INTRODUCTION

The developments in database connectivity, markup languages, Web server extensions, and the standardization efforts relating to such technologies are already revolutionizing the interactions between consumers and businesses as well as those between businesses. Theoretical advances in semi-structured data models are evolving and percolating through Web application software development in the integration of back end databases (and data warehouses) with front-end browsers to support decision-making, electronic commerce, and document management in all their glory (transaction management, concurrency control, versioning, document/information retrieval, security, etc.).

In this chapter, we provide a brief description of the development of markup languages, describe their role in the context of web application development and deployment, and finally provide an example of an XML application for a simple Web page displaying a typical sales invoice. Our objective in this example is to illustrate the relationship between the relational and XML representation of data that is required for the use of XML in the development of web-based applications in electronic commerce.
TEXT AS SEMISTRUCTURED DATA

Structured data such as those in relational (object, hierarchical or network) databases can be contrasted with text documents in that while the former have schema, the latter are schema-less or self-describing (see for example, Abiteboul et al., 2000). Since structured data has schema, there is a clear distinction between data and metadata; database stores only the data and the metadata describing the semantics of the data is embedded in the specification of the database (in, for example, the relational schemas in relational database systems). In case of text documents, on the other hand, the text content as well as the metadata pertaining to the document are stored together by serialization in byte-streams in the electronic text databases. While the model of data in relational (or other structured databases such as hierarchical, network, or object-oriented) is well understood, the study of modeling of text documents is more recent. Abiteboul et al., (2000) call such text-oriented data semi-structured in that there is no separate description of the structure and type of data.

The metadata describing the semantics of text documents consists of the form, structure and content of such documents, and the text document itself is a serialized byte-stream that contains data as well as such metadata. The metadata relating to form specifies the format of the text content and so governs the presentation of such content with respect to fonts (type, size, color, bold, italics, etc.), alignment, etc. The metadata relating to the structure, on the other hand, governs the presentation with respect to aspects such as headings (level), lists (ordered or unordered), etc. The metadata relating to the content, however, depends on the domain that the document originates and therefore requires an understanding of such underlying domain.

Early Markup Languages

During the early stages in the development of markup languages, the objective was to embed the metadata relating to the formatting of the document in such document itself. For example, in the earliest markup languages such as runoff, troff/nroff, and to a large extent in Donald Knuth’s (1992) TeX typesetting system that supported formatting of text, the tags almost exclusively had to do with the formatting of documents. The distribution of troff with the unix systems made it popular in scientific publishing since, in those early days, the delivery platform for such documents was primarily the print medium. The cost-free availability of the TeX/LaTeX system on diverse platforms (with the associated utilities for browsing, postscript/HTML conversion) has made it, since then and to this day, the platform of choice for
Related Content

A Measurement Ontology Generalizable for Emerging Domain Applications on the Semantic Web
www.igi-global.com/article/measurement-ontology-generalizable-emerging-domain/3365?camid=4v1a

Goal-Oriented Requirement Engineering for XML Document Warehouses
www.igi-global.com/chapter/goal-oriented-requirement-engineering-xml/7932?camid=4v1a
Ex Ante Evaluations of Alternate Data Structures for End User Queries: Theory and Experimental Test
[www.igi-global.com/article/ante-evaluations-alternate-data-structures/3320?camid=4v1a](www.igi-global.com/article/ante-evaluations-alternate-data-structures/3320?camid=4v1a)

Using DEMO and ORM in Concert: A Case Study
Jan L.G. Dietz and Terry Halpin (2004). *Advanced Topics in Database Research, Volume 3* (pp. 218-236).
[www.igi-global.com/chapter/using-demo-orm-concert/4361?camid=4v1a](www.igi-global.com/chapter/using-demo-orm-concert/4361?camid=4v1a)