Meta-data (or maybe metadata)

Lawrence Snyder
University of Washington, Seattle
Meta-data Is An Important Idea

- We have discussed tags before
  - HTML – describes page layout
  - Oxford English Dictionary – aided in look & look-up
  - XML – Today’s topic
    - Extensible Markup Language
    - Easy to learn because YOU make it up
    - Introduce the idea today
- Meta-data doesn’t REQUIRE tags, there are other ways of giving it, but they’re most common
Metadata – Specify with Tags

- **Metadata is information about information**

**byte** (baIt). *Computers.* [Arbitrary, prob. influenced by bit sb.\(^4\) and bite sb.] A group of eight consecutive bits operated on as a unit in a computer. 1964 Blaauw & Brooks in *IBM Systems Jnl.* Ill. 122 An 8-bit unit of information is fundamental to most of the formats [of the System/360]. A consecutive group of \(n\) such units constitutes a field of length \(n\). Fixed-length fields of length one, two, four, and eight are termed bytes, halfwords, words, and double words respectively. 1964 *IBM Jnl. Res. & Developm.* VIII. 97/1

When a byte of data appears from an I/O device, the CPU is seized, dumped, used and restored. 1967 *P. A. Stark Digital Computer Programming* xix. 351 The normal operations in fixed point are done on four bytes at a time. 1968 *Dataweek* 24 Jan. 1/1 Tape reading and writing is at from 34,160 to 192,000 bytes per second.

<e><hg><hw>byte</hw> <pr><ph>baIt</ph></pr></hg>. <la>Computers</la>. <etym>Arbitrary, prob. influenced by <xr><x>bit</x></xr> <ps>n.<hm>4</hm> </ps>and <xr><x>bite</x> <ps>n.</ps> </xr></etym> <s4>An 8-bit unit of information is fundamental to most of the formats *of* the System/360*. &es.A consecutive group of <i>n</i> such units constitutes a field of length <i>n</i>.&es.Fixed-length fields of length one, two, four, and eight are termed bytes, halfwords, words, and double words respectively. &es.When a byte of data appears from an I/O device, the CPU is seized, dumped, used and restored. &es.The normal operations in fixed point are done on four bytes at a time.</e>
Metadata Describes Data

- Metadata is data about data ... a description of what the data is
  - Knowing what the data is, as in the OED, allows us to process it better for users
  - Here’s an example: Search OED for def of “binary”
    - Without metadata, get 8,311 hits ... of which one is the definition
    - With metadata, get each definition in order ... how?

<e><hg><hw>binary</hw> ... </hg> ... <e>
Metadata Describes Data

- Metadata is data about data ... a description of what the data is
  - Knowing what the data is, as in the OED, allows us to process it better for users
  - Here’s an example: Search OED for def of “binary”
    - Without metadata, get 8,311 hits ... which one is the definition?
    - With metadata, get each definition in order ... how?

The Principle: We can program computers to better help us if we say what the content is
Metadata Separation

- Metadata describes what the data is, but because the tags can be distinguished from the content, it separates itself from the content – that’s smart

Separate the content and its tags entirely from the processing – produce a data-only file
The Advantage of Separating

- By separating the content from the processing it is possible to maximize expertise
  - The content expert (you) puts the data together
  - The processing expert (some programmers) write the processing code based on the tags

3/1/13 © 2011 Larry Snyder, CSE
The Extensible Markup Language (XML) the tool for defining metadata; YOU think up the tags ... it is a self-defining language!

