Uvod u veb i internet tehnologije

Matematički fakultet

vladaf@matf.bg.ac.rs

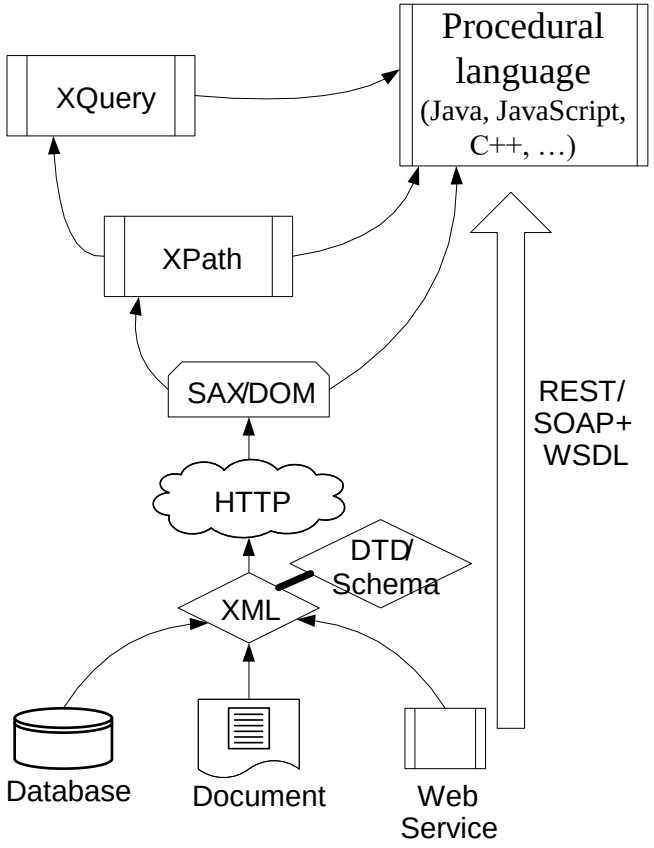
Jezici za obeležavanje XML

Matematički fakultet

vladaf@matf.bg.ac.rs

Šta je XML?

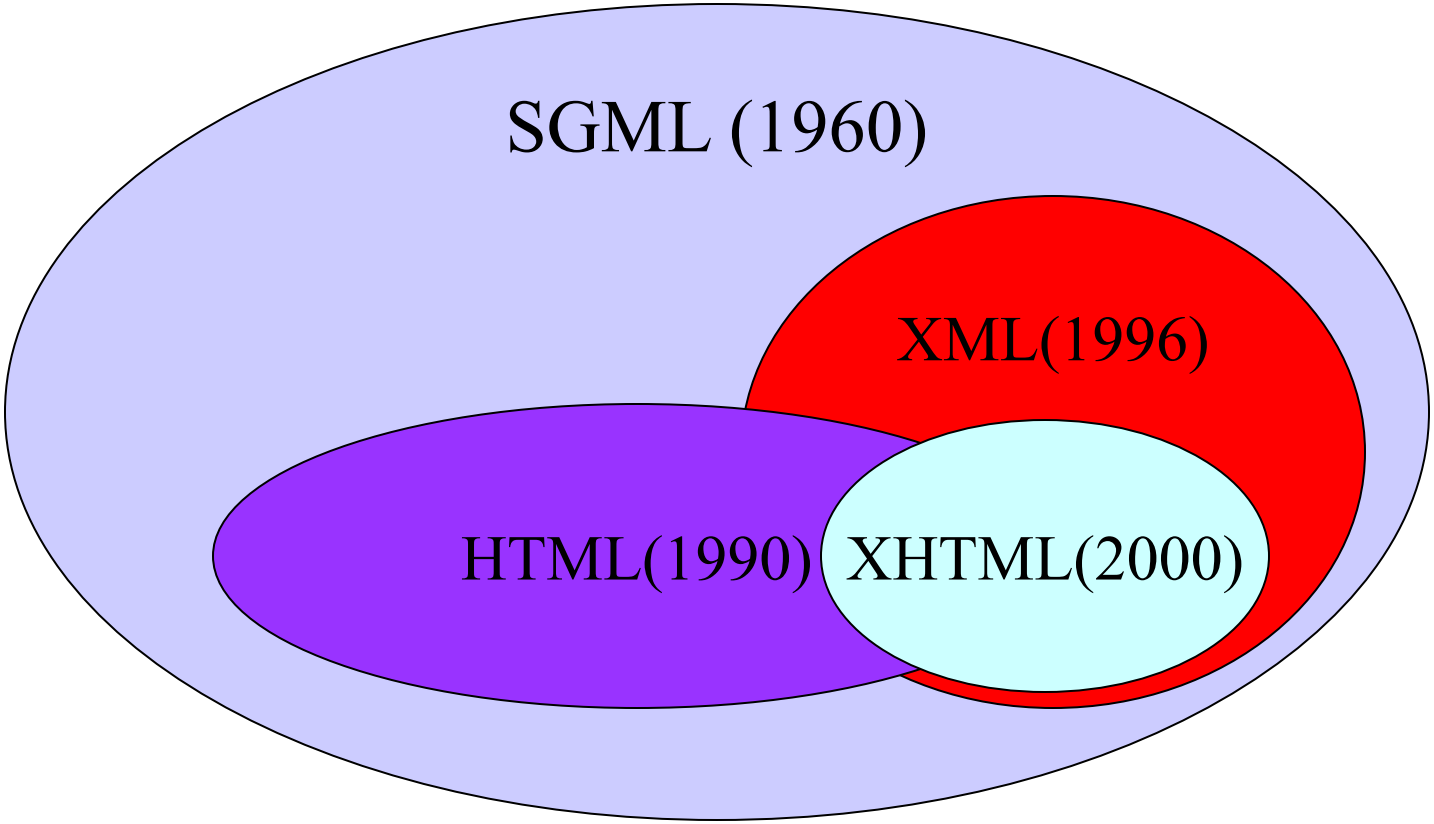
- Hijerarhijski format čitljiv za čoveka
  - Jezik “potomak” HTML-a, koji se uvek može parsirati
  - “Lingua franca” za podatke: služi za čuvanje dokumenata strukturisanih podataka
  - Smešani su podaci i struktura
- Jezgro šireg ekosistema
  - Podaci – XML
  - Shema – DTD i XML Shema
  - Programerski pristup – DOM i SAX
  - Upiti – XPath, XSLT, XQuery
  - Distribuisano programiranje – veb servisi



