Social Science Computer Review OnlineFirst, published on April 20, 2009 as doi:10.1177/0894439309332293

Social Science Computer Review
Volume 000 Number 00
Month 2009 1-15

10.1177/0894439309332293
http://ssc.sagepub.com
hosted at
http://online.sagepub.com

Supporting Systematic Reviews Using Text Mining

Sophia Ananiadou
Brian Rea
University of Manchester, United Kingdom
Naoaki Okazaki
University of Tokyo
Rob Procter
University of Manchester, United Kingdom
James Thomas
University of London

In this article, we describe how we are using text mining solutions to enhance the production of systematic reviews. The aims of this collaborative project are the development of a text mining framework to support systematic reviews and the provision of a service exemplar serving as a test bed for deriving requirements for the development of more generally applicable text mining tools and services.

Keywords: *text mining; terminology; summarisation; systematic reviews; document classification*

Like the natural sciences, the social sciences are facing a "data deluge" (Hey & Trefethen, 2003), which exceeds the capacity of current research methods and tools. One example is the challenge faced in literature surveys (systematic reviewing) by the rapid growth in the research literature. Another is the challenge posed by new sources of data such as the World Wide Web (news and corporate sites, wikis, blogs, etc.), digital communications (e-mail, newsgroups, speech, short message service [SMS]), and transactional records (purchases, etc.), which offer extremely rich resources for research. Equally, the emergence of research, learning, and teaching repositories in recent years containing textual data sources and materials offers the opportunity to analyze across multiple data collections in different locations. To deal with this data deluge, the social sciences are increasingly turning to powerful new technologies such as text mining. In practical terms, this requires the development of a set of interoperable text mining tools and services which can be integrated into different research practices and user communities.

Authors' Note: The ASSERT project is funded by the UK Joint Information Systems Committee (JISC) as part of its e-Infrastructure program. We thank Tingting Mu (NaCTeM) for her research on term clustering and Yutaka Sasaki (NaCTeM) for his work on document classification. Please address correspondence to Sophia Ananiadou, National Centre for Text Mining, M1B, University of Manchester, Manchester M1 7DN, UK; e-mail: sophia.ananiadou@manchester.ac.uk.

The ASSERT Project

In this article, we describe the ASSERT project (http://www.nactem.ac.uk/assert/) and how text mining has been used to enhance the production of systematic reviews. ASSERT is a collaborative project between the UK National Centre for Text Mining (http://www.nactem.ac.uk/), the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre),[1] and the National Centre for e-Social Science (NCeSS).[2] In this article, we discuss some aspects related with the application of text mining techniques for accelerating the process of systematic reviewing for the domain of *rehabilitation of people with mental health problems.*

Project Design and Development Methodology

A range of methods has been devised over the past 20 years to tackle the challenge of identifying user requirements and usability issues of IT systems and securing the effective involvement of users over the life time of a project (Jirotka & Gougen, 1994). Interviews, focus groups, ethnographic studies of work practice, and workshops all have their role to play, and we have been making use of these techniques in an iterative and user-driven process of work place studies, requirements gathering, rapid prototyping, evaluation, and refinement to ensure that user requirements are systematically identified and tracked over the course of the project. The key to our method, however, is to foster collaborative working between text mining tool developers and users, and thereby facilitate the "co-realization" of the system (Hartswood, Procter, Rouncefield, Slack, & Voss, 2008). This approach is critical if we are to understand how to embed text mining services within established routines of research practice and resource use, and how these may evolve as users begin to apply new tools in their work.

An Overview of Systematic Reviewing

Before undertaking any new policy, practice, or research, it is essential to find out what is already known about an issue in a fair and unbiased manner. However, the findings of individual research studies might alone be limited in their applicability and vulnerable to bias. A large number of people and organizations, such as the Cochrane Collaboration (http://www.cochrane.org/) and the EPPI-Centre, have developed methods for locating multiple studies and synthesizing them to inform decision making. The EPPI-Centre has thus developed ways of conducting literature reviews of social research in a systematic way, which provide users with a "shortcut" to relevant evidence.

Currently, systematic reviewing is performed mostly manually, consequently it is time-consuming. Because of the proliferation of textual information, the quantity of potentially relevant literature retrieved in the early stages of a review can become unmanageable—and with the literature expanding by several thousand papers per week, it is difficult even to manage bibliographic information (Hull, Pettifer, & Kell, 2008) let alone automatically extracting information from it.

Early on in a systematic review, reviewers usually undertake searches of electronic databases to retrieve relevant literature. Reviewers have been accustomed to sacrificing specificity in these searches to ensure they have not missed any relevant studies (because of poor indexing or use of general language thesauri which do not cover specific domains), leading to searches that yield large numbers of "hits." They then download the titles and abstracts and screen them manually. This is the most time-consuming part of the process and can involve the manual screening of tens of thousands of titles and abstracts. Complex systematic reviews can take more than a year to complete with up to half of that time being spent searching and screening hits. This is problematic because policy makers and practitioners often need to know the state of research evidence over a much shorter timescale than current methods allow. It can lessen the likelihood that research evidence will be used at all, with consequential dangers for people affected by policies or practices developed in the absence of a firm evidence base (Chalmers, 2003).

