

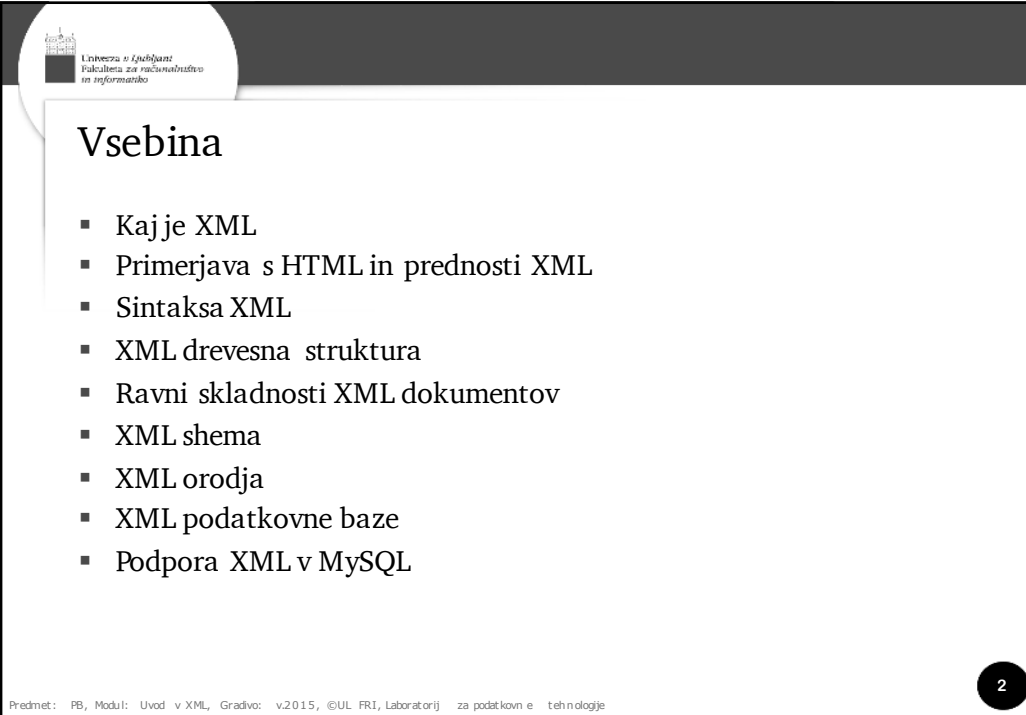
Univerza v Ljubljani
Fakulteta za računalništvo
in informatiko

Predmet:
Podatkovne baze

Modul:
Uvod v XML

Gradivo:
v.2015

20.5.
2016



Univerza v Ljubljani
Fakulteta za računalništvo
in informatiko

Vsebina

- Kaj je XML
- Primerjava s HTML in prednosti XML
- Sintaksa XML
- XML drevesna struktura
- Ravni skladnosti XML dokumentov
- XML shema
- XML orodja
- XML podatkovne baze
- Podpora XML v MySQL

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

2

Kaj je XML?...

- Razširljivi označevalni jezik (eXtensible Markup Language)
- Podoben jeziku HTML

Izsek primera HTML datoteke

```
<html>

<head>
  <title>My First Web Page</title>
</head>

<body>
  <h1>My First Web Page</h1>
  <p><b>Hello World Wide Web!</b></p>
  <p><i>Hello World Wide Web!</i></p>
  <p><u>Hello World Wide Web!</u></p>
  <p>This is my first web page.</p>
  <p>HTML tags can give <b><i>various</i></b>
  <u>looks and format</u> to the content of this web page.</p>
</body>

</html>
```

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

3

Kaj je XML?

- Namenjen prenosu, ne prikazu!
- Značke (tags) niso predpisane vnaprej.
- Je samopojasnjujoč.
- Priporočilo organizacije W3C (<http://www.w3.org/>).