Matematički fakultet

vladaf@matf.bg.ac.rs

Istorija: SGML vs. HTML vs. XML



Matematički fakultet

vladaf@matf.bg.ac.rs

Zašto XML

- Olakšava težnju da se „sadržaj“ razdvoji od „prezentacije”
  - Prezentacija obezbeđuje lepotu pri posmatranju
  - Sadžaj se može interpretirati od strane računara, a za računare prezentacija predstavlja hendikep
- Semantičko označavanje podataka
- XML je „polu-struktuiran“

```
<parents>
  <parent name="Jean" >
    <son>John</son>
    <daughter>Joan</daughter>
    <daughter>Jill</daughter>
  </parent>
  <parent name="Feng">
    <daughter>Ella</daughter>
  </parent>
...
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Zašto XML (2)

```
<book year="1967">
   <title>Politics of experience</title>
     <author>
              <firstname>Ronald</firstname>
              <lastname>Laing</lastname>
     </author>
</book>
```

Informacije o knjizi sačuvane u XML formatu

- Informacija je:
  1. razdvojena od prezentacije, pa
  2. isečena u male delove, i na kraju
  3. označena sa semantičkim značenjem
- Informacija u ovom formatu se lako može procesirati računarima
- XML opisuje samo sintaksu, a ne apstraktni logički model podataka

Matematički fakultet

vladaf@matf.bg.ac.rs

Ključni pojmovi XML-a

- To su:
  - Dokumenti
  - Elementi
  - Atributi
  - Deklaracije prostora imena
  - Tekst
  - Komentari
  - Instrukcije procesiranja
- Svi ovi pojmovi su nasleđeni iz SGML-a

Matematički fakultet

vladaf@matf.bg.ac.rs

Anatomija XML-a

Instrukcija procesiranja

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<dblp>
  <mastersthesis mdate="2002-01-03" key="ms/Brown92">
    <author>Kurt P. Brown</author>
    <title>PRPL: A Database Workload Specification
   Language</title>
    <year>1992</year>
    <school>Univ. of Wisconsin-Madison</school>
  </mastersthesis>
  <article mdate="2002-01-03" key="tr/dec/SRC1997-018">
    <editor>Paul R. McJones</editor>
    <title>The 1995 SQL Reunion</title>
    <journal>Digital System Research Center Report</journal>
    <volume>SRC1997-018</volume>
    <year>1997</year>
    <ee>db/labs/dec/SRC1997-018.html</ee>
    <ee>http://www.mcjones.org/System_R/SQL_Reunion_95/</ee>
  </article>
```

