ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Εισαγωγή στην Επιστήμη και Τεχνολογία των Υπηρεσιών

Ενότητα 5: Document Type Definitions (DTDs) - 2

Χρήστος Νικολάου
Τμήμα Επιστήμης Υπολογιστών

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται στην άδεια χρήσης **Creative Commons** και ειδικότερα

Αναφορά – Μη εμπορική Χρήση – Όχι Παράγωγο Έργο v. 3.0

(Attribution – Non Commercial – Non-derivatives)

- Εξαιρείται από την ως άνω άδεια υλικό που περιλαμβάνεται στις διαφάνειες του μαθήματος, και υπόκειται σε άλλου τύπου άδεια χρήσης. Η άδεια χρήσης στην οποία υπόκειται το υλικό αυτό αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

XML
Document Type Definitions (DTDs) - 2
605.444 / 635.444

David Silberberg

Lecture 4

ELEMENT Content

- Elements that contain child elements
- Examples of XML fragments
  - Example 1

```
<name title="Mr.">
        <first>Sam</first>
        <middle>Ta</middle>
        <middle>Chuan</middle>
        <last>Chu</last>
</name>
```

  - Example 2

```
<item item_no="I-5" supplier_id="S-1">
        <description>Ivory Soap</description>
        <in_stock>50</in_stock>
        <price>1.09</price>
        <cost>.28</cost>
</item>
```

ELEMENT Content (cont.)

- Text appears only within the element tags
- Only elements appear within the <name>…</name> and <item> … </item> tags
- DTD specifications
  - Example 1
    - <!ELEMENT name (first, middle, middle, last)>
  - Example 2
    - <!ELEMENT item (description, in_stock, price, cost)>
    - Note: attributes are dealt with later
  - Examples are simple cases

Sequence Lists

- List of all child elements in the sequence that they are to appear
- Comma separated elements
- Example

```
<name title="Mr.">
    <first>Sam</first>
    <middle>Ta</middle>
    <middle>Chuan</middle>
    <last>Chu</last>
</name>
```

- DTD specification

```
<!ELEMENT name (first, middle, middle, last)>
```

Sequence List Full Specification

<!ELEMENT name (first, middle, middle, last)>

<!ELEMENT first (#PCDATA)>
<!ELEMENT middle (#PCDATA)>
<!ELEMENT last (#PCDATA)>

Choice Lists

- Choice of child elements that may appear
- Bar "|" separated elements
- Example

```
<address>
        <street>123-A Kensington Circle</street>
        <city>London</city><country>England</country>
</address>
OR
<address>
        <street>11100 Johns Hopkins</street>
        <city>Baltimore</city><state>MD</state>
        <zip>21207</zip
</address>
```

- <!ELEMENT address (street, city, (country | (state, zip)))>

Full Choice List Specification

<!ELEMENT address (street, city, (country | (state, zip)))>

<!ELEMENT street (#PCDATA)>
<!ELEMENT city (#PCDATA)>
<!ELEMENT country (#PCDATA)>
<!ELEMENT state (#PCDATA)>
<!ELEMENT zip (#PCDATA)>

Cardinality Operators

- None
  - one and only one instance of the element is allowed
- **?**
  - 0 or 1 child element is allowed
- *****
  - 0 or many child elements are allowed
- **+**
  - 1 or many child elements are allowed
  - At least one must be specified

Cardinality Examples (1)

- Example 1

```
<customer>
    <name> … </name>
    <address> … </address>
    <purchase/>                    <!-- zero or more -->
</customer>
```

- DTD

```
<!ELEMENT customer  (name, address, purchase*)>

<!ELEMENT name (#PCDATA)>
<!ELEMENT address (#PCDATA)>
<!ELEMENT purchase (#PCDATA)>
```

Cardinality Examples (2)

- Example 2

```
<address>
    <street> … </street>        <!-- one or more -->
    <city> … </city>
    <country> … </ country >
</address>
```

- DTD

```
<!ELEMENT address (street+, city, (country | (state, zip)))>

<!ELEMENT street (#PCDATA)>
<!ELEMENT city (#PCDATA)>
<!ELEMENT country (#PCDATA)>
<!ELEMENT state (#PCDATA)>
<!ELEMENT zip (#PCDATA)>
```

Mixed Content

- PCDATA and elements are mixed
- Example (not from 'store' example)

```
<institution>
    The Johns Hopkins University is a premier institution
    <name>Johns Hopkins University</name>
    <state>Maryland</state>
</institution>
```

