Abstract

XML is recognized as a standard for data storage and exchange for web applications. This is because it has certain unique features like it is self-describing, extensible and it is stored in the form of text document. In spite of all these unique features XML has an inherent limitation of verbosity. Because of the strong presence of XML in database technology and its inherent verbosity there is ever increasing need to design compact storage for XML which can be effectively utilized for efficient indexing and querying of XML. The proposed technique creates a structure index which is a compact summarization of the XML document and data index which groups and stores the contents of all similar paths at one place. Based on this compact storage a novel query algorithm 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. The implementation of this technique and comparison with other techniques confirms our claim.

References

Efficient Querying of Structure and Contents for XML Documents

- Igor Totarinov, Stratis D Vigals, Kevin Beyer, Jayavel Shanmugasundram, Eugene Shekita, Chun Zhang, "Storing and Querying Ordered XML using a Relational Database System", in Proceeding of ACM SIGMOD Int'l Conference on anagement of Data, Madison Wisconsin USA, pp. 204-215, June 3-6 2002.
- A. Arion, A. Bonifati, G. Costa, S. D'aposta, A. Manolescu, and A. Pugliese, "XQueC: Pushing queries to compressed XML data", in Proceedings of the 29th International Conference of Very Large Data Bases (VLDB'03), 2003.
- Li Ying, MaJun Sun Yun, "Applying Dewey Encoding to Construct XML Index for Path and Keyword Query", in Proceeding of First International Workshop on Database Technology and Application 09, Wuhan, Hubei,China, pp553-556, 25-26 April 2009.
- Radha Senthilkumar, Priyaa Varshinee and A. Kannan. "Designing and Querying a

Index Terms
Computer Science Information Sciences

Keywords
Compact Storage Structure Index Content Index