


XML Schema

Recommendation, 2.5.2001

1

DTD:n puutteita

- ▼ ei noudata XML-merkkausta
- ▼ elementtien ja attribuuttien tietotyyppitys on puutteellista
 - elementille voidaan määrittää sisällöksi #PCDATA, EMPTY, ANY
 - elementeille ei voida antaa tarkkoja lukumäärätietoja
 - attribuutille voidaan määrittää tietotyyppi rajallisesta joukosta (CDATA, ID, IDREF, ...)
 - elementeille ei voida antaa oletusarvoja
 - omia tietotyyppisiä ei voida tehdä
 - tietotyyppisiä ei voida periyttää



2


- ▼ ei tue nimiavaruuksia
 - uudelleenkäytettävyys
 - laajennettavuus
- => DTD on suunniteltu tekstidokumenttien kirjoittamisen tarpeisiin
- ▼ XML:n käyttöalueita paljon muitakin esim.
 - elektroninen kaupankäynti (tilaus-, tuote- ja laskutustieto)
 - tietokantojen tiedonsiirto
- => näissä tiedon täsmällisyydelle tarkkoja vaatimuksia

3

XML Schema

- ▼ W3C:n kehittämä rakennekuvaus XML-dokumenteille
 - ▼ on kehitetty korjaamaan DTD:n puutteita
 - syntaksina XML
 - elementtien ja attribuuttien tietotyyppitys
 - omien tietotyyppien tekeminen
 - nimiavaruudet
 - periytyminen
- => on tarkoitettu tietokantamaisen tiedon rakenteen kuvaamiseen


4




XML Scheman tekeminen

5

XML Schema



- ▼ XML Schema dokumentti koostuu
 - elementtimäärittelyistä (element)
 - attribuuttimäärittelyistä (attribute)
 - monimutkaisten elementtirakenteiden määrittelyistä (complexType)
 - ▼ elementillä on alielementtejä
 - ▼ elementti on tyhjä
 - ▼ elementillä on attribuutteja
 - elementtien ja attribuuttien tietosisällön määrittelyistä (simpleType)
 - ▼ valmiiden tai niistä johdettujen perustietotyyppien arvojen rajoittaminen



6

Scheman liittäminen XML-dokumenttiin

XML-dokumentti:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<henkilosto
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="henkilosto.xsd">
  <henkilo></henkilo>
</henkilosto>
```

XML Schema (henkilosto.xsd):

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
<xs:schema>
```

7

- ▼ XML-dokumentissa käytetty XML Scheman nimiavaruus
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 - nimiavaruus XML-dokumentissa käytettäville XML Scheman määrittäville (xsi:type, xsi:nil, xsi:schemaLocation, xsi:noNamespaceSchemaLocation)
 - nimiavaruuden lyhenne (xsi) on vapaasti valittavissa, yleensä lyhenteenä on xsi

- ▼ XML Schemassa käytetty nimiavaruus
xmlns:xs="http://www.w3.org/2001/XMLSchema"
 - nimiavaruus XML Scheman sanastolle ja tietotyypeille
 - nimiavaruuden lyhenne (xs) on vapaasti valittavissa, yleensä lyhenteenä on xs tai xsd

=> nimiavaruuden avulla kerrotaan, mikä on XML Scheman sanastoa

8

Elementti

```
<xs:element name="Nimi" ref="Elem"  
  type="Tyyppi"  
  minOccurs="Min. määrä" maxOccurs="Maks. määrä"  
  default="Oletusarvo" nillable="true/false" />
```

<i>name</i>	elementin nimi
<i>ref</i>	viittaus globaaliin määrittelyyn
<i>type</i>	elementin tietotyyppi (oletusarvo anyType).
<i>minOccurs</i>	minimi esiintymismäärä (oletusarvo 1)
<i>maxOccurs</i>	maksimi esiintymismäärä (oletusarvo 1)
<i>default</i>	mikä on elementin oletusarvo
<i>nillable</i>	voiko elementti saada null arvon (oletusarvo false)

9

Huomioitavaa XML Schemassa

- vain name tai ref
- minOccurs="0", elementti ei ole pakollinen
- maxOccurs >= minOccurs
- maxOccurs="unbounded", yläraja määrittelemätön
- jos oletusarvo (default) on määritelty, jäsentäjä laittaa oletusarvon elementille vain, jos elementti on dokumentissa ilman tietosisältöä
- jos elementtiä ei ole dokumentissa, jäsentäjä ei laita dokumenttiin oletusarvoista elementtiä.
- jos nillable="true", elementti voidaan määritellä NULL-sisältöiseksi laittamalla elementille attribuutti xsi:nil="true"