Typically, the process of systematic reviewing follows the following stages:

1. *Searching*: Extensive searches are carried out to locate as much relevant research as possible according to a query. These searches include electronic databases, scanning references lists, and searching for published literature.
2. *Screening*: Narrows the scope of search by reducing the collection to only the relevant documents to a specific review. The aim is to highlight key evidence and results that may impact on the policy.
3. *Mapping*: The EPPI-Centre has pioneered the use of "maps" of research as a method both to understand research activity in a given area, and as a way of engaging stakeholders and identifying priorities for the focus of the review.
4. *Synthesizing*: Correlates evidence from a plethora of resources and summarizes the results.

Applying Text Mining to Systematic Reviewing

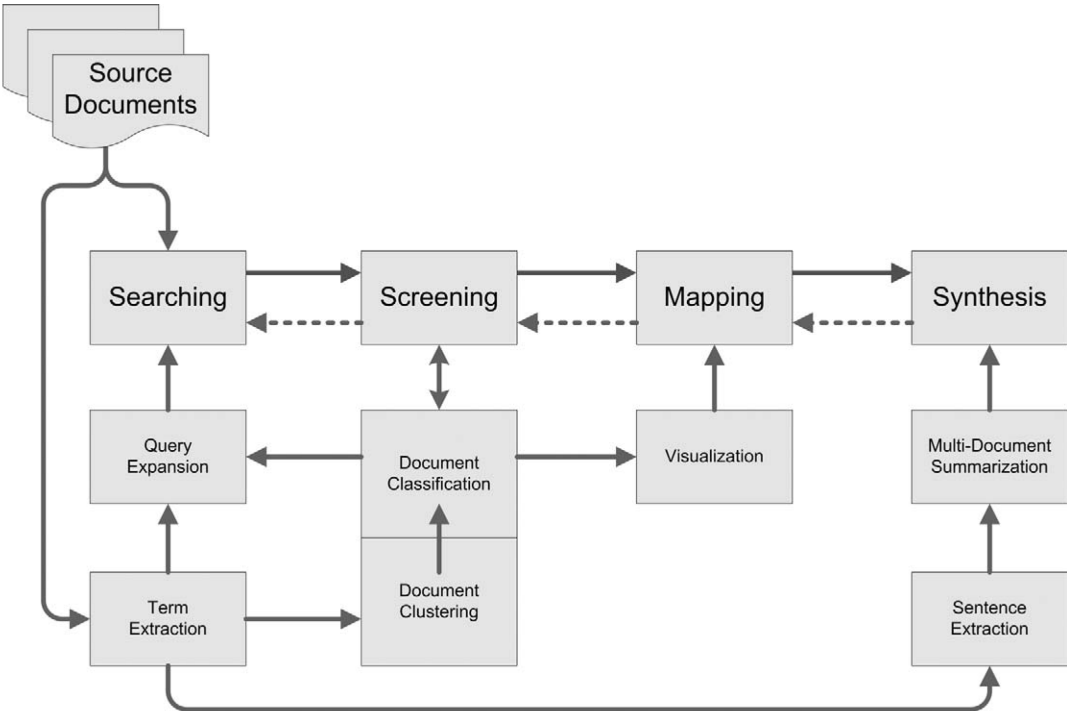
Informed by the study of systematic reviewing practices and requirements gathering, text mining techniques are being used to support these stages as follows (see Figure 1).

Improving the Search Strategy

Currently, searching is performed manually. Reviewers search bibliographic databases based on a defined search strategy, that is, an exhaustive list of keywords that is manually constructed by reviewers. The list details the issues important to the search. Reviewers use sets of *inclusion* and *exclusion criteria* to determine whether any given article is relevant to the review. To facilitate the search strategy, we have used the following text mining technologies.

Term extraction. Term extraction improves the search strategy by creating additional metadata that can improve its accuracy by automatically identifying key phrases and concepts, or technical terms, within the documents. Technical terms characterize the content of a document set. In our project, we extract the most significant terms in a collection of documents by using the TerMine service (`http://www.nactem.ac.uk/software/`

