

IR between Science and Engineering, and the Role of Experimentation

Norbert Fuhr

Department of Computer Science and Applied Cognitive Science,
Faculty of Engineering, University of Duisburg-Essen, 47048 Duisburg, Germany
`norbert.fuhr@uni-due.de`

Abstract. Evaluation has always played a major role in IR research, as a means for judging about the quality of competing models. Lately, however, we have seen an over-emphasis of experimental results, thus favoring engineering approaches aiming at tuning performance and neglecting other scientific criteria. A recent study investigated the validity of experimental results published at major conferences, showing that for 95% of the papers using standard test collections, the claimed improvements were only relative, and the resulting quality was inferior to that of the top performing systems [AMWZ09].

In this talk, it is claimed that IR is still in its scientific infancy. Despite the extensive efforts in evaluation initiatives, the scientific insights gained are still very limited – partly due to shortcomings in the design of the testbeds. From a general scientific standpoint, using test collections for evaluation only is a waste of resources. Instead, experimentation should be used for hypothesis generation and testing in general, in order to accumulate a better understanding of the retrieval process and to develop a broader theoretic foundation for the field.

Reference

- [AMWZ09] Armstrong, T.G., Moffat, A., Webber, W., Zobel, J.: Improvements that don't add up: ad-hoc retrieval results since 1998. In: Cheung, D.W.-L., Song, I.-Y., Chu, W.W., Hu, X., Lin, J.J. (eds.) CIKM, pp. 601–610. ACM, New York (2009)