The rapid progress of network and storage technologies has led to a huge amount of electronic data such as webpages and XML data being available on intra and internet. These electronic data are heterogeneous collection of ill-structured data that have no rigid structure, and are often called semi-structure data. These semi-structured data are stored in
large repositories (XML databases) and stored as a graph internally in database with tuple as nodes and relationships as edges. As there is ever-growing availability of semi-structured information on web and digital libraries, there is a need of effective keyword search in order to fetch the correct and proximal result on Semi-Structured Data. This paper conducts a survey on how keyword search can be performed on semi structure data, techniques involved in performing it, various result ranking strategies and result analysis techniques. It includes the analysis of various indexing schemes and different approaches for increase performance using caches for XML data in order to answer queries.

Reference

Survey on Information Retrieval in Semi Structured Data

- INEX. Initiative for the evaluation of xml retrieval. http://inex.is.informatik.uni-duisburg.de.

Index Terms

Computer Science           Information Retrieval

Key words

XML       ranking       Indexing       keyword

search