Otvarajuća etiketa

Element

Atribut

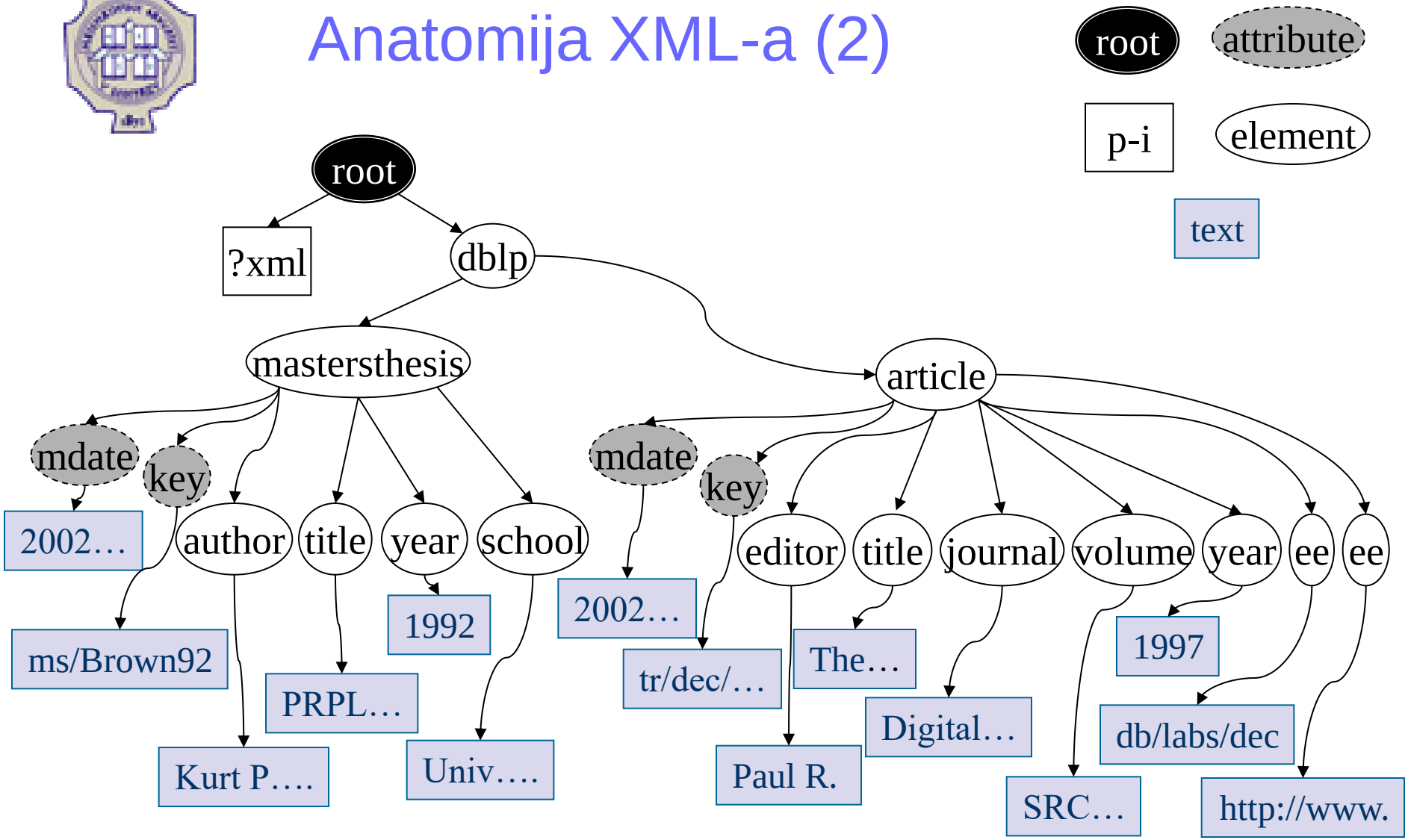
Zatvarajuća etiketa

https://dblp.uni-trier.de/faq/How+to+parse+dblp+xml.html

Matematički fakultet

vladaf@matf.bg.ac.rs

Anatomija XML-a (2)



Matematički fakultet

vladaf@matf.bg.ac.rs

Anatomija XML-a (3)

- XML lako čuva relacije
  Primer: Relacija student-course-grade

sid	cid	exp-grade
1	570103	B
23	550103	A

```
<student-course-grade>
   <tuple><sid>1</sid><cid>570103</cid>
        <exp-grade>B</exp-grade>
   </tuple>
   <tuple><sid>23</sid><cid>550103</cid>
        <exp-grade>A</exp-grade>
   </tuple>
</student-course-grade>
```

ili

```
<student-course-grade>
   <tuple sid="1" cid="570103" exp-grade="B"/>
   <tuple sid="23" cid="550103" exp-grade="A"/>
</student-course-grade>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Elementi

```
<book year=“1967”>
   <title>Politics of experience</title>
     <author>
               <firstname>Ronald</firstname>
               <lastname>Laing</lastname>
     </author>
</book>
```

Elemente karakteriše:

- Ugnježdena struktura
- Drvoidna struktura
- Redosled je važan
- Sadrži samo znakove, a ne cele brojeve, itd.

Matematički fakultet

vladaf@matf.bg.ac.rs

Elementi (2)

- Obuhvaćeni su etiketama
  - Otvarajuća etiketa: npr. `<bibliography>`
  - Zatvarajuća etiketa: npr. `</bibliography>`
  - Elementi bez sadržaja (prazni): npr. `<bibliography />` je skraćenica za `<bibliography> </bibliography>`

- Elementi mogu biti ugnježdeni
  `<bib> <book> Wilde Wutz </book> </bib>`
- Elementi koji su ugnježdeni mogu biti višečlani
  `<bib> <book> ... </book> <book> ... </book> </bib>`
- U tim slučajevima, redosled je veoma važan!

- Dokumenti moraju biti dobro formirani
  `<a> <b> </a> </b>` nije dopušteno!
  `<a> <b> </b>` nije dopušteno!

Matematički fakultet

vladaf@matf.bg.ac.rs

Atributi

- Atributi su pridruženi elementima
  Primer:
```
<book  price = “55“ year = “1967“ >
    <title> ... </title>
    <author> ... </author>
</book>
```
- Elementi mogu sadržavati samo atribute (unutar otvarajuće etikete)
  Primer: `<person name = “Wutz“ age = 33“/>`
- Imena atributa moraju biti jedinstvena!
  Primer: Nelegalna je sledeća konstrukcija
```
<person name = “Wilde“ name = “Wutz“/>
```
- Koja je razlika između umetnutog elementa i atributa?
  Da li su atributi korisni?
- Odluka pri modeliranju: da li da name bude atribut ili element ugnježden u element person?
  Šta da se radi sa elementom age?

Matematički fakultet

vladaf@matf.bg.ac.rs

Tekst i izmešani sadržaj

- Tekst se može javiti unutar sadržaja elementa

  Primer:

