

XML a DTD

Michal Kvasnica

Čo je XML?

- XML hovorí o:
 - dátach
 - štruktúre dát
 - opise takejto štruktúry
- XML nehovorí o:
 - spracovaní dát
 - spôsobe zobrazenia dát (zobrazenie je iba určitým druhom spracovania)
- XML je "iba" spôsob zápisu dát, nič viac

V čom je XML dobré

Meno	Mesto	Telefón	
Janko Mrkvicka	Bratislava	1234	
Mária Smutná	Brezno	5678	9012

???

Dáta

```
Janko  
  Mrkvicka  
  
  Pekna  
    12  
    Bratislava  
  
0909090909  
0901111111  
020102030405
```

Štruktúra dát

- Štruktúra definuje:
 - význam jednotlivých dát
 - súvis dát navzájom

Štruktúra dát

```
<zoznam>
    Janko
      Mrkvicka
        Pekna
        12
        Bratislava
          0909090909
          0901111111
          020102030405
</zoznam>
```

Štruktúra dát

```
<zoznam>
  <osoba>
    Janko
      Mrkvicka
        Pekna
        12
        Bratislava
          0909090909
          0901111111
          020102030405
  </osoba>
</zoznam>
```

Štruktúra dát

```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
      Pekna
      12
      Bratislava
        0909090909
        0901111111
        020102030405
  </osoba>
</zoznam>
```

Štruktúra dát

```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>
      Pekna
      12
      Bratislava
    </bydlisko>
      0909090909
      0901111111
      020102030405
  </osoba>
</zoznam>
```

Štruktúra dát

```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>
      <ulica>Pekna</ulica>
      <cislo>12</cislo>
      <mesto>Bratislava</mesto>
    </bydlisko>
      0909090909
      0901111111
      020102030405
  </osoba>
</zoznam>
```

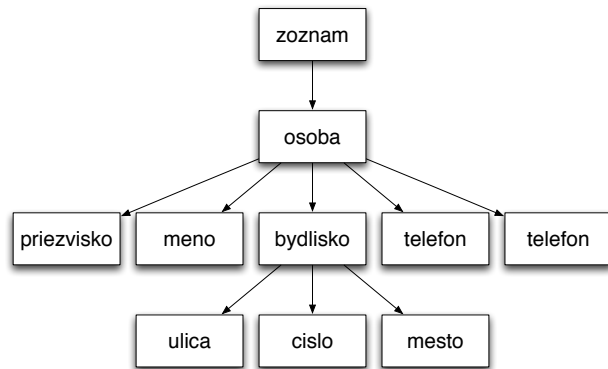
Štruktúra dát

```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>
      <ulica>Pekna</ulica>
      <cislo>12</cislo>
      <mesto>Bratislava</mesto>
    </bydlisko>
    <telefon typ="mobil">0909090909</telefon>
    <telefon typ="mobil">0901111111</telefon>
    <telefon typ="pevny">020102030405</telefon>
  </osoba>
</zoznam>
```

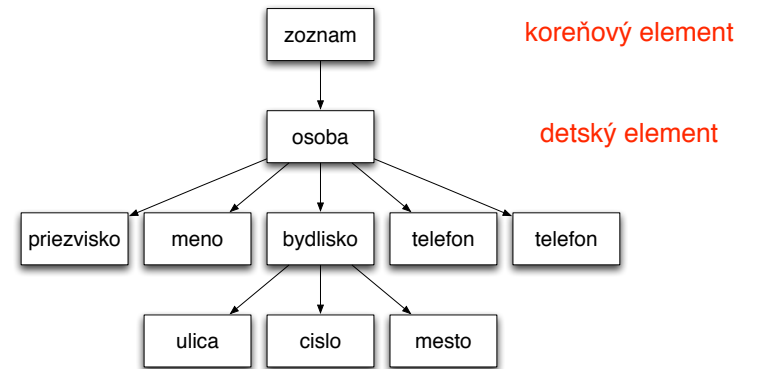
Samoopisujúci sa zápis

```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>
      <ulica>Pekna</ulica>
      <cislo>12</cislo>
      <mesto>Bratislava</mesto>
    </bydlisko>
    <telefon typ="mobil">0909090909</telefon>
    <telefon typ="mobil">0901111111</telefon>
    <telefon typ="pevny">020102030405</telefon>
  </osoba>
</zoznam>
```

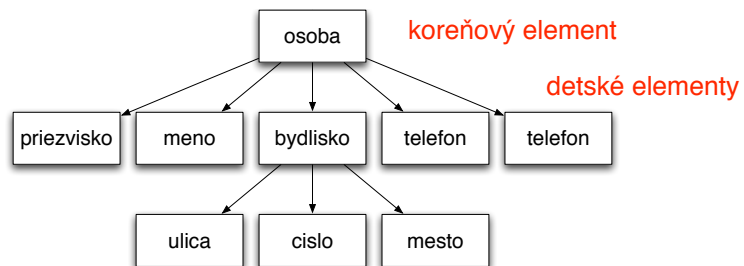
XML dokument je ako strom



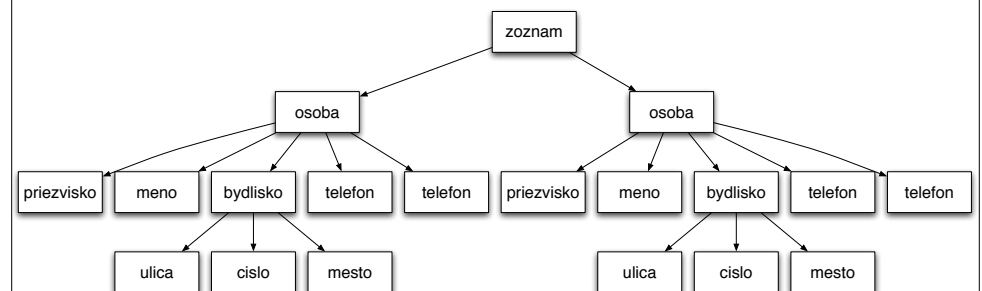
XML dokument je ako strom



Aj každá časť štruktúry je strom...



