

Evaluating a Text Mining Based Educational Search Portal

Sophia Ananiadou^{1,2}, John McNaught^{1,2}, James Thomas³, Mark Rickinson³, Sandy Oliver³

¹School of Computer Science, University of Manchester, UK

²National Centre for Text Mining, University of Manchester, UK

³Social Science Research Unit, Institute of Education, University of London

E-mail: Sophia.ananiadou@manchester.ac.uk, john.mcnaught@manchester.ac.uk, J.Thomas@ioe.ac.uk, S.Oliver@ioe.ac.uk

Abstract

In this paper, we present the main features of a text mining based search engine for the UK Educational Evidence Portal available at the UK National Centre for Text Mining (NaCTeM), together with a user-centred framework for the evaluation of the search engine. The framework is adapted from an existing proposal by the ISLE (EAGLES) Evaluation Working group. We introduce the metrics employed for the evaluation, and explain how these relate to the text mining based search engine. Following this, we describe how we applied the framework to the evaluation of a number of key text mining features of the search engine, namely the automatic clustering of search results, classification of search results according to a taxonomy, and identification of topics and other documents that are related to a chosen document. Finally, we present the results of the evaluation in terms of the strengths, weaknesses and improvements identified for each of these features.

1. Introduction

In education, as in many other professions, the Internet is becoming an increasingly important tool for providing the evidence required for practice and policy making. One of the problems faced by both professionals and lay people is that research findings can be dispersed among multiple sources, meaning that considerable time often has to be spent locating relevant resources. In response to this, the UK Education Evidence Portal (EEP) has been created, which draws on a wide range of documents from selected sources. By making these collections searchable from a central point of access, the portal aims to revolutionise work practices for the education community. The EEP provides a search engine that enables users to search the contents of the websites of 38 organizations and includes resources such as the Department for Children, Schools and Families, the National Foundation for Education Research, the Institute of Education, London, the General Teaching Council for England, the Higher Education Academy, HM Inspectorate of Education, Ofsted and the Scottish Executive. Whilst the EEP provides a convenient access point to relevant documents, the amount of data contained within their database is still very large. Based on web crawl results of the website content of the 38 organisations, the EEP database currently contains over 500,000 documents and, due to the data deluge problem, this number is likely to continue to grow significantly, with the literature expanding by several thousand papers every week. This means that, with only traditional search engine techniques at their disposal, users of the portal will often have to search through a large number of documents in order to locate information that is relevant to their needs. The aim of the JISC-funded ASSIST project has been to investigate the benefits of text mining in the social science disciplines (Ananiadou, 2009), in which textual information constitutes an important

source of research evidence. As part of the ASSIST project, a prototype web interface¹ has been developed, to demonstrate how the search facilities of the EEP portal could be expanded and enhanced through the inclusion of a number of text mining methods. A screenshot of the interface is shown in Figure 1. The enhanced interface would allow users to perform more focused searches than previously possible, and to allow them to locate relevant information within the retrieved documents in a more timely and efficient manner. Importantly, the project has adopted a user-centered design-build-evaluate approach, ensuring close interaction between developers and users from the earliest stages of the design, through to implementation and evaluation. The evaluation has focussed on examining the suitability of using new technologies integrated into an educational search portal as well as their implementation and acceptability to end users.

2. Text Mining Pipeline

Within the ASSIST project, several text mining technologies have been used to support the tasks expected by users of the EEP. Each technology either provides new features, or enhances the existing components of the education portal. The main new technologies employed are as follows:

- I. **Automatic classification of documents (browsable taxonomy)** – documents are indexed according to a domain-specific taxonomy, allowing users to browse documents related to particular taxonomy terms, as an alternative to the more traditional free text search. The hierarchical structure of the taxonomy allows searches to be more or less specific, according to the requirements of the user.
- II. **Automatic clustering of search results /generation of topics** - rather than relying on standard ranking o