  - <!ELEMENT institution (#PCDATA | name | state)* >
- List must contain only PCDATA and child elements
  - **#PCDATA must be the first item in the list**
- Not as expressive as you want
  - Cannot specify: state+, etc.
  - Cannot insure all elements are there

Avoid Mixed Content When Possible

- When possible, try this:

```
<institution>
    <introtext>
        The Johns Hopkins University is a premier institution
    </introtext>
    <name>Johns Hopkins University</name>
    <state>Maryland</state>
</institution>
```

- Define it this way:

```
<!ELEMENT institution (introtext, name, state) >
<!ELEMENT introtext (#PCDATA) >
```

Attribute Declarations

- Used to define the attributes associated with elements
- Attribute names must be legal NameChar
- DTDs allow attribute defaults to be specified
  - When attribute is optional, but not present

Attribute Types

- CDATA
  - Character data
- Enumerated data
  - List of valid values
- ID
  - Attribute value is unique identifier for the element
- IDREF
  - A reference to the element with and ID attribute with the same value
- IDREFS
  - A list of IDREFs delimited by white space

Attribute Types (cont.)

- NMTOKEN
  - A name token, which is a string that conforms to name rules
- NMTOKENS
  - List of NMTOKENs with white space separators
- ENTITY
  - Name of a predefined entity
- ENITITIES
  - List of entities with white space delimiters
- NOTATION
  - Notation type that is defined elsewhere in DTD

CDATA Attribute Type

- Syntax

  <!ATTLIST element_name attribute CDATA attr_default>

  – **element_name** - element name with which the attribute is associated

  – **attribute** - the attribute name

  – **CDATA** - character data

  – **attr_default**

    - #REQUIRED - attribute must always appear with element
    - #IMPLIED - attribute is optional
    - #FIXED - attribute is optional; if appears, it must equal default value. If not, parser may supply default value.
    - Default value(s) - attribute is optional; if appears, it must equal one of the default values. If not, parser may supply default value.

CDATA Example

- Example 1

  <purchase date="01-02-2002" … />

- #REQUIRED – attribute must be specified

  <!ELEMENT purchase EMPTY>

  <!ATTLIST purchase date CDATA #REQUIRED

  … … … …>

- #IMPLIED – attribute is optional

  <!ELEMENT purchase EMPTY>

  <!ATTLIST purchase date CDATA #IMPLIED

  … … … …>

- #FIXED – provides a default value

  <!ATTLIST compiler version CDATA #FIXED "2.02"

  … … … …>

- Default value – parser supplies the value if not there

  <!ATTLIST customer ctype CDATA "good">

Enumerated Values

- Example

```
<name title="Dr.">
    <first>Darsana</first>
    <last>Sudarsen</last>
</name>
```

- DTD

```
<!ELEMENT name (first, middle*, last) >
<!ATTLIST name title (Dr. | Mr. | Ms. | Miss) "Ms.">
<!ELEMENT first (#PCDATA)>
<!ELEMENT middle (#PCDATA)>
<!ELEMENT last (#PCDATA)>
```

ID / IDREF / IDREFS

- IDs are unique identifiers of elements
- Values must be NameChars
  - Starts with letter, “_”, or “:”
- Values must be unique within a document
- Attributes must be #REQUIRED or #IMPLIED
- There is a one ID per element restriction
  - Useful for implementing indexes (e.g., XQuery)
  - Unfortunately, there is no way to define other "indexed" attributes

ID

- Example

```
<supplier sid =“S-1”>
    <company>Proctor and Gamble</company>
    <telephone>1-800-PAMPERS</telephone>
</supplier>
```

- DTD

```
<!ELEMENT supplier (company, telephone) >
<!ATTLIST supplier sid ID #REQUIRED>
<!ELEMENT company (#PCDATA)>
<!ELEMENT telephone (#PCDATA)>
```

IDREF

- Example

```
<item item_no="I-34" supplier_id="S-1">
    <description>Bounty Paper Towels</description>
    <in_stock>112</in_stock>
    <price>1.48</price>
    <cost>.57</cost>
</item>
```

- DTD

```
<!ELEMENT item (description, in_stock, price, cost) >
<!ATTLIST item item_no ID #REQUIRED
                supplier_id IDREF #REQUIRED>
<!ELEMENT description (#PCDATA)>
<!ELEMENT in_stock (#PCDATA)>
<!ELEMENT price (#PCDATA)>
<!ELEMENT cost (#PCDATA)>
```

