CS 366 XML

About this course …

- This course is about DATA and VOCABULARY.
- This course is NOT a Web programming course.

- Why taking a course for which
  - plenty of free, excellent tutorials can be found on-line
  - a CS major probably can learn by himself / herself

- "Learning the fundamental of XML might take a programmer a week. Learning how to use XML effectively might take a lifetime."

  -- Elliotte Rusty Harold (an XML expert)

Chapter 1: What is XML?

Chapter 1 Objectives

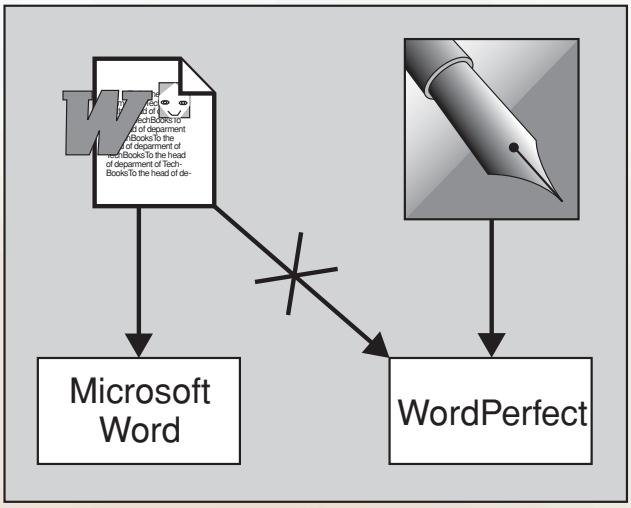
- The two major categories of computer file types—binary files and text files—and the advantages and disadvantages of each
- The history behind XML, including other markup languages such as SGML and HTML
- How XML documents are structured as hierarchies of information
- A brief introduction to some of the other technologies surrounding XML, which you will work with throughout the course
- A quick look at some areas where XML is useful

Of Data, Files, and Text

Binary Files

01010101010101010101

A binary file, at its simplest, is just a stream of bits.

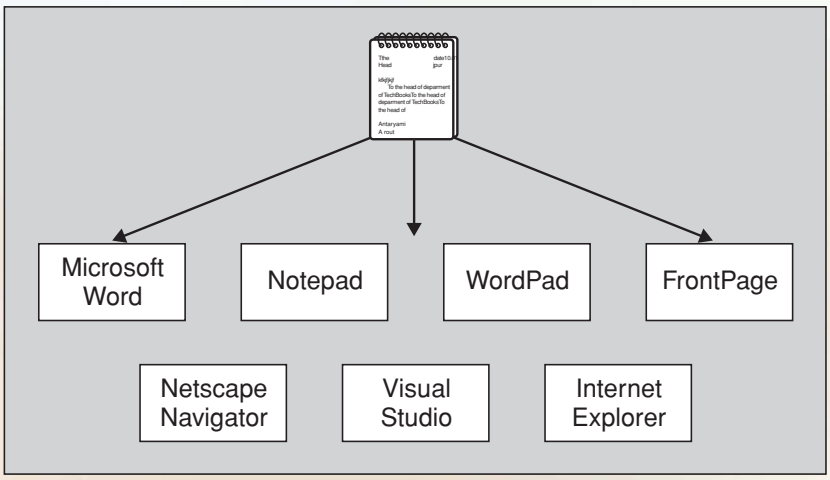


Of Data, Files, and Text

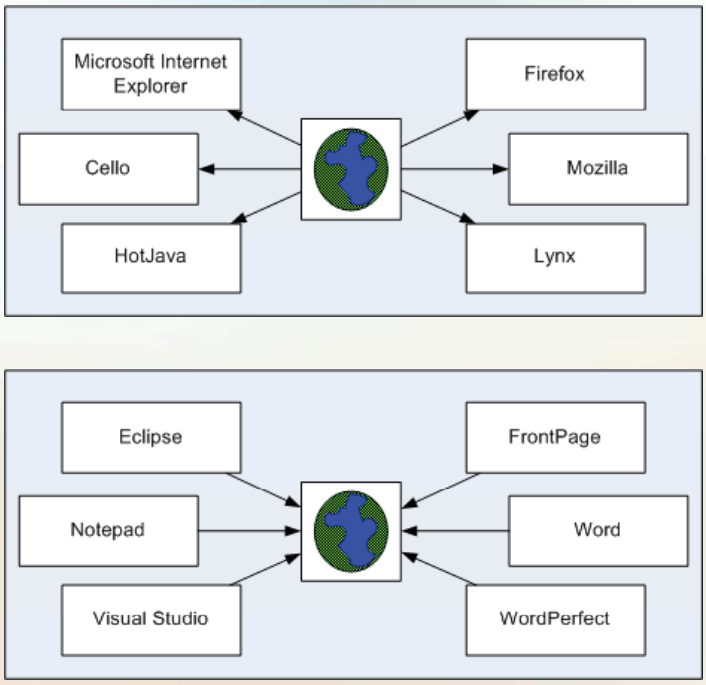
Text Files

1100001 = 97 = a

The streams of bits are encoded and have meaning.



