect: augmented browsing for the life scientist. Nat Biotechnol. 2009 Jun;27(6):508-10. PMID: 19513049.

103. Vazquez M, Carmona-Saez P, Nogales-Cadenas R, et al. SENT: semantic features in text. Nucleic Acids Res. 2009 Jul;37(Web Server issue):W153-9. PMID: 19458159.

104. Kuhn M, Campillos M, Letunic I, et al. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol. 2010;6:343. PMID: 20087340.

105. Marshall C, Brereton P, Kitchenham B. Tools to support systematic reviews in software engineering: a feature analysis. In: Conference Proceedings. EASE '14: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering; May 13, 2014; London.

106. Hernandes E, Zamboni A, Fabbri S, Di Thommazo A. Using GQM and TAM to evaluate StArt -- a tool that supports systematic review. CLEI Electron J. 2012 Apr;15(1):paper 2.

107. Fabbri S, Hernandes E, di Thommazo A, et al. Managing literature reviews information through visualization. In: Conference Proceedings. 14th International Conference on Enterprise Information Systems (ICEIS 2012); June 28, 2012. p. 36-45.

108. Bowes D, Hall T, Beecham S. SLuRp - a tool to help large complex systematic literature reviews deliver candid and rigorous results. In: Conference Proceedings. 2nd International Workshop on Evidential Assessment of Software Technologies, EAST 2012; September 22, 2012. p. 33-6.

109. Yang JJ, Cohen AM, Cohen A, et al. SYRIAC: The systematic review information automated collection system a data warehouse for facilitating automated biomedical text classification. AMIA Annu Symp Proc. 2008;825-9. PMID: 18999194.

110. Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. J Stat Software. 2008 Mar;25(5):1-54.

111. Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84(2):523-38.

112. Sarker A, Molla D, Paris C. Automatic evidence quality prediction to support evidence-based decision making. Artif Intell Med. 2015 Jun;64(2):89-103. PMID: 25983133.

113. Rebholz-Schuhmann D, Arregui M, Gaudan S, et al. Text processing through Web services: calling Whatizit. Bioinformatics. 2008 Jan 15;24(2):296-8. PMID: 18006544.

114. Perez-Iratxeta C, Perez AJ, Bork P, et al. Update on XplorMed: A web server for exploring scientific literature. Nucleic Acids Res. 2003 Jul 1;31(13):3866-8. PMID: 12824439.

Appendix A. Search Strategies

The following databases were searched:

1. SRC Methods Library
2. Inspec
3. LISA via ProQuest
4. Science Citation Index
5. Social Sciences Citation Index via Web of Knowledge
6. Library, Information Science & Technology Abstracts (LISTA) via EBSCO
7. HTAIvortal http://vortal.htai.org/?q=about/sure-info
8. Google Scholar
9. IEEE
10. Embase.com
11. PubMed (in process materials search)
12. PsycINFO
13. Cochrane Library

Other online sources searched were:

1. NacTEM website
2. Research Synthesis Methods TOC
3. PLOS text-mining collection: http://www.ploscollections.org/article/browseIssue.action?issue=info:doi/10.1371/issue.pcol.v01.i14
4. MillionShort.com
5. ACM digital library

The search syntax used in the database searches was tested in Embase.com. A sensitive search strategy was used in the title, abstract, and keyword fields (where available).

<table>
<tr><td rowspan="2">Text-mining</td><td>EMTREE

‘Machine learning’/exp

Ultimate parent is ‘information processing’/exp</td><td rowspan="2">Literature mining
Text mining
Text analysis

Machine NEAR/2 learning

(document OR Text) NEAR/2 (classif* OR cluster* OR characteri* OR categoriz*)

‘support vector machine’
SVM

Cluster NEAR/2 tool*</td></tr>
<tr><td>MeSH

Exp artificial intelligence/</td></tr>
<tr><td rowspan="2">Natural language processing</td><td>EMTREE

‘natural language processing’:de</td><td rowspan="2">‘Natural language processing’
NLP
‘Term recognition’
‘Word frequency analysis’</td></tr>
<tr><td>MESH

Included under artificial intelligence along with support vector machines</td></tr>
<tr><td></td><td>EMTREE
‘information retrieval’</td><td>(information OR knowledge OR text) NEAR/2 visual*</td></tr>
</table>

	MeSH Cluster analysis 'Information storage and retrieval'	'objectively derived' Summarization 'text retrieval' 'visual data exploration' automat* Semi NEAR/1 automat* Active learning Citation management Review management (article* OR citation* OR document*) NEAR/2 (identif* OR retrieval* OR screen*)
Specific tools		'Abstrackr' 'Aquad' 'carrott2' 'cassandre':ti,ab 'coding analysis toolkit' 'computer aided textual markup & analysis' 'EPPI-Reviewer 4' 'FreeQDA' 'Konstanz Information Miner' 'KH Coder' 'LibreQDA' 'linguamatics' 'Machine Learning for Language Toolkit' 'medsum' 'NLM Medical Text Indexer' 'Pubhub' 'Pubnet' 'PubReMiner' 'QCAmap' 'QDA Miner Lite' 'Qigga' 'RQDA' 'SAS on demand' 'Semantic Features In Text' 'SIDER 2' 'Text analysis markup system' 'text mining infrastructure in R' 'Weft QDA' 'WordStat' 'Termine' 'Carrot Lingo 3G' 'Bibexcel' 'Voyant'
Systematic review	'systematic review (topic)' 'systematic review' 'meta analysis' 'meta analysis (topic)'	systematic NEAR/2 review* (evidence OR research OR comprehensive) NEAR/2 (synthes* OR review) 'Meta analysis' Meta-analysis
Methods		Method* Technique* Algorithm*

Limits: 2005-2015, English language

Set Number	Concept	Search Statement	# Identified
1	Text-mining	‘machine learning’/exp OR (machine NEAR/2 learning)	71,160
2		(‘literature’ OR ‘text’) NEAR/2 mining	1,595
3		(document OR text) NEAR/2 (classif* OR cluster* OR characteri* OR categoriz*)	606
4		‘support vector machine’:de OR ‘support vector machine’ OR ‘SVM”	11,271
5		Cluster* NEAR/2 tool*	248
6		#1 OR #2 OR #3 OR #4 OR #5	75,134
7	Natural language processing	‘natural language processing’:de OR ‘natural language processing’ OR ‘NLP’ OR ‘term recognition’ OR ‘word frequency analysis’	2,876
8	Combine sets	#6 OR #7	77,011
9	Potential uses	‘information retrieval’:de	16,708
10		(information OR knowledge OR text) NEAR/2 visual*	7,341
11		‘objectively derived’ OR summarization OR ‘text retrieval’ OR ‘visual data exploration’ OR ‘visual data representation’ OR ‘data abstraction’	1,428
12		Automat* OR (semi NEAR/2 automat*)	136,711
13		(citation OR review) NEAR/2 manag*	6,575
14		(article* OR citation* OR document*) NEAR/2 (identif* OR retrieval* OR screen*)	13,197
15		#9 OR #10 OR #11 OR #12 OR #13 OR #14	177,247
16	Systematic reviews	‘systematic review’:de OR ‘systematic review (topic)’:de	94,565
17		Systematic NEAR/2 review*	123,057
18		(evidence OR research OR comprehensive OR critical OR Cochrane) NEAR/2 (synthes* OR review*)	70,436
19		‘meta analysis’:de OR ‘meta analysis (topic)’:de OR (meta NEAR/1 analy*)	109,864
20	Combine sets	#16 OR #17 OR #18 OR #19	225,345
21	Limiting concepts	#20 AND (method* OR technique* OR algorithm*)	123,113
22	Combine sets	#8 AND #15 AND #21	140
23	Specific programs	‘Abstrackr’ OR ‘Aquad’ OR ‘carrott2’ OR ‘cassandre’:ti,ab OR ‘coding analysis toolkit’ OR ‘computer aided textual markup & analysis’ OR ‘EPPI-Reviewer 4’ OR ‘FreeQDA’ OR ‘Konstanz Information Miner’ OR ‘KH Coder’ OR ‘LibreQDA’ OR ‘linguamatics’ OR ‘Machine Learning for Language Toolkit’ OR ‘medsum’ OR ‘NLM Medical Text Indexer’ OR ‘Pubhub’ OR ‘Pubnet’ OR ‘PubReMiner’ OR ‘QCAmap’ OR ‘QDA Miner Lite’ OR ‘Qigga’ OR ‘RQDA’ OR ‘SAS on demand’ OR ‘Semantic Features In Text’ OR ‘SIDER 2’ OR ‘Text analysis markup system’ OR ‘text mining infrastructure in R’ OR ‘Weft QDA’ OR ‘WordStat’ OR ‘Termine’ OR ‘Carrot Lingo 3G’ OR ‘Bibexcel’ OR ‘Voyant’	213
24	Combine sets	#21 AND #23	6
25	Combine sets	#22 OR #24	146