  ```
  <title>The politics of experience</title>
  ```

- Tekst može biti izmešan sa ostalim elementima ugnježdenim u dati element

  Primer:

  ```
  <title>The politics of <em>experience</em></title>
  ```

- Karakteristike izmešanog sadržaja:
  - Veoma je koristan za podatke u obliku dokumenata, tj. rečenica
  - Nema potreba za mešanim sadržajem u scenarijima „procesiranja podataka“, jer se tada obično obrađuju entiteti i relacije
  - Ljudi komuniciraju rečenicama, a ne entitetima i relacijama. XML omogućuje da se sačuva struktura prirodnog jezika, uz dodavanje semantičkih oznaka koje mogu računarski interpretirane

Matematički fakultet

Prelaz između prirodnog jezika, polu-struktuiranih i struktuiranih podataka

vladaf@matf.bg.ac.rs

1. Prirodni jezik:
   Dana said that the book entitled “The politics of experience“ is really excellent !
2. Polu-strukturisani podaci (tekst):
   <citation author=“Dana“> The book entitled “The politics of experience“ is really excellent ! </citation>
3. Polu-strukturisani podaci (mešani sadržaj):
   <citation author=“Dana“> The book entitled <title> The politics of experience</title> is really excellent ! </citation>
4. Strukturisani podaci:

```
<citation>
  <author>Dana</author>
   <aboutTitle>The politics of experience</aboutTitle>
   <rating> excellent</rating>
</citation>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Sekcija CDATA

- Ponekad treba sačuvati originalne znake, a ne interpretirati njihova označavanja
- Sekcija CDATA određuje da se sadržaj unutar nje ne parsira kao XML
- Primer (poruka Hello,world! je označena):

```
<message>
    <greeting>Hello,world!</greeting>
</message>
```

- Primer (označena poruka neće biti parsirana kao XML):

```
<message>
    <![CDATA[<greeting>Hello, world!</greeting>]]>
</message>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Komentari, instrukcije za procesiranje i prolog

- Komentar je tekst između <!-- i -->
  Primer: `<!-- ovo je komentar -->`
- Instrukcije za procesiranje
  One ne sadrže podatke, već ih interpretira procesor
  Sastoje se od para reči meta sadržaj, razdvojenih zarezom, kojima prethodi <?, a iza kojih sledi ?>
  Primer: U instrukciji za procesiranje <?pause 10 secs ?> pause je meta, a 10secs je sadržaj
  - Reč xml je rezervisana reč za metu, koja služi za označavanje prologa
- Prolog
  `<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>`
  Napomene:
  - Atribut standalone određuje da li postoji DTD
  - Encoding je obično Unicode.

Matematički fakultet

vladaf@matf.bg.ac.rs

Deklaracija korišćenja belina

- Beline predstavljaju neprekidnu sekvencu znakova Space, Tab i Return
- Za kontrolu korišćenja belina služi specijalan atribut xml:space
- Primer čitljivog XML-a (koji sadrži beline):
  <book xml:space="preserve" >
  <title>The politics of experience</title>
  <author>Ronald Laing</author>
  </book>
- Primer efikasnog (računarski čitljivog) XML-a:
  <book xml:space="default" ><title>The politics of experience</title><author>Ronald Laing</author></book>
- U drugom primeru su (u odnosu na prvi) performanse ubrzane sa faktorom 2

Matematički fakultet

vladaf@matf.bg.ac.rs

Prostori imena

- Omogućavaju integraciju podataka iz različitih izvora
- Omogućavaju integraciju različitih XML rečnika (tj. prostora imena)
- Svaki „rečnik“ ima jedinstven ključ, identifikovan URI-jem
- Isto lokalno ime iz različitih rečnika može imati
  - Različita značenja
  - Različite pridružene strukture
- Kvalifikovana imena (Qualified Names - Qname) služe za prigruživanje imena „rečniku“
- Kvalifikovana imena se odnose na sve čvoroveXML dokumenta koji imaju imena (atribute, elemente, Instrukcije za procesiranje)

Matematički fakultet

vladaf@matf.bg.ac.rs

Prostori imena (2)

- Način korišćenja
  - Povezivanje (tj. prefiks i URI) se uvode u otvarajućoj etiketi elementa
  - Kasnije se za opis koristi prefiks, a ne URI
  - Postoje podrazumevani prostori imena, pa je prefiks opcionalan
  - Prefiks se od lokalnog imena razdvaja dvotačkom, tj. znakom **:**
- Prostori imena se zapisuju slično atributima
  - Identifikuju se bilo sa “xmlns:prefix“, ili sa “xmlns“ (ako se radi o podrazumevanom prostoru imena)
  - Dati prefiks se, korišćenjem prostora imena, pozezuje sa URI-jem
- Opseg prostora imena je ceo elemenat u kome je taj prostor imena deklarisan – uključuje sam elemenat, njegove atribute i sve elemente koji su ugnježđeni u njega
- Primer:

```
<ns:a xmlns:ns=“someURI“ ns:b=“foo“>
    <ns:b>content</ns:b>
</ns:a>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Podrazumevani prostori imena

- Kad se specificira podrazumevani prostor imena, ne koristi se prefiks

```
<a  xmlns="someURI"  >
   <b/> <!-- a and b are in the someURI namespace! -->
</a>
```

- Podrazumevani prostor imena se odnosi samo na ugnježdene elemente, a ne na atribute

```
<a  xmlns="someURI"  c = "not in someURI  namespace">
   <b/> <!-- a and b are in the someURI namespace! -->
</a>
```

Matematički fakultet
vladaf@matf.bg.ac.rs

Primer rada sa prostorima imena

- Tanjiri iz servisa za ručavanje i satelitski „tanjiri“
  - Neka „rečnik“ DQ1 definiše dish for china
    - Diameter, Volume, Decor, ...
  - Neka „rečnik“ DQ2 definiše dish for satellites
    - Diameter, Frequency
  - Postavlja se pitanje: Koliko ovde ima „tanjira“?
  - To pitanje se svodi na jedno od sledeća dva pitanja:
    1. „How many dishes are there?“
       or
    2. „How many dishes are there?“

Matematički fakultet
vladaf@matf.bg.ac.rs

Primer rada sa prostorima imena (2)

XML opis „tanjira“ iz pribora za ručavanje:

```
<gs:dish  xmlns:gs = “http://china.com“ >
    <gs:dm gs:unit = “cm“>20</gs:dm>
    <gs:vol gs:unit = “l“>5</gs:vol>
    <gs:decor>Meissner</gs:decor>
</gs:dish>
```

XML opis „tanjira“ za prijem satelitskog signala:

```
<sat:dish xmlns:sat = “http://satelite.com“ >
    <sat:dm>200</sat:dm>
    <sat:freq>20-2000MHz</sat:freq>
</sat:dish>
```

Matematički fakultet
vladaf@matf.bg.ac.rs

Primer rada sa prostorima imena (3)

XML opis „tanjira“ za ručavanje, gde su merene jedinice iz drugog prostora imena:

```
<gs:dish  xmlns:gs = “http://china.com“ xmlns:uom =
  “http://units.com“>
   <gs:dm uom:unit = “cm“>20</gs:dm>
   <gs:vol  uom:unit = “l“>5</gs:vol>
   <gs:decor>Meissner</gs:decor>
   <comment>This is an unqualified element name</comment>
</gs:dish>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Primer rada sa prostorima imena (4)

- Kod prostora imena se razlikuju:
  - Vezivanje prostora imena sa URI-jem
  - Kvalifikovanje imena

Podrazumevani prostor imena se odnosi na sva imena koja nisu kvalifikovana

