



DTD (Document Type Definition) eli dokumentin rakennemäärittely

Rakennemäärittely
Validi XML-dokumentti
Elementtien määrittely
Attribuuttien määrittely
Entiteettien määrittely

1



Rakennemäärittely

Dokumentin rakennemäärittely eli DTD kertoo, mikä on sallittua XML-dokumentissa. DTD on joukko sääntöjä, jotka kertovat

- käytettävissä olevat elementit
- elementtien alielementit, tyhjät elementit
- elementtien attribuutit,
- entiteetit sekä
- notaatiot.

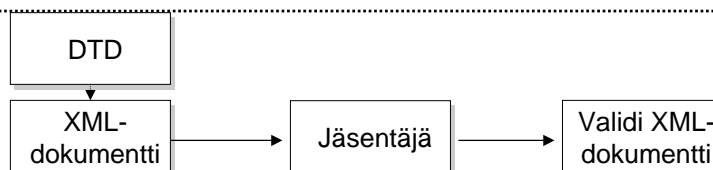
2

Tehtävä

- a. Onko XML-dokumentille pakko laatia DTD?
- b. Miksi XML-dokumentille laaditaan DTD?

3

Validi XML-dokumentti



XML-dokumentti on validi (valid), jos

- dokumentti on hyvin muodostunut (well-formed),
- dokumenttiin kuuluu rakennemäärittely ja
- dokumentti noudattaa rakennemäärittelyä.

XML Spy:ssä validointi tehdään F8-näppäimellä.

XML-dokumentti voidaan tehdä myös ilman DTD:tä. Dokumentti on silloin vain hyvin muodostunut (well-formed).

4

XML-dokumentti

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<!DOCTYPE henkilosto SYSTEM "henkilosto.dtd">
<henkilosto>
  <henkilo tunnus="mars">
    <hetu>123456-123A</hetu>
    <nimi>
      <sukunimi entinen="Lehtonen">Marttila</sukunimi>
      <etunimi>Sirpa</etunimi><etunimi>Anneli</etunimi>
    </nimi>
    <osoite>
      <postinumero>08680</postinumero>
      <postitoimipaikka>MUIJALA</postitoimipaikka>
    </osoite>
    <tiimi nimi="Tiedonhallinta" asema="jäsen"/>
    <email>sirpa.marttila@helia.fi</email>
    <puhelin type="työ">09-14890270</puhelin>
    <nimike>Lehtori</nimike>
    <tyohuone sijainti="Helia1">6036</tyohuone>
  </henkilo>
</henkilosto>
```

5

henkilosto.dtd DTD

<pre><!ELEMENT henkilosto (henkilo+)> <!ELEMENT henkilo (hetu, nimi, osoite?, tiimi?,email,puhelin*,nimike,tyohuone?)> <!ATTLIST henkilo tunnus ID #REQUIRED esimies IDREF #IMPLIED paikka NMTOKEN #FIXED "Helia" osaamisalue CDATA #IMPLIED> <!ELEMENT hetu (#PCDATA)> <!ELEMENT nimi ((sukunimi, etunimi+) (etunimi+, sukunimi))> <!ELEMENT sukunimi (#PCDATA)> <!ATTLIST sukunimi entinen NMTOKENS #IMPLIED> <!ELEMENT etunimi (#PCDATA)> <!ELEMENT osoite (lahiosoite?, postinumero, postitoimipaikka)></pre>	<pre><!ELEMENT lahiosoite (#PCDATA)> <!ELEMENT postinumero (#PCDATA)> <!ELEMENT postitoimipaikka (#PCDATA)> <!ELEMENT tiimi EMPTY> <!ATTLIST tiimi nimi NMTOKEN #REQUIRED asema (vetäjä jäsen) "jäsen"> <!ELEMENT email (#PCDATA)> <!ELEMENT puhelin (#PCDATA)> <!ATTLIST puhelin type (työ keskus gsm koti) #IMPLIED> <!ELEMENT nimike (#PCDATA)> <!ELEMENT tyohuone (#PCDATA)> <!ATTLIST tyohuone sijainti (Helia1 Helia2) "Helia1" ></pre>
---	---

6

DTD:n liittäminen XML-dokumenttiin

Ulkoinen DTD:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>  
<!DOCTYPE henkilosto SYSTEM "henkilosto.dtd">  
<henkilosto>  
</henkilosto>
```