Appendix B. Interview Guide

Introduction

The overall mission of the Agency for Healthcare Research and Quality's (AHRQ) Effective Health Care (EHC) Program is to provide evidence-based information to health care stakeholders that is relevant to their needs, timely, objective, scientifically rigorous in construct, and developed and presented with transparency. In the production of systematic reviews, we aim to answer questions about effectiveness of interventions and average population effects. We are aware that for certain conditions and behavioral interventions, these questions may miss important issues.

AHRQ engages stakeholders in all facets of their research enterprise, including producing systematic reviews, to ensure that research findings reflect the needs of diverse users, are relevant to their unique challenges, and are applicable in real-world situations.

Purpose of the Stakeholder Interview

The goal of our project is to understand which text-mining tools you have used, how you have used them, and the challenges you have encountered.

We are very interested in learning from your experience.

There are not right or wrong answers, so please feel free to share your thoughts openly.

We welcome any materials that you would like to share with us either before or after the interview session. Please send any materials to Robin.Paynter@va.gov.

Ground Rules for the Stakeholder Interview

The interview will be tape recorded, transcribed, and analyzed for overarching themes.

Although the report may list individuals who were interviewed, answers will not be identifiable to individuals or specific organizations.

You may refrain from answering any questions and are welcome to end the interview at any time.

Interview Guide – Senior Investigator/Systematic Review Organization Key Informant Questions:

1. Why did your organization decide to start using text-mining software?
2. Which text-mining software tool(s) has your organization used in past systematic review projects?
3. In which step or steps of the process has your organization used it?
4. What criteria were used to determine which software package to use?
5. Does utilizing text-mining software decrease or increase the length of time to complete the review process?
6. How well did it fit within your existing organizational workflows?
7. What were the expenses associated with its use in terms of staff training, software costs, etc.?
8. Were there or are there any issues for staff to adapt to its usage?
9. How long did it take for your organization to implement it?
10. Were there or are there technical facilitators or barriers to implementation?

11. Were there or are there organizational facilitators or barriers to implementation?
12. How do you evaluate the value of using text-mining software to the review?
13. How much confidence do you have in the results generated by the text-mining tools versus your previous process?
14. Do you report anything differently because you used text-mining in your review?
15. Would you recommend its use to other systematic review organizations?

Last question: Anything else you would like to add?

Interview Guide – Librarian Key Informant Questions:

1. Why did you begin using text-mining software to develop search strategies?
2. How long have you been using text-mining software, or in roughly how many search strategies have you used it?
3. Please describe how you utilize text-mining in your search process?
4. When using text-mining software as compared to not using it, are there gains or losses in efficiency (i.e., amount of time needed to develop) and/or completeness of the strategy?
5. Because search strategy topics vary widely, have you noted any types of questions for which text-mining software works particularly well or poorly?
6. How do you evaluate the value of using text-mining?
7. How much confidence do you have in the results generated by the text-mining tools versus your previous process?
8. How do you report using text-mining in your review?

If using keyword/synonym text-mining tools OR concept identification text-mining tools:

1. How did you evaluate this software to determine whether to use it?
2. Were there any loading or technical issues in setting up or using the software?
3. How long did it take to learn the software?
4. How easy is it to import/input data into the tool?
5. How do you use the output of the tool?
6. Would you recommend its use?

If using filter text-mining tools:

1. How did you evaluate the filter software to determine whether to use it?
2. Were there any loading or technical issues in setting up or using the software?
3. How long did it take to learn the software?
4. How do you assess whether the test set of articles is valid? Reliable?
5. Do you train on the citation/abstract or the full text or both?
6. How long does it take to train your software on average?
7. How easy is it to import/input data into the tool?
8. How do you use the output of the tool?
9. Would you recommend its use?

Last question: Anything else you would like to add?

Appendix C. Key Informant Interview Themes

Table C-1. Senior investigator/systematic review organization perspectives – interview themes

Themes	Exemplar Quote
Text-mining Overarching Comments	
Promise of technology	To help them screen things that have several hundred thousand citations to screen because they were doing something on social activity or something like that and there was just no terms that capture it properly within the different databases, but they were able to use a custom algorithm for computer learning to train it as they moved along. And it was able to cut down the several hundred thousand citations to only about 15,000 they actually screened in the end
	There's a gazillion ways to use these types of tools for screening, for example. You can use them to prioritize the order in which you will fully, manually screen everything. You can use them as the sole screener, so, for example, you may decide to rely completely on them for the second half of the review. You can use them only during the update that happens after the draft. You can use them as an anomaly detection system, meaning that you have your double manual screening, and then you train the classifiers and see if there are other papers that have been excluded by the humans that should be seen at the second—someone should take a second look at them
	This new technological platform is clearly pointing to the direction where this kind of thing will go, not least because the volume of scientific information being produced around the world is growing so fast and so exponentially, our conventional methods of 20 years or more ago are rapidly becoming out of date…. In broad terms, this kind of approach seems to ask to be, if not the future, certainly an important part of the future, and it's something that needs to be grasped with both hands
	There's one very important, though, proviso. The conventional systems, as we all know, start from the notion that a priori you define your research question, a priori you've got a pretty good sense of what's in the literature, and a priori you've got a pretty good sense of the direction of travel within that literature. The possibilities of text-mining, when you're facing a much broader literature, where there is a need to break out of the conventional ways of thinking, it allows you to search iteratively into the literature, and it allows for the problem formulation to emerge eminently in the literature and really, it turns on its head. At least it has the possibility to turn on its head the conventional paradigm
Issues to be aware of	Some papers in this area have extraordinarily good results, but that's because they have favorable datasets, so that with a nice degree of distinction between the relevance and the irrelevant classes with lots of jargon you can get great results. So if somebody sort of saw that and said oh, well great, well I'm going to use that now….I think that would be a naïve way of using it….It's just the applicability of research findings from one context to another due to the situation in which these tools were developed and tested, how close is it to mine, what can I do to check to see how well the text-mining tools are performing in my situation rather than just assuming that they're going to perform the same in any situation, because they don't
	So your average systematic reviewer won't find these particularly easy to use in sort of like to their native form; you have to sort of integrate and put a front end on them to make them useful in reviews
	You have to know about the strengths and the limitations of those things so that you know how to optimally use them. You cannot use them blindly, you cannot be shoddy or silly when you're training them…and you cannot rely on them as an excuse to do sloppy work, because these things are trained based on the labels that you provide to them, so if you are sloppy, thinking that the algorithm will do it, you are training a very bad algorithm, and in the end it will backfire
	[the tool] does offer you cutoff points according to sort of probabilities, but using the rules-based approach, when we applied the rules, there were 15,000 records that met the rules that the rules said were relevant and I think the reviewers looked through all 15,000, but didn't look through the remaining [65,000] records. I'm sure in and amongst those other records there were probably relevant records that didn't meet the rules, but the reviewers didn't have enough time to go through all of those a well. That's the question, it's sort of the rules are great in a sense, but you do worry about the records that didn't meet the rules, because there's bound to be some that were probably highly relevant, but just didn't meet the rules
Why Did Your Organization Decide to Start Using Text-mining Software?	
Process improvement	The reason was we were facing the big uphill task of undertaking two major reviews in areas where it was apparent that there was a very large amount of published material. That published material ranged over numerous different disciplines, had been generated using different methods and methodologies, and the conventional approach…would merely scratch the surface of what we wanted to know