Izsek primera XML datoteke

```
<person id="Big.Boss">
  <name>
    <family>Boss</family>
    <given>Big</given>
  </name>
  <email>chief@oxygenxml.com</email>
  <link subordinates="one.worker two.worker three.worker four.worker five.worker"/>
</person>
<person id="one.worker">
  <name>
    <family>Worker</family>
    <given>One</given>
  </name>
  <email>one@oxygenxml.com</email>
  <link manager="Big.Boss"/>
</person>
```

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

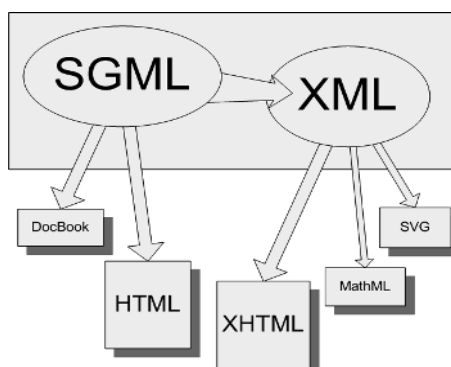
4

Razlike med HTML in XML

HTML	XML
vnaprej določen nabor značk	značke definiramo sami
značke namenjene določanju videza dokumenta	značke opisujejo pomen dokumenta
značke lahko izpuščamo	vse značke morajo biti prisotne
strani pogosto nepravilno zapisane – npr. napačne značke	dokumenti morajo biti „ustrezni“ – dobra definiranost

Uveljavljena tehnologija

- XML je podmnžica standarda SGML (Standardized Generalized Markup Language), ki je bil definiran že leta 1969



Sintaksa XML dokumenta

Univerza v Ljubljani
Fakulteta za računalništvo in informatiko

Deklaracija → `<?xml version="1.0" ?>`

Element → `<novice>`

Značke (Tags)

Začetna značka → `<avtor>Matej Huščič</avtor>`

Zaključna značka → `</besedilo>`

Atribut → `naslov="Google ustavi digitalizacijo starih časnikov"`

Vrednost atributa → `Google ustavi digitalizacijo starih časnikov`

Vsebina (CDATA) → `Google je končal digitalizacijo starih...`

```

<?xml version="1.0" ?>
<novice>
  <novica naslov="Google ustavi digitalizacijo starih časnikov">
    <kategorija1>omrežja</kategorija1>
    <kategorija2>internet</kategorija2>
    <datum>20.5.2011</datum>
    <vir>Heise</vir>
    <avtor>Matej Huščič</avtor>
    <besedilo>Google je končal digitalizacijo starih...
  </novica>
  <novica naslov="iPhone 7 kmalu v prodaji">
    <kategorija1>telefonija</kategorija1>
    ...
  </novica>
</novice>

```

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

10

XML dokument in drevesna struktura

Univerza v Ljubljani
Fakulteta za računalništvo in informatiko

Korenski element → Novice

Novica
Naslov="Google ustavi..."

Novica
Naslov="iPhone 7..."

Kategorija1
omrežja

Kategorija2
internet

Datum
20.5.2011

Vir
...

Besedilo
...

Kategorija1
...

CDATA → internet

```

graph TD
    Novice[Novice] --> Novica1[Novica  
Naslov="Google ustavi..."]
    Novice --> Novica2[Novica  
Naslov="iPhone 7..."]
    Novica1 --> Kat1[Kategorija1  
omrežja]
    Novica1 --> Kat2[Kategorija2  
internet]
    Novica1 --> Datum[Datum  
20.5.2011]
    Novica1 --> Vir[Vir  
...]
    Novica1 --> Besedilo[Besedilo  
...]
    Novica2 --> Kat3[Kategorija1  
...]

```

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

11



Deli XML dokumenta...

■ ELEMENTI

- Element = začetna značka + vsebina + končna značka
- Element lahko vsebuje besedilo ali druge elemente ali oboje (ali nič)

```
<KNJIGA>...vsebina...</KNJIGA>
```

element z vsebino

```
<KNJIGA> <NASLOV>...naslov...</NASLOV> </KNJIGA>
```

element, ki vsebuje element

```
<VRSTICA/>
```

kratka oznaka za prazen element

■ ATRIBUTI

- Elementom lahko pripišemo lastnosti oz. attribute
- Lastnosti zapišemo v začetne značke kot pare atribut = "vrednost"
 - Vrednost mora biti v enakih enojnih ali dvojnih narekovajih

```
<PRELOM stran="11"/>
```

```
<RAZDELEK številka="11" zvrst="podpoglavje">...</RAZDELEK>
```

12



Deli XML dokumenta...