```
<root xmlns="http://www.first.com/aspace" xmlns:otherns="…">
 <myns:tag xmlns:myns="http://www.fictitious.com/mypath">
  <thistag>is in the default namespace
          (www.first.com/aspace)</thistag>
  <myns:thistag>is in myns</myns:thistag>
  <otherns:thistag>is a different tag in
  otherns</otherns:thistag>
 </myns:tag>
</root>
```

Definiše "otherns" kvalifikator

Matematički fakultet

vladaf@matf.bg.ac.rs

Primeri XML podataka

- XHTML (pregledač/prezentacija)
- RSS (blogovi)
- UBL (univerzalni poslovni jezik)
- HealthCare Level 7 (medicinski podaci)
- XBRL (finansijski podaci)
- XMI (meta podaci)
- XQueryX (programo)
- XForms, FXML (forme)
- SOAP (poruke za komunikaciju)
- Microsoft ADO.Net (baze podataka)
- Microsoft Office, Powerpoint (dokumenti)

Matematički fakultet

vladaf@matf.bg.ac.rs

Struktuiranje XML-a

- Za razliku od drugih formata podataka, XML je veoma fleksibilan i elementi se mogu umetati na različite načine
- Može se početi sa pisanjem XML-a koji predstavlja podatke i bez prethodnog dizajniranja strukture
  - Tako se radi kod relacionih baza podataka ili kod Java klasa
- Međutim, strukturisanje ima veliki značaj:
  - Podspešuje pisanje aplikacija koje procesiraju podatke
  - Ograničava podatke na one koje su korektni za datu aplikaciju
  - Definiše „a priori“ protokol o podacima koji se razmenjuju između učesnika u komunikaciji
- Struktura XML-a modelira podatke i sadrži:
  - Definicije struktura
  - Definicije tipova
  - Podrazumevane vrednosti

Matematički fakultet

vladaf@matf.bg.ac.rs

Struktuiranje XML-a (2)

Karakteristike struktuiranja:

- Nije formalizovano na način kako je to urađeno kod relacionih baza podataka
  - Ono je obično zasnovano na strukturi koja se već nalazi u podacima, npr relacionoj SUBP ili raširenoj elektronskoj tabeli
- Šri struktuiranju se XML drvo orjentiše prema „centralnim” objektima
- Velika dilema: element ili atribut
  - Element se koristi za osobinu koja sadrži svoje osobine ili kada se može očekivati da će biti više takvih unutar elementa koji ih sadrži
  - Atribute se koristi kada se radi o jednoj osobini – mada je OK da se i tada koristi element!

Matematički fakultet
vladaf@matf.bg.ac.rs

Istorija i uloga jezika za opis strukture XML-a

- Postoji nekoliko standardnih jezika za opis strukture XML-a
  DTD-ovi, XML Shema, RelaxNG
- Jezici koji opisuju strukturu se definišu ortogonalno u odnosu na sam XML
- Kod XML-a su opis strukture i podaci potpuno razdvojeni
  - Podaci mogu postojati i uz opis strukture i bez njega
  - Podaci mogu postojati i uz više opisa struktura
  - Evolucija strukture veoma retko dovodi do evolucije podataka
  - Može se raditi tako što se struktura definiše pre podataka, a može i tako što se struktura ekstrakuje iz podataka
- Jezici za opis strukture čine da XML postaje pravi izbor za manipulaciju polu-strukturisanim podacima, podacima koji brzo evoluiraju ili podacima koji su podesivi u velikoj meri

Matematički fakultet
vladaf@matf.bg.ac.rs

Korektnost XML dokumenata

1. Dobro formirani dokumenti
   - Kod njih se verifikuju samo osnovna XML ograničenja, npr. <a>
2. Validni dokumenti
   - Kod njih se verifikuju dodatna ograničenja opisana u nekom od jezika za opis strukture (npr. DTD, XML shema)

- XML dokumenti koji nisu dobro formirani ne mogu biti procesirani
- XML dokumenti koji nisu validni ipak mogu biti procesirani (mogu se upitima izvlačiti podaci, mogu biti transformisani, itd.)

Matematički fakultet

vladaf@matf.bg.ac.rs

XML i DTD

Matematički fakultet

vladaf@matf.bg.ac.rs

DTD

- DTD je deo originalne XML 1.0 specifikacije
- DTD opisuje “gramatiku” za XML datoteku
  - Deklaracije elemenata: pravila i ograničenja koja opisuju dopuštene načine ugnježdavanja elemenata
  - Attributes lists: opisuje koji su atributi dopušteni nad kojim elementima
  - Dodatna ograničenje na vrednosti elemenata i atributa
  - Koji je elemenat koreni čvor XML strukture
- Provera strukturnih ograničenja pomoću DTD se naziva DTD validacija (određuje se da li je XML dokument validan ili invalidan)
- Zbog svojih ograničenja, DTD se sada relativno retko koristi za validaciju XML dokumenata

Matematički fakultet
vladaf@matf.bg.ac.rs

Referisanje na DTD u okviru XML-a

- Nema DTD-a (radi se o dobro formiranom XML dokumentu)
- DTD je unutar dokumenta:
  `<!DOCTYPE name [definition] >`
- Spoljašnji DTD, specificiran URI-jem:
  `<!DOCTYPE name SYSTEM “demo.dtd“>`
- Spoljašnji DTD, dato je ime i (opcionalno) URI:
  `<!DOCTYPE name PUBLIC “Demo“>`
  `<!DOCTYPE name PUBLIC “Demo“ “demo.dtd“>`
- DTD je unutrašnji + spoljašnji:
  `<!DOCTYPE name1 SYSTEM “demo.dtd” >`

Matematički fakultet

vladaf@matf.bg.ac.rs

Primeri XML validacije sa DTD

Primer DTD-a koji opisuje strukturu dblp sloga:

```
<!ELEMENT dblp((mastersthesis | article)*)>
<!ELEMENT
  mastersthesis(author,title,year,school,committeemember*)>
<!ATTLIST  mastersthesis(mdate   CDATA   #REQUIRED
   key            ID      #REQUIRED
   advisor        CDATA   #IMPLIED>
<!ELEMENT author(#PCDATA)>
…
```

Primer referisanja na DTD u okviru XML datoteke:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE dblp SYSTEM “my.dtd">
<dblp>…
```

Matematički fakultet
vladaf@matf.bg.ac.rs

Primeri XML validacije sa DTD (2)

Primer dela DTD koji opisuje strukturu knjige i indeksa:

```
<!ATTLIST book
        isbn    ID      #REQUIRED
        price   CDATA   #IMPLIED
        index  IDREFS   “” >
```

…

Primer dela XML-a koji opisuje knjige:

<book id=“1“ index=“2 3 “ >
<book id=“2“ index=“3“/>
<book id =“3“/>

Matematički fakultet
vladaf@matf.bg.ac.rs

Primeri XML validacije sa DTD (3)