¹ <http://nactem.mc.man.ac.uk/ASSIST-EPPI/>

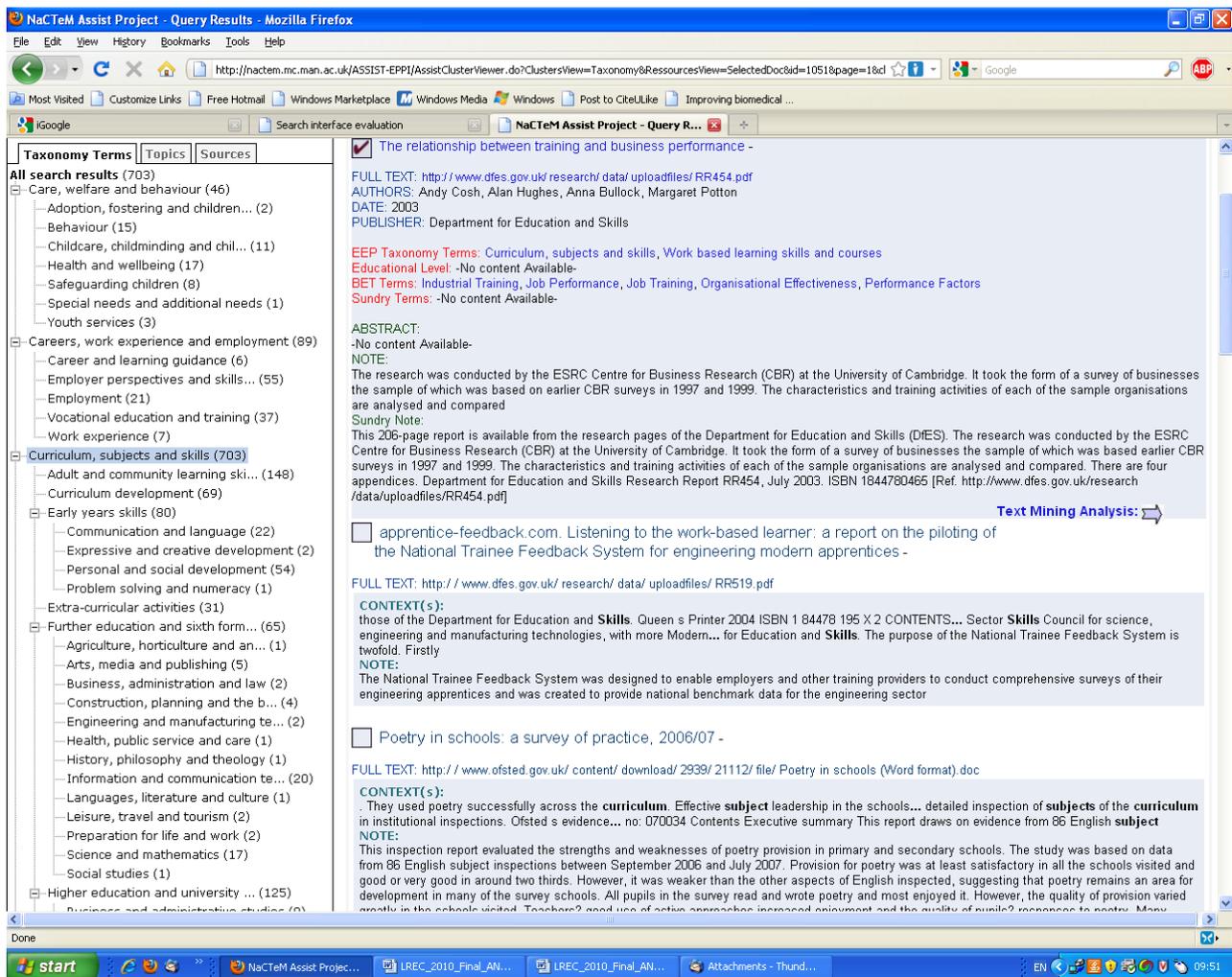


Figure 1: ASSIST Prototype Interface

documents using relative term frequency, results of queries are made more manageable for the user by automatic clustering of related documents and assignment of topic labels to the clusters.