XML súbor má iba 1 hlavný koreňový element



Ilustrácia

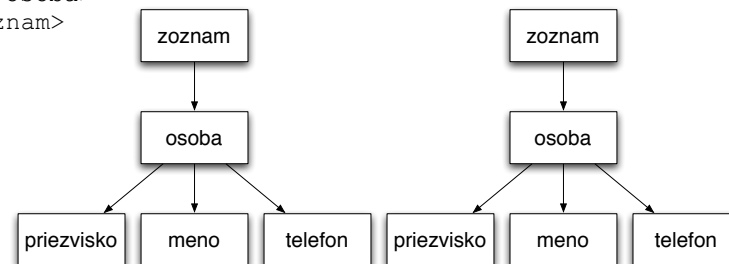
```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <riezvisko>Mrkvicka</riezvisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
</zoznam>
```



```
<zoznam>
  <osoba>
    <meno>Maria</meno>
    <riezvisko>Jednoducha</riezvisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

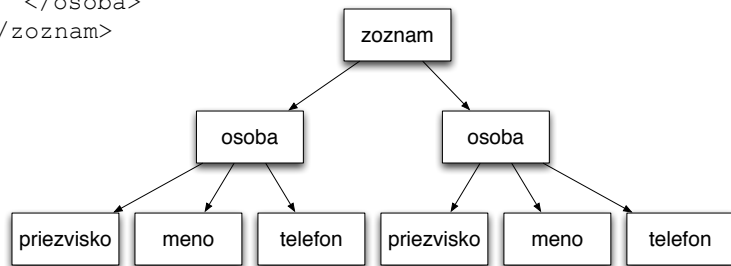
Nesprávne spojenie

```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <riezvisko>Mrkvicka</riezvisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
</zoznam>
<zoznam>
  <osoba>
    <meno>Maria</meno>
    <riezvisko>Jednoducha</riezvisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```



Správne spojenie

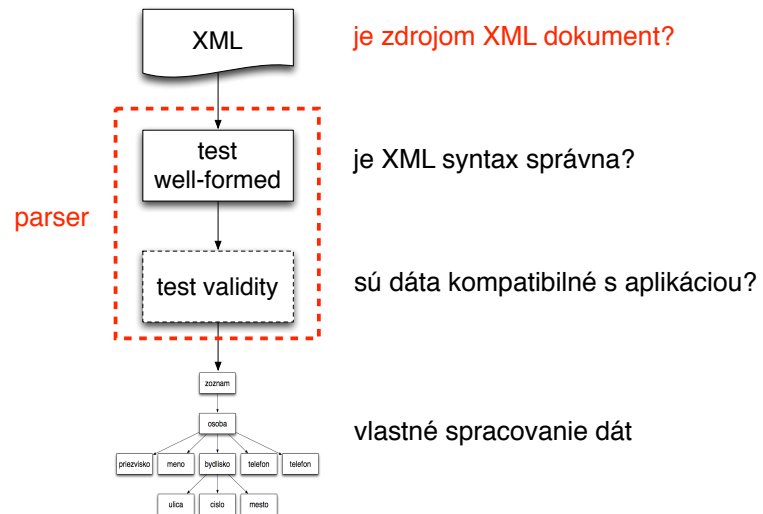
```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <riezvisko>Mrkvicka</riezvisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <riezvisko>Jednoducha</riezvisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```



Čo je XML?

- XML hovorí o:
 - dátach
 - štruktúre dát
 - opise takejto štruktúry
- XML nehovorí o:
 - spracovaní dát
 - spôsobe zobrazenia dát (zobrazenie je iba určitým druhom spracovania)
- XML je "iba" spôsob zápisu dát, nič viac

Spracovanie XML súborov



Ako vieme že ide o XML dokument?

telefonny_zoznam.xml

```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

- Windows svet:
 - typ dát sa určí podľa prípony názvu súboru
- Unix (Unix, Linux, Solaris, Mac OS X, ...)
 - typ dát je jednoznačne určený dátami samotnými

Unix príkaz "file"

```
$ file acc_2010.zip
acc_2010.zip: Zip archive data, at least v2.0 to extract

$ cp acc_2010.zip acc_2010.doc
$ file acc_2010.doc
acc_2010.doc: Zip archive data, at least v2.0 to extract
```

Unix príkaz "file"

```
$ file acc_2010.zip
acc_2010.zip: Zip archive data, at least v2.0 to extract

$ cp acc_2010.zip acc_2010.doc
$ file acc_2010.doc
acc_2010.doc: Zip archive data, at least v2.0 to extract

$ cat acc_2010.doc
PK          'à6;Å«Q[:
ñacc_2010/acc_2010_lightbulb.auxUT  πÊ][J~òfJUx
1•X-R,,[]}flØPmQu°Ód*!
```


Každý XML súbor musí byť označený ako XML

označenie špeciálneho elementu

```
<?xml version="1.0" encoding="UTF-8" ?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

```
$ file telefonny_zoznam.xml
telefonny_zoznam.xml: XML 1.0 document text
```

Každý XML súbor musí byť označený ako XML

názov elementu je potrebné dodržať

```
<?xml version="1.0" encoding="UTF-8" ?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

```
$ file telefonny_zoznam.xml
telefonny_zoznam.xml: XML 1.0 document text
```

Každý XML súbor musí byť označený ako XML

atribút označuje verziu formátu

```
<?xml version="1.0" encoding="UTF-8" ?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

```
$ file telefonny_zoznam.xml
telefonny_zoznam.xml: XML 1.0 document text
```

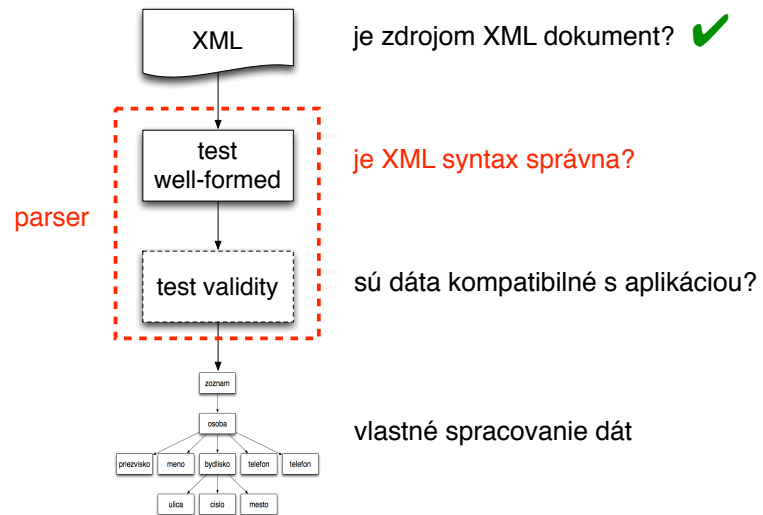
Každý XML súbor musí byť označený ako XML

a spôsob kódovania

