

B336 Advanced Internet Computing

Introducing XML Parsing and Processing

Learning Objectives

- Learn what XML parsers and processors are.
- Learn why we need XML parsers and processors.

Learning Objectives

- In the scheme of what we are doing in this unit:
 - We are studying how to use XML as an important set of Internet technologies to use as solutions in different areas (for example, the problem defined in Assignment 2).
 - A major part of developing an XML solution is to design and implement software to deal with the information in these XML documents. To do this, we must know how to parse and process XML documents.

XML Processors

- The ultimate aim of getting having an XML document is to do something useful with the information in it.
- To be able to do this with software, we must first be able to:
 - parse the XML document and associated DTD and return the information in a data structure.
 - Check the data structure for errors.
 - Have rich enough facilities in the data structure to manipulate the data to fit our needs.

XML Processors

- So a web browser that displays an XML document using HTML conventions by transforming it using XSLT can be considered an XML processor.

XML Parsers

- An XML parser is a programming API. It provides tools for a programmer to:
 - read and XML document and returns an appropriate data structure/object representing the information in the XML document.
 - check that the document is well-formed, and also valid according to its DTD.
- The term “XML parsers” also refers to a complete programs that parses XML documents and return whether the documents are well-formed and valid.

Validating and Non-validating Parsers

- Parsers which checks for both well-formedness and validity are called **validating parsers**.
- Parsers which only checks for well-formedness are called **non-validating parsers**.
 - All parsers MUST check for well-formedness - an XML document which is not well-formed is NOT an XML document at all.

XML Parsers

- An XML parser is a programming API. It provides tools for a programmer to:
 - read and XML document and returns an appropriate data structure/object representing the information in the XML document.
 - check the the well-formedness and validity status of the document.
- The term “XML parsers” also refers to a complete programs that parses XML documents and return the well-formedness and validity status of the documents.

In the next few lectures...

- A simple example stream-based XML parser in Perl.
- More sophisticated parsers (DOM, SAX) in Perl, Java and other languages.