

Information Retrieval in Digital Libraries: Dealing with Structure

Norbert Fuhr
Informatik VI
University of Dortmund
August-Schmidt-Strasse 12
D-44227 Dortmund
fuhr@cs.uni-dortmund.de

ABSTRACT

Whereas classical information retrieval (IR) methods have dealt with unstructured documents only, document structures in digital libraries (DL) pose new problems. In this talk, we present two IR approaches that address this issue from different perspectives.

From the system perspective, there is a need for query languages for retrieval in structured documents. As an alternative to XML query languages that are about to be standardized, we describe XIRQL, an extension of XQL with IR concepts such as weighting and ranking, data types with vague predicates and relevance-based selection of document parts.

From the user point of view, document and data structures in DLs offer a broad variety of new searching and browsing possibilities. In order to guide the user during the search, we adopt the idea of high-level search activities proposed by Bates. In addition to the low-level functions offered by typical DL systems, strategic support is given by means of so-called tactics and strategems. We have implemented an agent-based system as frontend to a federated digital library that supports several high-level search activities.