Table C-1.Senior investigator/systematic review organization perspectives – interview themes (continued)

Themes	Exemplar Quote
In Which Step(s) of the Process Has Your Organization Used?	
Searching	The information scientist, who puts together a lot of the searches, has used different tools to help think about search terms
Screening	The most heavy use we've made of it has been in terms of prioritizing the order in which citations are screened at the title and abstract stage, and we just routinely use text-mining for that now; it's sort of built into the system. We still haven't been able to let go of that usual sort of systematic reviewer obsession with looking at everything, so we still look at everything. But what we do is we prioritize the order, so we end up finding the most relevant citations very early in the screening process, which means we can then get on with the next stages of the review with those citations, retrieving the full text and screening them and getting on with the data abstraction and risk bias assessment and things, and that part gets moving while the screening is still going on the long tail of irrelevant studies
What Criteria Were Used to Determine Which Software Package(s) to Use?	
No general theme emerged, the quotes presented show their respective thoughts on the topic	One is performance, I mean do they actually do what we need them to do, and the other criterion would be integratability, for want of a better word, so can we actually integrate the tool within existing web processes and tools. We use these tools basically to save time and labor, so they need to be integrated, but within our current process, otherwise they just add to the workloads
	I've picked up developing software in VBA for the simple purpose that as part of the budget we had no specific budget for software, so a lot of people will use things like for managing their citations like Distiller SR....Some of the people will use EPPI-Reviewer. Others will use what is available to them and what they're trained on, so if they are trained on for example creating an access database, they will use that
	I'm not sure that we did apply criteria....So it wasn't done on the basis of an option to appraise it, rather, if you like ongoing developmental work arising out of a project in another university and the natural colleagueship between different colleagues working in this new area
Does Utilizing Text-Mining Software Decrease or Increase the Length of Time to Complete the Review Process?	
Systematic Review	We have not measured it formally in our shop
Scoping Review	It decreased it by eons. It would've been utterly impossible to undertake the scoping reviews that we did using conventional methods. The actual time spent in appraising the papers that were found and observing the linkages between them was probably undoubtedly less than a standard systematic reviewer would take using conventional search and find techniques. So very definitely, it's an aid to efficiency
How Well Did It Fit Within Your Existing Organizational Workflows?	
Integration with systems and work flows	We've integrated it within the software which we use all the time, because then there are no problems in terms of data loss and maintaining data integrity and all that sort of thing. There's no real learning curve to use a lot of this, because it's all just sitting there in the space that everyone's used to using anyway. Yes, so it fits well. But it does mean that you have to do some upfront work to actually get the integration
What Were the Expenses Associated with Its Use in Terms of Staff Training, Software Costs, etc.?	
Staff training	It's a difficult thing to answer for us, because people either have been involved with these things from the get-go or have been exposed to them for so long that it's—I don't know what the training cost would look like for someone who is using our tool de novo
Software	So there are costs associated with the tools and their implementation and their debugging and their maintenance, but it is very difficult to give you a breakdown
Were There or Are There Any Issues for Staff to Adapt to Its Usage?	
Searching	After creating the printout of the frequency distribution, etc., I circulate it to the different staff members, especially the content experts to discuss the different terms and different option and I have not really had any sort of feeling of any issues, any negative issues with understanding the concepts of it

Table C-1.Senior investigator/systematic review organization perspectives – interview themes (continued)

Themes	Exemplar Quote
Screening	We jokingly refer to it here as extreme reviewing. In other words, the volume of stuff that you generate like this is potentially overwhelming. The people who were appraising the papers that were generated by this way really had to pull out all the stops to do it…. I don't know if it's so much extra work, but it produces a hump of work, or a peak of work, which the team has to go full throttle to keep up with
Were There or Are There Organizational or Technological Facilitators or Barriers to Implementation?	
Facilitators	Once the decision was taken that that's what we were going to do by the senior members of the team, we did it. I don't think there were any significant implementation issues in terms of it taking time or anything like that. We've got a very committed staff who were very interested scientifically in how it would work, so they pulled out all the stops
	You clearly need good systems, and you need reliable systems....if you're trying to do this, or use this, in a system that was flaky, where connections didn't work, where systems were going down all the time, then it would be a major headache
Barriers	Although I'm a great enthusiast, and I see it as something helping university research and indeed systematic review organizations, there will, I think, be a bit of a war of attrition on the ground to resist some of this and to argue that machines can never do it as well as humans for all the reasons that people have been saying that since the invention of the spinning jenny and weaving cotton and so on
	Where the barriers came from was the specialist information scientists, who were the people who would help you to research strategies and all that kind of thing, who I think felt very threatened by text-mining's apparent ability to do what they spent their lives being training to do and they were professionally trained to do. I think that there was some considerable resistance to the research project from within the information specialists, information scientists, or at least some of them in the organization, who I think, with some justification actually, could see themselves being de-skilled by this technology
How Do You Evaluate the Value of Using Text-mining Software to the Review?	
Need to do more formal evaluation in future	We haven't done a formal evaluation
	I think that we need a lot of work, empirical work in terms of eliciting metrics from experts in order to answer that question of evaluation
How Much Confidence Do You Have in the Results Generated by the Text-mining Tools Versus Your Previous Process?	
No general theme emerged, the quotes presented show their respective thoughts on the topic	I think text-mining tends to work most effectively or historically has done so certainly in more clinical technical areas, because there's a nice little jargon for the text-mining to work off so it just finds that easier. I mean concepts such as healthy eating that kind of thing can be expressed in a number of different ways, and so the concepts of fuzzy area and it's more difficult for the text-mining to really sort of get a grip on what's going on
	All scientific findings of course are contingent, and ours are no different, and those generated by this method are no different. At the very least, I would say we had equal confidence in them from anything we've done before, and probably given the breadth of what we were able to look at, we may have even more confidence
	No one can guarantee you that these tools will find everything there is to find. There is always a risk, so if you use the tool the way that we have implemented it in our shop, which is to tell us in which order we will screen the literature. We still screen everything, but we start from the high, promising part and then move on to the other parts. You can trust it in the sense that you still complete the whole evidence base, but it has reordered the way that you do the full manual screening to make things move a bit faster….retrieving the full text and screening them and getting on with the data abstraction and risk of bias assessment and things…that part gets moving while the screening is still going on in the long tail of irrelevant studies

Table C-1.Senior investigator/systematic review organization perspectives – interview themes (continued)

Themes	Exemplar Quote
Do You Report Anything Differently Because You Used Text-mining in Your Review?	
No general theme emerged, the quotes presented show the variety of reporting strategies	…report as transparently as we can what we did. It is challenging, because not many reviews have used text-mining for work production. It's quite easy to do it for prioritizing screening, you're just changing the order in which things are screened… we just speeded things up a bit for the later stages of the review. For other purposes, because you can't just easily say we've used this method, you have to go into quite a lot of detail exactly about what was done and there's no established method for any of this yet, so if you really wanted to report in proper detail about what you did you need to get into the specifics about what kind of feature selection you did, whether you stemmed words, whether you had a stop word list, which stop word list you used, and all that kind of thing potentially if you wanted to be properly transparent about it. At the moment there's no real sort of accepted way in which people do report this, so the danger is you end up writing a great long treatise on text-mining rather than the other methods of the review and the actual topic of the review itself… I don't think that we really know exactly how much detail we should be putting in reports
	I don't report anything differently
	So the only thing we might write is that we use [tool name deleted to preserve anonymity] for this kind of thing, to do the screening, but manage the logistics of the screening. If we use the machine learning algorithms for a double check, sometimes we may mention it as a quality control, or sometimes we may not mention it at all, but we have not been doing anything like that systematically