Figure 1
Text Mining Techniques and Workflow for Supporting Systematic Reviewing

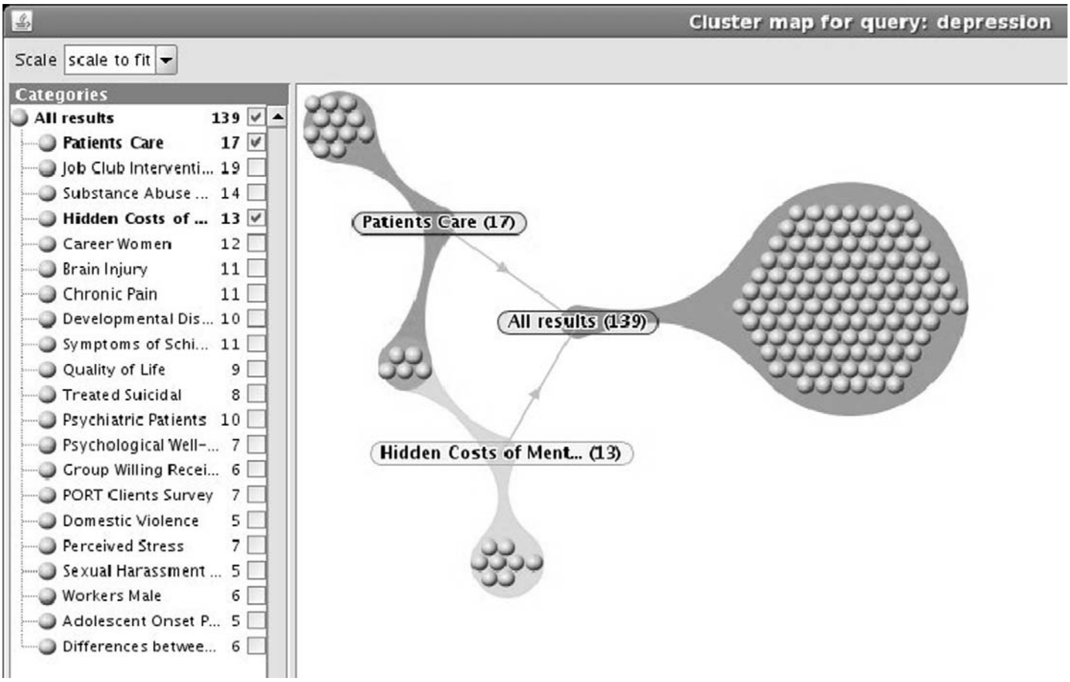


termine/). TerMine automatically extracts and ranks technical terms based on a hybrid term extraction technique *C* value (Frantzi, Ananiadou, & Mima, 2000). *C* value is a statistical measure used to evaluate how important a term is to a document or to a collection of documents. In this project, they are combined with the indexing capabilities of Lucene 2.2 (http://lucene.apache.org/index.html) for indexing and searching.

Document clustering. An important stage in systematic reviewing is evaluating the set of documents and their content against the original goal and scope of the review. This is often difficult to track, as a single document can often contain information useful in a number of areas, some of which may not become apparent until midway through the investigation. To assist in this process, we have used document clustering tools (Osiński & Weiss, 2005), which automatically assign documents to groups based on conceptual similarities. The groups or clusters are generated based on the content and underlying themes and differences.

Document similarities are calculated based on concept document vectors. Words and terms[3] from documents are grouped based on tf-idf[4] using matrix factorization. The most frequent terms are then used as concepts to represent documents. Clustering works better with larger document collections as these can reduce a lot of the noise and allow for a more complete view of the domain. Clustering generates human-readable descriptive labels,

Figure 2
Visualization of Clustered Documents for Mental Health Systematic Reviews



allowing the reviewer to gain a quick overview of the collection based on the variation of the labels.

The current version of ASSERT uses the open source document clustering engine Carrot2 (`http://project.carrot2.org/demos.html`). Figure 2 shows an example of visualization of document clustering[5] using Carrot2 for the domain of mental health. The clusters are "Patients Care" and "Hidden Costs of Mental Health," and the overlap between the sets is shown as a merged bubble. In this example, the documents in the overlap are: *recovery from depression, work productivity, and health care costs among primary care patients; cost effectiveness of practice-initiated quality improvement for depression; cost of treatment failure for major depression: direct costs of continued treatment; "Hidden" spending on community services* and *gender patterns in cost effectiveness of quality improvement for depression*. By adding more topics, we gain a better overview of the documents in the collection. Moreover, topic addition offers the user a quick method of selecting only the documents that are of interest and can be used multiple times, by passing a cluster back into the system to see how that is re-categorized.

Document clustering, while effective, is computationally intensive as each resource needs to be compared against every other resource. For this reason, clustering on larger document collections can take a considerable amount of time. For systematic reviews, this is generally not an issue, as the process would remain faster than the manual operations required. Overall, outcomes of a systematic review can be improved by running through multiple iterations of the searching and screening phases, with gradual improvement based on the partial results. With this in mind, it is nevertheless useful to look at ways in which this processing can be further reduced while preserving accuracy and balanced coverage.

Document classification. Although document clustering discovers groupings of documents based on the content, document classification identifies the underlying patterns and distinguishing features within documents that make them part of a defined grouping or class and uses this information to assign each new document to known classes (Joachims, 1998; Sebastiani, 2002).