- III. **Identification of key terms within documents and display of related documents/topics** – documents are automatically analysed with NaCTeM’s TerMine tool². This facilitates highlighting of key terms within documents, as well as two types of enhanced search functionality. Firstly, a set of “related documents” is displayed, based on the co-occurrence of terms. Secondly, a hyperlinked list of all terms found in the document is presented (“related topics”). The terms are ranked according to their importance in characterising the document.
- IV. **Advanced search capabilities** – additional operators may be specified within free text searches to allow searching on document metadata, such as titles, author names and keywords. This permits, for example, only those documents with a particular author to be retrieved by a search

3. The Evaluation Framework

There has been a large amount of research on the design and evaluation of user interfaces (e.g., Schneiderman, 2004; Hearst, 2009). Evaluation of search interfaces according to ISO standards (ISO, 1998) typically focuses on effectiveness, efficiency and satisfaction. In this study, we have adapted a user-centred evaluation framework based on the dimensions outlined in King (2007): *reliability, usability, efficiency, maintainability and portability*.

Each dimension includes positive and negative attributes established after user-oriented requirements analysis. As a result of a requirements analysis, we have characterized positive attributes in terms of three levels: expected or standard requirements, desirable requirements, and requirements that would provide the service with a competitive edge. Negative attributes are characterized by three aspects: cosmetic attributes, dysfunctional attributes and attributes that would preclude use, in other words, ‘show stoppers’. The portal aims to provide end users with quick, targeted access to high quality research evidence in education. Meeting the needs of users is critical, both in terms of encouraging

² <http://www.nactem.ac.uk/software/termine/>

people to use this service and also because the management group is concerned with promoting the use of evidence to inform people's educational practice. The search interface evaluation portal is accessible live at <http://www.eep.ac.uk/dnn2/Abouteep/Searchinterfaceevaluation/tabid/151/Default.aspx>

3.1 Evaluation Metrics for a Text Mining Based Search Portal

We have defined the following five dimensions, which were evaluated from the point of view of end users, partners and hosts of the portal.

3.1.1 Functionality

This dimension concerns whether the portal meets user needs (i.e., what the portal does and not how it does it). Examples of functional attributes from the point of view of end users include:

- negative attributes – e.g., site registration, payment required
- positive attributes – e.g., search for complex and new concepts, policies/curricula/terminology that change over time; search for different types of documents (policy, research, resources, news, pictures, blogs, curricula) rather than just content; search for UK/ date in search (not just search output); search for Higher Education, schools, key organisations; search for author, title; filter by concepts; free access to full text documents.

Within this dimension, the attributes of *suitability* (provision of an appropriate set of functions for specified tasks and user objectives), *accuracy* (whether the software conforms to the requirements), *interoperability* (embedding with other systems), *security* (protection of information and data) and *compliance to standards* have also been assessed by end users, partners and host organisation of the EEP education portal.

The text mining capabilities have been evaluated as part of the *accuracy* attribute. This evaluation component refers to the technical accuracy of the search tools and their perceived accuracy to users. The automatic document classification for the browseable taxonomy has been evaluated internally by measuring the accuracy of automated classification according to taxonomy terms, etc. Partners and end users have been asked to rate the following on an ordinal scale:

- a) the relevance of the terms identified for individual documents and related documents;
- b) the accuracy of the document clustering and usefulness of the labels attached;
- c) the accuracy of documents identified through the browsable taxonomy.

This metric has been compared with the same searches run on a version of a portal without text mining functionalities.

Overall, we are assessing how well the interface and

functionality match up with the expectations arising from the initial users' requirements and assessment.

3.1.2 Reliability (maintaining performance)

This dimension has been measured by the host organization (NaCTeM). The portal has been subjected to a series of 'stress tests' to simulate heavy and unpredictable user demand. As well as a simulated high frequency of search requests, the engine is being tested in terms of its capacity to deal with poorly formulated and potentially malicious queries. Taken together, these tests measure:

- a) Maturity (avoid failure as a result of faults in the software)
- b) Fault tolerance (maintain performance in spite of faults)
- c) Recoverability (re-establish performance and recover data in case of failure)

3.1.3 Usability

This dimension assesses how easily users can understand, learn, operate, and control the system.

Metrics for evaluating usability design include determining whether the design incorporates each positive attribute identified by end users and whether it avoids negative attributes.