Huomioitavaa XML-dokumentissa

- jos elementillä attribuutti xsi:nil="true", elementillä ei saa olla sisältöä

10

<code><sukunimi>Marttila</sukunimi></code>	<code><xs:element name="sukunimi" type="xs:string" /></code>
<code><etunimi>Sirpa</etunimi></code>	<code><xs:element name="etunimi" type="xs:string" maxOccurs="3"/></code>
<code><puhelin>050-123456</puhelin></code>	<code><xs:element name="puhelin" type="xs:string" minOccurs="0" maxOccurs="unbounded" default="09-148901" /></code>
<code><spaiva xsi:nil="true"></spaiva></code> <code><spaiva>2003-12-24</spaiva></code>	<code><xs:element name="spaiva" type="xs:date" minOccurs="0" nillable="true" /></code>

11

XML Scheman tietotyyppejä

string	merkkijono
boolean	true tai 1, false tai 0
float	32-bittinen liukuluku
double	64-bittinen liukuluku
decimal	desimaaliluku
byte	-128 - 128
unsignedByte	0 - 255
short	-32768 - 32767
unsignedShort	0 - 65535
long	-9223372036854775808 - 9223372036854775807
unsignedLong	0 - 1844674407309551615
int	-2147483648 - 2147483647
unsignedInt	0 - 4294967295

12

dateTime	2004-09-07MO3:20:00.000-05:00
date	CCYY-MM-DD Esim. 2004-09-07
time	hh:mm:ss.sss Esim. 10:15:00.000
gYear	Vuosi esim. 2004
anyURI	URI osoite esim. http://www.helia.fi
ID IDREFS NMTOKEN NMTOKENS ENTITY ENTITIES	Vastaa DTD:n attribuuttityyppejä. Voidaan käyttää vain attribuutin tietotyypin määrittelyyn.

13

Attribuutti

```
<xs:attribute name="Nimi" ref="Attribuutti"
  type="Tyyppi"
  use="Pakollisuus"
  default="Oletusarvo"
  fixed="Vakioarvo" />
```

name attribuutin nimi

ref viittaus globaaliin määrittelyyn

type attribuutin tietotyyppi (oletusarvo on anyType).

use onko attribuutti pakollinen (oletusarvo on optional). *Use* voi saada arvot *required* (pakollinen), *optional* (ei pakollinen), *prohibited* (käyttö kielletty)

default oletusarvo

fixed vakioarvo

14

▼ Huomioitavaa

- vain joko name tai ref
- jos default on määritelty, use voi olla ainoastaan "optional"

15

tunnus="marsi"

```
<xs:attribute name="tunnus"  
type="xs:string" use="required" />
```

esimies="honte"

```
<xs:attribute name="esimies"  
type="xs:string" use="optional" />
```

osaamisalue="XML"

```
<xs:attribute name="osaamisalue"  
type="xs:string" default="opettaminen"  
/>
```

paikka="Helia"

```
<xs:attribute name="paikka"  
type="xs:string" fixed="Helia" />
```

16

Lapsielementtejä sisältävä elementti

```
<xs:element name="Nimi">  
  <xs:complexType>  
    <xs:sequence>  
      <xs:element name="Alielementin_nimi" ... />  
    </xs:sequence>  
    <xs:attribute name="Attribuutin_nimi" ... />  
  </xs:complexType>  
</xs:element>
```

complexType elementti on monimutkaista tyyppiä, jos sillä on lapsielementtejä ja/tai attribuutteja.

sequence elementeillä on peräkkäisjärjestys siten, että schemassa ensimmäisenä oleva on myös XML-dokumentissa ensimmäisenä

choice valinta

17

```
<nimi>  
  <sukunimi>Marttila</sukunimi>  
  <etunimi>Sirpa</etunimi>  
  <etunimi>Anneli</etunimi>  
</nimi>
```

```
<xs:element name="nimi">  
  <xs:complexType>  
    <xs:sequence>  
      <xs:element name="sukunimi" type="xs:string"/>  
      <xs:element name="etunimi" type="xs:string" maxOccurs="3" />  
    </xs:sequence>  
  </xs:complexType>  
</xs:element>
```