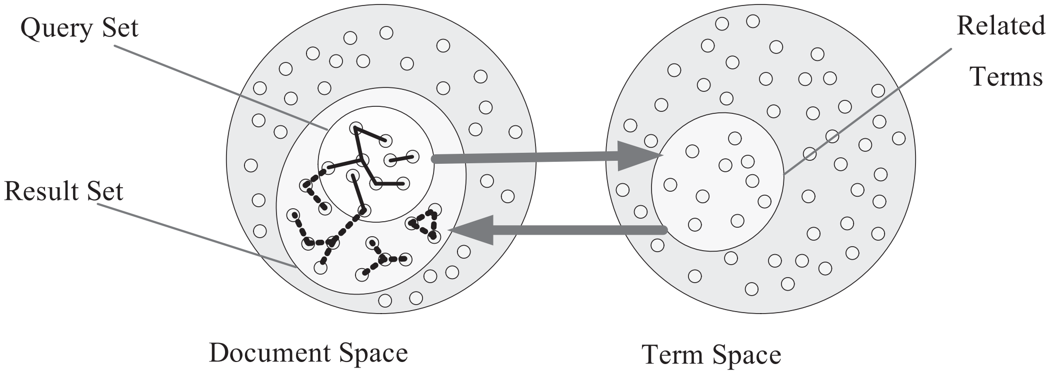
For the purposes of systematic reviews, this allows our system to generate clusters for a significant proportion of the best ranked documents and use these clusters as fixed classes for the purposes of training a document classifier. We then pass each of the remaining documents through the classifier to assign them to the original clusters. This approach is based on the assumption that the setoff higher ranked documents represent the major themes within the rest of the collection. From practical evaluation results, this appears to be the case during the initial iterations during which the document collections are the largest. Following several rounds of screening, the proportion of documents passing through the clustering process increases, providing richer results and thus leading to less reliance on the classification tools.

Document classification has been investigated by many researchers. In the late 1990s, machine learning techniques were successfully applied to topic classification (Dumais, Platt, Heckerman, & Sahami, 1998). After trials with various machine learning algorithms (Sasaki, Rea, & Ananiadou, 2007), we settled on using support vector machines (SVMs), as this approach yielded the best overall accuracy on the test domains. Within the current implementation, we have used tinySVM (`http://chasen.org/~taku/software/TinySVM/`) library for classification. Feature sets were chosen to match the other methods in use and included single words filtered by tf-idf and terms extracted by TerMine. Alternatively, word/term (feature) clustering can also be used for feature selection to reduce the dimensionality of text data. In future, distributional clustering will be applied for text categorization based on the distribution of class labels associated with each word/term.

Query expansion. To enhance further the search strategy, we have used query expansion. One of the major criticisms with current search engines is that queries are effective only when well crafted. A desirable feature is automatic query expansion according to the users' interests, but most search engines do not support this, beyond mapping selective query terms to ontology or thesaural headings (e.g., PubMed). Therefore, there are inevitable limitations of coverage in the typical case. To address this, we have used associative searching techniques to automatically identify important concepts and related documents. These concepts are then added to the original query, expanding the scope to include related keywords.

This process relies on the similarity between documents in the term space to discover sets of texts with similar content to a number of documents provided by the user. This is relatively straightforward for a single document, as we already have the document similarities calculated across the whole collection, but when multiple documents are used as a "query" this becomes more complicated.

Figure 3
Document-Term Mappings for Query Expansion



For this to work, we must examine the area of term space occupied by the query collection and use this to map back to the document space as shown in Figure 3. Document similarity can be used for this, but to find the best matching set of documents, it is essential that this be calculated differently—by comparing the combined term vector of the query collection to the rest of the documents.

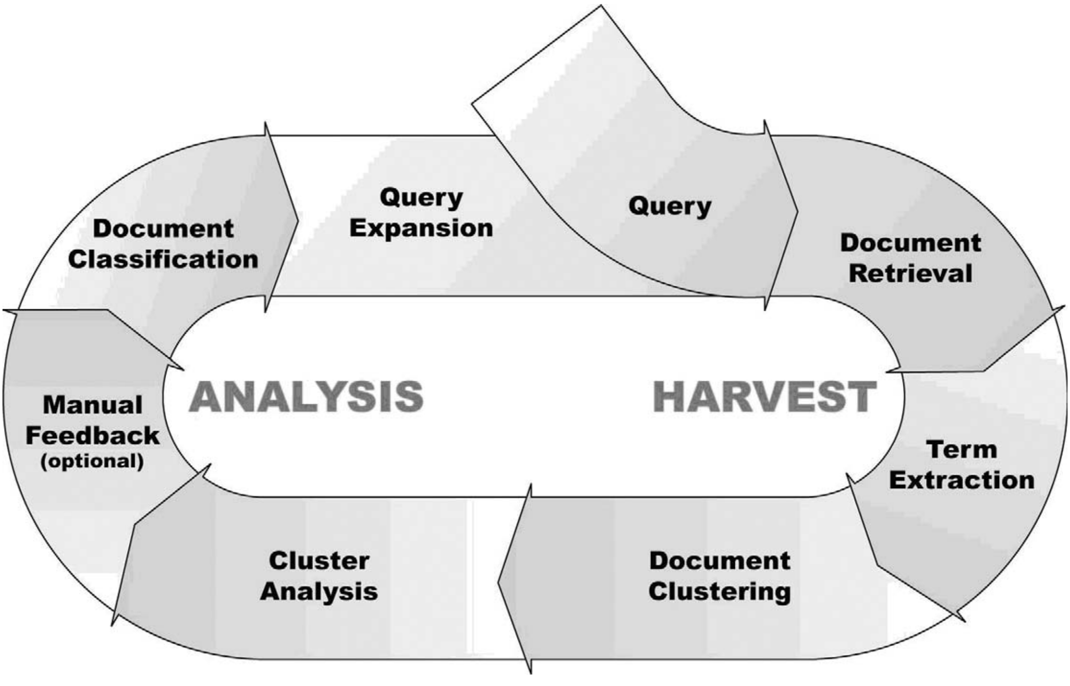
Each dot in Figure 3 represents a document in document space, or a term in term space. The lines connecting the dots represent a high level of similarity between the documents. As can be seen, this method not only returns directly similar documents but may also identify new areas that could be of interest, which would have otherwise gone undiscovered.