Table C-2. Librarian/review team member perspectives – interview theme

Themes	Exemplar Quote * **
Text-Mining: Overarching Comments	
[perform a lot of research in a short amount of time and cannot be experts on all of the different questions, so text-mining is helpful.... new people perform equally in searching even if they are not as experienced]	
[Tool in developing the strategy rather than the only way to develop a search]	
I find that if you use a search strategy that's created in that way [using a filter] it's often too sensitive and too non-specific and that you find too much noise if you do it. So sometimes I still resort to doing my own traditional way in which I'm rather fast and that's one of the downsides why I don't use the text-mining a lot because I'm so fast with creating a search strategy	
The ability to search across disciplines, the ability to search across paradigms and methods, and the possibility of finding connections which you would only ever come across serendipitously in using all the conventional approaches that we do – [Senior Investigator/Systematic Review Organization]	
Some of the possibilities for translational facilities that can be done within it, its ability to search into the gray literature and beyond, and now into the blogosphere and all that kind of thing, offer ways of making what is otherwise utterly unmanageable knowledge, modern Tower of Babel into something that you could get at forensically. Also very importantly, to find connections in ways that you didn't think possible or wouldn't have occurred to you - [Senior Investigator/Systematic Review Organization]	
Why Did You Begin Using Text-Mining Software to Develop Search Strategies?	
Objectivity	Text-mining allows me to do that in a very objective way and to identify this information without spending too much time as an individual, but rather let the computer run the algorithm and figure that out for me
How Do You Utilize Text-Mining in Your Search Process?	
Keyword / synonym tools	We will use text-mining to help us start to do the sort of scoping searches to find the search terms that might feature in the strategy. We use the text-mining to find synonyms, to find the subject headings, also to help us to see whether we can introduce new limits, so maybe geographic limits or whether there are particular literatures that we could try to safely exclude

Table C-2. Librarian/review team member perspectives – interview theme (continued)

Themes	Exemplar Quote * **
When Using Text-Mining Software as Compared to Not Using It, Are There Gains or Losses in Efficiency (i.e., Amount of Time Needed to Develop) and/or Completeness of the Strategy?	
Keyword / Synonym Tools	I think there are gains in efficiency, and there are gains in reliability. For example, if I'm running a quick search in PubReMiner, it makes me more confident that—I mean, PubReMiner can do frequency analysis and give me a list of terms from hundreds of records which I, personally, wouldn't have the time to look at. I couldn't look through 500 records and note down all the terms and make decisions about which ones are important for my strategy. PubReMiner gives me, for example, a nice table from which I can scan down and discuss with colleagues, and it gives me an objective feel for what those keywords are. It also lets me see the less frequently used terms which is going to be hard for me to spot if I'm just looking through a set of records by eye…Also by doing a series of small searches and getting to the point where I'm not finding extra terms, I can perhaps more quickly get to a sense of if I've got on top of the search terms that are available. That probably takes longer if I'm doing searches and looking through batches of records by eye
Filter Tools	To create a filter takes me maybe four hours if we have the gold standard strategy, so basically it often takes too long and I'm very happy with doing slightly less sensitive search strategies in a much shorter period than having to create filters for hours
Because Search Strategy Topics Vary Widely, Have You Noted Any Types of Questions for Which Text-Mining Software Works Particularly Well or Poorly?	
Works well	I think it does work well for these horribly fuzzy topics and for the multi-question reviews where you're trying to save time by doing one big search and then interrogating it from several different angles….But for public health topics and nasty complex questions, I think text-mining sort of really does come into its own
Works poorly	I tend to think that it's not so useful when you've got a very clear cut question. A lot of our reviews are for single drugs for a very specific condition, and so for questions like that, text-mining is probably not helpful. You just don't need it. You have the drug name and some alternatives, and it's fairly straightforward to either just search for the drug name or the drug name in combination with the condition. For topics like that, from our perspective, you don't use text-mining, because it's very easy to develop a traditional search strategy.
How Do You Evaluate the Value of Using Text Mining?	
No general theme emerged, the quotes presented show the variety of responses	I think there are possibly gains in efficiency there, but I've never actually measured it so I don't have any comparative information, but that does feel possibly more efficient than the usual way I'd go about finding terms by doing searches and looking at sets of records. If I'm looking at ten records, and deciding I've seen enough, you just get a chance to analyze more records than you could ever possibly scan by eye. That sort of makes me feel more confident, I guess, in those strategies I'm going to test out
	[Tried to look at different functions of the text-mining tool: word frequency, adjacency. Made the process longer but the search was more sensitive. … Exclusion analysis – a clear efficiency for later screening, for example a search in Medline on late diagnosis removed 4% of the studies (about 500 references) through using 'NOT']
	I guess what made us feel more reassured with the final review was that the review team also did lots of reference checking and looking at the cited references within relevant studies that they found. We sort of hoped that that would compensate in some way for the pragmatic decisions made elsewhere
How Much Confidence Do You Have in the Results Generated by the Text-Mining Tools Versus Your Previous Process?	
Collaboration	I collaborate with an experienced text-mining specialist to make sure that I'm not worried about the quality of the end product… They're very complex packages. You need to know what you're doing, you need to know the algorithms that you're trying to achieve, and obviously, with expertise you can do things faster than if you don't know
Keyword / synonym tools	I feel more confident that we've found a variety of search terms that reflect the literature, and it also can highlight areas where literatures are more complex
Filter tools	In some ways it's an extension of a manual process and gives a level of assurance. It can be problematic because you can be misled by the corpus that you use to train the tool

Table C-2. Librarian/review team member perspectives – interview theme (continued)

Themes	Exemplar Quote * **
Reporting Search Strategy Development in Systematic Reviews	
Some text-mining tools reported but not others	I consider it to the whole process of text-mining to be more of a background search. Again, it will be the same as reviewing citations or hand searching includes the references of included searches. I don't particularly cite it as anything particular, because it is not an automation in and of itself, it's only an indication of where the human interaction can be best used….for example if I was using beta filtering or if I was using some sort of computer learning algorithm, so therefore there is an added sensitivity/specificity to it. The accuracy can differ depending on your algorithm and that could affect your overall results. I believe that would be something that is needed to be cited, but I don't see that the text-mining, per se, it would be no different than me saying for example before writing up a full search strategy, I did a rapid scoping of the literature or I did some quick and dirty searches. Those would not be identified as part of the official search that was done. I consider the text-mining to be in the similar sort of the same ballpark as that
Keyword/Synonym Text-Mining Tools or Filter-Type Tools KI Comments	
How Did You Evaluate the Software to Determine Whether to Use It?	
No general theme emerged. The quotes presented show the variety of responses	I tend to evaluate them by the types of facilities they offer and by the flexibility for altering the options that they offer and also for getting the output out
	[So many tools for different functions. Tried to characterize them and choose best tool for the purpose. Considered accessibility at her desktop. They use Lingo 3G through EPPI-Reviewer. Cost is an important factor. "Trialability" is important as well]
	I don't know if I ever really evaluated it. I don't remember when I heard about PubReMiner. It's a long time. I think when I started working six years ago, it already existed and my colleagues used it and that's why I started using it
Were There Any Loading or Technical Issues in Setting Up or Using the Software?	
No issues	They're very straightforward, they're very reliable, and they tend to work every time
Issues	Problems with firewalls and software setting off local virus detection software, sometimes need to IT people to resolve the issue
	The text-mining packages I use, it's sort of that wide range, from things that anybody can use through to things where we actually need someone to help us use them
How Long Did It Take to Learn the Software?	
Keyword / synonym tools	I think PubReMiner, GoPubMed, programs like that are very rapid to learn. I mean you could learn them in half an hour; I think most people can get to grips with them very quickly. Once you move into the areas of more sophisticated text-mining packages like GATE or Mimir or even VOSviewer and SimStat then you need a lot of time to get the best out of the packages
How Do You Use the Output of the Tool?	
Strategy evaluation	[Inform search strategy, take a subset and re-analyze text to see what you don't want]

* Comments appearing in square brackets are derived from workgroup member notes from the two KI interviews that were not recorded due to technical issues.