While functionality determines whether or not the portal contains the technical potential to meet users' needs, this dimension is concerned with whether or not users are able to use those functions to achieve their objectives. Evaluation is carried out through a combination of observation – watching users interacting with the portal – and interview questions. The first three usability issues (*understandability*, *learnability* and *operability*) are addressed by giving users specific tasks to undertake using the portal (e.g., locating information about particular topics), and then asking them about how they approached the task. The other two features are *attractiveness* and compliance with appropriate W3C standards for search engines.

As text mining is a new technology for end users, *understandability* assesses if users understand whether the software is suitable, how it can be used for particular tasks, and what the conditions are for its use. Correct assessment and evaluation will ensure uptake of the new technology by users.

After a period of acclimatisation and active use of the portal, users are asked if they understand the text mining search functions i.e. document clustering, automatic extraction of terms and classification of related documents based on terms. The interviewer reassures users that they are not being tested on their individual knowledge, but on how well the portal helps them to understand its functionality. This evaluation metric provides a score which represents how closely each user describes each feature.

We recognise that there will be considerable variability between evaluation participants in terms of:

- their use of, and exposure to, evidence for

- informing their decisions;
- their experience and knowledge of the EEP portal;
- their general computer literacy;
- in particular, their information retrieval skills.

Thus, *understandability* was evaluated for given user groups, taking into account different levels of experience, knowledge and expectations that shape how ‘understandable’ the portal was to them.

3.1.4 Efficiency

This dimension of the evaluation relates to both the efficiency of the portal, in terms of how quickly users are able to accomplish their objectives, as well as the demands placed by the system on the computational infrastructure of the host organisation. During requirements assessment, end users identified fast searches as being of key importance. Evaluation of both time behaviour (response and processing time) and resource utilization by the host institution are carried out. The latter include physical hardware (what specification of server is required); bandwidth (given that the portal includes a web crawler); maintenance (how much maintenance the system requires; to what extent it can be relied upon to run without intervention; processing power and memory required per user.)

3.1.5 Maintainability

This dimension concerns the ability to correct, improve or adapt the software. The majority of this part of the evaluation will be carried out internally and relates to error analysis, modification of portal functionalities based on integration of new text mining tools, stability of portal etc.

4. EEP Taxonomy

As an alternative to Google-style free text search, it was decided that documents in the EEP database would be classified according to the EEP taxonomy. Users then have the option of browsing through the taxonomy to help locate the documents of most relevance to them. Evaluation of existing educational taxonomies showed that they were too narrow in scope, and thus a new taxonomy was developed by EEP partners and evaluated by the wider education community. The final taxonomy consists of 108 concept categories. The taxonomy is organised into broad topics with a shallow hierarchy, due to the specific purpose for which it has been designed. Rather than being highly detailed and exhaustive, containing thousands of terms (like, for example, the British Education Thesaurus), the intention is that the taxonomy complements the other tools within the EEP. After selecting a term, users are presented with a list of conceptual clusters that are generated on demand by the portal. These clusters can assist the user in identifying which areas within their chosen topic they would like to explore in more depth. It is also possible to conduct free-text searches within a given area of the taxonomy.

The taxonomy has a fairly flat structure with usually only one level beneath the top level terms. At the top level, the taxonomy consists of:

- Curriculum, subjects and skills
- Teaching and learning
- Performance, assessment and quality improvement
- Careers, work experience and employment
- Management, governance and finance
- Teachers and staff
- Families, community and society
- Care welfare and behaviour
- Research methods and use of evidence

5. Results

The findings emerging from the user evaluation can be summarised in terms of what has been learned about the strengths, weaknesses and potential improvements of the three key features of the portal (automatically generated topics, related topics/documents, and browseable taxonomy). The evaluation was conducted according to the criteria defined above, but is summarised here for clarity.

5.1 Automatic Generation of Topics (Document Clustering)

5.1.1 Strengths

When asked about the utility of the automatically generated topics, there was definite endorsement from almost three quarters of the survey respondents (74.2% agreeing or strongly agreeing) (Figure 2). Similar views were expressed by interviewees who also liked being able to refine their search results using the topics function: “I see that ‘topics’ enables you to sub-select something like ‘basic skills’ which would be useful”. It was also clear, however, that the usefulness of the topics tabs depended on: (i) a desire to browse (“I can’t see I’d use the topics much because I tend to come to databases with a specific task, not to browse generally”), and (ii) a recognition of one’s interests in the topics available.