Primer dela XML-a sa identifikatorima i referencama na identifikatore:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE graph SYSTEM “special.dtd">
<graph>
   <author id=“author1”>
        <name>John Smith</name>
   </author>
   <article>
        <author ref=“author1” /> <title>Paper1</title>
   </article>
   <article>
        <author ref=“author1” /> <title>Paper2</title>
   </article>
…
```

Pretpostavimo da je definisano da ovo bude tipa ID

Pretpostavimo da je ovo tipa IDREF

Matematički fakultet

vladaf@matf.bg.ac.rs

XML Sheme

Matematički fakultet

vladaf@matf.bg.ac.rs

Ograničenja DTD-ova

- DTD opisuje samo „gramatiku“XML datoteke, a ne detaljnu strukutru niti tipove
- Tako, na primer, preko DTD se ne može iskazati da:
  - elemenat “length” mora sadržavati nenegativan ceo broj *(ograničenje koje se odnosi a tip vrednosti elementa ili atributa)*
  - elemenat “unit” treba da bude dopušten samo onda kada je prisutan elemenat “amount” *(ograničenje koje se odnosi na zajedničko pojavljivanje)*
  - elemenat “comment” može da se pojavi na bilo kom mestu *(fleksibilnost sheme)*
  - DTD-ov ID nije preterano dobra implementacija za vrednost ključa
  - Ne postoji podrška za nasleđivanje kao kod objetktno-orjentisanih jezika
  - Sintaksa koja je bliska XML-u, ali nije XML nije pogodna da se na toj osnovi razvijaju alati

Matematički fakultet

vladaf@matf.bg.ac.rs

Principi dizajna za sheme

Jezik XML shema treba da bude:

1. Izražajniji od XML DTD-ova
2. Izražen pomoću XML-a
3. Samo-opisiv
4. Pogodan za korišćenje za širok opseg aplikacija koje koriste XML
5. Direktno pogodan za korišćenje na Internetu
6. Optimizovan za interoperabilnost
7. Dovoljno jednostavan da se može implementirati na skromnim resursima
8. Usaglašen sa relevantnim W3C specifikacijama

Matematički fakultet

vladaf@matf.bg.ac.rs

Osnove XML sheme

Kreirana tako da prevaziđe probleme sa DTD-ovima

- Ima XML sintaksu
- Može definisati ključeve korišćenjem XPath konstrukcija
- Tip korenog elementa za dokument je globalno naniže komptibilan sa DTD
- Prostori imena su deo XML shema
- Podržano je nasleđivanje tipova, koje uključuje i ograničavanje opsega
  - Nasleđivanje proširivanjem (by extension) kojim se dodaju novi podaci
  - Nasleđivanje ograničavanjem (by restriction) kojim se dodaju nova ograničenja
- Podržani su domeni i predefinisani tipovi podataka

Matematički fakultet

vladaf@matf.bg.ac.rs

Osnove XML sheme (2)

- Prosti tipovi predstavljaju način restrikcije domena na skalarne vrednosti
  - Tako se može definisati prosti tip zanosvan na celobrojnom tipu, pri čemu su vrednosti tog prostog tipa utar zadatog opsega
- Složeni tipovi su način definisanja struktura element/atribut
  - U osnovi ekvivalentni sa !ELEMENT kod DTD-a, ali moćnije
  - Specificira bilo sekvencu, bilo izbor među elementima - potomcima
  - Specificira minimalni i maksimalni broj pojavljivanja (minOccurs i maxOccurs), pri čemu je podrazumevana vrednost 1
- Elementima se može pridružiti složeni tip ili prosti tip
- Atributima se može pridružiti samo prosti tip
- Tipovi mogu biti predefinisani ili korisnički
  - Složeni tipovi mogu biti samo korisnički, ne mogu biti predefinisani

Matematički fakultet

vladaf@matf.bg.ac.rs

Struktura sheme

Primer dela sheme koji opisuje strukturu knjige:

```
<?xml version="1.0" ?>
<xsd:schema xmlns:xsd="http://w3.org/2001/XMLSchema">
    <xsd:element name=„book" type="BookType"/>
    <xsd:complexType name="BookType">
        <xsd:sequence>
            <xsd:element name="title" type="xsd:string"/>
            <xsd:element name="author" type="PersonType"
                minOccurs="1" maxOccurs="unbounded"/>
            <xsd:complexType name=„PersonType">
              <xsd:sequence> ... <xsd:sequence>
            </xsd:complexType>
            <xsd:element name="publisher" type="xsd:anyType"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Prolog sheme

Primer prologa sheme:

```
<?xml version=“1.0“ ?>
<xsd:schema xmlns:xsd=“http://w3.org/2001/XMLSchema“>
    ...
</xsd:schema>
```

- Napomene:
  - Shema se obično čuva u odvojenom XML dokumentu
  - Rečnik za shemu se definiše u specijalnom prostoru imena, a kao prefiks se obično koristi xsd
  - Postoji shema koja opisuje XML sheme
  - Elemenat schema je uvek koren za XML shemu

Matematički fakultet

vladaf@matf.bg.ac.rs

Globalne deklaracije

- Instance (tj. primerci) globalne deklaracije elementa predstavljaju potencijalne korene elemente za date XML dokumente

- Globalne deklaracije se mogu referencirati na sledeći način:

```
<xsd:schema xmlns:xsd="...">
   <xsd:element name="book" type="BookType"/>
   <xsd:element name="comment"  type="xsd:string"/>
   <xsd:ComplexType name="BookType">
   ...   <xsd:element ref="comment" minOccurs="0"/>
...
```