■ BESEDILO (CDATA sekcija):

- Lahko se nahaja znotraj elementov

```
<KNJIGA>...vsebina...</KNJIGA>
```

- Lahko je podano eksplicitno kot

```
<![CDATA[To je CDATA sekcija, ki lahko vsebuje tudi <'> posebne znake]]>
```

■ POSEBNI ZNAKI

- Dokument XML lahko vsebuje posebne znake, ki se ob procesiranju nadomestijo z nečim drugim.
- Sklic na poseben znak se začne z znakom & in konča s podpičjem.

```
<formula>1 &lt; 2</formula>
```

```
<podjetje>Procter &amp; Gamble</podjetje>
```

1 < 2

Procter & Gamble

```
&lt; = <
&gt; = >
&amp; = &
&apos; = '
&quot; = "
```

13

Deli XML dokumenta

- XML deklaracija

```
<?xml version="1.0"?>
```

- Navodila za procesiranje (Pis): `<? ...?>`

```
<?xml-stylesheet type="text/css" href="mys.css"?>
```

- Komentarji `<!-- ... -->`

```
<!-- comment text -->
```

Ravni skladnosti XML dokumentov

- Pravilno strukturiran (well formed): zadošča vsem WC3 sintaktičnim pravilom za XML
 - Pravila za poimenovanje, gnezdenje, označevanje atributov...
 - Dobra definiranost je za XML dokumente obvezna
- Veljaven (valid) dokument je dokument, ki je skladen s shemo
 - Formalnost je za XML dokument opcijska;
 - Lahko preverjamo z XML shemo;

Pravila dobro definiranih XML dokumentov...

- XML standard zahteva, da se v dokumentih upoštevajo naslednja pravila:

- Obstajati mora en element, ki vsebuje vse ostale
- Značke morajo biti uravnotežene

```
<BOOK>...</BOOK>
<BOOK />
```

- Gnezdenje značk mora biti izvedeno pravilno.

```
<BOOK> <LINE> to je pravilno </LINE> </BOOK>
<LINE> <BOOK> to </LINE> gotovo ni </BOOK> pravilno
```

- Tekst značk je občutljiv na velike in male črke

```
<P> različna začetna in končna značka - XML tega ne dopušča </p>
```

Pravila dobro definiranih XML dokumentov

- Atributi v značkah (tags) morajo biti znotraj narekovajev.

```
< ITEM Category="Home and Garden" Name="hoe-matic t500">
```

- Komentiranje je dovoljeno.

```
<!-- They are done just as in HTML... -->
```

- Dokument se mora začeti z ustreznim nizom (XML deklaracija).

```
<?xml version='1.0' ?>
```

- Posebni znaki se morajo pričeti z ubežnim znakom (&)

```
<formula> x &lt; y+2x </formula>
<cd title="&quot; music &quot;">
```



XML shema

- Shema je ločen dokument, ki definira elemente, attribute in strukturo XML dokumenta
- Shemo določimo tako, da
 1. definiramo slovar in
 2. določimo postavitev ter število elementov in atributov v označevalnem jeziku
- Shema definira točno določen razred dokumentov.



XML Shema - Primer

Author.xsd

Shema Author.XSD
definira strukturo
elementa Author

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="Author">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="FirstName" type="xs:string" />
        <xs:element name="LastName" type="xs:string" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

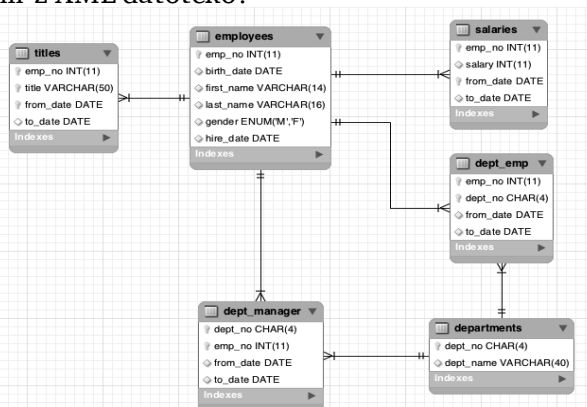
MarkTwain.xml

MarkTwain.XML je
instanca oz. primerek
zgoraj definirane sheme

```
<?xml version="1.0"?>
<Author xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="Author.xsd">
  <FirstName>Mark</FirstName>
  <LastName>Twain</LastName>
</Author>
```


Vaja

- Kako bi podatke, ki ustrezajo prikazanemu podatkovnemu modelu predstavili z XML datoteko?



Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

23

Pretvorba relacijske sheme v XML...

- Povezava 1 : 1

Relacija A

A1	A2
a11	a21
a12	a22

Relacija B

B1	B2	*A1
b11	b21	a11
b12	b22	a12



```

<A A1="a11" A2="a21">
  <B B1="b11" B2="b21"></B>
</A>

<A A1="a12" A2="a22">
  <B B1="b12" B2="b22"></B>
</A>
    
```

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

24



Pretvorba relacijske sheme v XML...

▪ Povezava 1 : n

Relacija A

A1	A2
a11	a21
a12	a22

Relacija B

B1	B2	*A1
b11	b21	a11
b12	b22	a12
b13	b23	a12



```

<A A1="a11" A2="a21">
  <B B1="b11" B2="b21"></B>
</A>

<A A1="a12" A2="a22">
  <B B1="b12" B2="b22"></B>
  <B B1="b13" B2="b23"></B>
</A>
  
```



Pretvorba relacijske sheme v XML...

▪ Povezava m : n

Relacija A

A1	A2
a11	a21
a12	a22

Relacija B

B1	B2
b11	b21
b12	b22

Relacija R

*A1	*B1
a11	b11
a12	b12



```

<A A1="a11" A2="a21" A_id="1"></A>
<B B1="b11" B2="b21" B_id="2"></B>
<R A_idref="1" B_idref="2"></R>

<A A1="a12" A2="a22" A_id="3"></A>
<B B1="b12" B2="b22" B_id="4"></B>
<R A_idref="3" B_idref="4"></R>
  
```

XML orodja...

- XML urejevalniki:
 - Orodja, ki omogočajo izdelavo, urejanje in validacijo XML dokumentov.
- Orodja za obdelavo XML:
 - Orodja, ki omogočajo transformacijo XML podatkov (dokumentov) v druge standardizirane oblike... npr HTML.
- Orodja/standardi, ki omogočajo obdelavo XML iz programskih jezikov:
 - XML razčlenjevalci, ki omogočajo razčlenjevanje, validacijo, serializacijo in manipulacijo XML. Npr. Apache Xerces.
 - XML API-ji, ki omogočajo posodabljanje in branje XML podatkov potem, ko so ti že v podatkovnih strukturah nekega programskega jezika (DOM, SAX).

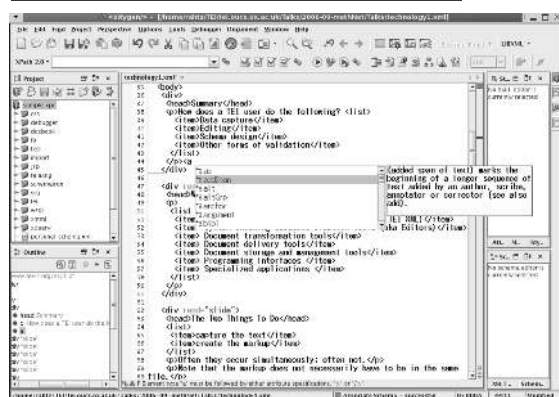
Izdelava XML dokumentov

- Primeri XML urejevalnikov:
 - EditiX – <http://www.editix.com>
 - XMLSpy – <http://www.altova.com>
 - Jedit - <http://www.jedit.org/>
 - Oxygen - <http://www.oxygenxml.com/>
 - XML Marker - <http://symbolclick.com/>
 - Sublime text - <http://www.sublimetext.com/>
 - Notepad

DEMO:



- Oxygen – XML urejevalnik
- <http://www.oxygenxml.com/videos>



30

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

XML tehnologije

- Xpath – omogoča identifikacijo podsklopov XML dokumentov.
- DTDs in XML Schema – omogoča definicijo XML dokumenta.
- Xquery – omogoča poizvedovanje po XML dokumentih.
- Xlink, Xpointer – omogoča povezovanje in referenciranje podatkov znotraj XML dokumenta.
- XML-RPC, SOAP – omogoča oddaljene klice procedur (RPC) prek XML.

