FAKYR - a Method Base System for Education and Research in Information Retrieval

P. Bollmann, E. Konrad, H. Zuse
Technische Universität Berlin

Abstract

An information retrieval system FAKYR is described which incorporates a variety of methods for organizing and retrieving information and for evaluating retrieval effectiveness. The system has been developed in order to support education and research in the area of information retrieval. With respect to this purpose FAKYR is a comfortable and large method data base.

1 Introduction

This paper is concerned with an information retrieval system FAKYR (Fachbereich-Kybernetik-Retrievalsystem) which has been developed at the Technische Universität Berlin. The FAKYR project was initiated 1972 in the spirit of the work of Salton (/7/). Similar to the SMART system FAKYR is an experimental document retrieval system incorporating a variety of methods for organizing and retrieving information and for testing hypotheses. Different from SMART the FAKYR system is working on-line in a time-sharing environment; furthermore it has a fuzzy retrieval component and allows automatically generating dendrograms and plotting evaluation graphs.
The FAKYR system is running on an ITEL AS 5-703\textsuperscript{1} (4 MB main memory, 4 GM secondary memory, 1.4 M/PS) under the operating system VM 370. It consists of 869 PL/I programs (136 609 lines PL/I code) and 30 assembler programs. The working set is 1 MB. FAKYR calculates ca. 2000 correlations per second. The response time is 10 seconds on the average for 1 query posed to 15 000 documents. Up to now, 45 man years of work have been done for designing and implementing FAKYR. This large system is documented by applying the ISAC method.

FAKYR is a method base system (/5/) containing the following components:

a) Data base management:
   direct and inverted files, thesaurus organization, trivial word dictionary, statistical data.

b) Retrieval:
   Boolean retrieval, ranked retrieval by 43 association measures, fuzzy retrieval, modified retrieval by classes.

c) Relevance feedback:
   modification of queries, modification of documents, hybrid strategies.

d) Classification:
   generation of connected components, cliques, stars; single pass algorithm; automatic generation of dendrograms for hierarchical classification.

e) Automatic evaluation:
   calculation of recall, precision and fallout; plotting of recall-precision and recall-fallout graphs.

f) Automatic generation of document spaces on a plotter.