A Brief History of Markup



So What Is XML?

- A way to describe our data
- A subset of SGML
- A standard for creating languages
- Extensible

"Extensible," not "eXtensible."

Try It Out

Opening an XML File in Web Browser

What Does XML Buy Us?

- XML Parsers
  - Many specialized parsers available
- Extensible
  - We can shape it anyway we want
    - SVG
    - MathML
    - CML
      - And some to be invented by you

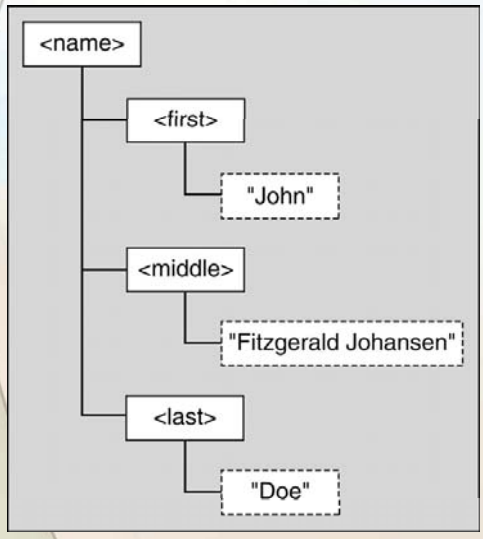
So What is XML?

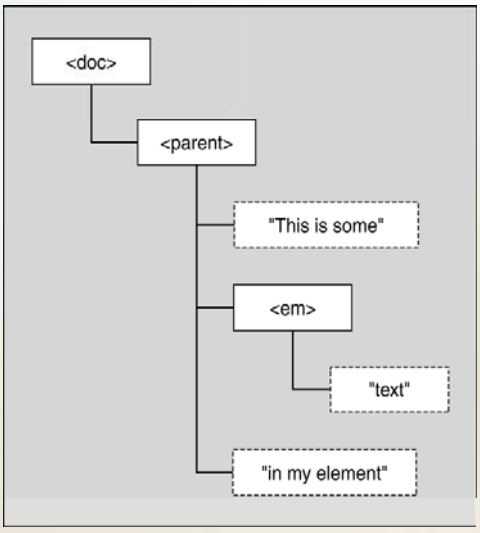
HTML and XML:

- HTML is to *display* as XML is to *data*
  - XML separates *data* and its *presentation**
- HTML is for *humans* to consume as XML is for *machines* to consume
  - XML is self describing
  - XML makes *data exchange among applications* easier

So What Is XML?

Hierarchies of Information





So What is XML?

What's a Document Type?

- **An attempt to describe what the document is in terms of *structuring information***
- ***Vocabulary* is another term to try and describe what the document is in terms of recognizable terms**

What is the World Wide Consortium?

Also known as the W3C

- Well organized *Think Tank*
- Defining standards and specifications since 1994
- Resides at http://www.w3.org
- A place to develop standards
- Anyone can participate

Its a think tank full of talent.

What is the World Wide Consortium?

What Are the Pieces That Make Up XML?

- XML 1.0
- DTDs and Schemas
- Namespaces
- XPath
- CSS, XSL, XSLT
- HTML, XHTML
- Specialized parsers

This is the short list

Where Is XML Used, and Where Can It Be Used?

- Web site content
- Distributed computing
- Semi-structured data modeling
- e-Commerce
- Document engineering
- Data communication protocol design
- Bioinformatics
- Mathematics, chemistry, …
- Semantic web
- Data mashup
- and many more …

Major Advantages of XML

- Extensibility
- Built-in document validation provision
- Both human and machine readable
- Platform independent
- Public standard
- Large and growing set of tools
- Working well with Internet
- Global Unicode support

Disadvantages and Issues

- Major disadvantage – verbosity!
- Issues
  - Not a cure-all for data integration
  - No guarantee for unified format
  - Learning curve