

IT350 Web and Internet Programming

Fall 2006

SlideSet #16: XML and Semantic Web

(XML from Chapter 20 of text)

Outline

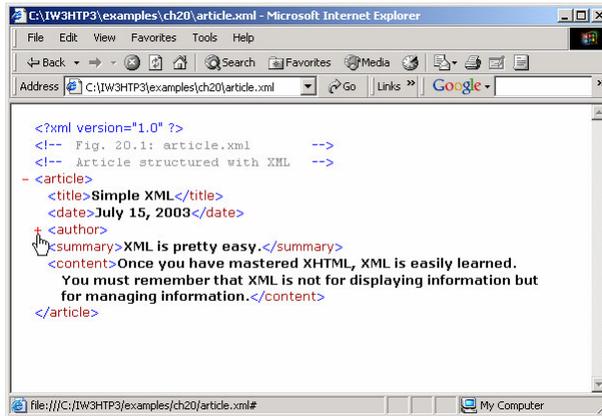
- Why Structured Data?
- Types of Structured Data
 - XML and Friends
 - RDF and Semantic Web

Structured Data

- Overall Goal – add more “structure” to our otherwise freeform data
 - Why?
-
- HTML/XHTML primarily designed to be rendered for _____ viewing. We want something that can be exploited by _____.

XHTML → XML

```
1 <?xml version = "1.0"?>
2
3 <!-- Fig. 20.1: article.xml -->
4 <!-- Article structured with XML -->
5
6 <article>
7
8   <title>Simple XML</title>
9
10  <date>July 15, 2003</date>
11
12  <author>
13    <firstName>Carpenter</firstName>
14    <lastName>Ca1</lastName>
15  </author>
16
17  <summary>XML is pretty easy.</summary>
18
19  <content>Once you have mastered XHTML, XML is easily
20    learned. You must remember that XML is not for
21    displaying information but for managing information.
22  </content>
23
24 </article>
```



Another example

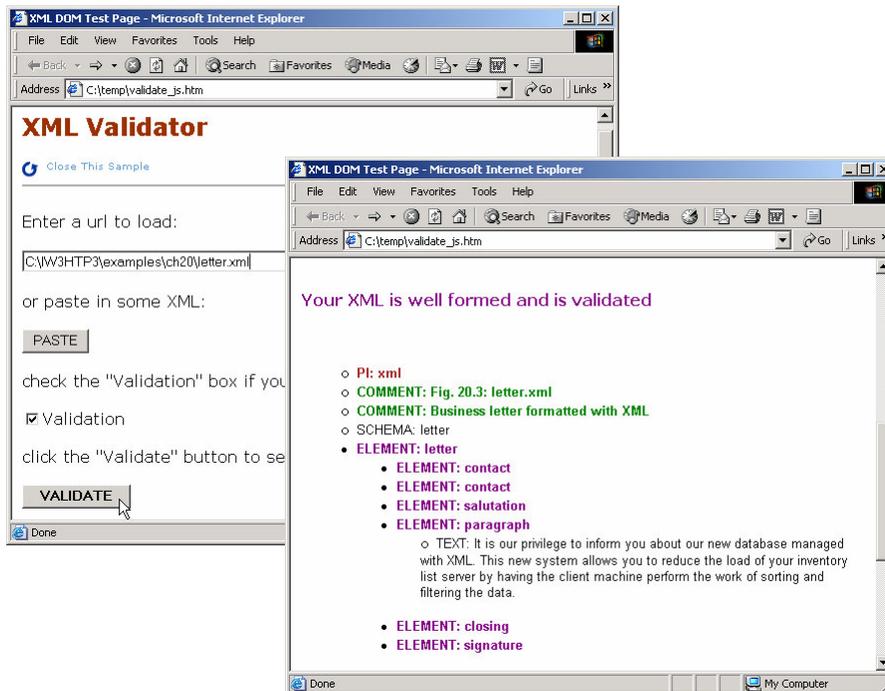
```

1 <?xml version = "1.0"?>
2
3 <!-- Fig. 20.3: letter.xml -->
4 <!-- Business letter formatted with XML -->
5
6 <!DOCTYPE letter SYSTEM "letter.dtd">
7
8 <letter>
9
10 <contact type = "from">
11   <name>John Doe</name>
12   <address1>123 Main St.</address1>
13   <address2></address2>
14   <city>Anytown</city>
15   <state>Anystate</state>
16   <zip>12345</zip>
17   <phone>555-1234</phone>
18   <flag gender = "M"/>
19 </contact>
20
21 <contact type = "to">
22   <name>Joe Schmoe</name>
23   <address1>Box 12345</address1>
24   <address2>15 Any Ave.</address2>
25   <city>Othertown</city>