5.1.2 Weaknesses

Users reported difficulties with automatically generated topics in two main areas: accuracy and understandability. With respect to accuracy, evaluation participants reported that 51% of the automatically-generated topics were useful but 12% (standard deviation 10) thought the topics were not relevant.

More of an issue for users, however, was the fact that the ‘topics’ function was not easy to understand. Even after using it, interviewees were not able to explain the exact function and nature of topics. Typical responses were “I’m not sure where they come from”, “are they linked to the taxonomy terms?” and “another way of sorting information”.

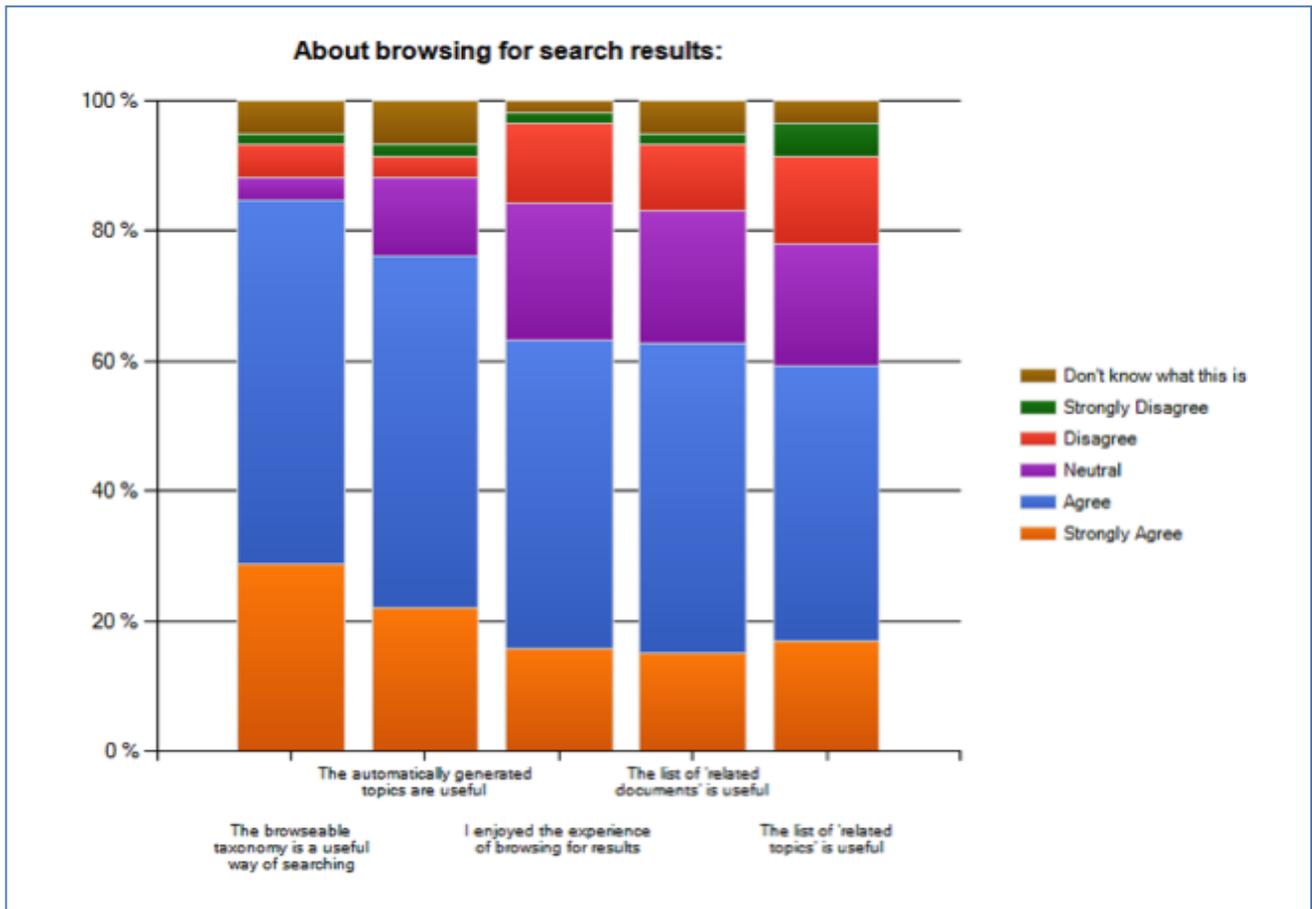


Figure 2: Survey Responses Concerning Browsing for Search Results

5.1.3 Improvements

The following features were identified as desirable :