18

```
<henkilo tunnus="mars" esimies="honte" >
  <!-- Lapsielementit -->
</henkilo>
```

```
<xs:element name="henkilosto">
<xs:complexType>
  <xs:sequence>
    <!-- Lapsielementtien esittelyt -->
  </xs:sequence>
  <xs:attribute name="tunnus" type="xs:string" use="required"/>
  <xs:attribute name="esimies" type="xs:string"/>
</xs:complexType>
</xs:element>
```

19

Mixed Content

```
<xs:element name="Nimi">
  <xs:complexType mixed="true">
    ...
  </xs:complexType>
</xs:element>
```

```
<muistio>Tänään <paiva>2005-12-24</paiva> päätettiin</muistio>
```

```
<xs:element name="muistio">
  <xs:complexType mixed="true">
    <xs:sequence>
      <xs:element name="paiva" type="xs:date" minOccurs="0"
maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

20

Tyhjä elementti

```
<xs:element name="Nimi">  
  <xs:complexType>  
    <xs:attribute name="Nimi" ... />  
  </xs:complexType>  
</xs:element>
```

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

```
<xs:element name="tiimi" minOccurs="0">  
  <xs:complexType>  
    <xs:attribute name="nimi" type="xs:string" use="required" />  
    <xs:attribute name="asema" type="xs:string" use="required" />  
  </xs:complexType>  
</xs:element>
```

21


Attribuutillinen elementti

```
<xs:element name="Nimi">  
  <xs:complexType>  
    <xs:simpleContent>  
      <xs:extension base="Periyttävä_tyyppi">  
        <xs:attribute name="Nimi" ... />  
      </xs:extension>  
    </xs:simpleContent>  
  </xs:complexType>  
</xs:element>
```

simpleContent elementti on tyypiltään yksinkertainen eli sillä ei ole lapsielementtejä.


extension periytetään uusi tyyppi laajentamalla *base* attribuutilla kerrottua tyyppiä.

22




```
<tyohuone sijainti="R1">P024</tyohuone>
```

```
<xs:element name="tyohuone" minOccurs="0">  
  <xs:complexType>  
    <xs:simpleContent>  
      <xs:extension base="xs:string">  
        <xs:attribute name="sijainti" type="xs:string"  
          use="required" />  
      </xs:extension>  
    </xs:simpleContent>  
  </xs:complexType>  
</xs:element>
```



23



Omien tietotyyppien tekeminen


Yksinkertaisten omien tietotyyppien tekeminen

- Arvojen rajoittaminen
- Luettelo
- Sisältömalli

Monimutkaisten omien tietotyyppien tekeminen

- Lapsielementtejä ja attribuutteja sisältävä oma tietotyyppi
- Oma tietotyyppi tyhjälle elementille
- Oma tietotyyppi attribuutilliselle elementille

24



Yksinkertaisten omien tietotyyppien tekeminen

Yksinkertaisella tietotyyppillä rajoitetaan perustietotyyppien sisältöä, joten yksinkertainen tietotyyppi voi sisältää vain merkkitietoa ei elementtejä eikä attribuutteja.

25



Yksinkertainen oma tietotyyppi

Yksinkertainen oma tietotyyppi rajoittaa elementin tai attribuutin sisältöä. Tietotyyppi voidaan periyttää sekä scheman valmiista että itse määritellyistä yksinkertaisista tietotyypeistä.

```
<xs:simpleType name="Tietotyypin_nimi">  
  <xs:restriction base="Periyttävä_tietotyyppi">  
    arvon rajoittaminen,  
    luettelo,  
    sisältömalli  
  </xs:restriction>  
</xs:simpleType>
```

restriction ilmoittaa, mistä tietotyypistä uusi tyyppi periytetään.



26

Lukualueiden rajoittaminen

```
<ika>31</ika>
```

```
<xs:element name="ika" type="IkaType" minOccurs="0" />
```

```
<xs:simpleType name="IkaType">
```

```
  <xs:restriction base="xs:short">
```

```
    <xs:minInclusive value="0" />
```

```
    <xs:maxInclusive value="150" />
```

```
  </xs:restriction>
```

```
</xs:simpleType>
```

minInclusive value="0" -> minimiarvoon kuuluu 0

maxInclusive value="100" -> maksimiarvoon kuuluu 100

minExclusive value="0" -> minimiarvoon ei kuulu nolla

maxExclusive value="100" -> maksimiarvoon ei kuulu 100

27

