

The most important integrity constraints in XML are primary keys and foreign keys. In general, keys are essential in understanding both the structure and properties of data. They provide an instrument by which values from a given set of attributes uniquely identify tuples in a database. As a result, keys are important to main database operations. Since XML becomes lingua franca for data exchange on the web, it is widely accepted as a model of real world data. Because XML documents in general can appear in any semi-structured form, structural constraints (including keys) are often imposed on the data that are to be modified or processed. These constraints are formally defined in a schema. Unfortunately, in spite of the obvious advantages, the presence of a schema is not mandatory and many XML documents are not joined with any. Consequently, no integrity constraints are specified in those documents, neither. This thesis is mainly focused on the inference of primary and foreign keys from XML documents.