Sisäinen DTD:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>  
<!DOCTYPE henkilosto [  
<!-- Tähän kirjoitetaan DTD -->  
>  
<henkilosto>  
</henkilosto>
```

7

Sekä sisäinen että ulkoinen DTD:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>  
<!DOCTYPE henkilosto SYSTEM "henkilosto.dtd" [  
<!-- Tähän täydennetään henkilosto.dtd:hen  
kirjoitettua DTD -->  
>  
<henkilosto>  
</henkilosto>
```

8

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
```

- ▼ standalone="no", kun
 - elementti käyttää oletusarvoista attribuuttia
 - attribuutin arvo muuttuu normalisoinnin seurauksena (esim. attribuutin tyyppinä on muuta kuin CDATA ja sisältönä merkityksettömiä merkkejä)
 - elementti sisältää alielementtejä ja elementin ja sen alielementtien välissä on white space -merkkejä
 - dokumentissa käytetään muuta kuin viittä valmiiksi määriteltyä entiteettiä

9

Elementtien määrittely

```
<!ELEMENT elementin_nimi sisältö>
```

elementistä määritellään sen nimi ja sisältö

sisältö voi olla

- alielementtejä (children content),
- alielementtejä sekä jäsenettyä merkkietoa (mixed content)
- jäsenettyä merkkietoa (PCDATA),
- tyhjä (EMPTY).

10

Alielementtejä

`<!ELEMENT elementin_nimi sisältö>`

alielementeille määritellään esiintymisjärjestys sekä lukumäärät

(A,B,C) esiintymisjärjestys
(A, (B | C)) vaihtoehtoinen esiintymisjärjestys

lukumäärät

+	1-n
*	0-n
?	0-1

11

`<henkilosto>`
`<henkilo>`
`</henkilo>`

`<henkilo>`
`</henkilo>`
`</henkilosto>`

`<henkilo>`
`<nimi>..</nimi>`
`<email>..</email>`
`<puhelin>..</puhelin>`
`<nimike>..</nimike>`
`<tyohuone>..</tyohuone>`
`</henkilo>`

`<!ELEMENT henkilosto (henkilo+)>`

`<!ELEMENT henkilo (nimi, osoite?,
tiimi?, email, puhelin*, nimike,
tyohuone?, ika?)>`

12

Tehtävä

a. Määrittele henkilötiedot-elementti.

```
<henkilotiedot>  
  <nimi>  
    <sukunimi>Majava</sukunimi>  
    <etunimi>Matti</etunimi>  
    <etunimi>Juhani</etunimi>  
  </nimi>  
</henkilotiedot>
```

b. Määrittele osio-elementti.

```
<osio>  
  <laajuus>2</laajuus>  
  <arvosana>4</arvosana>  
  <opettaja>Marttila Sirpa</opettaja>  
</osio>
```

13

Tekstisisältö, tyhjä

```
<!ELEMENT elementin_nimi sisältö>
```

#PCDATA tekstiä (merkkejä, merkkiviittauksia, entiteettejä)

EMPTY tyhjä