- i. Clear guidance as to the differences between using 'topics' and 'taxonomy terms' to refine search results
- ii. The ability for users to select several topics simultaneously.

5.2 Related Documents and Topics

5.2.1 Strengths

While users questioned the usefulness and accuracy of 'related documents' and 'related topics' (see below), these functions seemed relatively easy for people to use and understand, at least at a basic level. Having experimented with 'related documents', for example, most interviewees had a sense of what it was providing, if not a clear understanding of how it worked. One interviewee compared this feature to "a report bibliography i.e., giving further/other references".

5.2.2 Weaknesses

As shown in Figure 2, users were slightly less positive about the usefulness of the lists of related documents and related topics, compared with other portal features. This was echoed by the interviewees, several of whom felt

that these may cause "information overload" and "may end up pulling users away from their search". That said, some users made the point that they could be useful "if you don't know what you are looking for" or "want to browse for related peripheral documents".

Other issues were also raised about their accuracy. On average, interviewees felt that 53% (standard deviation 33) of the 'related documents' were appropriate and about half of the 'related topics' (49% SD 24) were useful. Several interviewees were unimpressed that titles from their original search results list were suggested again under 'related documents'. In addition, the 'related topics' sometimes gave the appearance of unreliability e.g. with irrelevant terms and capitalisation confusing the engine.

Some users found it difficult to see the link between a selected document and the related topics and documents (since the user would often have to scroll to see them). As shown in Figure 3, a document is selected for viewing further details, but it is necessary to scroll quite a way before the related topics and documents appear on the left below the taxonomy. During interviews, there were also complaints about the related topics and documents being in "very small text" and their hover help boxes being "confusing" and "irritating".

Furthermore, it was clear that some users did not find it easy to understand the purpose of 'related topics'. Some said they simply "can't understand these" or made fairly

vague references to “pulling out related items”.

5.2.3 Improvements

In the light of the issues raised above, there were calls for:

- i. Clear signposting for both ‘related documents’ and ‘related topics’
- ii. Better explanation about the purpose and basis of these functions.

5.3 Browseable Taxonomy (Document Classification)

5.3.1 Strengths

The browseable taxonomy received the strongest support in terms of being a useful feature of the portal (84.7% agreeing or strongly agreeing, as shown in Figure 1). This was backed up by feedback from interviewees:

“The taxonomy terms are very useful in giving me a way to break down the search results by other areas. I like that as the taxonomy terms are gateways into the information”.

In terms of *understandability*, most interviewees could figure out what the taxonomy terms are, what they do and how they can be used. Descriptions included “nine key terms under which documents will be clustered” and “pre-defined terms that allow things to be put in categories”. However, it should be noted that familiarity with “the sub-divisions of Children’s Services” and “library-type classification systems” seemed to help with understanding the taxonomy.

Accuracy was judged in terms of interviewees, on average, reporting that 59% (SD 25) of the automatically assigned taxonomy terms were correct.

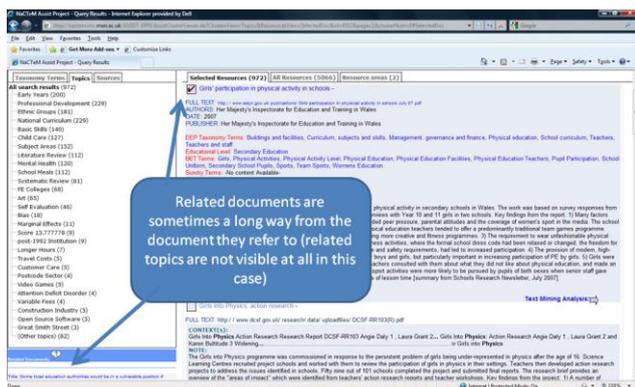


Figure 3 : Problems with viewing ‘related documents and topics’

