

German Indexing and Retrieval Test Data Base (GIRT)

Some Results of the Pre-test

Michael Kluck

InformationsZentrum Sozialwissenschaften
LennEstr. 30
D-53113 Bonn, Germany
mkl@bonn.iz-soz.de

Abstract

The Project German Indexing and Retrieval Test Data base (GIRT) provides a framework for the comparison of indexing and retrieval systems using German scientific information which is based on the social sciences. The aims and methods of the test environment are described. Some results of the pre-test, which was carried out with the retrieval systems Messenger and freeWAISf, are shown and the consequences are discussed. The results are shown with respect to the precision and recall values, to the intersection of hits and the distribution of the hits on both systems, to the transformation of the queries into searches by the test persons.

1 General Aspects of GIRT

The Project German Indexing and Retrieval Test Data base (GIRT) provides a framework for a comprehensive comparison of indexing and retrieval systems. Including intelligent indexing and software ergonomics the power of the systems can be judged in contrast to customary systems.

Indexing and retrieval systems, which already exist or which are being developed, are to be tested with respect to their capacity and usability specially for the area of scientific information. On the part of retrieval the superiority of quantitative statistical systems over purely Boolean methods is accepted by common consent inside the information science community. This is approved by several tests, specially in the context of TREC (Text Retrieval Conference): Experiments have shown that vector space models and probabilistic models (i.e. such models, which possess a ranking functionality in the output) have better results than exact match models.¹ But the results of these systems still are insufficient because they are unsatisfactory with respect particularly to the qualitative results (mainly the low overlap of the results of different systems)². Therefore the quantitative statistical systems still are critical discussed.

The present discussion concerning the future perspective of intellectual indexing and referring to the power of automatic indexing and retrieval systems relies on the surveys and the experiments that were carried out in the English speaking world, on top within the above mentioned TREC initiative. But the results cannot be directly transformed to the situation of German language scientific information (and within that to the IZ): In the framework of TREC their are test data which mainly consist of English text and of newsletter and newswire text which make other demands on the search process than the reference retrieval in a scientific data base and in German texts. Comparable test results, which rely on research conducted with data bases out of the scientific information area and which reflect the specific problems of domain specific terminology at the same time, do not exist. Surveys on German materials have still not often been carried out, but it is necessary to test the existing TREC results affecting linguistic components of automatic systems with German language data.

With the construction of the GIRT test data base and the offer of test facilities we want to remedy those two deficits and to give a solid basis for the comparison of automatic and intellectual indexing. The necessity of optimising the own resources and of orienting on the actual technical developments and the actual research makes it worth that information centres like the IZ gather information on an empirical basis and stimulate the development of such systems. It is necessary to recognise the advantages and the problems of the different systems during practical tests and then to derive the criteria for the selection and the combination of the different methods or modules.

¹ Womser-Hacker 1996, p. 19 [translated from German]

² Shaw/Burgin/Howell 1997 state: Retrieval performance in traditional and TREC test collections is generally mediocre; recall, precision, and effectiveness are rarely greater than 0.50.