```
<sukunimi>Marttila  
</sukunimi>
```

```
<!ELEMENT sukunimi (#PCDATA)>
```

```
<tiimi nimi= "toteemi" />
```

```
<!ELEMENT tiimi EMPTY>
```

14

Tehtävä

```
<opintosuoritukset>
<oppilas>
  <henkilotiedot>
    <nimi>
      <sukunimi>Majava</sukunimi>
      <etunimi>Matti</etunimi>
      <etunimi>Juhani</etunimi>
    </nimi>
  </henkilotiedot>
  <suoritukset>
    <suoritus>
      <laajuus>12</laajuus>
      <arvosana>5</arvosana>
      <paivays />
    <osiot>
      <osio>
        <laajuus>2</laajuus>
        <arvosana>4</arvosana>
        <opettaja>Marttila S</opettaja>
      </osio>
```

```
<osio>
  <laajuus>6</laajuus>
  <arvosana>5</arvosana>
  <opettaja>Virkki O</opettaja>
</osio>
<osio>
  <laajuus>4</laajuus>
  <arvosana>5</arvosana>
  <opettaja>Marttila S</opettaja>
</osio>
</suoritukset>
</oppilas>
</opintosuoritukset>
```

- Täydennä DTD valmiiksi elementtien osalta.
- Miten voit liittää DTD:n XML-dokumenttiin?

15

Mixed content

```
<!ELEMENT elementin_nimi sisältö>
```

- elementti sisältää jäsenettyä merkkietoa ja/tai toisia elementtejä
- määrittelyn täytyy alkaa PCDATA:lla ja mahdolliset elementit yhdistetään tai-sanalla (|) ja lopussa täytyy olla *
- elementtien lukumäärä ja järjestystä ei voida määrätä, kuten ei merkkijonojenkaan määrää ja kirjoituskohtaa

```
<muistio>Kokouksessa
<paiva>1.4.2005</paiva>
päätettiin ...</muistio>
```

```
<!ELEMENT muistio (#PCDATA |
paiva)*>
```

16

Attribuuttien määrittely

```
<!ATTLIST elementin_nimi  
attribuutin_nimi tyyppi pakollisuus>
```

tyyppi voi olla

- CDATA, ID, IDREF, IDREFS, NMTOKEN, NMTOKENS, luettelo mahdollista arvoista, ENTITY, ENTITIES, NOTATION

pakollisuus voi olla

- #REQUIRED, #IMPLIED, oletusarvo, #FIXED

17

Attribuutin tyyppi

```
<!ATTLIST elementin_nimi  
attribuutin_nimi tyyppi pakollisuus>
```

CDATA	merkkijono
NMTOKEN	yksi sana
NMTOKENS	usea sana
ID	tunniste
IDREF	viittaus tunnisteeseen
IDREFS	viittaus useaan tunnisteeseen
luettelo mahdollista arvoista	
ENTITY	entiteetti
ENTITIES	usea entiteetti

18

<code><henkilo osaamisalue="XML, Java"></code>	<code><!ATTLIST henkilo osaamisalue CDATA #IMPLIED></code>
<p>▼ CDATA</p> <ul style="list-style-type: none">– attribuutin arvona on merkkijono, mikä voi olla tyhjä ("")	
<code><henkilo osaamisalue="XML"></code>	<code><!ATTLIST henkilo osaamisalue NMTOKEN #IMPLIED></code>
<p>▼ NMTOKEN</p> <ul style="list-style-type: none">– arvo on yksi sana, mikä ei voi olla tyhjä ("")– arvon noudatettava <i>nmtoken</i>-sääntöä: voi olla kirjaimia, numeroita, alaviivoja (_), tavuviivoja (-), pisteitä (.) tai kaksoispisteitä (:)– arvo ei voi sisältää tyhjiä merkkejä	

19

<code><henkilo osaamisalue="XML HTML"></code>	<code><!ATTLIST henkilo osaamisalue NMTOKENS #IMPLIED></code>
<p>▼ NMTOKENS</p> <ul style="list-style-type: none">– arvo on yksi tai useampi sana välilyönnillä erotettuna– arvon noudatettava <i>nmtoken</i>-sääntöä	

20

Tehtävä

- a. Määrittele opettaja -elementin tunnus-attribuutti.