IDREFS

- Example

  <purchase date="01-12-2002" items="I-34 I-62 I-15" … />

- DTD

  <!ELEMENT purchase EMPTY >

  <!ATTLIST purchase

  date CDATA #IMPLIED

  items IDREFS #REQUIRED

  …>

NMTOKEN / NMTOKENS

- Types of data constraints
  - CDATA - any string is valid
  - enumerated type - only the set of enumerated values are valid
- NMTOKEN provides something in between
  - Permits only NameChar characters
  - Not limited to finite specifications
  - Multiple NMTOKENs are specified by NMTOKENS
  - NMTOKENS values are white space delimited

NMTOKENS

- Example

  <purchase date="01-12-2002" items="I-34 I-62 I-15" qty="3 1 1" />

- DTD

  <!ELEMENT purchase EMPTY >

  <!ATTLIST purchase

  date CDATA #IMPLIED

  items IDREFS #REQUIRED

  qty NMTOKENS #REQUIRED >

Entities

- There are different types of entities in XML
  - XML document entities
  - Parameter entities
    - Defined in DTDS
    - Can be internal or external
    - Two types
      - Parsed entities - internal or external
      - Unparsed entities - external only
- Parsed entities
  - <!ENTITY JHU “The Johns Hopkins University”>
  - In XML file:

    <description>University attending: &JHU;</description>

Entity References

- Parsed entities

  <!ENTITY JHU “The Johns Hopkins University”>

  – In XML file:

    <description>University attending: &JHU;</description>

  – In DTD file:

    <!ENTITY JHUAddress “&JHU; 11100 Johns Hopkins Road”>

  – Built-in values

    - & - ampersand
    - < - less than
    - > - greater than

  – Character references

    <!ENTITY copy “©”>

    <!ENTITY copyright “© All rights reserved”>

Bad References

- Self references are not allowed

  <!ENTITY bad1 “Cannot refer to itself with &bad1;”>

- Indirect self references

  <!ENTITY bad2 “Cannot refer to another &bad3; that refers back to itself” >

  <!ENTITY bad3 “Cannot refer to another &bad2; that refers back to itself” >

  – Forward (acyclic) references are allowed

Parameter Entities

- Used for parsed entities that are parameters
- Example

```
<!ENTITY % RequiredNameToken “NMTOKENS #REQUIRED” >

<!ELEMENT purchase EMPTY >
<!ATTLIST purchase
                date CDATA #IMPLIED
                items IDREFS #REQUIRED
                qty %RequiredNameToken; >
```

store.dtd (1)

```
<!ELEMENT store (cust_list, inventory_list,
   supplier_list) >

<!ELEMENT cust_list (customer*) >

<!ELEMENT customer (name, address, purchase*) >
<!ATTLIST customer
   cid    ID        #REQUIRED
   ctype  NMTOKEN  #IMPLIED>
```

store.dtd (2)

```
<!ELEMENT name (first, middle*, last) >

<!ATTLIST name title (Dr. | Mr. | Ms. | Miss)
   "Ms.">
<!ELEMENT first (#PCDATA) >
<!ELEMENT middle (#PCDATA) >
<!ELEMENT last (#PCDATA) >

<!ELEMENT address (street+, city, (country |
   (state, zip)))>
```

store.dtd (3)

```
<!ELEMENT street   (#PCDATA)>
<!ELEMENT city     (#PCDATA)>
<!ELEMENT country  (#PCDATA)>
<!ELEMENT state    (#PCDATA)>
<!ELEMENT zip      (#PCDATA)>

<!ELEMENT purchase EMPTY >
<!ATTLIST purchase
   date   CDATA      #IMPLIED
   items  IDREFS     #REQUIRED
   qty    NMTOKENS   #REQUIRED >
```

store.dtd (4)

```
<!ELEMENT inventory_list (item*) >

<!ELEMENT item (description, in_stock, price,
   cost) >
<!ATTLIST item
   item_no       ID      #REQUIRED
   supplier_id  IDREF  #REQUIRED >

<!ELEMENT description (#PCDATA)>
<!ELEMENT in_stock    (#PCDATA)>
<!ELEMENT price       (#PCDATA)>
<!ELEMENT cost        (#PCDATA)>
```

store.dtd (5)

```
<!ELEMENT supplier_list (supplier*) >

<!ELEMENT supplier (company, telephone) >
<!ATTLIST supplier
   sid ID      #REQUIRED>

<!ELEMENT company    (#PCDATA)>
<!ELEMENT telephone  (#PCDATA)>
```

Τέλος Ενότητας