

Data and text mining

Response to comment on 'MeSH-up: effective MeSH text classification for improved document retrieval'Dolf Trieschnigg^{1,2,*}, Piotr Pezik¹, Vivian Lee¹, Franciska de Jong², Wessel Kraaij³ and Dietrich Rebholz-Schuhmann¹¹European Bioinformatics Institute, Hinxton, UK, ²HMI, University of Twente, Enschede and ³TNO ICT, Delft, The Netherlands

Received and revised on August 5, 2009; accepted on August 7, 2009

Advance Access publication August 24, 2009

Associate Editor: Jonathan Wren

Contact: trieschn@ewi.utwente.nl; dolf@trieschnigg.nl

As developers and primary users of MTI and MetaMap, Névél *et al.* made a number of interesting comments on our recent publication in *Bioinformatics*. However, some of the results and conclusions found in the reply seem premature and lack proper clarification.

1 TASK AND EVALUATION

In response to the methodological considerations, we emphasize that in our paper we compare different MeSH classification systems on two tasks: (i) reproducing manual MeSH recommendations (referred to as *indexing* by Névél *et al.*) and (ii) translating a textual query to an additional MeSH representation (referred to as *query expansion*).

We show that the approach we propose works well on both tasks and compare it to methods which serve a similar purpose. *Recommending* indexing terms is clearly different from automatically assigning of terms to citations and these two tasks require different evaluation measures. Therefore, we would like to re-emphasize that our evaluation of MTI measured its performance as an automatic indexing system rather than a recommender tool. Far from being 'abstract performance scores', the reported measures simply indicate the performance of the task at hand.

2 EXPERIMENTAL RESULTS

We were surprised by the experimental results produced by Névél *et al.*. The difference in MetaMap results can be explained, the other results reported in the letter require further investigation. Nevertheless, we believe that the results provided by Névél *et al.* do in fact confirm our original conclusions.

The difference in MetaMap results could be attributed to a different usage of the tool. We used MetaMap Transfer (MMTx) (<http://mmtx.nlm.nih.gov/>), which offers a parameter¹ to output MeSH terms only, and we assumed this to be a valid way to obtain MeSH terms from MetaMap. We suggest the removal of this parameter if it should not be used.

Névél *et al.* also report a difference in the MTI scores. We used the secured web interface of MTI, with default settings to obtain the results. After consulting the NLM on how to use MTI for this evaluation, we submitted title, abstract and PMID for each Medline record to the system.

*To whom correspondence should be addressed.

¹The class MMTxAPI can be created with the parameter -R = MSH

The results obtained with a re-implementation of our KNN method are disappointing. We cannot judge the quality of this implementation, but the results obtained for the PRC algorithm strongly suggest that a mistake has been made. We find it unfortunate that the authors did not contact us to sort out the difference and promptly concluded 'the results were difficult to reproduce'.

The reported results on PRC by Névél *et al.* are interesting. Again, the comment lacks implementation details, but it should be noted that Névél *et al.* refer to PRC as a 'modified KNN' system. We assume that the PRC algorithm is used similarly to our language model IR system to obtain related citations, and that from these related citations MeSH terms are suggested. Therefore, the large difference between the two KNN approaches is surprising. In any case, they only confirm our conclusion that a KNN method 'clearly outperforms the other published approaches'. From this perspective, Névél *et al.*'s comments do not add much to the results reported in our paper. Névél *et al.* mainly manage to demonstrate that another KNN system clearly outperforms MTI for the task at hand, regardless of the MTI's primary application and irrespective of Névél *et al.*'s lack of success in reproducing our KNN system.

In the concluding section, Névél *et al.* briefly state that our results 'did not confirm previous work in the field' and we assume this refers to the usefulness of MeSH for improving IR. Since the advent of biomedical full-text retrieval systems, there has been an ongoing debate whether domain-specific thesauri can be used to improve retrieval performance. We strongly emphasize that the usefulness of query expansion should always be considered within the context of the retrieval model (see, e.g. Abdou and Savoy (2008)). We have shown that in the context of language model IR, the contribution of using query expansion seems to be related to the performance of document indexing.

As a concluding remark, we mention that our KNN system is available as a web service found at <http://www.ebi.ac.uk/Rebholz-srv/MeshUP/>². Additionally, we offer assistance and expertise in reproducing our system and other results.

Conflict of Interest: none declared.

REFERENCES

Abdou,S. and Savoy,J. (2008) Searching in medline: query expansion and manual indexing evaluation. *Inf. Process. Manage.*, **44**, 781–789.

²We do note that this tool cannot be used directly for reproducing the results in the original paper because as a production system it does not distinguish between training and test sets of Medline abstracts.