```
<opettaja tunnus="marsu">Marttila Sirpa</opettaja>
```

- b. Määrittele opintosuoritus-elementin oppilaitos- ja päiväys-attribuutit.

```
<opintosuoritus oppilaitos="Helia" päiväys="8.3.2006">
```

- c. Määrittele suoritus-elementin tunnus- ja nimi-attribuutit.

```
<suoritus tunnus="ICT03D" nimi="Tieto ja tiedon  
varastointi">
```

21

```
<puhelin type="gsm">  
040-7654321</puhelin>
```

```
<!ATTLIST puhelin  
type (työ|keskus|gsm|koti) #IMPLIED>
```

▼ luettelo mahdollista arvoista

- sopii rajoitetun arvojoukon määrittelyyn
- luettelossa oleva arvo on yksi sana, joka noudattaa *nmtoken*-sääntöä
- sama arvo ei saa esiintyä samassa luettelossa useita kertoja

22

Tehtävä

- a. Määrittele laajuus-elementin yksikko-attribuutti luettelotyyppiseksi.

```
<laajuus yksikko="op">2</laajuus>
```

23

```
<henkilo tunnus="marsii"  
esimies="honte">  
<henkilo tunnus="honte">
```

```
<!ATTLIST henkilo  
tunnus ID #REQUIRED  
esimies IDREF #IMPLIED>
```

▼ ID

- yksilöi elementin
- elementillä voi olla vain yksi ID-tyyppinen attribuutti
- arvon noudatettava *name*-sääntöä: arvon täytyy alkaa kirjaimella tai alaviivalla (_). Muut merkit voivat olla kirjaimia, numeroita, alaviivoja (_), tavuviivoja (-), pisteitä (.) tai kaksoispisteitä (:)
- arvon täytyy olla yksikäsitteinen dokumentissa määritellyn elementin joukossa (yllä henkilö-elementtien joukossa)
- arvo ei voi olla tyhjä
- pakollisuus voi olla vain #REQUIRED tai #IMPLIED

24

<pre><henkilo tunnus="mars" esimies="honte"> <henkilo tunnus="honte"></pre>	<pre><!ATTLIST henkilo tunnus ID #REQUIRED esimies IDREF #IMPLIED></pre>
---	--

▼ **IDREF**

- viittaa samassa dokumentissa olevaan ID:n arvoon
- arvon noudatettava *name*-sääntöä

▼ **IDREFS**

- viittaa samassa dokumentissa olevaan yhteen tai useaan ID:n arvoon välilyönnillä erotettuna
- arvojen noudatettava *name*-sääntöä

25

Tehtävä

XML-dokumentissa voi olla usean oppilaan opintosuorituksia (alla yhden oppilaan).

```
<oppilas hetu="123456-1234">
  <suoritus tunnus="ICT03D" nimi="Tieto ja tiedon varastointi">
    <osiot>
      <osio tunnus="Johdanto">
        <opettaja tunnus="mars">Marttila Sirpa</opettaja>
      </osio>
      <osio tunnus="RDBMS">
        <opettaja tunnus="virou">Virkki Outi</opettaja>
      </osio>
      <osio tunnus="XML">
        <opettaja tunnus="mars">Marttila Sirpa</opettaja>
      </osio>
    </osiot>
  </suoritus>
</oppilas>
```

- Voiko hetu olla ID-tyyppinen?
- Voiko osion tunnus-attribuutti olla ID-tyyppinen?
- Voiko opettajan tunnus-attribuutti olla ID-tyyppinen?
- Voiko suorituksen tunnus-attribuutti olla ID-tyyppinen?

26

Attribuutin pakollisuus