**Some comments included in this table are from senior investigators, [Senior Investigator/Systematic Review Organization] was added to each quote to clarify its source.

Appendix D. Key Informant Comments on Specific Text-Mining Tools

Table D-1. Librarian comments on specific text-mining tools

Name	Comment*
AntConc	[Which terms are next to each other to identify phrases and words next to each other] I've recommending AntConc to some people designing and creating filters….if you really want to have a good objective way of creating a filter, you need more than just PubReMiner, then you need AntConc …If I do a search strategy, what I often use is proximity, so that words can be in close proximity within the certain set of words and that's one thing that I miss with the current text-mining programs. They cannot identify words that I could use in proximity. They only identify phrases and that's one thing that I would like to see in the future that maybe AntConc can solve [this problem] for me… [Problems with firewalls and software setting off local virus detection software, sometimes need to IT people to resolve the issue] AntConc is something that you download and you don't even need administrator rights, so you can just download it on [indiscernible] and save it anywhere, so no technical issues in it
EndNote	We use EndNote as well, particularly for the databases other than PubMed and MEDLINE where you don't have all the nice interfaces that people have built. The frequency analysis in EndNote we use for analyzing records from Embase and PsycINFO and so on. We can analyze the subject headings there [EndNote used to analyze keywords because they are easy to use. Shows frequency of keywords regardless of subheadings]
EPPI-Reviewer	[Eppi-Reviewer tools didn't take long [to learn how to use]]
GATE	For very complex questions, we've used the sort of more sophisticated software such as GATE Once you've got them loaded on your machine you're fine; I'm mean VOSviewer doesn't take long to load at all, even I can load that. But things like Mimir and GATE, it's not so much loading the software it's the sequence. You have to sort of bolt together a series of tasks within the software, and its knowing the order that you need to do those and the impact of each of the tasks. They're more of a tool box than a ready to run package, and you need to know the tools you want to select. They're not as user-friendly and intuitive as things like PubReMiner With GATE, I don't feel confident at all. I wouldn't drive that myself, I would leave it to somebody far more experienced who actually knew what they were doing
GoPubMed	More recently have been using software interfaces through PubMed such as PubReMiner and GoPubMed which again are frequency analysis packages to help us develop our strategies, but they perhaps are not as sophisticated as some text-mining packages GoPubMed looks nicer, but doesn't have a lot of the flexibility that I like in PubReMiner. It's nice for some of the nice graphics for showing the review team certain things and it's useful for graphical presentation of offered groupings, but it doesn't have a lot of the analyses that I actually value most
KNALIJ	PubReMiner and GoPubMed and KNALIJ and some of the other interfaces to PubMed, there people tend to do quick and clean searches or quick and dirty searches There are other packages out there, ones like KNALIJ....This is supposed to provide you with graphical presentations of batches of search results. It looks nice when it works, but it's very unpredictable about whether it'll actually work
Lingo3G	Lingo3G and Termine - good for quick overviews

Table D-1. Librarian comments on specific text-mining tools (continued)

Name	Comment*
PubReMiner	PubReMiner is easy to use and I've recommended it very frequently indeed. Also to lay people, the people that just started learning to create more systematic searches than just typing in four words in PubMed…. They can understand what you can gain out of PubReMiner I personally prefer PubReMiner compared to GoPubMed, because PubReMiner will give you an analysis of the words in the title and abstract. It also seems to give you a bit more control over the analyses you can undertake compared to GoPubMed, but GoPubMed now has …introduced sort of semantic analysis. It's not genuinely semantic analysis, but they've introduced some semantic analysis options as well which means it can do things that PubReMiner can't do PubReMiner will allow you to export the frequency tables as Excel spreadsheets, so that's really handy both for records and for sharing with people. Again, that's sort of a plus on the side of PubReMiner I think PubReMiner, GoPubMed, programs like that are very rapid to learn. I mean you could learn them in half an hour; I think most people can get to grips with them very quickly [PubReMiner is fast & easy to use. [It] has keyword limitations – can only analyze keywords together. Have to delete subheadings to get an accurate count. Is a good starting tool but doesn't do overrepresented terms] One of the downsides of … PubReMiner is [it is] only on PubMed, and we use MEDLINE … It's why I would not use it as often as I would if I used PubMed as my basic database
SimStat	Once you move into the areas of more sophisticated text-mining packages like GATE or Mimir or even VOSviewer and SimStat then you need a lot of time to get the best out of the packages
TerMine	Lingo3G and Termine - good for quick overviews
Text Mining Infrastructure in R	[Initially used R statistical program with text-mining module. It was free but it didn't have an interface so you needed to use a syntax language. Their colleagues had difficulty using this because they weren't programmers. They had syntax-related errors]
VOSviewer	Sometimes doing things like visual presentations, like using VOSviewer to show us the makeup of the literature visually can help us to see particular concepts that we're actually going to want to exclude later on …we've been working with a specialist in text-mining ….using more sophisticated text-mining software such as GATE and VOSviewer… These have been more serious text-mining packages that are not usually usable by most information professionals, so we've needed an intermediary to help us use that type of software I've used VOSviewer. It took me no time at all. Actually it took me probably a couple of hours reading the manual and then trying a few things out. In half a day I was feeling reasonably confident with VOSviewer You can export 10,000 results and load them very easily say into VOSviewer in a matter of moments just from exporting 10,000 records from PubMed in the PubMed format. Most of these sophisticated packages are built to take bibliographic records including abstracts in other fields
WordStat	[They switched to WordStat because it is focused just on text-mining. It isn't free, but it had a good user interface. Easier to use and fewer user errors. They could import from PubMed and calculate the overrepresentation]

*Comments appearing in square brackets are from workgroup member notes of the two KI interviews that were not recorded due to technical issues.

Appendix E. Tools Catalog

Table E-1. Tools evaluation table

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
Abstrackr http://abstrackr.cebm.brown.edu/account/login	FR 0	1	NA	Brown University	1	1	14	52	IE	1	0	ND	ND	1	ND
Anne O'tate http://arrowsmith.psych.uic.edu/cgi-bin/arrowsmith_uic/AnneOTate.cgi	FR 0	1	NA	Department of Psychiatry and Psychiatric Institute, University of Illinois at Chicago	0	0	16 34 64	64	TR FI	1	0	IR DCL ATM TSM	ND	0	ND
AntCont http://www.laurenceanthony.net/software/antconc/	FR 0	0	MULTI	Laurence Anthony	UC	1	ND	ND	TR FI	1	1	ATM DCL	1	ND	Requires installation
AQ21 machine learning software http://www.mli.gmu.edu/software	FR 0	0	MULTI	Janusz Wojtusiak	UC	0	65	ND	ND	1	ND	ND	1	0	Requires installation
Aquad http://www.aquad.de/en/	FR 0	0	MULTI	Gunter L. Huber	UC	0	ND	ND	ND	1	ND	ND	ND	ND	Requires installation
askMEDLINE http://askmedline.nlm.nih.gov/ask/ask.php	FR 0	1	NA	National Library of Medicine	0	0	16	66	TR FI	1	0	NLP IR	ND	ND	ND

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
Automated Sequence Annotation Pipeline (ASAP) http://bioinformatics.fccc.edu/software/OpenSource/ASAP/ASAP.shtml	FR 0	0	MULTI	Fox Chase Cancer Center	UC	0	54	67	IE	1	2	DCL NLP	1	0	ND
BioIE http://www.bioinf.manchester.ac.uk/dbbrowser/bioie/	FR 0	1	NA	The University of Manchester, Bioinformatics	0	0	ND	68	TR FI	1	0	ND	ND	0	ND
BioTextQuest	FR 0	1	NA	University of Crete	0	0	34	69	TR FI	1	0	DCL NLP VIZ	ND	0	Web site cannot be displayed
carrott2 http://project.carrot2.org/download-workbench-win32-64bit.html	FR 0	0	MULTI	David Weiss and Stanislaw Osinksi	0	1	70 34	ND	DCT	1	ND	ND	1	ND	Requires installation
Cassandre http://cassandra.apache.org/	FR 0	0	MULTI	Apache Software Foundation	ND	0	ND	ND	ND	ND	2	ND	ND	ND	Probably out of scope
Chilibot http://www.chilibot.net/	FR 0	1	NA	University of Tennessee Health Science Center	0	0	16	71	TR FI	1	0	NLP IR VIZ ATM	1	0	ND
Coding Analysis Toolkit (CAT) http://cat.ucsur.pitt.edu/	FR 0	0	MULTI	Qualitative Data Analysis Program, University of Pittsburgh	0	0	ND	ND	IE EX	1	2	DCL IEX	1	ND	Requires installation

