

STRATEGIC WEBSITE TECHNOLOGIES

LECTURE 11 (XML)

What is XML?

- eXtensible Markup Language
- Derived from SGML (Standard Generalized Markup Language)
 - Mid 1980s
 - An international standard (ISO 8879)
 - Considered as the “grand daddy” of all mark up languages.
 - Captures the document structure rather than the presentation information.
 - XML - addresses SGML's shortfalls and removes complications.
- W3C Standard
- Text File

Why XML?

- Structured Data.
- Searching Sites.
- Data Exchange.
- Information content can be richer and easier to use.

XML vs. HTML

| XML | HTML |
|--|--|
| <ul style="list-style-type: none">- Semantics.- User-defined tags.- Content and Style are separated.- Uses browser-side scripts.- Multi-directional and powerful links.- Optional user-defined DTD to validate. | <ul style="list-style-type: none">- Style.- Pre-defined tags.- Content and Style are in the same file.- Uses server-side scripts.- Unidirectional links. |

XML Syntax

Plain text file:

```
Turan Soylemez
```

No meta information.

HTML file:

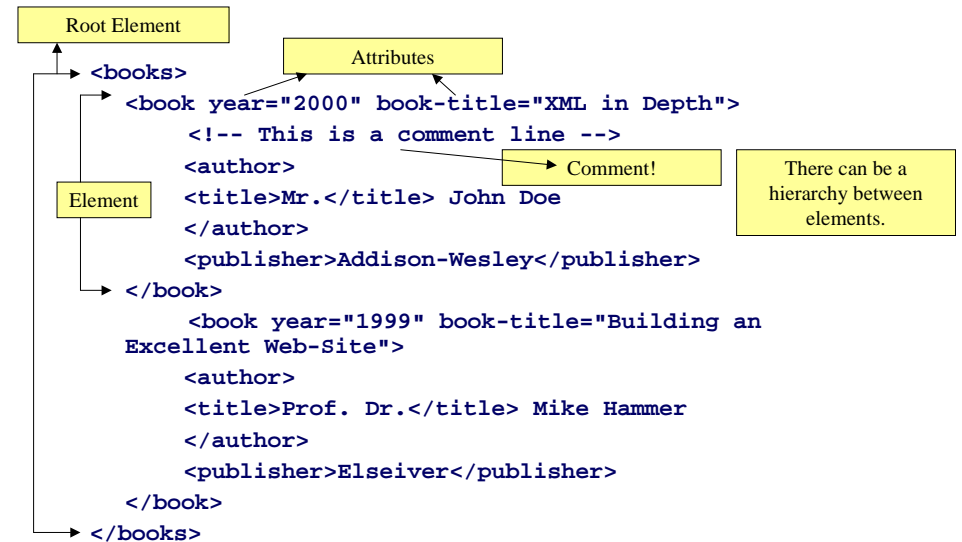
```
<HTML> <HEAD>  
<TITLE> Name </TITLE>  
</HEAD> <BODY>  
<P> Turan Soylemez </P>  
</BODY>
```

Restricted meta information using pre-defined tags.

XML file:

```
<name>  
  <first> Turan </first>  
  <last> Soylemez </last>  
</name>
```

Free use of tags.



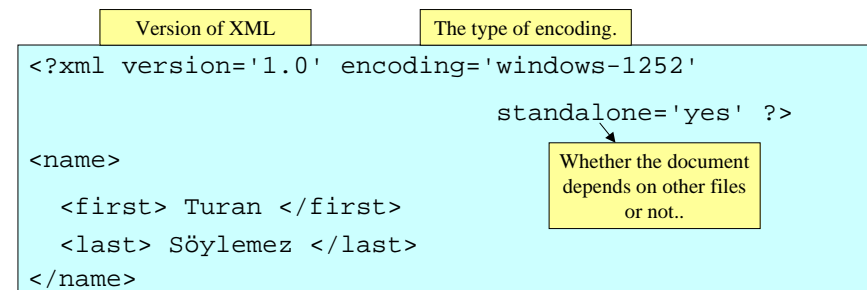
There can be a hierarchy between elements.

Rules (Well-Formed XML)

- Every start-tag must have an end-tag.
- Tags can't overlap.
- Everything between start-tag and end-tag (including the tags) is called an element.
- Elements can have a hierarchy.
- XML documents can have only one root element.
- Element names must obey XML naming conventions.
- Attributes must have values (in **quotes**).
- XML is case sensitive.
- XML will keep white space in your text. (This cannot be seen through IE5).

