

ABSTRACT

Schema-Based XML Compression

Marvin Steve Virgil

XML has become a popular means for the storage and interchanging of data. However, the textual self-describing nature of XML, results in verbose, relatively large documents. Over the past few years, many methods of addressing this issue have been proposed.

In this report, several of these existing compression methods are surveyed. Additionally, a new method is proposed, which compresses XML documents with the assistance of information extracted from their associated Schemas. Experimental evaluation has shown this method to provide reasonable performance, while only requiring a single pass over the XML document.

Keywords: Extensible Markup Language, XML, Schema, Data Compression, Metadata, Arithmetic Encoding