```
<?xml version="1.0" encoding="UTF-8" ?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

```
$ file telefonny_zoznam.xml
telefonny_zoznam.xml: XML 1.0 document text
```


Spracovanie XML súborov



Well-formedness

- Well-formed document = dokument, ktorý je zapísaný správnou XML syntaxou
- Zznaky správnej syntaxe:
 - dokument obsahuje aspoň jeden element
 - dokument obsahuje unikátny koreňový element
 - tagy musia byť správne otvorené a uzatvorené
 - tagy musia mať kompatibilné názvy
 - elementy musia byť správne vnorené
 - atribúty musia byť správne zapísané

Príklady

Well-formed	Nie je well-formed
<code><meno>Janko</meno></code>	<code>"Janko"</code>
<pre> <osoby> <meno>Janko</meno> <meno>Jozko</meno> </osoby> </pre>	<pre> <meno>Janko</meno> <meno>Jozko</meno> </pre>
<pre> <meno>Janko</meno> <Meno>Jozko</Meno> </pre>	<pre> (meno>Janko<meno) <meno>Janko{ /meno> <meno>Janko<meno> <Meno>Janko</meno> <MENO>Janko</Meno> </pre>

Príklady

Well-formed	Nie je well-formed
<pre> <osoba> <meno>Janko</meno> </osoba> </pre>	<pre> <osoba> <meno>Janko </osoba> </meno> </pre>
<pre> <osoba id="123"> </pre>	<pre> <osoba id=123> <osoba id="123> <osoba id=123'> </pre>

Ako overit' well-formedness?

- Webové služby:
 - http://www.w3schools.com/Dom/dom_validate.asp
 - <http://www.w3.org/RDF/Validator/>
 - google "XML validator"

http://www.w3schools.com/Dom/dom_validate.asp

Syntax-check your XML

To help you syntax-check your xml, we have created an XML validator.

Paste your XML in the text area below, and syntax-check it by pressing the "Validate" button.

```
<?xml version="1.0" ?>
<Note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```



The page at <http://www.w3schools.com> says:

XML Parsing Error: mismatched tag. Expected: </Note>.
Location: http://www.w3schools.com/Dom/dom_validate.asp
Line Number 7, Column 3:

Validate

OK

Syntax-check

You can also syntax-check your XML file by typing the URL of your file into the input field below, and then press the "Validate" button

Filename:

http://www.w3schools.com/dom/note_error.xml

Validate

<http://www.w3.org/RDF/Validator/>

Check by Direct Input

```
<meno>Maria</meno>
<priezvisko>Jednoducha</priezvisko>
<bydlisko>
  <ulica>Strma</ulica>
  <cislo>2</cislo>
  <mesto>Brezno</mesto>
</bydlisko>
<telefon typ="mobil">0808080808</telefon>
</osoba>
```

Parse RDF

Restore the original example

Clear the textarea

Fatal Error Messages

FatalError: The element type "Ulica" must be terminated by the matching end-tag "</Ulica>". [Line = 19, Column = 18]

Triples of the Data Model

Number	Subject	Predicate	Object
--------	---------	-----------	--------

The original RDF/XML document

```
1: <?xml version="1.0"?>
2: <zoznam>
3:   <osoba>
4:     <meno>Janko</meno>
5:     <priezvisko>Mrkvicka</priezvisko>
6:     <bydlisko>
7:       <ulica>Pekna</ulica>
8:       <cislo>12</cislo>
9:       <mesto>Bratislava</mesto>
10:    </bydlisko>
11:    <telefon typ="mobil">0909090909</telefon>
12:    <telefon typ="mobil">0901111111</telefon>
13:    <telefon typ="pevny">020102030405</telefon>
14:  </osoba>
15:  <osoba>
16:    <meno>Maria</meno>
17:    <priezvisko>Jednoducha</priezvisko>
18:    <bydlisko>
19:      <Ulica>Strma</ulica>
20:      <cislo>2</cislo>
21:      <mesto>Brezno</mesto>
22:    </bydlisko>
23:    <telefon typ="mobil">0808080808</telefon>
24:  </osoba>
25: </zoznam>
```

Ako overiť well-formedness?

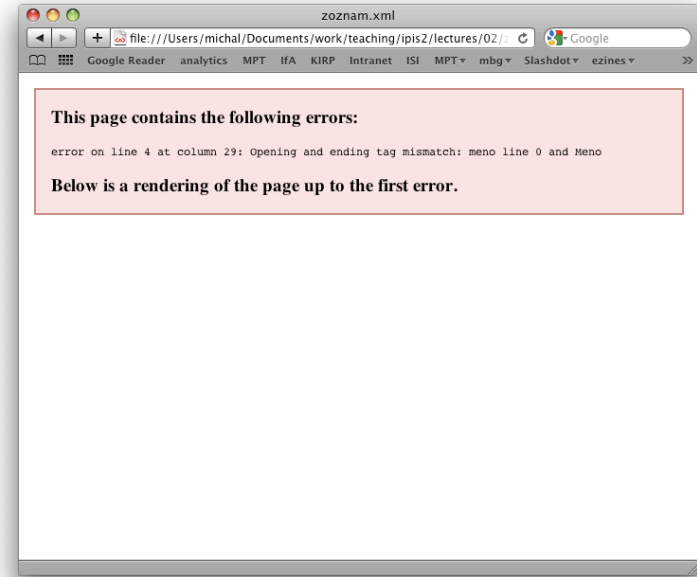
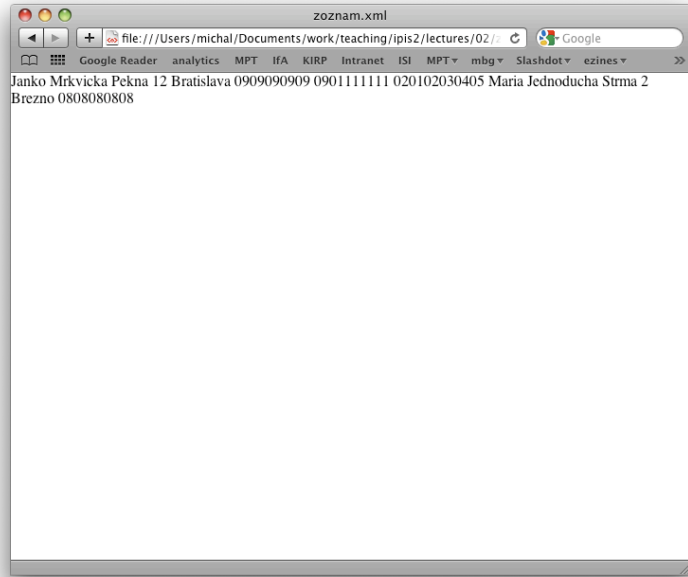
- Webové služby:
 - http://www.w3schools.com/Dom/dom_validate.asp
 - <http://www.w3.org/RDF/Validator/>
 - google "XML validator"
- XML nástroje:
 - XML Copy Editor (<http://xml-copy-editor.sourceforge.net/>)
 - Syntex Serna (<http://www.syntex.com/products/serna/index.html>)
 - Jaxe (<http://jaxe.sourceforge.net/>)
 - oXygen Editor (<http://jaxe.sourceforge.net/>)
- Webový prehliadač