XML Declaration

In order to state that a document is written in XML <?XML tag is used in the first line.



Whether the document depends on other files or not..

The escape character (&)

The escape character is & in XML. For example,

- & → the & character
- < → the < character
- > → the > character
- ' → the ' character
- " → the " character
- © → the © character

Unicode numbers for some of the Turkish letters.

350 → Ş
351 → ş
286 → Ğ
287 → ğ
304 → İ
305 → ı

Cascading Style Sheets and XML

It is possible to use CSS with XML.

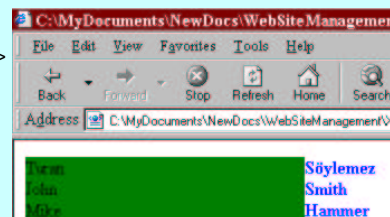
We use `<?xml-stylesheet type=mimetype href=stylesheetURL>` tag for this.

This should be after XML declaration line and before the first element of the XML document.

The CSS can be written like it is written for HTML. One important difference, however, there will not be any class definitions (like `H1.warning`) in here.

```
<?xml version='1.0' encoding='windows-1252'
      standalone='yes' ?>
<?xml-stylesheet type="text/css" href="players1.css" ?>
<players>
  <player>
    <firstname> Turan </firstname>
    <lastname> Söylemez </lastname>
  </player>
  <player>
    <firstname> John </firstname>
    <lastname> Smith </lastname>
  </player>
  <player>
    <firstname> Mike </firstname>
    <lastname> Hammer </lastname>
  </player>
</players>
```

Players1.xml



Players1.css

```
firstname {color:black; width:250px; background:green}
lastname  {color:blue; font:bold; width:250px}
```

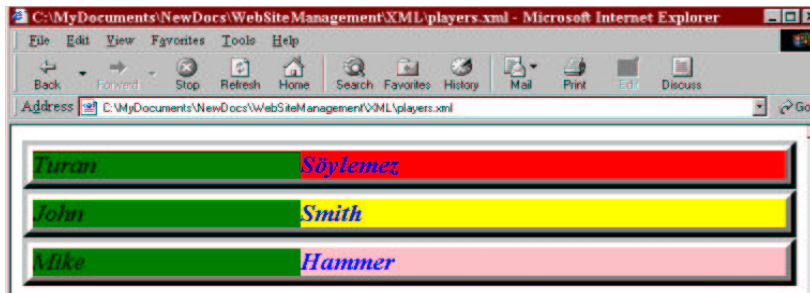
```
<?xml version='1.0' encoding='windows-1252'
      standalone='yes' ?>
<?xml-stylesheet type="text/css" href="players.css" ?>
<players>
  <captain>
    <player>
      <firstname> Turan </firstname>
      <lastname> Söylemez </lastname>
    </player>
  </captain>
  <player>
    <firstname> John </firstname>
    <lastname> Smith </lastname>
  </player>
  <goalkeeper>
    <player>
      <firstname> Mike </firstname>
      <lastname> Hammer </lastname>
    </player>
  </goalkeeper>
</players>
```

Players.xml

Players.css

Each player is represented in a different block (paragraph).

```
players {display:block; background:yellow}
player {display:block; border-width:10px;
        border-style:outset; font-style:italic;
        font-size:150%;}
captain player {display:block; background:red}
goalkeeper {display:block; background:pink}
firstname {color:black; width:250px;
           background:green}
lastname {color:blue; font:bold; width:250px}
```



Extensible Style Sheet Language (XSL)

XSL can be used as an alternative to CSS.

XSL is more flexible and conforming with XML standards.

2 types of XSL exist:

1. XSL Formatting Objects (XSL-FO)
2. XSL Transformation Language (XSLT)

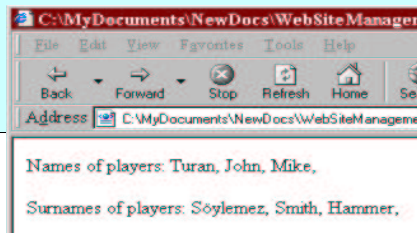
By using XSL Transformation (XSLT) it is possible to transform documents easily. (This is specially useful in B2B applications and presenting the same document for different purposes).