```
<palkka>156.87</palkka>
```

```
<xs:element name="palkka" type="PalkkaType" />
```

```
<xs:simpleType name="PalkkaType">
```

```
  <xs:restriction base="xs:decimal">
```

```
    <xs:totalDigits value="7" />
```

```
    <xs:fractionDigits value="2" />
```

```
  </xs:restriction>
```

```
</xs:simpleType>
```

totalDigits -> luvun kokonaispituus mukaan lukien etumerkki,
desimaalipiste ja desimaalit esim. value="7"
kelvollisia arvoja ovat 1234567, 1234.56 ja
+0.0005

fractionDigits -> kuinka monta desimaalia korkeintaan voi olla

28

Merkkijonojen pituuden rajoittaminen

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

```
<xs:element name="sukunimi" type="NimiType" />
```

```
<xs:simpleType name="NimiType">
```

```
  <xs:restriction base="xs:string">
```

```
    <xs:minLength value="3" />
```

```
    <xs:maxLength value="30" />
```

```
  </xs:restriction>
```

```
</xs:simpleType>
```

minLength -> minimipituus
maxLength -> maksimipituus
length -> vakiopituus

29

Luettelo

```
sijainti="R1"
```

```
<paikka>R1</paikka>
```

```
<xs:attribute name="sijainti" type="SijaintiType" use="required" />
```

```
<xs:element name="paikka" type="SijaintiType" />
```

```
<xs:simpleType name="SijaintiType">
```

```
  <xs:restriction base="xs:string">
```

```
    <xs:enumeration value="R1" />
```

```
    <xs:enumeration value="R2" />
```

```
    <xs:enumeration value="P" />
```

```
    <xs:enumeration value="S" />
```

```
  </xs:restriction>
```

```
</xs:simpleType>
```

30

Sisältömalli

```
<onumero>a1234567</onumero>
```

```
<xs:element name="onumero" type="ONumeroType" />
```

```
<xs:simpleType name="ONumeroType">
```

```
  <xs:restriction base="xs:string">
```

```
    <xs:pattern value="a\d{7}" />
```

```
  </xs:restriction>
```

```
</xs:simpleType>
```

31

Merkkiryhmät

\w	Kirjain, numero tai alaviiva
\W	Muu kuin kirjain, numero tai alaviiva
\d	Numero
\D	Muu kuin numero
\s	Whitespace
\S	Muu kuin whitespace

Toisto ja ryhmittely

*	0-n
+	1-n
?	0-1
{N}	Tasan N
{N,}	Vähintään N
{N,M}	Vähintään N, enintään M
	Vaihtoehto
()	Määrittysten ryhmittely

32

Yksittäiset merkkihahmot		
a	Merkki	a
.	Mikä tahansa merkki paitsi rivinvaihto	
[abc]	Mikä tahansa hakasulkujen välissä oleva merkki	
[a-z0-9]	Merkki väliltä a-z tai 0-9	
[^bd]	Ei merkkejä b tai d	
([a-zääö])	Merkki väliltä a-ö	
\	merkin erikoismerkityksen poisto	
	Alla olevien merkkien edessä on oltava kenoviiva (\), jos tarkoitetaan merkkiä itseään eikä merkin säännöllisen lauseen merkitystä:	
	\ . - ? * + { } [] ()	
	* => tarkoittaa merkkiä *	
	* -> tarkoittaa lukumäärää 0-n	

33

Esimerkkejä		
Lauseke	Esimerkki	Selitys
\d{2}	12	numero 2 kertaa
\d{2,}	1234	numero vähintään 2 kertaa
\d{2,3}	123	numero 2-3 kertaa
\d*		numero 0-n kertaa
\d+	1	numero 1-n kertaa
\d?		numero 0-1 kertaa
\w{0,5}	s_1_m	numero, kirjain tai alaviiva 0-5 kertaa
[abc]	c	jokin kirjaimista a, b tai c
[a-zääö-]{2,}	sirpa	kirjain a-ö tai - vähintään kaksi kertaa
(a A)\d{7}	A1234567	kirjain a tai A ja 7 numeroa
\d{6}[\+ -A]\d{3}[A-Z0-9]	123456-123A	suomalainen hetu

34

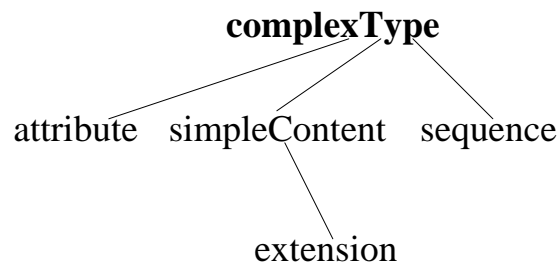
Tietotyyppi ja sen mahdollinen rajoitus

- ▼ string, anyURI
 - length, minLength, maxLength, pattern, enumeration
- ▼ decimal, byte, short, long, int
 - totalDigit, fractionDigits, minInclusive, maxInclusive, minExclusive, maxExclusive, pattern, enumeration
- ▼ float, double, datetime, date, time, gYear
 - minInclusive, maxInclusive, minExclusive, maxExclusive, pattern, enumeration

35

Monimutkaisten omien tietotyyppien tekeminen

Monimutkainen tietotyyppi sisältää elementtejä ja/tai attribuutteja.



36

Oma tietotyyppi tyhjälle elementille

