

Document Type Definiton (DTD)



Sadržaj

- **Razlozi za korišćenje DTD-a**
- **Dobro formiran nasuprot valjanom dokumentu.**
- **Gramatička pravila za XML dokument koristeći DTD.**
- **Ograničenja DTD-a.**



Dobro formiran XML

- Sadrži opcionu prvu liniju, obavezna ukoliko kodiranje znakova (atribut encoding) nije UTF-8 ili UTF-16.
`<?xml version="1.0"?>`
- Svaki početni kontrolni kod mora da ima završni kontrolni kod.
 - `<tag>data</tag>`
- Definisanje atributa pridruženih početnom kontrolnom kodu sa obaveznim dodeljenim vrednostima koje su pod navodnicima.
`<tag attribute="x">data</tag>`
- Pravilno ugnježdavanje elemenata.
 - `<employee>`
 - `<name>John Smith</name>`
 - `<id>X04913</id>`
 - `</employee>`
- ... i jedinstven osnovni element kao i XML-ove konvencije o imenima elemenata.
- Ovo su osnovna pravila koja se tiču structure XML dokumenta.



Zašto nam je potreban DTD?

- Šta ako uvedemo dodatna ograničenja.
`<message urgent="yes">`
 `<greeting>hi</greeting>`
 `<farewell>bye</farewell>`
`</message>`
- Mogu postojati samo 2 specijalna elementa (deca elementi): "greeting" i "farewell".
- Dete element "greeting" mora prethoditi "farewell" elementu.
- Poruka može sadržati opcioni "urgent" atribut.
- Šta ako hoćemo da definišemo i objavimo strukturu na koju se XML dokument odnosi?
- Šta ako hoćemo da računar "proveri" da li XML dokument ispunjava postavljena ograničenja?



Šta je DTD?

- **Nacrt strukture dokumenta.**
Sadrži niz deklaracija.
- **DTD**
 - može biti zaseban fajl XML dokumenta.
 - može biti ugrađen u XML fajl.
- **DTD definiše:**
 - elemente koji se mogu ili moraju pojaviti.
 - koliko često se elementi mogu pojaviti.
 - kako se ugnježdavaju elementi.
 - dozvoljene, zahtevane I standardne attribute.
- **Pažnja: Upotreba DTD-a nije obavezna.**
- **Za XML koji poštuje pravila DTD-a se kaže da je valjan.**



Šta je dozvoljeno u DTD-u?

- **Deklaracija tipa elementa <!ELEMENT...>**
 - Sintaksa koja formalno opisuje tip elementa i kog tipa su podaci koje može da sadrži. Osnovni oblik je <!ELEMENT ime (model_sadržaja)>, gde je "ime" tip elementa, a "model_sadržaja" je tip podatka koji se može pojaviti u datom elementu.
- **Lista atributa <!ATTLIST...>**
 - Lista atributa elementa omogućava grupisanje svih srodnih atributa elementa. Svi elementi moraju imati svoje attribute nabrojane u listi atributa.
 - Deklaracija atributa – sintaksa za formalno deklarisanje tipa atributa i kog tipa su podaci koje može da sadrži. Osnovni oblik je <attribute ime tip-podataka default> gde je tip-podataka, tip podatka koji atribut može sadržati i default je podrazumevana vrednost atributa. Definicije za svaki atribut iz liste atributa moraju da se navedu.



XML i DTD primer

```
<?xml version='1.0'?>
<address>
  <name>
    <title>Mrs.</title>
    <first-name>Mary</first-name>
    <last-name>McGoon</last-
name></name>
    <street>1401 MainStreet</street
><city>Sheboygan</city>
    <state>WI</state>
    <zip>38472</zip>
    <country>USA</country>
  </address>
```

```
<!ELEMENT address (name+, street+,
                    city, state, zip?, country)>
<!ELEMENT name (title?, first-name,
                last-name)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT first-name (#PCDATA)>
<!ELEMENT last-name (#PCDATA)>
<!ELEMENT street (#PCDATA)>
<!ELEMENT city (#PCDATA)>
<!ELEMENT state (#PCDATA)>
<!ELEMENT zip (#PCDATA)>
<!ELEMENT country (#PCDATA)>
```



Šta je dozvoljeno u deklarisanju tipa Element

- **Syntaksa:**
<!ELEMENT elementName (contentModel)**>**
- **Deklaracija tipa element u DTD-u i odgovarajući element u XML dokumentu:**
- **Deklaracija (DTD):**
<!ELEMENT greeting (#PCDATA)**>**
- **Odgovarajući validni XML fragmenti:**
<greeting>Hello, World!**</greeting>**
<greeting>
 <![CDATA[G'day!]]>
</greeting>



Modeli sadržaja elementa

- **Sadržaj elementa je opisan u specifikaciji sadržaja**
- **Tipovi DTD modela sadržaja**
 - EMPTY
 - ANY
 - Elements
 - Mixed
 - Text content (#PCDATA)



EMPTY model sadržaja

- Element ne sme da sadrži nikakve podatke (ni tekst ni elemente decu). Dozvoljeni su samo atributi elemenata.
- Deklaracija (DTD):
<!ELEMENT placeholder EMPTY>
- Valjani XML:

<placeholder></placeholder>

ili

<placeholder/>



ANY model sadržaja

- Može da sadrži "ANY" (bilo koje) dobro formirane XML podatke. Elementi koje koristimo moraju biti deklarirani u DTD-u.
- Deklaracija (DTD):
<!ELEMENT universe ANY>
<!ELEMENT galaxy (#PCDATA)>
- Valjani XML fragmenti:
<universe/> or **<universe></universe>**
<universe>the whole universe</universe>
<universe>
<galaxy>galaxy1</galaxy>
</universe>



Elements model sadržaja

- U “particles” modelu sadržaja su imena elemenata predstavljena sa a i b, a njihova interakcija ispod:

<!ELEMENT name (particle structure) >

- | Particle | Sintaksa |
|----------------|--|
| sekvenca | <code><!ELEMENT name (a,b)></code> |
| izbor | <code><!ELEMENT name (a b)></code> |
| jedan | <code><!ELEMENT name (a)></code> |
| jedan ili više | <code><!ELEMENT name (a)+></code> |
| nula ili više | <code><!ELEMENT name (a)*></code> |
| nula ili jedan | <code><!ELEMENT name (a)?></code> |

- Pažnja: a ili b mogu biti složena struktura, kao na primer:

a = (c,d)



Elements model sadržaja – Primeri (1 of 3)

- Deklaracija:
<!ELEMENT person((fname,lname) | (lname,fname))**>**
<!ELEMENT fname (#PCDATA)**>**
<!ELEMENT lname (#PCDATA)**>**
- Valjani XML:
<person>
 <lname>Smith**</lname>**
 <fname>John**</fname>**
</person>
- Takođe valjani XML:
<person>
 <fname>John**</fname>**
 <lname>Smith**</lname>**
</person>



Elements model sadržaja – Primeri (2 of 3)

- Deklaracija:

<!ELEMENT order (order-item+,delivery-address,order-date?)>

<!--Deca elementi definisani tako da sadrže #PCDATA **-->**

- Valjani XML fragmenti:

<order>

<order-item>item1**</order-item>**

<delivery-address>123 State Street**</delivery-address>**

</order>

<order>

<order-item>item3**</order-item>**

<order-item>item4**</order-item>**

<delivery-address>123 State Street**</delivery-address>**

</order>

<order>

<order-item>item5**</order-item>**

<order-item>item6**</order-item>**

<delivery-address>123 State Street**</delivery-address>**

<order-date>July 5, 2001**</order-date>**

</order>



Elements model sadržaja – Primeri (3 of 3)

- Deklaracija:

```
<!ELEMENT phonebook (page)+>  
<!ELEMENT page (heading, (entry | advert)+)>  
<!ELEMENT heading (#PCDATA)>  
<!ELEMENT entry (#PCDATA)>  
<!ELEMENT advert (#PCDATA)>
```

- Valjani XML fragment:

```
<phonebook>  
  <page>  
    <heading>The whole town</heading>  
    <entry>John Smith, 555-1212</entry>  
    <advert>Fred's Fish n' Chips - 123-4567</advert>  
  </page>  
</phonebook>
```

- Nevažeci XML fragmenti:

```
<phonebook><page><entry/><entry/></page></phonebook>  
<phonebook><page/></phonebook>
```



Mixed (mešani) model sadržaja

- Mešani sadržaj: elementi koji imaju znakovni sadržaj kombinovan sa decom elementima.
- Dva primera deklaracije:
<!ELEMENT product(#PCDATA)> ← znakovni sadržaj
<!ELEMENT review (#PCDATA|product)*>
znakovni sadržaj + dete element
- Valjani XML fragmenti:
<review>review text goes here**</review>**
<review>This is a review of some **<product>**car**</product>** that goes on for pages of regular text.**</review>**



Šta je dozvoljeno u deklarisanju atributa

- Lista parametara atributa
 - Koristi se za deklaraciju atributa elementa
- Sintaksa:

Opcija 1:

```
<!ATTLIST elementName  
        attributeName attributeType defaultDecl>
```

Opcija 2:

```
<!ATTLIST elementName  
        attributeName attributeType defaultDecl  
  
        attributeName attributeType defaultDecl>
```



Tipovi atributa

Tipovi atributa	Opis
Znakovni tip	
CDATA	Koristi se za deklarisanje atributa čija je vrednost proizvoljan znakovni sadržaj. Ovo je jedini tip atributa koji dozvoljava vrednosti atributa koje se ne nalaze u pravilima za imena u XML 1.0 gramatici.
Tokenizovan tip	
NMTOKEN	Koristi se za deklarisanje atributa čija definicija mora da se podudara sa definicijom za imena u XML 1.0.
NMTOKENS	Dozvoljava više NMTOKEN-a razdvojenih razmacima.
ID	Koristi se za deklarisanje atributa čija vrednost mora biti jedinstvena unutar XML dokumenta.
IDREF, IDREFS	Vrednosti atributa se moraju odnositi na ID vrednosti deklarisanu na drugom mestu u dokumentu. IDREFS? Pogledaj NMTOKENS
ENTITY	Koristi se za deklaraciju atributa čije vrednosti moraju odgovarati imenu deklarisanog entiteta (ENTITY).
ENTITIES	Dozvoljava više ENTITY imena razdvojenih razmacima.
Nabrojiv tip	
NOTATION	Odnosi se na <!NOTATION> deklaraciju u DTD-u.



