Abstract

XML has gained prominence as data storage and exchange format for web applications. This is because there are certain features which are unique to XML like self descriptivism, extensibility and non proprietary text document storage. In spite of all these unique features XML has an inherent limitation of verbosity. This size problem of XML should be dealt with efficiently so that a good compression is achieved and at the same time the compressed data is directly queriable i.e. it should not require decompression at the time of querying. The proposed technique creates a new query engine based on novel three dimensional indexes consisting of structure, attribute and content index. The structure index consists of all unique root to leaf paths of the XML document, the content index stores the contents path wise i.e. all the contents of one particular type of path class is stored in one file and attribute index is created in manner similar to that of content index. Based on this three dimensional compact storage a new query engine is proposed which can answer xpath queries very efficiently. This approach dramatically reduces the storage requirement for XML coupled with efficient processing of xpath queries.
A New Query Engine using Novel Three Dimensional Index for XML Documents

- A. Arion, A. Bonifati, G. Costa, S. Dapos;Aguanno, I. Manolescu, and A. Pugliese,
"XQueC: Pushing queries to compressed XML data," in Proceedings of the 29th International Conference on Very Large Data Bases (VLDB'03), 2003.

**Index Terms**

Computer Science      Query Processing

**Keywords**

Structure index  content index  attribute index.