

Flexible Representation and Querying of Heterogeneous Structured Documents.

Gloria Bordogna; Gabriella Pasi

Abstract: In this paper we present a fuzzy model for representing documents having a hierarchical structure and possibly containing multimedia information. We consider an archive containing documents with distinct (heterogeneous) logical structures. We also propose a flexible query language for expressing soft selection conditions on the structured documents. The documents' content is organized into thematic (topical) sections where the index terms play a distinct role. The proposed document representation is adaptive to the user, who can indicate the preferred sections of documents, i. e. those which they estimate to bear the most interesting information, and can linguistically quantify the number of sections which determine the global potential interest of the documents. Linguistic quantifiers in the query specify the approximate number of the sections in which the query terms should appear.

Keywords:

AMS Subject Classification: 68T;