For the purposes of searching, this offers systematic reviewers an efficient automated method for identifying conceptually related documents. Within the ASSERT project, we use this for two purposes. First, we can present users with a ranked list of documents that are similar to the one currently being viewed: this has proved useful for finding specific information in a narrowly defined area. Second, we can extend the notion to take into account *terms in multiple documents*. The associative search routines can then be considered query expansion tools, finding documents that contain conceptually similar content, related across document sets. In this case, we set the defined boundaries of input to be the generated clusters of documents discussed above—or the set of all relevant documents.

By multiple searches and through directed expansions using these three sizes of input data, the user can zoom in or out of the document collection at various levels of detail. The set of relevant documents gives the most expansive option, but this can bring in additional noisy terms. Expansion at the cluster level allows focused searching across particular topics or strands of research. The single document expansion generated the most detailed and focused expansion for identifying sources containing specific information or themes.

Additional strength is given to the system through the use of the terms extracted by TerMine. Instead of focusing on individual words, we can bias the expansion toward specific areas. The weight of this bias is generated by the *C* value score, meaning that terms more important to the document or collection carry more sway in the expansion process. This reduces the noise generated by loosely associated terms and strengthens the conceptual

Figure 4
Iterations of Searching and Screening Within ASSERT



similarity across the expansion. The multi-objective analysis of these weights allows for a flexible granularity of expansion of the significant terms by topic or by similarity to the individual texts while maintaining the overall scope of the query. The critical stage of this process involves a combination of term variation techniques (Tsuruoka, McNaught, & Ananiadou, 2008) with term clustering technology to ensure a more complete and thorough expansion at a semantic level. Expansion can happen across multiple iterations to gradually improve the results, with each stage having an optional interaction phase to examine the results and/or manually tweak the expansion process (see Figure 4).

Improving the Screening Strategy

One of the aims of screening is to narrow the scope of search, thus reducing the collection to only the relevant documents of a specific review. A large proportion of this screening will have already taken place in the components discussed in the searching phase. The interactive approach to searching espoused here, along with the optional focused expansions, means that a great deal of the irrelevant documents otherwise generated in the equivalent manual procedures will not be included. Additional filters are further put in place, derived from any stored metadata, which allow documents to be removed based on information such as publication date or location. Finally, text mining can provide further support for screening through alternative use of the classification software.

During the interactive searching phase, the user can remove documents from the results by attaching details of the exclusion criteria. Removed documents continue to be tracked by the system, and classifiers are trained on each collection using the same approach carried out with the document clusters. As new documents are added through the expansion process or an additional search, they are passed through the classifier and presented to the user with quantified predictions for the categories. This then acts as a recommender for reviewers, by highlighting the documents most appropriate for inclusion or exclusion from the review. Over a number of iterations, as more documents are added to the training sets, the results show greater accuracy and, as trust in the results is built up, reviewers can spend more time in analyzing the more ambiguous documents. We thus achieve a useful balance between man and machine, relieve the human of tedious and error-prone manual processing, and provide greater opportunity for the human to attack the more interesting aspects of the task.

An additional benefit of the SVM component is the possibility to investigate the underlying features that define the classes. The features in these experiments were the set of TerMine terms and single words used as input to the classifier. In the system described above, these features are presented in a ranked list according to how they contribute to the overall classification result. By examining these data closely, it is possible to gain an insight into how the terms and topics are related suggesting to the reviewer areas that may be appropriate for further investigation and also potentially identifying areas where further exploration may not be as fruitful.

The combination of these results can be exploited to provide an audit trail for the reviewer to appraise quantified evidence of the classification and the original document source, should a result ever be questioned. This is a key feature in technology-assisted systematic reviews, as trust in the results is vital for an effective and accurate review. Moreover, the direct linking between evidence and source can speed up any synthesis.

Improving the Mapping and Synthesis Strategy

The final stage of systematic reviews combines mapping and synthesis. We improve on manual methods here by using an adaptable multi-document summarization component, driven by user-defined *viewpoints* (Bollegala, Okazaki, & Ishizuka, 2006; Lin & Hovy, 2002; Okazaki, Matsuo, & Ishizuka, 2004). The literature is first broken down into relevant topics and areas of research during the mapping process, and then summary reports are written to inform policy and practice. The summaries of research that are produced in this *systematic* way are then used to help users of research make evidence-informed decisions. Synthesis is a complex process that involves both the description of the research identified, an assessment of its reliability, and the combination of its findings. We see summarization as facilitating the description of research activity and the identification of relevant information for reviewers to assess the quality of that information efficiently.