Change second line of players.xml as follows:

```
<?xml-stylesheet type="text/xsl" href="players.xsl" ?>
```

Players.xsl

```
<?xml version="1.0"?>
<x:stylesheet xmlns:x="http://www.w3.org/TR/WD-xsl">
  <x:template match="/">
    <p> Names of players:
      <x:for-each select="/players//player/firstname">
        <x:value-of /> ,
      </x:for-each>
    </p>
    Surnames of players:
    <x:for-each select="/players//player/lastname">
      <x:value-of /> ,
    </x:for-each>
  </x:template>
</x:stylesheet>
```



Reaching inside XML (DOM and SAX)

It is possible to reach information contained in an XML documents using the Document Object Model (DOM).

We can write JavaScript or VBScript code that processes a given XML document and changing it if necessary.

There is also an API called SAX which allows writing applications (specifically in Java) that can process XML documents.

(This practically means we do not have to write XML parsers)

DTD

- Document Type Definition.
- Basic use: Validate XML Document against it.
- What you can include in your document.
- Defines structure and vocabulary.
- Optional.
- A DTD can be
 - Internal.
 - External.

DTD Example

```
<?XML encoding="UTF-8"?>
<!-- This is an XML DTD -->
<!ELEMENT books (book*)>

<!ELEMENT book (author|(co-author,co-author+),publisher)>
<!ATTLIST book year CDATA #REQUIRED
              book-title CDATA #REQUIRED>
<!ELEMENT author (title?,#PCDATA)>

<!ELEMENT co-author (title?,#PCDATA)>

<!ELEMENT publisher (#PCDATA)>

<!ENTITY pub "Prentice Hall">
```

Linking XML Documents (XLink and XPointer)

By using the XLink and XPointer technologies it is possible to link XML documents to other documents.

The link can be simple link or it can be a multi-resource multi-directional link.

XLink and XPointer are not fully complete yet but several specifications have been given by the W3C.

```
<?xml version='1.0' encoding='windows-1252' standalone='no' ?>
<!DOCTYPE name SYSTEM "name3.dtd">
<person xmlns:xlink="http://www.w3.org/1999/xlink">
  <name>Turan Soylemez</name>
  <picture xlink:type="simple" xlink:href="resim.gif"
           xlink:actuate="onRequest"
           xlink:show="embed" xlink:tite="resim">
    Click me to see a picture
  </picture>
</person>
```

XML-Related Technologies

- X-Link and X-Pointer.
- XHTML.
- X-Path.
- XML Query Languages:
 - XML-QL.
 - XQL.
 - Quilt.
- DOM.
- XSL and XSLT.

An Example: Using ConText

- XML document 'InsuranceClaim'

```
<?xml version="1.0"?>
<InsuranceClaim>
  <ClaimID>12345</ClaimID>
  <Settlements>
    <Payment>
      <Payee>Borden Real Estate</Payee>
      <Date>12-OCT-1998</Date>
      <Amount>200000</Amount>
      <Approver>JCOX</Approver>
    </Payment>
  </Settlements>
  <DamageReport>
    A massive <Cause>Fire</Cause> ravaged the
    building and <Casualties>12</Casualties> people
    were killed. Early FBI reports indicate that
    <Motive>arson</Motive> is suspected.
  </DamageReport>
  ...
</InsuranceClaim>
```

An Example: Using ConText (Cont.)

- Query: "How much money has Jim Cox approved to date in settlement payments for arson-related fire claims?"

```
SELECT SUM(Amount)
FROM Claim_Header ch,
     Claim_Settlements cs,
     Claim_Settlement_Payments csp
WHERE csp.Approver='JCOX'
AND
CONTAINS (DamageReport, 'Arson WITHIN Motive') > 0
AND
CONTAINS (DamageReport, 'Fire WITHIN Cause' ) > 0
AND . . . /* Join Clauses */
```

Conclusion

- XML is a very powerful tool on the web.
- XML is fairly new but growing fast.
- Lots of recommendations and drafts but few of them are fully implemented.

References

- XML tutorial <http://msdn.microsoft.com/xml/tutorial/default.asp>
- XML FAQ <http://www.ucc.ie/xml/#FAQ-ACRO>
- General XML info page <http://www.oasis-open.org>
- WWW consortium <http://www.w3.org>
- XML dtd <http://www.dtd.com>.
- Xml-ql <http://www.w3.org/TR/NOTE-xml-ql/>
- Quilt http://www.almaden.ibm.com/cs/people/chamberlin/quilt_euro.html
- XQL FAQ <http://www.ibiblio.org/xql/>
- XML links <http://www.xml.com/>