Abstract

Worldwide standard for data definition is XML. For developing SOA based applications XML is extensively used. SOA based applications contains many different applications which are integrated to each other. For solving the problem of interoperability XML documents are used. XML is widely used for a variety of tasks, including configuration files, protocols, and web services. XML has problem with processing. It is verbose nature. Simple messages can be quite large, containing very small information. In XML documents lots of information are duplicated, which take more computing resources and thus performance of web services decreases. Lots of research is going on regarding how to process XML, so that web services’ performance can
increase. We present an algorithm for compressing XML documents using Document Type Definition (DTD) specifications. Our algorithm is based on lossless compression technique. The model used for compression and decompression is generated automatically from the DTD, and is used in conjunction with an arithmetic encoder to produce a compressed XML document. Our compression technique is on-line, that is, it can compress the document as it is being read. We have implemented the compressor generator, and we have mentioned the results of our experiments performed with XML documents created from Oracle database. The average compression is better than that of XMLPPM and XMill. The processor, XPrFAST, is able to compress large documents where XMLPPM failed to work as it ran out of memory. The technique we have proposed is simple and effective and we have compared it with XMLPPM and XMill.

References

- Bzip2: (http://www.bzip.org)
- DBLP: (http://www.informatik.uni-trier.de/?ley/db)
- Roza Leyderman, Oracle Database Sample Schemas, 11g Release 1 (11.1), B28328-03, Oracle, 2008.
- UniProt: (http://www.ebi.uniprot.org)
- XMLZip, http://www.xmls.com

Index Terms

Computer Science  
Information Technology

Keywords

Arithmetic coding  compression ratio  DTD  DFA  XMLPrFAST