```
<!ATTLIST elementin_nimi
  attribuutin_nimi tyyppi pakollisuus>
```

#REQUIRED	pakollinen
#IMPLIED	ei pakollinen
"oletusarvoinen"	

<pre><puhelin type="gsm"> 040-7654321</puhelin></pre>	<pre><!ATTLIST puhelin type (työ keskus gsm koti) #IMPLIED></pre>
<pre><tiimi nimi="Tiedonhallinta"/></pre>	<pre><!ATTLIST tiimi nimi NMTOKEN #REQUIRED asema (vetäjä jäsen) "jäsen"></pre>

27

<pre><tiimi nimi="Tiedonhallinta"/></pre>	<pre><!ATTLIST tiimi nimi NMTOKEN #REQUIRED asema (vetäjä jäsen) "jäsen"></pre>
<pre><tiimi nimi="Tiedonhallinta" asema="jäsen"/></pre>	

28

Tehtävä

Määrittele opintosuorituksiasi kuvaavalle XML dokumentille DTD.

```
<?xml version="1.0" encoding="ISO-8859-15"?>
<opintosuoritukset oppilaitos="Helia"
paivays="8.3.2006">
  <oppilas hetu="123456-1234">
    <henkilotiedot>
      <nimi>
        <sukunimi>Majava</sukunimi>
        <etunimi>Matti</etunimi>
        <etunimi>Juhani</etunimi>
      </nimi>
    </henkilotiedot>
    <suoritukset>
      <suoritus tunnus="ICT03D" nimi="Tieto
ja tiedon varastointi">
        <laajuus yksikko="op">12</laajuus>
        <arvosana>5</arvosana>
        <paivays paiva="20" kuukausi="12"
vuosi="2005"/>


```

```
<osiot>
  <osio tunnus="Johdanto">
    <laajuus yksikko="op">2</laajuus>
    <arvosana>4</arvosana>
    <opettaja tunnus="marsi">Marttila
Sirpa</opettaja>
  </osio>
  <osio tunnus="RDBMS">
    <laajuus yksikko="op">6</laajuus>
    <arvosana>5</arvosana>
    <opettaja tunnus="virou">Virkki
Outi</opettaja>
  </osio>
  <osio tunnus="XML">
    <laajuus yksikko="op">4</laajuus>
    <arvosana>5</arvosana>
    <opettaja tunnus="marsi">Marttila
Sirpa</opettaja>
  </osio>
</osiot>
</suoritus>
```

29

```
<suoritus tunnus="ATKA3D" nimi="Web-ohjelmointi">
  <laajuus yksikko="op">3</laajuus>
  <arvosana>3</arvosana>
  <paivays paiva="20" kuukausi="5" vuosi="2005"/>
  <opettaja tunnus="marsi">Marttila Sirpa</opettaja>
</suoritus>
</suoritukset>
</oppilas>
<oppilas hetu="123456-4321">
  <henkilotiedot>
    <nimi>
      <sukunimi>Jänis</sukunimi>
      <etunimi>Liisa</etunimi>
    </nimi>
  </henkilotiedot>
  <suoritukset>
    <suoritus tunnus="ATK28D" nimi="XML-dokumenttien
käsittely">
      <laajuus yksikko="op">4,5</laajuus>
      <arvosana>4</arvosana>
      <paivays paiva="20" kuukausi="5" vuosi="2005"/>
      <opettaja tunnus="marsi">Marttila Sirpa</opettaja>
    </suoritus>
  </suoritukset>
</oppilas>
</opintosuoritukset>
```

30



Entiteettien määrittely

Parametrientiteetti

31



Parametrientiteetti

```
<!ENTITY % nimi "sisältö">
```

- käytetään korvaustekstinä DTD:ssä
- parametrientiteettejä voidaan käyttää vain DTD:ssä
- parametrientiteettejä ei saa laittaa sisäiseen DTD:hen

```
<!ENTITY % PC "(#PCDATA)">  
<!ELEMENT sukunimi %PC;>
```



32

Tehtävä

Miten voisi hyödyntää suoritus.xml-dokumentin DTD:ssä parametrientiteettejä?