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
Computer Aided Textual Markup & Analysis (CATMA) http://www.catma.de/	FR 0	1	NA	University of Hamburg	ND	0	ND	ND	ND	ND	ND	ND	1	1	Literary analysis
Concordance http://www.concordancesoftware.co.uk/	FRT 0	0	MULTI	R. J. C. Watt	0	0	[37] [36]	ND	ND	ND	ND	ND	ND	ND	Requires installation
CoPub http://services.nbic.nl/copub/portal/	FR 0	1	NA	Computational Drug Discovery Group of Radboud University Nijmegen Medical Centre	0	0	[16]	[72]	TR FI	1	0	NLP	ND	0	ND
COREMINE medical http://www.coremine.com/medical/#search	FR 0	1	NA	PubGene AS	0	0	[16]	ND	TR FI	1	0	NLP ATM IR VIZ	1	0	ND
Dbpedia http://wiki.dbpedia.org/	FR 0	0	MULTI	Dbpedia	0	0	ND	[73]	ND	ND	ND	ND	ND	ND	Requires installation
EndNote http://endnote.com/	CP varies	1	MULTI	Thomas-Reuters	1	1	[46]	[46]	ND	1	ND	ND	ND	ND	ND
EPPI-Reviewer 4 (EPPI) http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=1913	FRT UC	1	NA	Evidence for Policy and Practice Information and Coordinating Centre, UCL Institute of Education, University of London	1	1	[14]	ND	IE EX	1	1	ATM DCT DCL	1	1	ND

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
ETBLAST http://etest.vbi.vt.edu/etblast3/	FR 0	1	NA	Virginia Bioinformatics Institute	0	0	16	74	TR FI	1	0	NLP	1I	0	ND
ExaCT Not found	ND ND	ND	ND	ND	ND	0	ND	75	EX	ND	ND	ND	ND	ND	ND
FACTA+ http://www.nactem.ac.uk/facta/	FR 0	1	NA	National Centre for Text Mining, School of Computer Science, The University of Manchester	0	0	16	76	TR FI EX	1	0	NLP IR IEX	1	0	ND
FreeQDA https://github.com/produnis/FreeQDA/downloads	FR 0	0	ND	ND	0	0	ND	ND	ND	ND	ND	ND	ND	ND	Requires installation
GAPScreener http://www.hugenavigator.net/HuGENavigator/HNDescription/opensource_GAP.htm	FR 0	0	MULTI	Wei Yu, National Office of Public Health Genomics, Centers for Disease Control and Prevention	UC	1	ND	77	TR FI IE	1	2	IEX	ND	ND	Requires installation; human genome epidemiology network
GATE https://gate.ac.uk/	FR 0	0	MULTI	University of Sheffield	UC	1	78	60	TR FI IE	1	2	TCL	1	ND	Requires installation
Gopubmed http://www.gopubmed.org/web/gopubmed/	FR 0	1	NA	Transinsight GmbH	0	1	16 34 35	79	TR FI	1	0	NLP	1	1	ND
Hierarchical Cluster Explorer (HCE) http://www.cs.umd.edu/hcil/hce/	FR 0	0	WIN	Human-Computer Interaction Lab, University of Maryland	UC	0	78	80	ND	ND	ND	ND	ND	ND	Requires installation

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
HubMed http://git.macropus.org/hubmed/	FR 0	1	NA	ND	0	0	ND	[81]	TR FI	1	0	NLP	1	0	
HuGENet http://www.hugenavigator.net/HuGENavigator/HNDescription/opensource_GAP.htm	FR 0	1	NA	ND	ND	1	ND	ND	ND	ND	ND	ND	ND	ND	ND
Idealist Unavailable	ND ND	ND	ND	ND	ND	0	[36]	ND	ND	ND	ND	ND	ND	ND	Blackwell bibliographic software
KH Coder http://khc.sourceforge.net/en/	FR 0	0	MULTI	Koichi Higuchi	UC	0	ND	[82]	ND	ND	ND	ND	ND	ND	Requires installation
Konstanz Information Miner (KNIME) https://www.knime.org/	FR 0	0	NA	KNIME	UC	0	ND	ND	ND	ND	ND	ND	ND	1	Requires installation
LEXIMANCER http://info.leximancer.com/	FRT varies	1	NA	Leximancer	1	0	[41]	ND	ND	ND	ND	ND	ND	ND	Requires installation
libSVM http://ntucsu.csie.ntu.edu.tw/~cjlin/libsvm/	FR 0	0	MULTI	Chih-Chung Chang and Chih-Jen Lin	UC	0	[40]	ND	IE	1	2	DCL	ND	ND	Requires installation
Lingo3G http://carrotsearch.com/lingo3g-overview	FRT varies	0	MULTI	Carrot	UC	1	[83]	ND	ND	ND	ND	TCL VIZ	ND	ND	Requires installation

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
LingPipe http://alias-i.com/lingpipe/	FR 0	0	WIN	Alias-i	UC	0	[84]	ND	EX	1	2	IEX	1	ND	Requires installation
linguamatics (I2E) http://www.linguamatics.com/welcome/software/I2E.html	CP varies	1	NA	Linguamatics	UC	0	ND	[85]	ND	ND	ND	NLP	ND	ND	Trial includes subset of data
Machine Learning for Language Toolkit (MALLET) http://mallet.cs.umass.edu/	FR 0	0	MULTI	Andrew McCallum, University of Massachusetts	UC	0	ND	ND	ND	1	2	NLP DCL IEX	1	ND	Requires installation
MEDIE http://www.nactem.ac.uk/medie/	FR 0	1	NA	Tsujii Laboratory	0	0	[16]	[86]	TR FI	1	0	NLP IR ATM	0	ND	ND
Medline Ranker http://cbdm-01.zdv.uni-mainz.de/~jfontain/cms/?page_id=4	FR 0	1	NA	Computational Biology and Data Mining group, Miguel Andrade-Navarro, Johannes Gutenberg University Mainz	0	0	[16] [34]	ND	TR FI	1	0	ATM TCL	1	ND	ND
medsum http://webtools.mf.uni-lj.si/public/medsum.html	FR 0	ND	ND	ND	ND	0	[34]	[87]	ND	ND	ND	ND	ND	ND	ND
MeSHY http://tools.bat.infspire.org/meshy/	FR 0	1	NA	Bioinformatics Analysis Team	0	0	[87]	[88]	TR FI	1	1	IR TCL	ND	ND	ND

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
MetaMap http://metamap.nlm.nih.gov/	FR 0	0	ND	Alan Aronson, National Library of Medicine	ND	0	38 64	89	ND	ND	ND	ND	1	ND	Out of scope
Meva http://www.med-ai.com/meva/	FR 0	1	ND	Institute for Medical Statistics and Epidemiology, Technical University of Munich	0	0	ND	90	TR FI	1	1	ATM TCL	1	ND	ND
Mscanner http://mscanner.stanford.edu/	FR 0	1	NA	Graham Poulter	0	0	43	91	TR FI	1	0	DCL IR	1	ND	ND
MyMiner http://myminer.armi.monash.edu.au/	FR 0	1	NA	Australian Regenerative Medicine Institute, Monash University	0	0	NA	92	TR FI	1	1	TCA	1	ND	ND
NLM Medical Text Indexer (MTI) http://ii.nlm.nih.gov/MTI/	FR 0	ND	ND	ND	ND	0	ND	93	TR FI	ND	ND	NLP	ND	ND	ND
Nvivo http://www.nvivo10.com/en/whats-new.html	CP varies	0	MULTI	QSR	UC	0	15	ND	ND	1	1	TCA	1	ND	Requires installation
OpenNLP https://opennlp.apache.org/	FR 0	0	MULTI	Apache Software Foundation	UC	0	54	ND	IE	1	1	TCA	ND	ND	ND

