INFSCI2560 – Web Technologies

Degan Hao

deh95@pitt.edu

* **Description: What is being standardized?**
* XML, as a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.
* The W3C's XML 1.0 Specification and several other related specifications—all of them free open standards—define XML.
* The design goals of XML emphasize simplicity, generality, and usability across the Internet.
* It is a textual data format with strong support via Unicode for different human languages. Although the design of XML focuses on documents, the language is widely used for the representation of arbitrary data structures[6] such as those used in web services.
* Several schema systems exist to aid in the definition of XML-based languages, while programmers have developed many application programming interfaces (APIs) to aid the processing of XML data.
  + **Assessment: What are the related standards and how is this different?**
    - Related standards such as Document Style Semantics and Specification Language (DSSSL), which is is an international standard developed to provide a style sheet for SGML documents.
    - The difference between DSSSL and XML
    - Another related standard is HTML.
    - The weakness of HTML can be, for example, HTML has very little internal structure, which means that you can easily write valid HTML that does not make sense at all when you consider the semantics of the elements.
    - The main difference between XML and HTML is that, by defining our own markup language, we can encode the information of your documents much more precisely than is possible with HTML.
* **Description: What is the goal of the standard and why is it necessary?**

The goals of the XML are as follows:

* XML shall be straightforwardly usable over the Internet.
* XML shall support a wide variety of applications.
* It shall be easy to write programs which process XML documents.
* XML documents should be human-legible and reasonably clear.
* XML shall be compatible with SGML.

The reason why it is necessary:

The historical reason for this goal is that the complexity and difficulty of SGML was greatly increased by its use of minimization, i.e. the omission of pieces of markup, in the interest of terseness. In the case of XML, whenever there was a conflict between conciseness and clarity, clarity won.

Moreover, GML parsers were a pain to write. An SGML processor was really a compiler compiler where you could change delimiters, keywords and a whole lot of different behaviors. Note that process here is a defined term: a XML processor is the parser and support utilities. This goal does not state that it is against XML's goals to write complicated programs that use XML data!

* + **Assessment: Does this standard achieve the goal and solve the need?**
    - Yes.
    - Simplicity: Information coded in XML is easy to read and understand, plus it can be processed easily by computers.
    - Openness: XML is a W3C standard, endorsed by software industry market leaders.
    - Extensibility: There is no fixed set of tags. New tags can be created as they are needed.
    - Supports multilingual documents and Unicode. And this is important for the internationalization of applications.
    - Can embed existing data: Mapping existing data structures like file systems or relational databases to XML is simple.
* **Description: How is the standard being standardized?**
* Character: An XML document is a string of characters. Almost every legal Unicode character may appear in an XML document.
* Processor and application: The processor analyzes the markup and passes structured information to an application. The specification places requirements on what an XML processor must do and not do, but the application is outside its scope. The processor (as the specification calls it) is often referred to colloquially as an XML parser.
* Markup and content: The characters making up an XML document are divided into markup and content, which may be distinguished by the application of simple syntactic rules. In addition, whitespace before and after the outermost element is classified as markup.
* Tag: A tag is a markup construct that begins with < and ends with >.
* Element: An element is a logical document component that either begins with a start-tag and ends with a matching end-tag or consists only of an empty-element tag.
  + **Assessment: Does the standardization mechanism limit its effectiveness or adoption?**
    - Although XML has been widely used internationally, there are still criticisms involving the limitation of its effectiveness and simplicity.
    - XML and its extensions have regularly been criticized for verbosity and complexity.
    - Mapping the basic tree model of XML to type systems of programming languages or databases can be difficult, especially when XML is used for exchanging highly structured data between applications, which was not its primary design goal.
    - Other criticisms attempt to refute the claim that XML is a self-describing language (though the XML specification itself makes no such claim).
    - JSON, YAML, and S-Expressions are frequently proposed as alternatives (see Comparison of data serialization formats);
    - The focus on representing highly structured data rather than documents, which may contain both highly structured and relatively unstructured content.
* **Description: Who are the prime movers and/or contributors to the standard?**
* The XML standards was developed by‎: ‎[World Wide Web Consortium](https://en.wikipedia.org/wiki/World_Wide_Web_Consortium)
* IETF RFC 7303 gives rules for the construction of Internet Media Types for use when sending XML.
* XML was compiled by a working group of eleven members, supported by a (roughly) 150-member Interest Group.
* A record of design decisions and their rationales was compiled by Michael Sperberg-McQueen on December 4, 1997.
* James Clark served as Technical Lead of the Working Group, notably contributing the empty-element "<empty />" syntax and the name "XML".
* The co-editors of the specification were originally Tim Bray and Michael Sperberg-McQueen.
* Halfway through the project Bray accepted a consulting engagement with Netscape, provoking vociferous protests from Microsoft.
* Bray was temporarily asked to resign the editorship. This led to intense dispute in the Working Group, eventually solved by the appointment of Microsoft's Jean Paoli as a third co-editor.
  + **Assessment: What specific consequences (good or bad) result from the prime mover contributors.**
    - Neutral: The name XML was created by James Clark and has been used until today.
    - Good: Microsoft’s involvement especially using XML as configuration file is helpful for building project.
    - Bad: XML and its extensions have regularly been criticized for verbosity and complexity.
    - Good: Almost any Unicode code point can be used in the character data and attribute values of an XML 1.0 or 1.1 document
    - Good: The notion of well-formedness as opposed to validity (which enables parsing without a schema) was first formalized in XML.
* **Description: What impact does the standard have on the intended target audience?**
* As a standard, XML eliminates many of the data representation problems that plague HTML
* With XML, the presentation of data is handled separately with either Cascading Style Sheets (CSS) or the Extensible Style sheet Language (XSL), both of which are also W3C standards.
* Data Base: XML allows the creation of an indefinitely large number of new data labels or tags. This means that with XML, information publishers are free to invent their own tags as needed for their particular applications, or to work together with other organizations to define shared sets of tags that promote interoperability.
* Applications: XML documents can be associated with a Document Type Description (DTD), which defines a structure for conforming applications. This allows applications that import data to validate that data for conformation to the DTD.
* Data Sharing: XML provides some assistance with distribution by supporting mechanisms for remote function invocation across the web.
  + **Assessment: What is your view of the importance and implications of the standard?**
    - Standard are backbones of Communication.
    - The Standard provides some syntax that can be shared by people working in different areas just like how language works.
    - The very beginning period when standard forms is crucial to the development of the whole industry that is guided by that standard.
    - When designing standards, we should take into consideration some basic principles such as least astonishment and the balance of simplicity and informative.
    - It is very important to comply with web standards to Complying with Web standards can give your Web pages greater visibility in Web searches.