```
- <zoznam>
- <osoba>
  <meno>Janko</meno>
  <riezvisko>Mrkvicka</riezvisko>
  - <bydlisko>
    <ulica>Pekna</ulica>
    <cislo>12</cislo>
    <mesto>Bratislava</mesto>
  </bydlisko>
  <telefon typ="mobil">0909090909</telefon>
  <telefon typ="mobil">0901111111</telefon>
  <telefon typ="pevny">020102030405</telefon>
</osoba>
- <osoba>
  <meno>Maria</meno>
  <riezvisko>Jednoducha</riezvisko>
  - <bydlisko>
    <ulica>Strma</ulica>
    <cislo>2</cislo>
    <mesto>Brezno</mesto>
  </bydlisko>
  <telefon typ="mobil">0808080808</telefon>
</osoba>
</zoznam>
```

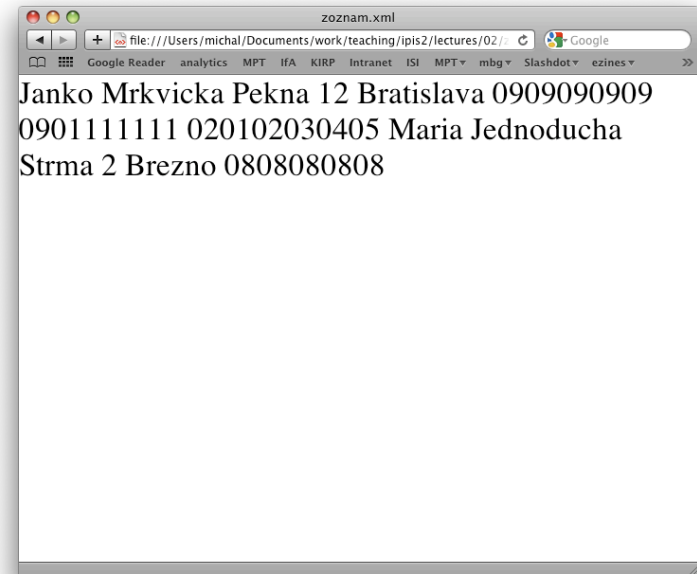
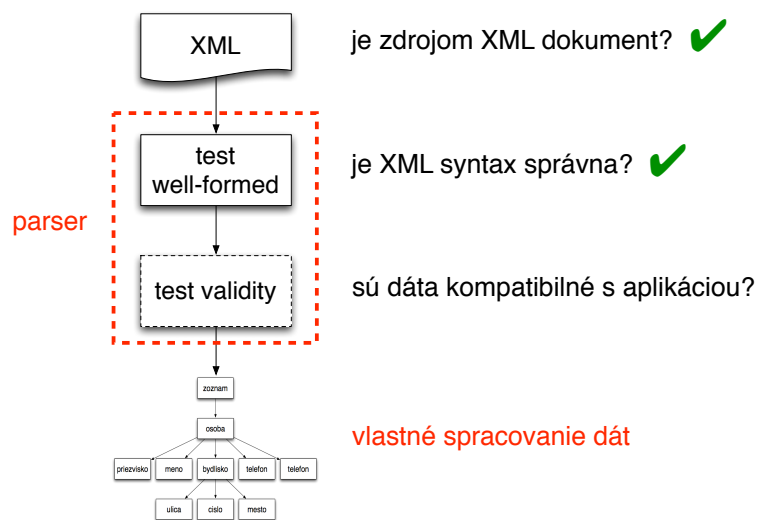
```
- <zoznam>
+ <osoba></osoba>
- <osoba>
  <meno>Maria</meno>
  <riezvisko>Jednoducha</riezvisko>
  + <bydlisko></bydlisko>
  <telefon typ="mobil">0808080808</telefon>
</osoba>
+ <osoba></osoba>
</zoznam>
```

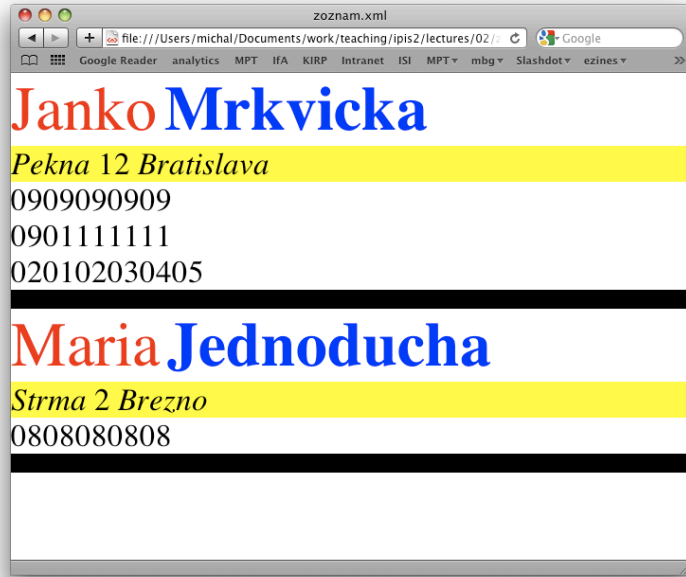
XML Parsing Error: mismatched tag. Expected: </meno>.
Location: file:///Users/michal/Documents/work/teaching/ipsis2/lectures/02/zoznam.xml
Line Number 4, Column 16:

```
<meno>Janko</MENO>
```



Spracovanie XML súborov





Čo je XML?

- XML hovorí o:
 - dátach
 - štruktúre dát
 - opise takejto štruktúry
- XML nehovorí o:
 - spracovaní dát
 - spôsobe zobrazenia dát (zobrazenie je iba určitým druhom spracovania)
- XML je "iba" spôsob zápisu dát, nič viac

XML štýly

```
<?xml version="1.0" encoding="UTF-8" ?>
<?xml-stylesheet type="text/css" href="zoznam.css"?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

XML štýly

označenie špeciálneho elementu

```
<?xml version="1.0" encoding="UTF-8" ?>
<?xml-stylesheet type="text/css" href="zoznam.css"?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

XML štýly

názov špeciálneho elementu

```
<?xml version="1.0" encoding="UTF-8" ?>
<?xml-stylesheet type="text/css" href="zoznam.css"?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

XML štýly

štýl bude typu CSS

```
<?xml version="1.0" encoding="UTF-8" ?>
<?xml-stylesheet type="text/css" href="zoznam.css"?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

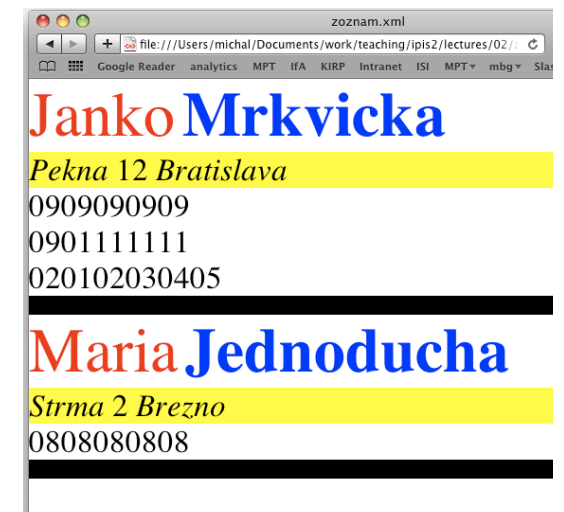
XML štýly

a nájdeš ho v tomto súbore

```
<?xml version="1.0" encoding="UTF-8" ?>
<?xml-stylesheet type="text/css" href="zoznam.css"?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0909090909</telefon>
  </osoba>
  <osoba>
    <meno>Maria</meno>
    <priezvisko>Jednoducha</priezvisko>
    <bydlisko>...</bydlisko>
    <telefon typ="mobil">0808080808</telefon>
  </osoba>
</zoznam>
```

zoznam.css

```
osoba {
  display: block;
  border-bottom: 10px solid black;
}
meno {
  font-size: 2em;
  color: red;
}
priezvisko {
  font-size: 2em;
  color: blue;
  font-weight: bold;
}
bydlisko {
  display: block;
  background: yellow;
}
ulica, mesto {
  font-style: italic;
}
telefon {
  display: block;
}
```



Čo sa stane po zmodikovaní XML súboru?

```
osoba {
  display: block;
  border-bottom: 10px solid black;
}
meno {
  font-size: 2em;
  color: red;
}
priezvisko {
  font-size: 2em;
  color: blue;
  font-weight: bold;
}
bydlisko {
  display: block;
  background: yellow;
}
ulica, mesto {
  font-style: italic;
}
telefon {
  display: block;
}
```



```
<zoznam>
<osoba>
  <meno>Janko</meno>
  <priezvisko>Mrkvicka</priezvisko>
  <bydlisko>...</bydlisko>
  <telefon typ="mobil">0909090909</telefon>
</osoba>
</zoznam>
```

Čo sa stane po zmodikovaní XML súboru?

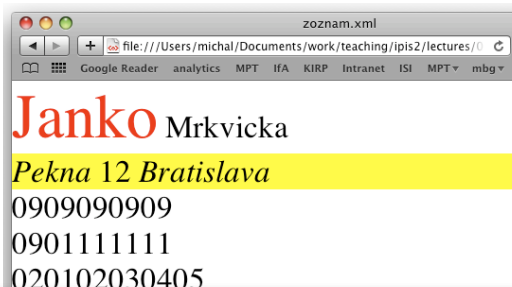
```
osoba {
  display: block;
  border-bottom: 10px solid black;
}
meno {
  font-size: 2em;
  color: red;
}
priezvisko {
  font-size: 2em;
  color: blue;
  font-weight: bold;
}
bydlisko {
  display: block;
  background: yellow;
}
ulica, mesto {
  font-style: italic;
}
telefon {
  display: block;
}
```



```
<zoznam>
<osoba>
  <meno>Janko</meno>
  <Priezvisko>Mrkvicka</Priezvisko>
  <bydlisko>...</bydlisko>
  <telefon typ="mobil">0909090909</telefon>
</osoba>
</zoznam>
```

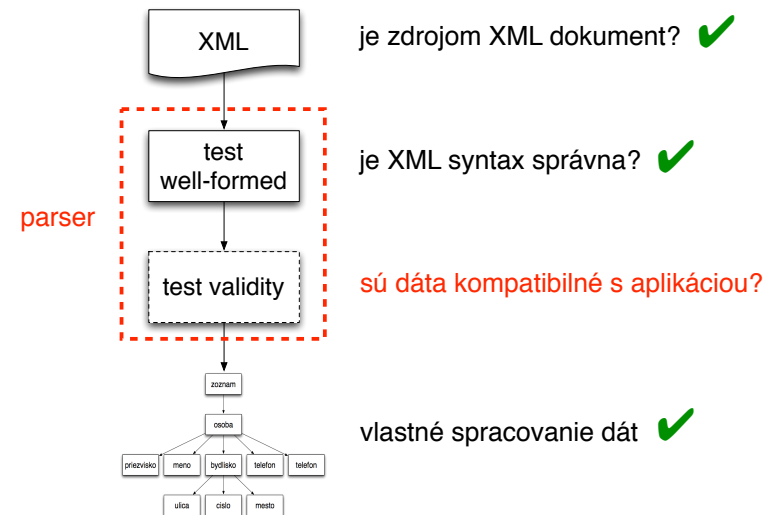
Čo sa stane po zmodikovaní XML súboru?