31

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

XML podatkovne baze

- XML PB - nudijo upravljanje s podatki v XML formatu:
 - Shranjevanje
 - Poizvedovanje
 - Pretvorbe v različne formate
 - ...
- Tipično temeljijo na dokumentnih bazah.
- Razlog za pojav: XML standard za prenos podatkov. Nepotrebno pretvarjanje med formati...

Vrste XML PB

- PB, ki podpirajo XML dokumente:
 - Pretvarjajo XML v npr. relacije in na izhodu nazaj v XML...
 - Večina nvejših SUPB podpira posebne XML podatkovne tipe.
- Naravne XML PB:
 - Interni podatkovni model temelji na XML - XML dokumenti osnovna enota shranjevanja.

PB, ki podpirajo XML

- Navadno ponujajo enega izmed naslednjih pristopov za shranjevanje XML podatkov v relacijsko bazo:
 - XML shranjen v CLOB (Character large object)
 - XML pretvorjen v množico tabel na osnovi sheme;
 - XML shranjen v naravnem XML formatu (XML tip, ISO standard)
- Relacijske PB, ki podpirajo ISO XML tip:
 - IBM DB2;
 - Microsoft SQL Server;
 - Oracle Database;
 - PostgreSQL.

ORACLE
DATABASE

IBM
DB2

Microsoft
SQL Server

PostgreSQL

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

34

Podpora XML v MySQL...

- Preusmerjanje izhoda poizvedb v XML

```
mysql -u root -p --xml;
mysql> select * from employees limit 1;

mysql> select * from employees limit 1;
<?xml version="1.0"?>

<resultset statement="select * from employees limit 1;" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <row>
    <field name="emp_no">10001</field>
    <field name="birth_date">1953-09-02</field>
    <field name="first_name">Georgi</field>
    <field name="last_name">Facello</field>
    <field name="gender">M</field>
    <field name="hire_date">1986-06-26</field>
  </row>
</resultset>
1 row in set (0.00 sec)
```

Predmet: PB, Modul: Uvod v XML, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

35



Podpora XML v MySQL...

▪ Branje XML datotek

person.xml

```
<?xml version="1.0"?>
<list>
  <person person_id="1" fname="Pekka" lname="Nousiainen"/>
  <person person_id="2" fname="Jonas" lname="Oreland"/>
  <person person_id="3"><fname>Mikael</fname><lname>Ronström</lname></person>
  <person person_id="4"><fname>Lars</fname><lname>Thalmann</lname></person>
  <person><field name="person_id">5</field><field name="fname">Tomas</field>
  <field name="lname">Ulin</field></person>
  <person><field name="person_id">6</field><field name="fname">Martin</field>
  <field name="lname">Sköld</field></person>
</list>
```

```
mysql> LOAD XML LOCAL INFILE 'person.xml'
-> INTO TABLE person
-> ROWS IDENTIFIED BY '<person>';
```

```
mysql> SELECT * FROM person;
+-----+-----+-----+-----+
| person_id | fname | lname | created |
+-----+-----+-----+-----+
| 1 | Pekka | Nousiainen | 2007-07-13 16:18:47 |
| 2 | Jonas | Oreland | 2007-07-13 16:18:47 |
| 3 | Mikael | Ronström | 2007-07-13 16:18:47 |
| 4 | Lars | Thalmann | 2007-07-13 16:18:47 |
| 5 | Tomas | Ulin | 2007-07-13 16:18:47 |
| 6 | Martin | Sköld | 2007-07-13 16:18:47 |
+-----+-----+-----+-----+
6 rows in set (0.00 sec)
```



Podpora XML v MySQL...

▪ Iskanje po XML datotekah

```
mysql> SET @xml = '<a><b>X</b><b>Y</b></a>';
Query OK, 0 rows affected (0.00 sec)

mysql> SET @i = 1, @j = 2;
Query OK, 0 rows affected (0.00 sec)

mysql> SELECT @i, ExtractValue(@xml, '//b[@i]');
+-----+-----+
| @i | ExtractValue(@xml, '//b[@i]') |
+-----+-----+
| 1 | X |
+-----+-----+
1 row in set (0.00 sec)
```