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
Perceptron classifier http://reference.wolfram.com/applications/neuralnetworks/NeuralNetworkTheory/2.4.0.html	NA NA	NA	NA	NA	NA	1	[94]	ND	ND	NA	2	ND	NA	ND	Classifier algorithm
PICO http://pubmedhh.nlm.nih.gov/nlmd/pico/piconew.php	FR 0	1	NA	National Library of Medicine	0	0	[16]	ND	TR FI	1	0	IR	ND	ND	ND
Pimiento http://erabaki.ehu.es/jjga/pimiento/	FR 0	0	MULTI	Juan Jose Garcia Adeva	UC	0	[14]	[95]	ND	1	2	DCT TSM	ND	ND	Requires installation
PPIinterFinder http://www.biominingbu.org/ppinterfinder/about.html	FR 0	1	NA	Data Mining and Text Mining Laboratory, Department of Bioinformatics, Bharathiar University	0	0	ND	ND	ND	ND	ND	ND	ND	ND	Protein-protein interactions
Projection Explorer (PEx) http://infoserver.lcad.icmc.usp.br/infovis2/PEx	FR 0	0	MULTI	Fernando Vieira Paulovich, University of Sao Paulo	UC	0	[96]	ND	TR	UC	2	VIZ	ND	ND	ND
PubCrawler http://pubcrawler.gen.tcd.ie/	FR 0	1	NA	Ken Wolfe Lab, Genetics Department Trinity College	0	0	[16]	[97]	TR FI	1	0	IR	1	1	ND
PubGet http://pubget.com/	FR 0	1	NA	ND	ND	0	[16]	ND	ND	ND	ND	ND	ND	ND	ND

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
PubMatrix http://pubmatrix.grc.nia.nih.gov/	FR 0	1	NA	National Institute of Health	0	0	16 34	98	TR FI	1	1	IR	ND	1	Registered but unable to login
PubNet http://pubnet.gersteinlab.org/	FR 0	1	NA	Gerstein Lab, Department of Molecular Biophysics and Biochemistry, Yale University	0	0	16	99	TR FI	1	1	IR VIZ	1	ND	ND
PubReMiner http://hgserver2.amc.nl/cgi-bin/miner/miner2.cgi	FR 0	1	NA	Jan Koster, Academic Medical Center	0	1	42	ND	TR FI	1	0	TCL IR	1	ND	Standalone for searching
PubViz https://github.com/aha01/PubViz/	FR 0	0	MULTI	Matthias Fabi, Andrea Haberson, San Rasul, Elisabeth Schnaitt, Paul Theisen	0	0	16	ND	TR FI	1	1	VIZ	1	ND	Requires installation
QCAmap http://www.qualitative-content-analysis.aau.at/software/	FR 0	1	NA	Philipp Mayring, Thomas Fenzl	0	0	ND	ND	ND	ND	ND	ND	1	yes	Qualitative content analysis
QDA Miner http://provalisresearch.com/products/qualitative-data-analysis-software/	FR 0	0	WIN	Provalis Research	1	0	ND	ND	ND	1	1	ND	ND	ND	Qualitative content analysis
Qiqqa http://www.qiqqa.com/Download	FR 0	1	WIN ADR	ND	0	0	ND	82	TR FI EX	1	1	DCL CLA	1	ND	Premium features with paid subscription

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
Quetzal https://www.quetzal-search.info/loginpage	FRT varies	1	NA	Quertle LLC	UC	0	[16]	[100]	TR FI EX	1	1	IR DCL	ND	1	Key concepts available with paid subscription; previously Quertle
r Simple Learner with Iterative Pruning to Produce Error Reduction (SLIPPER)	ND ND	ND	ND	ND	ND	0	[94]	[101]	ND	ND	ND	ND	ND	ND	Rule learning algorithm
RapidMiner https://rapidminer.com/	CP varies	1	NA	RapidMiner	UC	0	[14]	ND	ND	1	1	ND	ND	ND	Requires installation
Rayyan http://rayyan.qcri.org/	FR 0	1	NA	QCRI	1	0	ND	ND	IE	1	0	ND	ND	1	ND
RCT Tagger http://arrowsmith.psych.uic.edu/cgi-bin/arrowsmith_uic/RCT_Tagger.cgi	FR 0	1	NA	Cohen et al.	0	0	[56]	[102]	FI IE	1	0	DCL	ND	0	ND
Reflect http://reflect.embl.de/	FR 0	1	NA	European Molecular Biology Laboratory	0	0	[16]	[53]	ND	1	0	ND	ND	ND	Gene and protein tagging tool
RobotReviewer http://vortext.systems/robotreviewer	FR 0	1	MULTI	ND	UC	0	[53]	ND	EX	1	1	IEX	ND	ND	Web version for testing; requires installation
RQDA http://rqda.r-forge.r-project.org/	FR 0	0	MULTI	ND	UC	0	ND	[103]	ND	ND	ND	ND	ND	ND	Requires installation; qualitative content analysis

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
Semantic Medline http://skr3.nlm.nih.gov/SemMed/index.html	FR 0	1	MULTI	Fiszman, Marcelo	0	0	ND	[104]	ND	1	0	NLP	1	1	Requires registration; extract semantic predication
SIDER 2 http://sideeffects.embl.de/drugs/4158/	FR 0	1	NA	Michael Kuhn, Structural and Computational Biology Unit, European Molecular Biology Laboratory	ND	0	ND	ND	ND	1	0	IEX	ND	1	ND
SimStat http://provalisresearch.com/products/simstat/	CP varies	0	MULTI	Provalis Research	UC	1	ND	ND	ND	1	2	ND	ND	ND	Requires installation; qualitative and quantitative statistical package
Site Content Analyzer http://www.cleverstat.com/en/sca-website-analysis-software-index.htm	OTH 79	0	WIN	CleverStat	UC	0	[78]	ND	ND	UC	UC	UC	1	ND	Requires installation; website content parser
SLRTOOL http://slrtool.org/v0	FR 0	1	NA	ND	1	0	[105]	ND	ND	0	1	UC	0	1	ND
SPSS http://www-01.ibm.com/software/analytics/spss/	CP varies	0	MULTI	IBM	UC	0	[36]	[106]	ND	1	1	ND	1	ND	Requires installation

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
State of the Art through systematic review (StArt) http://lapes.dc.ufscar.br/tools/start_tool	FR 0	0	MULTI	Laboratory of Research on Software Engineering, Computing Department of the Federal University of São Carlos	UC	0	107 78 105	ND	FI EX	1	2	IEX TCL	1	ND	Requires installation
STRING http://string-db.org/	FR 0	1	NA	Swiss Institute of Bioinformatics	0	0	16	108	ND	ND	ND	ND	1	ND	Protein-protein interactions
Systematic Literature unified Review Program (SLuRp) https://codefeedback.cs.herts.ac.uk/SLuRp/	FR 0	0	MULTI	David Bowes, Science and Technology Research Institute University of Hertfordshire	1	0	78 105	109	FI IE EX	1	2	IEX	ND	ND	ND
Systematic review supported by visual analytics (REVIS) http://ccsl.icmc.usp.br/pt-br/projects/revis	FR 0	0	UC	University of Sao Paulo	UC	0	14	47	ND	1	ND	VIZ	ND	ND	Requires installation
TerMine http://www.nactem.ac.uk/software/termine/	FR 0	1	NA	Sophia Ananiadou, National Centre for Text Mining	UC	1	40	ND	IE	1	1	ATM	1	ND	Web version for demonstration; register for batch service
Text Analysis Markup System (TAMS) http://tamsys.sourceforge.net/	FR 0	0	MAC	Matthew Weinstein	UC	0	ND	110	ND	1	2	TCL	1	ND	Requires installation; qualitative content analysis