Parametar podrazumevane vrednosti atributa

Parametar	Opis
#REQUIRED	Atribut mora biti prisutan
#IMPLIED	Atribut nije obavezan
attribute-value	(podrazumevana vrednost) ako vrednost atributa nije prisutna, procesor moze da ponudi podrazumevanu vrednost
#FIXED attribute-value	Ako je atribut prisutan, mora da ima navedenu podrazumevanu vrednost



Parametar podrazumevane vrednosti atributa PRIMERI

- Deklaracije:

<!ELEMENT shirt (#PCDATA)>

<!ATTLIST shirt type CDATA #REQUIRED>

<!ATTLIST shirt collar CDATA #IMPLIED>

<!ATTLIST shirt size (small | medium | large) "large">

<!ATTLIST shirt manufacturer CDATA #FIXED "Levi">

- Valjani XML:

<shirt type="short">cotton</shirt>

<shirt type="short" size="large">wool</shirt>

<shirt type="short" manufacturer="Levi">denim</shirt>

<shirt type="short sleeve" collar="button-down"></shirt>

- Neispravan XML:

<shirt></shirt>

<shirt type="short" size="medium large">cardigan</shirt>

<shirt type="short" manufacturer="Gap">designer</shirt>



Parametar opcione vrednosti atributa

- Ista informacija predstavljena korišćenjem drugog oblika parametra vrednosti atributa.

<!ELEMENT shirt (#PCDATA)>

**<!ATTLIST shirt size (small | medium | large)
"large"**

collar CDATA #IMPLIED

type CDATA #REQUIRED

manufacturer CDATA #FIXED "Levi">



Tipovi atributa: Tokenizovan tip IDREFS Primer

- Sintaksa:

<!ATTLIST elementName attributeName IDREF defaultDecl>

Deklaracija:

<!ELEMENT employee (#PCDATA)>

<!ATTLIST employee serialNumber ID #REQUIRED>

<!ATTLIST employee manager1 IDREF #IMPLIED>

<!ATTLIST employee manager2 IDREFS #IMPLIED>

- Valjani XML fragment:

<employee serialNumber="e00001">Joe Smith</employee>

<employee serialNumber="e00002">Bill Smith</employee>

<employee serialNumber="e00003" manager1="e00001">John Smith</employee>

<employee serialNumber="e00004" manager1="e00001" manager2="e00002 e00001">John Smith</employee>

- Neispravan XML fragment: **manager2="e00001 e00005"** ako e00005 nije element u okviru dokumenta.



DTD kao eksterni podskup

DTD i XML kao zasebni fajlovi

Ime fajla: hello.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE message SYSTEM "message.dtd">
<message>
  <greeting>Hello, World!</greeting>
  <farewell>Goodbye, World!</farewell>
</message>
```

Ime fajla: message.dtd

```
<!ELEMENT message (greeting, farewell)>
<!ELEMENT greeting (#PCDATA)>
<!ELEMENT farewell (#PCDATA)>
```



DTD kao interni podskup

DTD i XML kao združeni fajl:

Ime fajla: hello.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE message [
  <!ELEMENT message (greeting, farewell)>
  <!ELEMENT greeting (#PCDATA)>
  <!ELEMENT farewell (#PCDATA)>
]>
<message>
  <greeting>Hello, World!</greeting>
  <farewell>Goodbye, World!</farewell>
</message>
```



Provera validnosti dokumenta

- **Potrebno smestiti dokument i DTD-ove na javno dostupan web server**
- **Dostupni analizatori na mreži:**
- <http://www.stg.brown.edu/service/xmlvalid/>
- <http://www.cogsci.ed.ac.uk/~richard/xml-check.html>



Šta nije u redu sa DTD-om?

- **Veoma slaba tipiziranost podataka.**
 - **#PCDATA** može biti bilo koji tekstualni string (osim kontrolnog koda)
- **Sintaksna pravila DTD-a nisu usklađena sa sintaksnim pravilima XML-a.**
 - **<!ELEMENT zip (#PCDATA)>**
- **Postoje neka ograničenja koja DTD ne može tako lako da izrazi:**
 - **Element x može da se pojavi od 4 do 17 puta.**
- **XML schema (XML šema) rešava mnoga ograničenja DTD-a.**
 - **XML schema je sada pod preporukom W3C-a.**
 - **Podrška za W3C Schema je nešto novo.**
 - **Osnovne odlike:**
 - **XML sintaksa, restrikcije**



Prednosti XML Scheme

- **Podrška proste i složene tipove podataka**
- **Mogućnost izvođenja i nasleđivanja tipova**
- **Ograničenja zastupljenosti elemenata**
- **Deklaracije elemenata i atributa koje vode računa o prostorima imena (namespaces)**