- The usual rules for tags apply
  - Enclose in < and > and use lowercase ONLY
  - Start tag `<mynewtag>` and End tag `</mynewtag>`
  - Tags must always be matched or self-terminated
  - Tags can have attributes (think those up, too) of form `attributename="valueInQuotes"`
  - Use `.xml` as the file extension
  - Always start with “standard text” (shown later)
Example of XML

- Suppose I want to record information about this class; using XML, I might write:

```xml
<class dept="cse">
  <catalog qsr="true" credits="5">
    <num>120</num>
    <lec len="50" num="3">M, W, F</lec>
    <lab len="50" num="2">Tu, Th</lab>
    <descrip>
      Must-know computing knowledge for the 21st century
    </descrip>
  </catalog>
  <teach>L. Snyder</teach>
</class>
```

I invented the tags; they make sense to me, and I could write software to process such descriptions.
Learning XML

- Since we think up the tags ourselves, it’s the easiest language in the world to learn, right?
- Right.
- It’s trivial?!
- Not quite ... there is a little technique, and we’ll do that now

- Tags can serve in three roles ...
Ways To Use Tags

- **Identity** – tag it so you know what it is
  
  `<name>George Washington</name>`

- **Affinity** – all properties of a thing should be grouped together

  `<personal>`
  
  `<name>George Washington</name>
  `<height>6’ 2”</height>
  `<teeth>Wooden</teeth>
  `<home>Mount Vernon</home>`

  `</personal>`
Ways To Use Tags (continued)

- **Collection** – enclose a group of items of the same type in a collective tag

```xml
<presidents>
  <prez num="1"><personal><name>George ...</name></prez>
  <prez num="2"><personal><name>John ...</name></prez>
  <prez num="3"><personal><name>Thomas ...</name></prez>
  ...
  <prez num="44"><personal><name>Barack ...</name></prez>
</presidents>
```

- These uses become intuitive quickly
Ex:

Classify tag types:
Identity
Affinity
Collection
Here is a simple picture viewer I got off the net…

Two parts
- Setup XML & pics
- View

http://www.flashxml.net/components/
Illustrative Example

- Here is a simple picture viewer I got off the net...
- Two parts
  - Create XML file
  - View

http://www.flashxml.net/components/

Just Do IT
How Photo Gallery Works

Slideshows

3D banner

banner rotator

media slideshow

page flip

accordion

photo stack

photo rotator

banner navigator

focus slider

slide viewer

product viewer

image list
How The Software Works

- The software comes with a brief explanation

```
Name
assets
big.xml
holder
images
  big
  thumbs
ImageScrollerFX
index.html
PhotoGalleryFX.swf
readme.txt
settings.xml
swfobject.js
thumbs.xml
video
```

- The XML metadata for each item
- Folder of picture files
- Folder of thumbnail picture files
- The Web page of the display
- Implementing software
- XML for all customization
- Implementing software
- The XML for the thumbnail picture files
Check out the content, tags

```xml
<images>
  <photo image="images/big/larry.jpg">
    <![CDATA[<head>Larry</head><body>Relaxing</body>]]>
  </photo>

  <photo image="images/big/judy.jpg">
    <![CDATA[<head>Judy</head><body>Pixelated</body>]]>
  </photo>

  <photo image="images/big/jeff.jpg">
    <![CDATA[<head>Jeff</head><body>Ready To Go</body>]]>
  </photo>

  <photo image="images/big/alain.jpg">
    <![CDATA[<head>Alain and Mi-Sug</head><body>Versailles</body>]]>
  </photo>

  ...

  <photo image="images/big/tohru.jpg">
    <![CDATA[<head>Tohru</head><body>At Grandad's</body>]]>
  </photo>

</images>
```
Check out the content, tags

I created this file manually, but there is software to help, or you can write a Processing program for it, too.
settings.xml Is More of the Same

- Software on the FX site help you configure this file interactively

```xml
<settings>
  <General_Properties>
    <position value="bottom"/>
    <hide value="slide"/>
    <slideDirection value="leftright"/>
    <time value="1"/>
    <motion value="Cubic"/>
    <ease value="easeOut"/>
    <useButton value="false"/>
    <arrow value="assets/show.gif"/>
    <margin value="0"/>
    <path value="ImageScrollerFX"/>
    <showAbout value="true"/>
  </General_Properties>
</settings>
```
Summary

- Metadata is data about data
- Tags are a common form of metadata
- XML is main technology for metadata spec.
- Three roles for tags to fill ... you’re building a tree
- By separating data from processing, expertise can be exploited, flexibility, wide usage
- We used metadata to add an image