Exhibit 37

FUZZY SET THEORETICAL APPROACH TO DOCUMENT RETRIEVAL†

TADEUSZ RADECKI

Wrocław Technical University, Main Library and Scientific Information Centre, Wybrzeże St. Wyspiańskiego 27, 50-370 Wrocław, Poland

Abstract—The aim of a document retrieval system is to issue documents which contain the information needed by a given user of an information system. The process of retrieving documents in response to a given query is carried out by means of the search patterns of these documents and the query. It is thus clear that the quality of this process, i.e. the pertinence of the information system response to the information need of a given user depends on the degree of accuracy in which document and query contents are represented by their search patterns. It seems obvious that the weighting of descriptors entering document search patterns improves the quality of the document retrieval process.

A mathematical apparatus which takes into consideration, in a natural manner, the fact that the grades of importance of the descriptors in document search patterns are of the continuum type, that is an apparatus adequate to the description of a retrieval system of documents indexed by weighted descriptors is—among known mathematical methods—the theory of fuzzy sets, formulated by L. A. Zadeh.

It is the aim of this paper to present a new method of document retrieval based on the fundamental operations of the fuzzy set theory. We start by introducing basic notions, then the syntax and semantics of the proposed language for document retrieval will be given and an algorithm allocating documents to particular queries will be described and its properties discussed.

The basic advantage of the use of the fuzzy set theory for document retrieval system description is that it takes into consideration, in a simple way, the differentiation of the importance of descriptors in document search patterns and the differentiation of the formal relevance grades of particular documents of an information system to a given query. Documents of the highest grades (in the given information system) of formal relevance to the given query may be retrieved by means of the application of simple operations of the fuzzy set theory.

1. INTRODUCTION

The aim of a document retrieval system is the issue of documents which contain information required by a given user of the system. The process of document retrieval is such that in response to a given query those documents are issued whose search patterns correspond best to the query search pattern, according to a given criterion. Most usually document search patterns are sets of index-terms representing the ideas contained in the subject-matter of the documents, or sets of index-terms with numerical weights assigned to the terms according to their importance. Sometimes the search patterns of queries are created in a similar way but more usually they are constructed by using the Boolean operators AND, OR and NOT to connect the index-terms of queries. Often while creating the search patterns of documents and queries one uses a thesaurus which is a set of terms on which specific kinds of relations are defined, e.g. the relation of synonymity, the relation of hierarchy, the relation of affinity. It is intuitively obvious that the weights assigned to index-terms and the use of a thesaurus during the creation of search patterns of documents and queries improve this process as well as the effectiveness of the document retrieval process. This has also been confirmed experimentally (see e.g.[1]).

---

†A preliminary version of this paper was presented at the Sixth Cranfield International Conference on Mechanised Information Storage and Retrieval Systems (26–29 July 1977).