```
osoba {
  display: block;
  border-bottom: 10px solid black;
}
meno {
  font-size: 2em;
  color: red;
}
priezvisko {
  font-size: 2em;
  color: blue;
  font-weight: bold;
}
bydlisko {
  display: block;
  background: yellow;
}
ulica, mesto {
  font-style: italic;
}
telefon {
  display: block;
}
```



```
<zoznam>
<osoba>
  <meno>Janko</meno>
  <Priezvisko>Mrkvicka</Priezvisko>
  <bydlisko>...</bydlisko>
  <telefon typ="mobil">0909090909</telefon>
</osoba>
</zoznam>
```

Spracovanie XML súborov



Test validity

- Načo potrebujeme validitu?
 - aby sme ochránili aplikáciu pred dátami, ktoré nedokáže spracovať
 - napr. ak v telefónnom čísle uvedieme iné znaky ako čísla
 - predstavuje "zmluvu", ktorej ak sa držíme, máme istotu, že môžeme aplikácii poslať dáta a tá ich aj spracuje
- Čo sa chápe validáciou? Overenie, že...
 - v dokumente sa nachádzajú všetky povinné údaje
 - existuje závislosť (ak máme "meno", musí byť aj "priezvisko")
 - údaje sú uložené v potrebnej štruktúre (bydlisko: ulica, číslo, mesto)
 - údaje majú určitý typ (reťazec, číslo, dátum, ...)
 - atribúty sú definované správne
 - ...

Spôsoby overenia validity

- Prostredníctvom externého súboru, ktorý definuje, ako dokument môže vyzeráť a čo môže obsahovať
 - čo nie je povolené, je zakázané!
- Viacero "štandardných" techník:
 - Document type definition (DTD)
 - XML Schema (DTD tak, ako má byť)
 - Relax NG
 - Schematron

Príklad

```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>
      <ulica>Pekna</ulica>
      <cislo>12</cislo>
      <mesto>Bratislava</mesto>
    </bydlisko>
    <telefon typ="mobil">0909090909</telefon>
    <telefon typ="mobil">0901111111</telefon>
    <telefon typ="pevny">020102030405</telefon>
  </osoba>
</zoznam>
```

zoznam obsahuje aspoň jednu alebo viac osôb
pričom každá osoba má definované:

- jedno meno
- jedno priezvisko
- jedno bydlisko, ktoré má
 - jednu ulicu (nepovinný údaj)
 - jedno číslo domu
 - jedno mesto
- jeden alebo viac telefónov

Príslušné DTD

zoznam obsahuje aspoň jednu alebo viac osôb
pričom každá osoba má definované:

- jedno meno
- jedno priezvisko
- jedno bydlisko, ktoré má
 - jednu ulicu (nepovinný údaj)
 - jedno číslo domu
 - jedno mesto
- jeden alebo viac telefónov

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT zoznam (osoba+)>
<!ELEMENT osoba (meno,priezvisko,bydlisko,telefon+)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT bydlisko (ulica?,cislo,mesto)>
<!ELEMENT ulica (#PCDATA)>
<!ELEMENT cislo (#PCDATA)>
<!ELEMENT mesto (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```


Príslušné DTD

DTD je tiež len XML dokument, musí byť teda príslušne označený

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT zoznam (osoba+)>
<!ELEMENT osoba (meno,priezvisko,bydlisko,telefon)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT bydlisko (ulica?,cislo,mesto)>
<!ELEMENT ulica (#PCDATA)>
<!ELEMENT cislo (#PCDATA)>
<!ELEMENT mesto (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

Príslušné DTD

DTD definuje elementy, ktoré budú v dokumente povolené

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT zoznam (osoba+)>
<!ELEMENT osoba (meno,priezvisko,bydlisko,telefon)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT bydlisko (ulica?,cislo,mesto)>
<!ELEMENT ulica (#PCDATA)>
<!ELEMENT cislo (#PCDATA)>
<!ELEMENT mesto (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

Príslušné DTD

názov povoleného elementu

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT zoznam (osoba+)>
<!ELEMENT osoba (meno,priezvisko,bydlisko,telefon)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT bydlisko (ulica?,cislo,mesto)>
<!ELEMENT ulica (#PCDATA)>
<!ELEMENT cislo (#PCDATA)>
<!ELEMENT mesto (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

Príslušné DTD

definícia obsahu, ktorý sa môže v konkrétnom elemente vyskytovať

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT zoznam (osoba+)>
<!ELEMENT osoba (meno,priezvisko,bydlisko,telefon)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT bydlisko (ulica?,cislo,mesto)>
<!ELEMENT ulica (#PCDATA)>
<!ELEMENT cislo (#PCDATA)>
<!ELEMENT mesto (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

Príslušné DTD

definícia sekvencie detských elementov

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT zoznam (osoba+)>
<!ELEMENT osoba (meno, priezvisko, bydlisko, telefon+)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT bydlisko (ulica?, cislo, mesto)>
<!ELEMENT ulica (#PCDATA)>
<!ELEMENT cislo (#PCDATA)>
<!ELEMENT mesto (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

Príslušné DTD

a zároveň definuje počet opakovaní

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT zoznam (osoba+)>
<!ELEMENT osoba (meno, priezvisko, bydlisko, telefon+)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT bydlisko (ulica?, cislo, mesto)>
<!ELEMENT ulica (#PCDATA)>
<!ELEMENT cislo (#PCDATA)>
<!ELEMENT mesto (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

Obsah typu "#PCDATA"

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

Valid	Invalid
<pre><A>Nejaky text</pre>	<pre><A> <A><C></C> <A>Nejaky text</pre>

Obsah typu "detský element"