```

Document Type Definitions

```
1 <!-- Fig. 20.6: letter.dtd -->
2 <!-- DTD document for letter.xml -->
3
4 <!ELEMENT letter ( contact+, salutation, paragraph+,
5   closing, signature )>
6
7 <!ELEMENT contact ( name, address1, address2, city, state,
8   zip, phone, flag )>
9 <!ATTLIST contact type CDATA #IMPLIED>
10
11 <!ELEMENT name ( #PCDATA )>
12 <!ELEMENT address1 ( #PCDATA )>
13 <!ELEMENT address2 ( #PCDATA )>
14 <!ELEMENT city ( #PCDATA )>
15 <!ELEMENT state ( #PCDATA )>
16 <!ELEMENT zip ( #PCDATA )>
17 <!ELEMENT phone ( #PCDATA )>
18 <!ELEMENT flag EMPTY>
19 <!ATTLIST flag gender ( M | F ) "M">
20
21 <!ELEMENT salutation ( #PCDATA )>
22 <!ELEMENT closing ( #PCDATA )>
23 <!ELEMENT paragraph ( #PCDATA )>
24 <!ELEMENT signature ( #PCDATA )>
```



Additional XML Technology

- XML Schema – like DTD, but in XML and more sophisticated
- XPath – way of identifying a particular element inside a document
- XSL – render XML for viewing in browser
- XSLT – transform one XML doc to another

What's Missing With XML?

- XML is centered around *documents*
 - DTDs/Schemas say what is legal
 - Software can use documents with DTDs it “knows about”
- RDF is centered around *statements*
 - Statement is (subject, predicate, object)
 - Example: (<it350> , location, “MI 223”)
 - Example:(<it350> , type, course)
 - XML-based syntax:

```
<course about= http://www.cs.usna.edu/~it350/F06 >
  <location>MI 223</location>
</course>
```
 - RDF =

