



HTML's History and Versions

- The “War” Years
- The Clean-up Effort

S. Prabhavathi
AP/IT



HTML: History



- 1990: HTML invented by Tim Berners-Lee
- Elements used: title, paragraph, hyperlinks, headings, simple lists, glossaries, and address blocks (containing information about the document author, and typically italicized).
- No facility for producing tables or fill-in forms, much less for including images within a document.



HTML:The “War” Years



- 1993: Mosaic browser adds support for images, sound, video to HTML
- “Browser wars” between Netscape and Microsoft.
- 1993~1997: HTML was defined operationally by the elements that browser support developers chose to implement and the ways in which their browsers responded to these elements.
- ~1994: Tim Berners-Lee launched the World Wide Web Consortium (W3C), in part with the goal of producing standards for HTML.
- “browser wars” slowed and the standards community caught up.



The Clean-up Effort



- Change in HTML standards then come from the standards community more than from browser developers.
- Defining a language such as HTML (or any computer language) involves two aspects: **its syntax and its semantics.**
- The ***syntax*** of a computer language defines which strings of characters represent a document that conforms to the language and which do not.

Ex: For a programming language such as Java, a program that compiles is a syntactically correct document.

- The ***semantics*** of a language is a description of what the various elements of a syntactically correct document mean.

For example, a syntactically correct assignment statement in Java has a certain meaning: a variable is associated with a value that can later be referenced by the variable's name.

Similarly, the `p` element in HTML 4.01 also has a certain meaning: its content is to be displayed as a paragraph.



The Clean-up Effort



- The ***semantics*** of the elements and attributes in HTML 4.01 are defined using ***natural language***
- The ***syntax*** for a computer language is defined using some other language specially designed for the purpose of defining language syntax.
- A language used to describe the syntax of other languages is sometimes referred to as a ***metalanguage***.
- The metalanguage commonly used to describe the syntax of programming languages such as Java is called Backus-Naur Form (BNF) notation. In fact, BNF notation could also be used to define the syntax for HTML.



The Clean-up Effort



- The metalanguage used to define the syntax for HTML 4.01 is SGML, (**Standard Generalized Markup Language.**)
- SGML's generality increases the difficulty of parsing is its feature allowing certain tags to be omitted.
- Parsing a HTML document involves creating an *abstract syntax tree* or *parse tree*.
- It is obviously more difficult to write a parser that allows for omitted tags.
- In 1998, W3C introduced the **Extensible Markup Language (XML)**, a restricted version of SGML.
- A hypertext markup language whose syntax is defined using XML rather than SGML is called **an XHTML language**.
- **XHTML 1.0** (Jan 2000) syntax defined using Extensible Markup Language (XML)



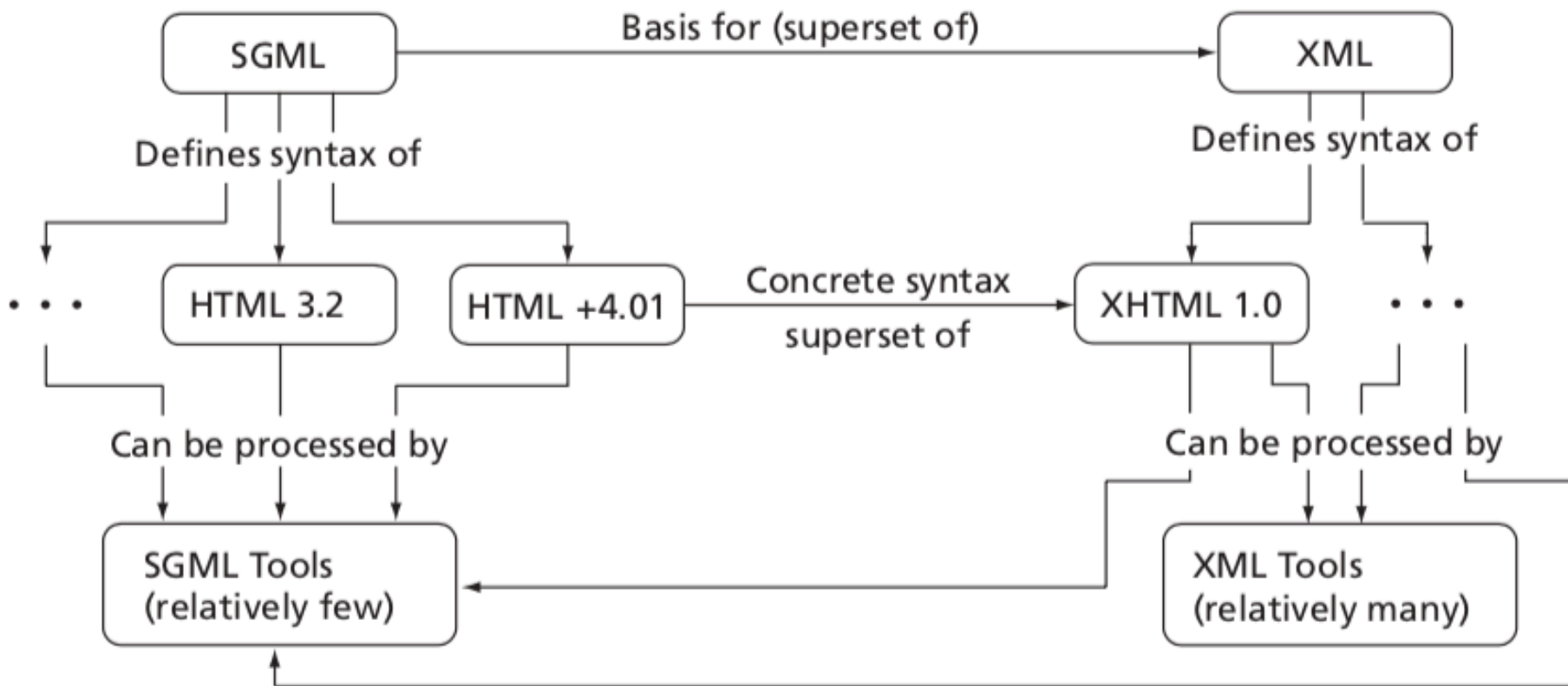
The Clean-up Effort



- Primary differences:
 - HTML allows some **tag omissions** (e.g., end tags)
 - XHTML element and attribute names are **lower case** (HTML names are case-insensitive)
 - XHTML requires that attribute **values** be **quoted**
- **XHTML Basic 1.0**, designed for use with limited devices such as cell phones.
- **XHTML 1.1**, defined using a grammar that is more modular (and somewhat more complicated) than the grammar used to define **XHTML 1.0**.
- A **grammar** is the collection of rules (XML-based rules in the case of these languages) defining the syntax of a language.



Relationships between SGML, XML, HTML and XHTML





Thank you