```
<!ELEMENT A (B)>
```

Valid	Invalid
<pre><A>... <A> ... </pre>	<pre><A> <A>Nejaky text <A>Text... <A><C>...</C> <C>...</C></pre>

Obsah typu "sekvencia detských elementov"

<!ELEMENT A (B,D,C)>

Valid	Invalid
<pre><A> ... <D>...</D> <C>...</C> </pre>	<pre><A> <A>Nejaky text <A> ... <C>...</C> <D>...</D> <A> ... <D>...</D> </pre>

? = operátor "žaden alebo jeden"

<!ELEMENT A (B?)>

Valid	Invalid
<pre><A> <A> ... </pre>	<pre><A>Nejaky text <A><C>...</C> <A> <A> ... <D>...</D> </pre>

+ = operátor "jeden alebo viac"

<!ELEMENT A (B+)>

Valid	Invalid
<pre><A> ... <A> </pre>	<pre><A> <A>Nejaky text <A><C>...</C> <A> ... <D>...</D> ... </pre>

* = operátor "žaden, jeden alebo viac"

<!ELEMENT A (B*)>

Valid	Invalid
<pre><A> <A> ... <A> </pre>	<pre><A>Nejaky text <A><C>...</C> <A> ... <D>...</D> ... </pre>

I = operátor “bud’, alebo”

<!ELEMENT A (B|C)>

Valid	Invalid
<pre><A> ... <A> <C>...</C> </pre>	<pre><A> <A> <A> ... <C>...</C> <A> ... <D>...</D> </pre>

I = operátor “bud’, alebo”

<!ELEMENT A (#PCDATA|B)>

Valid	Invalid
<pre><A>Nejaky text <A> ... </pre>	<pre><A>Nejaky text... <A> <C>...</C> <A> </pre>

() = operátor “zoskupenie”

<!ELEMENT A ((B|C),D)>

Valid	Invalid
<pre><A> ... <D>...</D> <A> <C>...</C> <D>...</D> </pre>	<pre><A> <A>Nejaky text <A> <C>...</C> <A> ... <C>...</C> </pre>

Ľubovoľná kombinácia

<!ELEMENT A (B*|C)>

Valid	Invalid
<pre><A> <A>... <A> <A> <C>...</C> </pre>	<pre><A><D>...</D> <A> ... <C>...</C> <A> <C>...</C> <C>...</C> </pre>

Ľubovoľná kombinácia

<!ELEMENT A ((B*|C*),D?,E+)>

Valid	Invalid
<pre><A> <E>...</E> <E>...</E> <E>...</E> <A> <C>...</C> <D>...</D> <E>...</E> </pre>	<pre><A>Nejaký text <A> ... <C>...</C> <A> <C>...</C> <D>...</D> <D>...</D> <E>...</E> </pre>

Pozor na umiestnenie modifikátora

<!ELEMENT A (B | C)*>

<!ELEMENT A (B* | C*)>

<!ELEMENT A (B,C)*>

<!ELEMENT A (B*,C*)>

Pozor na umiestnenie modifikátora

<!ELEMENT A ((B | C)*)>

Valid	Invalid
<pre><A> ... <C>...</C> <C>...</C> <C>...</C> <C>...</C> </pre>	<pre><A>Nejaký text <A> <D>...</D> </pre>

Pozor na umiestnenie modifikátora

<!ELEMENT A ((B* | C*)*)>

Valid	Invalid
<pre><A> <A> <C>...</C> <C>...</C> <C>...</C> </pre>	<pre><A> ... <C>...</C> <A> <C>...</C> </pre>

Pozor na umiestnenie modifikátora

<!ELEMENT A ((B,C)*)>

Valid	Invalid
<pre><A> ... <C>...</C> <A> ... <C>...</C> ... <C>...</C> ... <C>...</C> </pre>	<pre><A> ... <A> <C>...</C> <A> ... <C>...</C> ... </pre>

Pozor na umiestnenie modifikátora

<!ELEMENT A ((B*,C*))>

Valid	Invalid
<pre><A> ... <A> <A> <C>...</C> <C>...</C> </pre>	<pre><A> ... <C>...</C> ... <C>...</C> </pre>

DTD pre telefónny zoznam

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT zoznam (osoba+)>
<!ELEMENT osoba (meno,priezvisko,bydlisko,telefon+)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT bydlisko (ulica?,cislo,mesto)>
<!ELEMENT ulica (#PCDATA)>
<!ELEMENT cislo (#PCDATA)>
<!ELEMENT mesto (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

```
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>
      <cislo>12</cislo>
      <mesto>Bratislava</mesto>
    </bydlisko>
  </osoba>
</zoznam>
```

DTD pre zoznam

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT zoznam (osoba+)>
<!ELEMENT osoba (meno,priezvisko,bydlisko,telefon+)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT bydlisko (ulica?,cislo,mesto)>
<!ELEMENT ulica (#PCDATA)>
<!ELEMENT cislo (#PCDATA)>
<!ELEMENT mesto (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

```
<?xml version="1.0" encoding="UTF-8" ?>
<zoznam>
  <osoba>
    <meno>Janko</meno>
    <priezvisko>Mrkvicka</priezvisko>
    <bydlisko>
      <cislo>12</cislo>
      <mesto>Bratislava</mesto>
    </bydlisko>
  </osoba>
</zoznam>
```

DTD: sumarizácia

- DTD = spôsob vynútenia si istej štruktúry XML dokumentu
- DTD validuje správnosť daného XML (zmluva)
- Dva pohľady na DTD:
 - tvorca aplikácie: vytvára a publikuje DTD na to, aby iní mohli poskytnúť našej aplikácii kompatibilné dáta
 - tvorca dát: číta DTD a na základe neho vytvorí vhodný XML dokument
- Nevýhody DTD:
 - obsahom elementu môžu byť iba ďalšie elementy alebo #PCDATA (#PCDATA nerozlišuje medzi reťazcom, číslom, dátumom, ...)
 - iba 4 úrovne opakovania:
 - práve raz
 - 0 alebo 1 raz
 - jedenkrát alebo viac
 - 0, 1 alebo viackrát

Na ďalších prednáškach...

- Prostredníctvom externého súboru, ktorý definuje, ako dokument môže vyzeráť a čo môže obsahovať
 - čo nie je povolené, je zakázané!
- Viacero “štandardných” techník:
 - Document type definition (DTD)
 - XML Schema (DTD tak, ako má byť)
 - Relax NG
 - Schematron

Matica

- Požiadavky na DTD:
 - hlavný koreňový element: <matica>
 - matica obsahuje aspoň jeden alebo viac riadkov
 - každý riadok obsahuje aspoň jeden alebo viac stĺpcov
 - každý stĺpec obsahuje reťazec

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT matica (riadok*)>
<!ELEMENT riadok (stlpec*)>
<!ELEMENT stlpec (#PCDATA)>
```

```
<?xml version="1.0" encoding="UTF-8" ?>
<matica>
  <riadok>
    <stlpec>1.2</stlpec><stlpec>3.4</stlpec>
  </riadok>
  <riadok>
    <stlpec>5.6</stlpec>
  </riadok>
</matica>
```