More Complete RDF Example

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://www.cs.washington.edu/semweb/rdf-equiv-ns#">
  <rdf:Description>
    <course>
      <rdf:Description>
        <name>Networking Seminar</name>
        <time>1 p.m.</time>
        <location>Sieg 134</location>
        <instructor>
          <rdf:Description>
            <value>Prof. John Fitz</value>
            <office hours>Tue 3-4 p.m.</office hours>
          </rdf:Description>
        </instructor>
        <instructor>
          <rdf:Description>
            <value>Prof. Helen Randolph</value>
            <office hours>Fri 9-10 a.m.</office hours>
          </rdf:Description>
        </instructor>
        ...
      </rdf:Description>
    </course>
  </rdf:Description>
</rdf:RDF>
```

RDF → Semantic Web?

- XML forces nesting of tags in certain order
 - Vocabulary and structure fixed by the DTD
- With RDF, can place statements anywhere, in any order
 - And can mix vocabulary (types, predicates) from many different documents
- RDF Schema lets us define different schemas for different domains
 - University domain, sailing domain, animal domain, ...
 - Software that understands <sailing:knots> can utilize this info in any document that it finds it
 - Crawlers can collect knowledge → RDF databases
- Semantic == meaning
 - “Semantic Web” – collection of data for which software can “know” meaning of many of the terms – and exploit this
 - Aggregate data from many different sites (prices, ratings)
 - Search for info where answer not found on any one page
 - Personal agents to reason about such info to buy/schedule/search for you

Example Semantic Services – Calendar

May 23 - 29, 2004

Current Filter Selection: All Non-Course Events

	Mon May 24	Tue May 25	Wed May 26	Thu May 27	Fri May 28
9:30am					
10:30am				> Research Seminar (590 ED)	
11:30am		> General Examination: Alex Yates	> Research Seminar (590 NL)		12:00pm> Lunch with the Chair
12:30pm		> Research Seminar (590 ST)			
1:30pm	> Research Seminar (590 NS)	> Research Seminar (590 DM)			
2:30pm	3:00pm> Research Seminar (590 O)	> Research Seminar (590 CB)			
3:30pm	> Research Seminar (590 DM)	> Grad Stu	http://www.cs.washington.edu/education/grad/exams.html		> Final Examination: Yung-Yu Chuang
4:30pm			> Matt Mohebbi, Robocup	> Research Seminar (590 SY)	

General Examination: Alex Yates
Time: 11:30am
Description:
 • Presenter: Alex Yates
 • Remarks: Advisor: Oren Etzioni
 • Topic: Noun Phrase Coreference in English: Paradigms and Methods
 • Location: Allen Center, CSE 678
 • Source:
<http://www.cs.washington.edu/education/grad/exams.html>

(<http://abstract.cs.washington.edu/~semweb/public/webcal/week.php>)

Example Semantic Services – “Who’s Who”

Research Interests: high-level languages, and programming systems support for such languages

Advisor: Craig Chambers

Julie Letchner - (Graduate Student)

Portrait:

Office: Allen 390

Year Of Study: First year

Advisor: Dieter Fox

Birthday: June 17

Hobbies: Dancing (ballet, modern, jazz), jogging, reading

Alan L. Liu - {aliu_at_cs.washington.edu} - (Graduate Student)

Portrait:

Office: CSE 490

Work Phone: x3-4149

Year Of Study: 1st year grad

Undergrad University: UC Berkeley

Degree Completed: B.S. in [EECS](#)

Research Interests: context-aware, ubiquitous computing systems

Birthday: May 7

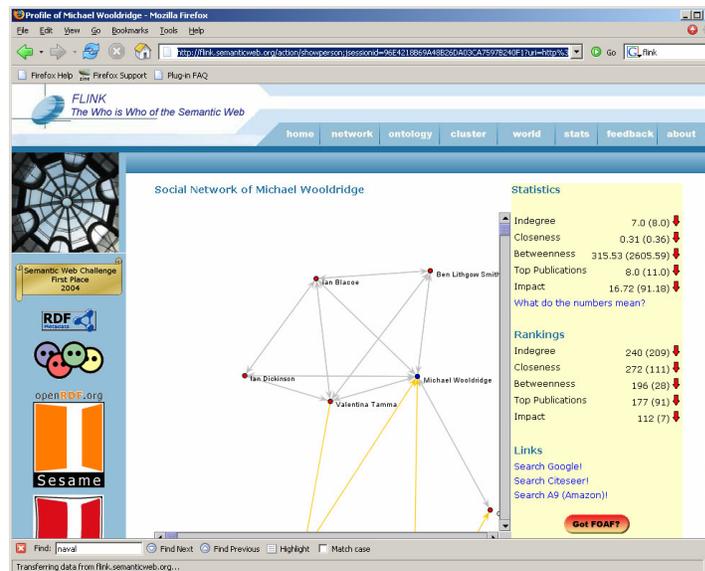
Sports Played:

Hobbies:

Other Interests:

(see <http://www.cs.washington.edu/research/semweb/index.html>)

Example Semantic Services – “Flink”



(see <http://flink.semanticweb.org/index.jsp>)

Conclusion

- HTML/XHTML has humans in mind
- Both XML and RDF add _____ to data, making more amenable to _____ interpretation
- Any one application could always just use XML with a fixed Schema
 - But RDF allows a much more flexible approach
 - Hopefully enabling data to be re-used in unexpected ways
 - “Semantic Web” is all about making smart apps to exploit this data
- Key Challenge – who will make the RDF?
 - Idea 1 – make it rewarding for normal people
 - Idea 2 – smart software to “scrape” it off the web