```
<xs:complexType name="Tyypin_nimi">  
  <xs:attribute name="Attribuutin_nimi" ... />  
</xs:complexType>
```

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

```
<xs:element name="tiimi" type="TiimiType" minOccurs="0"/>
```

```
<xs:complexType name="TiimiType">  
  <xs:attribute name="nimi" type="xs:string" use="required"/>  
  <xs:attribute name="asema" type="xs:string" default="jäsen"/>  
</xs:complexType>
```

37

Oma tietotyyppi attribuutilliselle elementille

```
<xs:complexType name="Tietotyypin_nimi">  
  <xs:simpleContent>  
    <xs:extension base="Periyttävä_tyyppi">  
      <xs:attribute name="Attribuutin_nimi" ... />  
    </xs:extension>  
  </xs:simpleContent>  
</xs:complexType>
```

simpleContent elementti on tyypiltään yksinkertainen eli sillä ei ole lapsielementtejä.

extension periytetään uusi tyyppi laajentamalla *base* attribuutilla kerrottua tyyppiä.

38

```
<tyohuone sijainti="R1">P024</tyohuone>
```

```
<xs:element name="tyohuone" type="TyohuoneType"
minOccurs="0"/>
```

```
<xs:complexType name="TyohuoneType">
  <xs:simpleContent>
    <xs:extension base="xs:string">
      <xs:attribute name="sijainti" type="xs:string" default="R1"/>
    </xs:extension>
  </xs:simpleContent>
</xs:complexType>
```

39

Lapsielementtejä sisältävä oma tietotyyppi

```
<xs:complexType name="Tietotyypin_nimi">
  <xs:sequence>
    <xs:element name="Elementin_nimi" ... />
  </xs:sequence>
  <xs:attribute ... />
</xs:complexType>
```

complexType tietotyyppi on monimutkaista tyyppiä, jos sillä on lapsielementtejä ja/tai attribuutteja.

sequence elementeillä on peräkkäisjärjestys siten, että schemassa ensimmäisenä oleva on myös XML-dokumentissa ensimmäisenä jne.

40

```
<nimi>
  <sukunimi>Marttila</sukunimi>
  <etunimi>Sirpa</etunimi>
  <etunimi>Anneli</etunimi>
</nimi>
```

```
<xs:element name="nimi" type="NimiType" />
<xs:complexType name="NimiType">
  <xs:sequence>
    <xs:element name="sukunimi" type="xs:string" />
    <xs:element name="etunimi" type="xs:string" maxOccurs="3" />
  </xs:sequence>
</xs:complexType>
```

41

Yhteenveto

- ▼ Schemalla voidaan kuvata tietokantamaista tietoa riittävän tarkalla tasolla.
- ▼ Schemassa voidaan antaa elementeille ja attribuuteille tietotyypit sekä oletusarvot sekä elementeille tarkat lukumäärätiedot.
- ▼ Tietotyypit voivat olla scheman valmiita tai itse määriteltyjä tietotyyppejä.
- ▼ Schema tekniikka on yrityssovelluksia ajatellen XSL-tekniikan lisäksi tärkein XML:n liitännäistekniikoista.

42