Prechodová charakteristika

- Požiadavky na DTD:
 - hlavný koreňový element: <prech>
 - obsahuje jeden element <system> a jeden element <matica>
 - systém je buď “prenos” alebo “stavovy_opis”
 - každý riadok obsahuje aspoň jeden alebo viac stĺpcov
 - pre prenos sa definuje “citateľ” a “menovateľ” ako “matica”
 - pre stavový opis sa definujú matice “A”, “B”, “C”, “D”

- Požiadavky na DTD:

- hlavný koreňový element: <prech>
- obsahuje jeden element <system> a jeden element <matica>
- systém je buď "prenos" alebo "stavovy_opis"
- každý riadok obsahuje aspoň jeden alebo viac stĺpcov
- pre prenos sa definuje "citateľ" a "menovateľ" ako "matica"
- pre stavový opis sa definujú matice "A", "B", "C", "D"

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT prech (system,matica)>
<!ELEMENT system (prenos | stavovy_opis)>
<!ELEMENT prenos (citateľ,menovateľ)>
<!ELEMENT citateľ (matica)>
<!ELEMENT menovateľ (matica)>
<!ELEMENT stavovy_opis (A,B,C,D)>
<!ELEMENT A (matica)>
<!ELEMENT B (matica)>
<!ELEMENT C (matica)>
<!ELEMENT D (matica)>
<!ELEMENT matica (riadok*)>
<!ELEMENT riadok (stĺpec*)>
<!ELEMENT stĺpec (#PCDATA)>
```

Databáza kníh a vydavateľov

- Požiadavky na DTD:

- hlavný koreňový element: <db>
- môže obsahovať jeden alebo viac kníh a jeden alebo viac vydavateľov
- kniha má elementy: titulok, rok publikovania, vydavateľstvo (všetky povinné)
- vydavateľ je buď reťazec (meno+priezvisko), alebo "meno" a "priezvisko" zvlášť, alebo záznam v tvare "firma"
- "firma" povinne obsahuje názov, sídlo a alternatívne aj telefónny kontakt (všetko ako reťazce)

- Požiadavky na DTD:

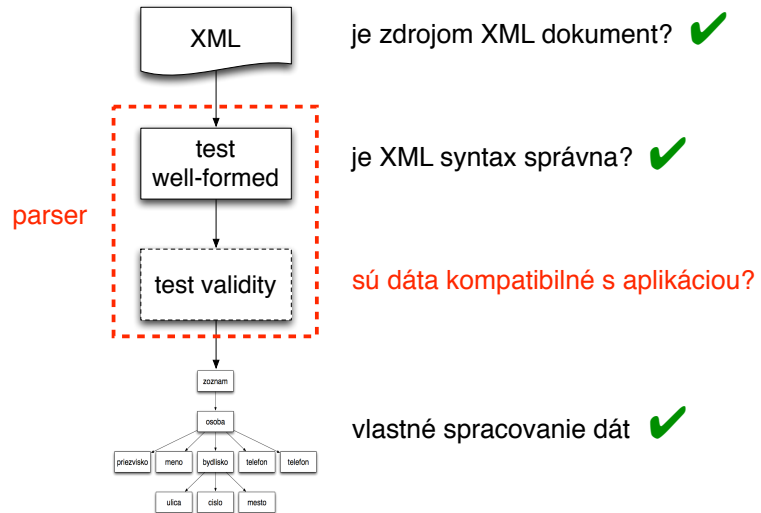
- hlavný koreňový element: <db>
- môže obsahovať jeden alebo viac kníh a jeden alebo viac vydavateľov
- kniha má elementy: titulok, rok publikovania, vydavateľstvo (všetky povinné)
- vydavateľ je buď reťazec (meno+priezvisko), alebo "meno" a "priezvisko" zvlášť, alebo záznam v tvare "firma"
- "firma" povinne obsahuje názov, sídlo a alternatívne aj telefónny kontakt (reťazce)

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT db (kniha+,vydavateľ+)>
<!ELEMENT kniha (titulka,rok,vydavateľstvo)>
<!ELEMENT titulka (#PCDATA)>
<!ELEMENT rok (#PCDATA)>
<!ELEMENT vydavateľstvo (#PCDATA)>
<!ELEMENT vydavateľ (#PCDATA|(meno,priezvisko)|firma)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT firma (nazov,sídlo,telefon?)>
<!ELEMENT nazov (#PCDATA)>
<!ELEMENT sídlo (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

```
<?xml version="1.0" encoding="UTF-8" ?>
<!ELEMENT db (kniha+,vydavateľ+)>
<!ELEMENT kniha (titulka,rok,vydavateľstvo)>
<!ELEMENT titulka (#PCDATA)>
<!ELEMENT rok (#PCDATA)>
<!ELEMENT vydavateľstvo (#PCDATA)>
<!ELEMENT vydavateľ (#PCDATA|(meno,priezvisko)|firma)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
<!ELEMENT firma (nazov,sídlo,telefon?)>
<!ELEMENT nazov (#PCDATA)>
<!ELEMENT sídlo (#PCDATA)>
<!ELEMENT telefon (#PCDATA)>
```

```
<?xml version="1.0" encoding="UTF-8" ?>
<db>
  <vydavateľ>Jan Knihovy</vydavateľ>
  <vydavateľ>
    <meno>Jan</meno>
    <priezvisko>Knihovy</priezvisko>
  </vydavateľ>
  <vydavateľ>
    <firma>
      <nazov>Knihy noveho milenia</nazov>
      <sídlo>Knizna 24, Bratislava</sídlo>
    </firma>
  </vydavateľ>
</db>
```


Spracovanie XML súborov



Ako XML súbor povie, voči ktorému DTD sa má validovať?

verifikator_matica.dtd

```
<?xml version="1.0" encoding="UTF-8" ?>  
<!ELEMENT matica (riadok*)>  
<!ELEMENT riadok (stlpec*)>  
<!ELEMENT stlpec (#PCDATA)>
```

data.xml

```
<?xml version="1.0" encoding="UTF-8" ?>  
<!DOCTYPE matica SYSTEM "verifikator_matica.dtd">  
<matica>  
  <riadok>  
    <stlpec>1.2</stlpec><stlpec>3.4</stlpec>  
  </riadok>  
  <riadok>  
    <stlpec>5.6</stlpec>  
  </riadok>  
</matica>
```

Spracovanie XML súborov