In more detail, first, the scope and range of relevant research is described (mapping). This is particularly important when reviewing social science literature, because, for example, interventions, populations, and outcomes may differ between documents in the same review, whereas in more clinical areas, the research included is usually more homogenous. Standard frameworks for describing research in key areas (i.e., population characteristics, etc.) are usually used. These are similar to the *inclusion* and *exclusion* criteria detailed earlier, in that they are categorical methods of classifying research. Thus, similar text mining tools can be used to assist in the process—that is, the classifiers described above, although we do not foresee the use of text mining to completely replace the manual work in this stage, as current technology is no match for human insight.

During the screening phase, the clustering technology was used to help maintain the scope of the review and filter down the resources to contain only the most relevant evidence. For the mapping phase, we can use the same tool, combined with the visualization interface to assist in sectioning the evidence into important strands, representing the core interests featured in the inclusion criteria. These strands are originally suggested by the ASSERT tools, based on the terminological profiles. Through an interactive process, they can then be adapted or enhanced by the user, saving both time and effort in the mapping stage and potentially suggesting alternative views of the evidence that the user may not have first considered.

The summarization component used in the ASSERT project (Okazaki, Matsuo, & Ishizuka, 2004) takes as input the documents representing identified strands of evidence during the mapping stage. The overall strategy for summarization is as follows:

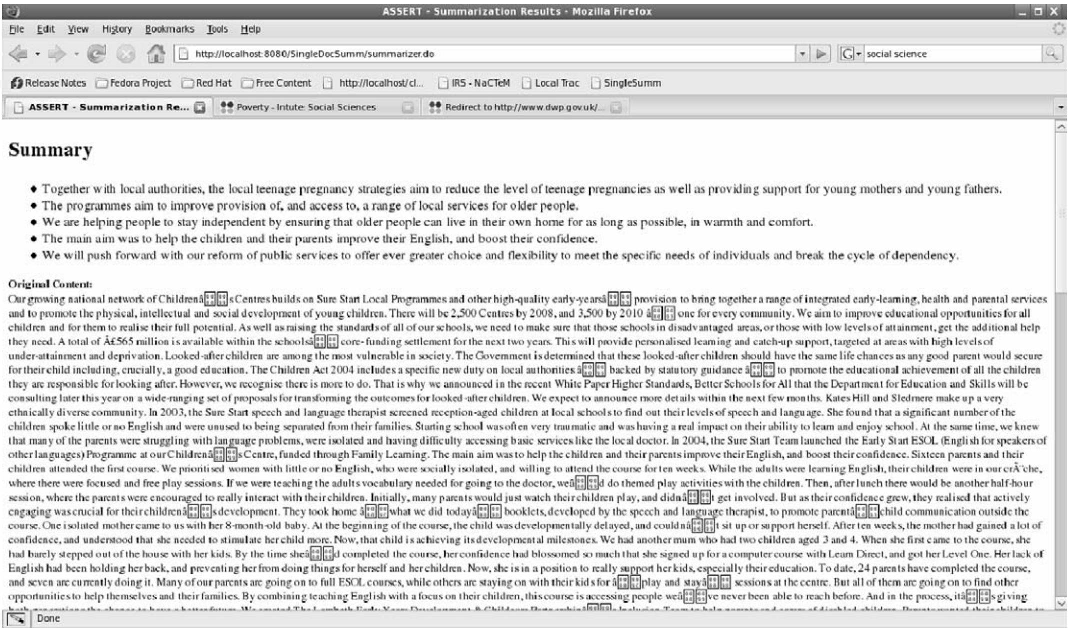
- Score significant terms in retrieved documents based on statistics generated within the subset of documents, that is, significance within this strand.
- Choose a set of sentences (extracts) that contain significant information, excluding potentially redundant information scattered across the input documents.

We adopt a practical solution to summarization as it is still very difficult to generate comprehensible summaries from an internal linguistic representation. In addition, domain-specific documents use a number of technical terms (and variants) for describing the same concept. Hence, it is crucial to perform carefully the statistical analysis to improve the quality of a summary, incorporating terminological variations such as synonyms, acronyms, and so on (Okazaki & Ananiadou, 2006). Our summarization system is based on a systematic terminological analysis, which is important for domain-specific areas, whose key characteristic is that they are rich in terminology.

The synthesis component extracts salient sentences based on the assumptions that: a human reader breaks each sentence into a set of *information fragments* to which the sentence is referring; information fragments are mutually independent; and an information fragment has an importance score (see Figure 5). Among various sentence representations (e.g., bag-of-words and n-grams), the system uses terms extracted via the *C* value method, which identify important technical terms for specific domains. To emphasize terms appearing specifically in a cluster obtained from the previous stage, each information fragment (term) has a weight value computed by the frequency of occurrence of the term, inverse document frequency, and inverse cluster frequency. The synthesis component solves a combinatorial optimization problem that determines a set of sentences containing as many important information fragments and as few redundant information fragments as possible, under the constraint of the summary length specified by the user.

The redundancy reduction technique works particularly well for single documents. However, because of the nature of the approach, it can produce suboptimal results when used for summarizing multiple documents at once. Care must therefore be taken to ensure that there

Figure 5
Results of the Summarization Engine of ASSERT



is a primary focus within the collection of documents or the number of nonoverlapping concepts can lead to an extractive summary that while technically accurate becomes difficult to read. It must therefore be appreciated that, as a stand-alone tool, summarization may yield limited results. However, through integration into the mapping stage of a systematic reviewing process, the above-mentioned primary focus can be assured, leading to useful results and representative summaries of the core strands of evidence. The overall benefit of this process then is to provide evidence-rich representations of the chosen strands, which can be used as a starting point for construction of the review report, rather than automatic generation of the report itself.