5.3.2 Weaknesses

One potential problem with the functionality of the browseable taxonomy was identified: selecting a taxonomy term also causes the documents within that branch of the taxonomy to be clustered and assigned topics by the Lingo3G clustering engine. As there are limits to the number of documents that can be processed

by the engine within a reasonable time, a limit of 1000 documents has been set. This limit works admirably for free text searches, where the most relevant material is assumed to be at the ‘top’ of the list. However, the results of browsing the taxonomy tree cannot be ordered in terms of their relevance. There is therefore a danger that the search is truncated arbitrarily and a possibility that results that are relevant to the user are lost. Users also questioned the accessibility of the taxonomy to non-specialist users. There were frequent statements along the lines of: “the phrase ‘taxonomy terms’ may well be confusing for a non-specialist/practitioner user”. Indeed, even from experienced database users it was not uncommon to hear statements such as “The browsable taxonomy is useful but I didn’t understand it at first”. Furthermore, in some cases, an additional issue was the fact that the taxonomy terms on the search results screen took “a long time to load”.

5.3.3 Improvements

Suggested improvements included:

- i. displaying the taxonomy terms in an un-expanded format
- ii. ensuring better explanation of the relationship (if any) between the ‘taxonomy terms’ and the free text search
- iii. providing more guidance on the basis and function of the taxonomy.

6. Conclusion

Text mining services have been used to enhance search and discovery options for the UK Education Evidence Portal. Combinations of metadata enhancement, improved browsing and navigation, alongside alternative views of resources, have all strengthened the overall proposition of the portal. This is building towards a competitive framework for collection management and analysis. Regular and continued user engagement during the lifetime of the project has led to a significant service exemplar of the applications and benefits of text mining within the social sciences. Rigorous quality assurance and comprehensive evaluation strategies are being used to ensure the tools meet the needs of the EEP stakeholders. We anticipate that this will be extended further to support a framework for the wider evaluation of text mining components. As we further expand upon the work of the ASSIST project, opportunities to reflect upon the outputs of EEP and related projects have highlighted several strands of potential development. Other domains could clearly benefit from services similar to those that have been created as extensions to the EEP portal, either involving further subject customisation or through the integration of complementary components. Given the wider themes surrounding repositories and metadata in the community, the role of text mining is destined to become increasingly important.

7. Acknowledgements

The work described has been carried out as part of the ASSIST project, which was funded by JISC. The project and its evaluation was also supported by the Educational Evidence Portal Development Group. We would like to thank Davy Weissenbacher, Paul Thompson, Brian Rea, Yutaka Sasaki, Bill Black (NaCTeM) and Ruth Stewart and Claire Stansfield (EPPI-Centre) for their valuable contributions to the work described in this paper.

8. References

- Ananiadou, S., Rea, B., Okazaki, N., Procter, R. & Thomas, J. (2009) Supporting Systematic Reviews Using Text Mining. *Social Science Computer Review* 27, pp.509--523.
- 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* 589(1), pp. 22--40.
- 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(2), pp. 115--130.
- Hearst, M. A. 2009 *Search User Interfaces*. Cambridge: Cambridge University Press
- ISO 9241-11 I (1998) Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.
- King, M. (2007) General Principles of User-Oriented Evaluation. In Dybkjær, L., Hemsén, H. and Minker, W. (Eds.) *Evaluation of Text and Speech Systems*, pp. 125-161. Springer
- Shneiderman, B. & Plaisant, C. (2004) *Designing the user interface: strategies for effective human-computer interaction*. Boston, MA: Addison Wesley.
- Tsuruoka, Y., Tateishi, Y., Kim, J. D., Ohta, T., McNaught, J., Ananiadou, S. & Tsujii, J. (2005) Developing a Robust Part-of-Speech Tagger for Biomedical Text. *Advances in Informatics*, pp. 382-392.
- White, J. (2000) Toward an Automated, Task-Based MT Evaluation Strategy. In *Proceedings of the Workshop on the Evaluation of Machine Translation, Third International Conference on Language Resources and Evaluation*.