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
Text Mining Infrastructure in R (tm in R) http://tm.r-forge.r-project.org/	FR 0	0	MULTI	ND	ND	1	32	ND	ND	1	2	DTM	ND	ND	ND
Textpresso http://www.textpresso.org/about.html	FR 0	ND	ND	ND	ND	0	34	ND	ND	ND	ND	TCA	ND	ND	Biological literature; entity recognition
TinySVM http://chasen.org/~taku/software/TinySVM/	FR 0	0	MULTI	ND	UC	0	70	ND	FI IE EX	1	2	ND	ND	ND	Learning algorithms
Ultimate Research Assistant http://www.ultimate-research-assistant.com/GenerateRes earchReport.asp	ND ND	ND	ND	ND	ND	0	34	54	ND	ND	ND	ND	ND	ND	ND
Unstructured Information Management Application Framework (UIMA) http://uima.apache.org/	FR 0	0	WIN	IBM; currently at Apache Software Foundation	UC	0	54	ND	ND	1	2	DCL NLP	0	ND	Requires validation workstation; out of scope
VOSviewer http://www.vosviewer.com/Home	FR 0	0	MULTI	Centre for Science and Technology Studies, Leiden University	ND	1	111	ND	ND	ND	1	VIZ TCL	ND	ND	ND
Voyant Tools http://voyant-tools.org/	FR 0	1	NA	Stefan Sinclair and Geoffrey Rockwell	0	1	ND	ND	ND	1	ND	VIZ TCL	0	ND	ND

Table E-1. Tools evaluation table (continued)

Name Acronym URL	Availability[a] Cost	Web Based[b]	Platform	Developer	Multiuser[b]	KI[b]	Lit Review [a] Citation(s)	Publication(s)[a]	SR Support[a]	Deploy Status[a,b]	Tech Level[a]	Features[a]	Help[b]	Registration[a,b]	Comments
Weka http://www.cs.waikato.ac.nz/ml/weka/	FR 0	0	MULTI	University of Waikato	ND	0	112	113	ND	ND	ND	ND	ND	ND	Machine learning
Whatizit https://www.ebi.ac.uk/webservices/whatizit/info.jsf	FR 0	1	NA	European Bioinformatics Institute	0	0	16	46	TR FI	1	0	TCL	ND	ND	ND
WordStat http://www.stata.com/news/wordstat-for-stata/	CP varies	0	MULTI	Provalis Research	ND	1	46	ND	ND	ND	ND	ND	ND	ND	ND
xPDF http://www.foolabs.com/xpdf/home.html	FR 0	0	MULTI	Foolabs	0	0	53	114	ND	ND	0	TRV	ND	ND	PDF to text utility; not relevant
XplorMed http://xplormed.ogic.ca/	FR 0	1	NA	Carolina Perez-Iratxeta, Peer Bork, and Miguel A. Andrade	0	0	16	ND	TR FI	1	0	TCL IR	1	ND	ND

Notes: [a] See Table E-2 for explanation of fields and answer codes. [b] "0"= No; "1"=Yes.

ATM = automatic term recognition; CP = commercial product; DCL = document classification; DCT = document clustering; DTM = document term matrix; EX = data extraction or synthesis; FI = searching collecting, retrieving literature or data; FR = free; FRT = free trial; IE = screening or eligibility assessment; IEX = information extraction; IR = information retrieval; MULTI = multiple; NA = not applicable; ND = no data; NLP = national language processing; OTH = other; TCA = text categorization; TCL = text cluster; TR = topic refinement; TSM = text summarization; TRV = text retrieval; UC = unclear; VIZ = visualization

Table E-1. Tools evaluation table (continued)

Note: Fourteen tools were reviewed but not included in the table of characteristics. The tool name, URL, and reason for exclusion:

Aggregator	Not found	Cannot locate; possibly forthcoming; not publicly available
Ali Baba	http://alibaba.informatik.hu-berlin.de/	Retired
ClusterMed	http://compbio.dfci.harvard.edu/compbio/tools/clustermed	Retired; appears that it is no longer supported or functional
ConceptLink	http://project.cis.drexel.edu/conceptlink	Unavailable; cannot access URL
KNALIJ	http://knalij.com	Unavailable; cannot access URL
LibreQDA	Not found	Cannot locate
Literature and Genetic Electronic Resource and Catalogue	http://ligercat.ubio.org/	Unavailable; cannot access URL
MiSearch	http://misearch.ncibi.org/	Not working; site indicates "down for maintenance"
PubFocus	http://www.pubfocus.com/	Not working; site available but not functional
Pubhub	Not found	Cannot locate
Semantic Features In Text	http://sent.dacya.ucm.es./	Unavailable; cannot access URL
Spa	Not found	Cannot locate
SYstematic Review Information Automated Collection	Not found	Cannot locate; possibly not publicly available
Weft QDA	http://www.pressure.to/qda/	Retired; site indicates "(Jun 2014) Since Weft QDA was developed, free (but not open source) versions of some commercial software have emerged (e.g., QDA Miner Lite)."

Table E-2. Evaluation components and choices

Item	Description	Choices
Name	The full name of the tool or resource	Any
Acronym	Acronym or alternate name of the tool or resource	Any
URL	The URL to the web-based tool or resource or to the file download	Any
Availability	Availability, either freely available to all users or proprietary or commercial product	FR, Free| FRT, Free limited time trial| CP, Commercial Product| UC, Unclear| OTH, Other| NA, Not Applicable
Cost	Cost of the tool or resource or subscription to services in U.S. dollars	Any
Web-Based	Is the tool accessible for use via the internet using a URL?	0, No| 1, Yes| UC, Unclear
Platform	Supported operating system for tools or resources that require download or installation	WIN, Windows| MAC, Macintosh |LIN, Linux| MULTI, Multiple| OTH, Other| UC, Unclear| NA, Not Applicable
Developer	Company, institution, or individual credited as the developer of the tool or resource	Any
Multiuser	Does the tool or resource support multiple concurrent users for collaboration on an individual project	0, No| 1, Yes| UC, Unclear
Literature	Has literature been identified that describes the development, validation, or evaluation of the tool or resource?	0, No| 1, Yes| UC, Unclear
Citation(s)	List the publication(s) that describe development, validation, or	Any/ link to PubMed UI

Table E-2. Evaluation components and choices (continued)

Item	Description	Choices
	evaluation of the tool or resource.	
Applicability	Is the tool or resource designed to support, or could it be easily modified to support, the conduct of a systematic reviews?	0, No| 1, Yes| UC, Unclear
Prior Use in a Systematic Review	Has the tool or resource previously been used to support a systematic reviews?	0, No| 1, Yes| UC, Unclear
Systematic Review Support	How the tool has been or could be used to support conduct of the systematic review.	TR, topic refinement, scope, or question development| FI, searching, collecting, retrieving literature or data| IE, screening or eligibility assessment| EX, data extraction or synthesis| OTH, other| UC, unclear
Deploy Status	Is the tool or resource fully functional and available for use (i.e., not in development or pilot status)?	0, No| 1, Yes| UC, Unclear
Tech Level	Degree of technical expertise or support expected for an average user to install, customize and/or run the tool or resource.	0, Low| 1, Medium| 2, High| UC, Unclear
Features	Broad list of functions or features specific to text mining.	ATM, Automatic Term Recognition| DCT, Document Clustering| DCL, Document Classification| DTM, Document Term Matrix| IEX, Information Extraction| IR, Information Retrieval| NLP, Natural Language Processing| TRV, Text Retrieval | TCA, Text Categorization| TCL, Text Cluster| TSM, Text Summarization| VIZ, Visualization| WFA, Word Frequency Analysis
Help	Availability of instructions or other help documentation for use and installation of the tool	Any, link to help documentation or manual
Comments	Comments	Any
Registration	Is registration or an account required?	0, No| 1, Yes| UC, Unclear