

Introduction

The Automatic Summarisation for Systematic Reviews using Text Mining (ASSERT) project centres around providing for broader institutional involvement in text mining and related activities. As part of this we are developing an exemplar service for the social sciences domain, looking at how text mining techniques can support the systematic reviewing process. For this we are working closely with the EPPI-Centre, an organisation concerned with providing unbiased reviews of literature to support policy making. The background work developing this exemplar service will provide a foundation for the work relating to the two new projects funded by a community call organised jointly with JISC in April 2007. We look forward to working with the successful projects in the near future to provide further examples and case studies of how text mining can be beneficial to current and future projects.



Last Published: Wed Jul 25 10:55:00 BST 2007

Israel hears Arab peace proposal

Ministers from Egypt and Jordan are in Israel to present the Arab League's peace plan for the Middle East.

The foreign ministers of Egypt and Jordan are in Israel to present the Arab League's Middle East peace plan.

The two ministers are meeting Prime Minister Ehud Olmert and other Israeli officials.

They are trying to push forward an Arab peace plan that Israel has in the past treated coolly but which commentators say it may now be willing to discuss.

New Middle East envoy Tony Blair said after visiting the West Bank he saw a "moment of opportunity" for peace.

Related Stories

- Arabs woo Israel over peace plan
- Envoys Blair heads to Middle East
- Hamas urges peace summit boycott
- Rice urges push on Mid-East peace
- Palestinian PM holds Israel talks

Related Topics LOW MED HIGH

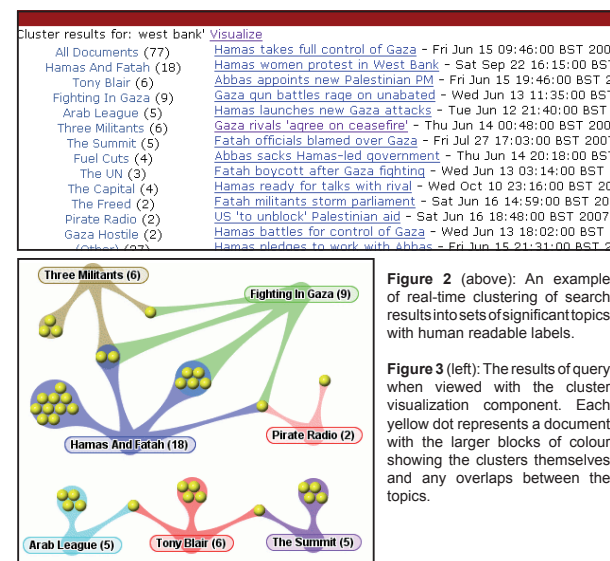
- Arab League
- Middle East peace
- West Bank
- Middle East
- Arab League Secretary General Amr Moussa
- Israeli foreign ministry spokesman Mark Regev
- New Middle East envoy Tony Blair
- nonbound Middle East peace process
- former UK prime minister

Figure 1: Sample document view from the BBC data. Key topics and terms are highlighted in pink. The sections shown on the right hand side present similar documents and a ranked list of topics, both offering links to further explore the document collection.

Within this project we are focussing upon the social sciences communities where there is currently very little use of text mining. Through this we aim to show that there are potentially huge advantages to using text mining methods to:

- save time for researchers
- open up new areas of research
- encourage new ways of doing research

Given the recent success of the early releases of ASSERT on social science documents and BBC news articles (shown in figures 1-3) it is appropriate to examine potential future integration with other NaCTeM tools to leverage their proven functionality in the biomedical arena. As well as increasing awareness of text mining in the social sciences we aim that the tools and services that are developed will eventually be applicable across the whole range of projects currently being funded in the UK academic community particularly in an environment where e-research is becoming increasingly important.



Cluster results for: west bank¹ Visualize

- All Documents (77)
- Hamas And Fatah (18)
- Tony Blair (6)
- Fighting In Gaza (9)
- Arab League (5)
- Three Militants (6)
- The Summit (5)
- Fuel Cuts (4)
- The UN (3)
- The Capital (4)
- The Freed (2)
- Pirate Radio (2)
- Gaza Hostile (2)
- (24) (1) (2)

Figure 2 (above): An example of real-time clustering of search results into sets of significant topics with human readable labels.

Figure 3 (left): The results of query when viewed with the cluster visualization component. Each yellow dot represents a document with the larger blocks of colour showing the clusters themselves and any overlaps between the topics.

Text Mining Services to Support Research Workflows

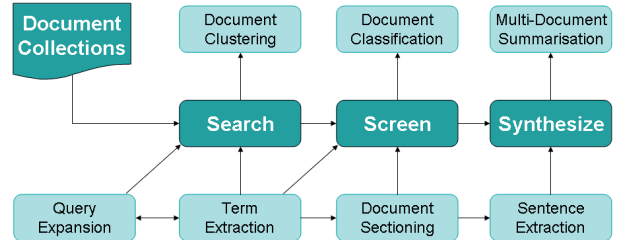


Figure 4: Overview of systematic review pipeline and related text mining technologies used in the ASSERT project.

Searching extensive searches are carried out in order to locate as much relevant research as possible according to a query. These searches include electronic databases, scanning references lists and searching for unpublished literature. This process is improved by using query expansion techniques based on the most important concepts (terms), similarities among terms but also ontologies and thesauri.

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. This process is improved by using document clustering which groups documents into clusters that correspond to a single topic that is shared by all the documents they contain and by no other document in the collection. Visualisation allows the reviewer to see the associations between documents and topics. By selecting topics the user obtains an overview of the documents in the sub-collection and is able to browse visually for alternative categories (see screenshots). Document classification automatically assigns documents into existing categories, generating subsets of documents focused on a specific topic, allowing for more efficient and accurate analysis during subsequent stages of information filtering. Multi-topic classification is useful for systematic reviewing as single documents may be relevant to multiple review topics.

Synthesizing correlates evidence from a plethora of resources and summarises the results. This process is improved by using multi-document summarisation driven by user defined viewpoints. Source documents for a summary may have been written by different authors, and have different text styles, dates, etc therefore arranging the salient sentences in a coherent manner is important. We select sentences from each document based upon the significance of its terms which are combined with classification techniques to discover the most relevant passages within the important sections of a document such as introduction, background, methodology, results, conclusions. This technique provides a more informative overview of the document than a traditional abstract.

References

- Ananiadou, S., Procter, R., Rea, B., Sasaki, Y and Thomas, J. (2007). Supporting Systematic Reviews using Text Mining, in 3rd International Conference on e-Social Science, Ann Arbor.
- Chalmers, I. (2003). Trying to do more Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations. The ANNALS of the American Academy of Political and Social Science, Vol. 589(1), 22-40.
- Frantzi, K., Ananiadou, S. and Mima, H. (2000). Automatic Recognition of Multi-word Terms. International Journal of Digital Libraries 3(2), 117-132.
- Osiński, S. and Weiss, D. (2005a). A Concept-Driven Algorithm for Clustering Search Results. IEEE Intelligent Systems, May/June, 3 (vol. 20), 2005, pp. 48-54.
- Sasaki, Y, Rea, B. and S. Ananiadou, S. (2007). Multi-Topic Aspects in Clinical Text Classification. IEEE International Conference on Bioinformatics and BioMedicine, Silicon Valley, November.