Conclusions

Through use of semiautomated techniques to perform some of the more time-consuming tasks of systematic reviewing, reviews can be completed more quickly, and importantly, more systematically, than heretofore, as more evidence can be harvested, filtered, and summarized. Such gains have already been achieved in practice. In addition, searching, screening, and synthesizing can become more customized, focusing on pertinent terms, retrieving relevant documents, and synthesizing salient information fragments. Critical aspects for the uptake of text mining technologies and tools in systematic reviewing are the existence of robust, scalable, efficient, and rapidly responsive services for very large collections and the need to consult large-scale resources to support effective processing (corpora, thesauri). Equally important is the question of what is the right balance between automation of the process and user intervention and control. Pursuing a "co-realization" approach with EPPI-Centre ensures that these issues have been thoroughly investigated, though we accept that this may differ slightly between other communities of practice.

In recent years, developments in e-Infrastructure have opened up new opportunities for the application of text mining applications and services (Carroll, Evans, & Klein, 2005). Computationally intensive tools have previously only been usable on small-scale tasks but are now being developed for much larger scale tasks, thanks to alternative models of processing including Grid computing along with improved storage and data distribution models. This allows us to expand on current tools to take into account the additional information available in full text documents and not just relatively small abstracts. With recent research showing that abstracts alone contain less than half of the overall information content of a article, this is a significant boost for the analysis of documents. Combining large scale document repositories with Web crawling technologies to provide access to the increasing amount of grey literature can offer vital insights into current research, potentially months before publication through traditional routes.

In all, this provides growing opportunities for the application of text mining in systematic reviewing and in the social sciences in general. Text mining techniques have the potential to revolutionize the way we approach research synthesis, but our longer term interest is to understand how we can apply these techniques more widely in the social sciences. To achieve this, we are using systematic reviewing to demonstrate the potential of text mining for the social science research community and building established requirements for a generic toolkit of text mining services, which can be integrated into different research practices.

This work provides its own set of issues for development in terms of interoperability, not only with techniques or software currently used in the systematic review activity but also with other text mining tools and services used by the social science community. For example, a researcher investigating the role of new media in politics could be interested in combining the toolset with Internet news feed or blog readers, their own evidence tracking systems, or even other tools for carrying out opinion analysis. We need to ensure that our tools are therefore flexible and robust enough to allow for this while providing sufficient functionality to ensure interoperability between the many formats and standards that this would entail.

Future Work

In addition to its aims of delivering more powerful tools for systematic reviewing, we are using the ASSERT project as a test bed for exploring the requirements for the application of text mining tools in support of a wide range of social science needs. To this end, we have already had discussions with UK research service providers such as the Economic and Social Data Service (ESDS, `http://www.esds.ac.uk/`), which provides access and support for a range of social science quantitative and qualitative datasets, Intute (`http://www.intute.ac.uk/irs/`), a nationally funded service made up of a consortium of seven universities providing access to resources for education and research, and the International Bibliography of the Social Sciences (IBSS, `http://www.lse.ac.uk/collections/IBSS/`), an online bibliographic database whose mission is to provide easy access to international and interdisciplinary social science research literature. Possible areas identified for further investigation include the use of text mining as an aid to the retrieval and indexing of online documents and alternative ways of clustering and presenting collections of online materials, and the uses of text mining to facilitate and enhance the work of subject specialists.

Our wider goal is to investigate and develop applications of text mining for qualitative social science research. There already exists, of course, a range of computer-assisted qualitative data analysis (CAQDAS)[6] packages to assist social scientists with manual document coding and analysis. These tools, however, are not sufficiently sophisticated to cope with the challenges of the social science data deluge. The manual processes involved simply do not scale when presented with the larger collections now becoming available through large scale initiatives.

To accomplish this wider goal, we have been hosting a series of user workshops for the wider social science research community, using demonstrations of the ASSERT prototype, brainstorming, and focus groups to explore the requirements of different research communities. Based on our findings to date, we recognize the highly diverse nature of qualitative social science research methods. It is clear from this that a strategy that relies on attempting to create the "killer app" for qualitative social science research is bound to fail because it will not be sufficiently well matched to any one community's requirements. Similarly, a strategy based on creating bespoke solutions for each community will be unsustainable because it does not scale.

The way forward, we would argue, is to encourage the development of a text mining toolkit, based on open standards and interoperable services, which will enable researchers to select from different components, adapt them to match their specific needs, and compose them into sharable "workflows" that are appropriate for any given research method. To this end, we have recently begun work to investigate text mining toolkit requirements, using analysis of news media as our initial social science research exemplar.[7]

Notes

1. Institute of Education, University of London. See `http://ioewebserver.ioe.ac.uk/ioe/`
2. University of Manchester. See `http://www.ncess.ac.uk/`
3. Extracted by TerMine.
4. Term frequency–inverse document frequency.
5. See `http://www.aduna-software.com/technologies/clustermap/overview.view`
6. See `http://caqdas.soc.surrey.ac.uk/`
7. See `http://www.nactem.ac.uk/assist/`

References

Bollegala, D., Okazaki, N., & Ishizuka, M. (2006). A bottom-up approach to sentence ordering for multi-document summarization. *Proceedings from the 21st International Conference on Computational Linguistics*, 385-392.

