Flexible Querying of Semi-Structured Information
(Invited Talk)

Gabriella Pasi
CNR-ITC, via Bassini 16, 20133 Milano, Italy.
gabriella.pasi@itc.cnr.it

Abstract

The problem of dealing with semi-structured information is an increasingly studied topic by both communities of researchers in Database Management and in Information Retrieval. DBMSs and IRSs are the two main types of systems aimed at representing, storing and retrieving information. When semi-structured information is considered, the main problems addressed in the design of these systems are to represent and to inquiry semi-structured data and structured documents respectively.

In DBMSs by semi-structured data it is meant data which do not conform to a strict database schema, i.e. whose structure is not regular [1]. In Information Retrieval Systems by structured documents it is meant documents sharing a basic structure, constituted by sub-components (or document sections).

With the development of the World Wide Web the diffusion of the de-facto standards for the definition of structured documents such XML, witnesses the tendency of producing documents in which the information is organized into (often hierarchical) components. Two main aspects have received a strong emphasis In recent years: the integration of data having different structures and the management of documents having heterogeneous structures.

In particular, XML is increasingly gaining importance as a standard format for information interchange on the WWW. XML has been employed as a basic model for describing semi-structured data, and it constitutes the basic standard for representing structured documents in IR.

In order to inquiry semi-structured information the need of flexible query languages has soon emerged. In the context of semi-structured databases, by flexible query languages it is substantially meant languages that take into account the lack of a rigid schema of the database, thus allowing to enquiry both data and the type/schema [1,3]. In the context of IRSs, modelling flexibility means to take into account the possibility to explicitate a non-uniform structure of the documents when formulating queries.

In the talk, the problem of defining flexible query languages for inquiring XML documents is considered. In particular, the use of fuzzy set theory to define flexible query languages to the aim of expressing soft selection conditions on both the documents’ structure and contents is presented.

Some References