Carroll, J., Evans, R., & Klein, E. (2005). Supporting text mining for e-science: The challenges for grid-enabled natural language processing. *Proceedings from the UK e-Science All Hands Meeting*, Nottingham, UK.
Chalmers, I. (2003). Trying to do more good than harm in policy and practice: The role of rigorous, transparent, up-to-date evaluations. *Annals of the American Academy of Political and Social Science, 589,* 22-40.
Dumais, S., Platt, J., Heckerman, D., & Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. In G. Gardarin (Ed.), *Information and knowledge management* (pp. 148-155). Bethesda, MD: ACM Press.
Frantzi, K., Ananiadou, S., & Mima, H. (2000). Automatic recognition of multi-word terms: The C-value/NC-value method. *International Journal on Digital Libraries, 3,* 115-130.
Hartswood, M., Procter, R., Rouncefield, M, Slack, R., & Voss, A. (2008). Co-realisation: Evolving IT artefacts by design. In M. Ackerman, T. Erickson, C. Halverson, and W. Kellog (Eds.), *Resources, co-evolution and artifacts.* Springer.
Hey, T., & Trefethen, A. (2003). e-Science and its implications. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences, 361,* 1809-1825.
Hull, D., Pettifer, S. R., & Kell, D. B. (2008). Defrosting the digital library: Bibliographic tools for the next generation web. *PLoS Computational Biology, 4(10),* e1000204.
Jirotka, M., & Goguen, J. (Eds.). (1994). *Requirements engineering: Social and technical issues*. London: Academic Press.
Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Lecture Notes in Computer Science, 1398,* 137-142.
Lin, C. Y., & Hovy, E. (2002). From single to multi-document summarization: A prototype system and its evaluation. *Proceedings from the 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, PA, 457-464.
Okazaki, N., & Ananiadou, S. (2006). A term recognition approach to acronym recognition. *Proceedings from the 21st International Conference on Computational Linguistics*, Sydney, Australia, 643-650.
Okazaki, N., Matsuo, Y., & Ishizuka, M. (2004). TISS: An integrated summarization system for TSC-3. *Proceedings from NTCIR-4*, Tokyo.
Osiński, S., & Weiss, D. (Eds.). (2005). Carrot2: Design of a flexible and efficient web information retrieval framework. *Lecture notes in computer science*, 3528, 439-444.
Sebastiani, F. (2002). Machine learning in automated text categorization. *Computing Surveys, 34,* 1-47.
Sasaki, Y., Rea, B., & Ananiadou, S. (2007). Multi-topic aspects in clinical text classification. *Proceedings from the IEEE International Conference on BioInformation and BioMedicine*.
Tsuruoka, Y., McNaught, J., & Ananiadou, S. (2008). Normalizing biomedical terms by minimizing ambiguity and variability. *BMC Bioinformatics, 9*(Suppl. 3), S2.

Sophia Ananiadou is reader in Text Mining in Computer Science, University of Manchester, and Director of the National Centre for Text Mining (NaCTeM). Her current research includes building terminologies for the life sciences, integrating text mining to biochemical networks, and text mining for systematic reviews and media analysis. She is recipient of the 2004 Daiwa Adrian prize for her research in Knowledge Mining for Biology, and for three consecutive years of the IBM UIMA innovation award for leading work on the interoperability of text mining (TM) tools. E-mail: sophia.ananiadou@manchester.ac.uk.

Brian Rea is the project manager at NaCTeM and has carried out the technical development and integration for the ASSERT project. His previous experience in the digital libraries community has recently been focused toward text mining and collection level analysis for the social sciences. E-mail: brian.rea@manchester.ac.uk.

Naoaki Okazaki was a research fellow of National Centre for Text Mining (NaCTeM) from 2005 and became a researcher in the University of Tokyo in 2007. His research interest includes automatic text summarization, terminology extraction, and machine translation. He was a developer of TerMine and the summarization service used in the ASSERT project. E-mail: okazaki@is.s.u-tokyo.ac.jp.

Rob Procter is professor and research director at the UK National Centre for e-Social Science at the University of Manchester. His research interests include socio-technical issues in the design, implementation, evaluation, adoption, and use of interactive computer systems, with emphasis on ethnographic studies of work practices, computer-supported cooperative work, and participatory design. E-mail: rob.procter@manchester.ac.uk.

James Thomas is an assistant director of the Social Science Research Unit, Institute of Education, London, and an associate director of the EPPI-Centre. He has experience of a wide range of research projects within education and health, systematic reviewing, user involvement, historical research, and information technology. He has written extensively on research synthesis, including meta-analysis and methods for combining qualitative and quantitative research in "mixed method" reviews. E-mail: j.thomas@ioe.ac.uk.