- Ograničenja
  - `ref` se ne može koristiti u globalnoj deklaraciji
  - Ni `minOccurs`, `maxOccurs` se ne mogu koristiti u globalnoj deklaraciji

Matematički fakultet

vladaf@matf.bg.ac.rs

Deklaracija globalnog elementa

Primer definisanja elementa u XML shemi:

```
<xsd:element name="book" type="BookType"/>
```

- Napomene:
  - Etiketa element služi za deklarisanje elemenata
  - Atribut name služi za imenovanje elementa
  - Atribut type definiše tip elementa
    - Primer:
      U prethodnom slučaju, tip elementa book je tip BookType, koji je definisan u nastavku
  - Declaracije direktno unutar elementa schema su **globalne**
    - Samo oni elementi čije su deklaracije globalne mogu doći u obzir da budu koreni za XML shemu
    - Primer:
      U prethodnom slučaju, jedini globalni element je book, pa stoga koren za validni XML dokument opisano ovm shemom mora biti book

Matematički fakultet

vladaf@matf.bg.ac.rs

Deklaracija globalnog tipa

Primer definisanja složenog tipa u XML shemi:

```
<xsd:complexType name="BookType">
      <xsd:sequence>
            ...
      </xsd:sequence>
</xsd:complexType>
```

- Napomene:
  - Ovaj složen tip je definisan kao sekvenca elemenata
  - Atribut name određuje ime tipa
  - Ova definicija tipa je **globalna** (nalazi se direktno unutar elementa schema), pa se ovako definisan tip može koristiti u ma kojoj drugoj definiciji

Matematički fakultet

vladaf@matf.bg.ac.rs

Lokalni elemenat

Primer definisanja lokalnog elementa u XML shemi:

```
<xsd:sequence>
            <xsd:element name="title" type="xsd:string"/>
</xsd:sequence>
```

- Napomene:
  - Ovo je lokalni elemenat, jer se ne nalazi direktno u elementu schema, već unutar kompleksnog tipa Dakle, elemenat title ne može biti koreni elemenat dokumenta
  - Atribut name služi za imenovanje elementa
  - Atribut type definiše tip elementa
  - Tip xsd:string je predefinisani tip za XML shemu

Matematički fakultet

vladaf@matf.bg.ac.rs

Deklaracija lokalnog elementa

Primer definisanja lokalnog elementa author:

```
<xsd:element name="author" type="PersonType"
        minOccurs="1" maxOccurs="unbounded"/>
```

- Napomene:
  - Ako se analizira cela XML shema, jasno je da se radi o deklaraciji lokalnog elementa
  - Tip PersonType je korisnički definisan tip
  - Atributi minOccurs i maxOccurs određuju kardinalnost elementa author u tipu BookType
  - Ovakva vrsta definisanja atributa se u žargonu naziva „brušenje“ („facet“);
  - Postoji 15 različitih vrsta brišenja, npr. minExclusive, totalDigits, itd.
  - Podrazumevane vrednosti za ove atribute su minOccurs=1, maxOccurs=1

Matematički fakultet

vladaf@matf.bg.ac.rs

Definicija lokalnog tipa

Primer definisanja lokalnog tipa PersonType:

```
<xsd:complexType name=„PersonType“>
    <xsd:sequence>
            <xsd:element name=„first“ type=„xsd:string“/>
            <xsd:element name=„last“ type=„xsd:string“/>
    <xsd:sequence>
</xsd:complexType>
```

- Napomene:
  - S obzirom da je definisan u okviru definicije tipa BookType, tip PersonType je **lokalan** i može biti korišćen samo unutar opsega definicije tipa BookType
  - Sintaksa ovog tipa je slična sintaksi tipa BookType

Matematički fakultet

vladaf@matf.bg.ac.rs

Definicija lokalnog elementa

Primer definisanja lokalnog elementa publisher:

```
<xsd:element name=”publisher“ type=“xsd:anyType“/>
```

- Napomene:
  - S obzirom da je definisan u okviru definicije tipa BookType, ovaj element je **lokalan**
  - Svaka knjiga ima tačno jedan elemenat publisher
  - Brušenja minOccurs, maxOccurs imaju podrazumevanu vrednost 1
  - Tip anyType je predefinisani tip koji dozvoljava bilo kakav sadržaj
  - Tip anyType je podrazumevan – ako se ništa ne napiše, onda se radi o tom tipu
    Definicija koja sledi je ekvivalentna definciji iz primera:

```
<xsd:element name=“publisher“ />
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Primer sheme

Shema koji opisuje strukturu knjige:

```
<?xml version="1.0" ?>
<xsd:schema xmlns:xsd="http://w3.org/2001/XMLSchema">
    <xsd:element name=„book" type="BookType"/>
    <xsd:complexType name="BookType">
        <xsd:sequence>
            <xsd:element name="title" type="xsd:string"/>
            <xsd:element name="author" type="PersonType"
                minOccurs="1" maxOccurs="unbounded"/>
            <xsd:complexType name=„PersonType">
              <xsd:sequence> ... <xsd:sequence>
            </xsd:complexType>
            <xsd:element name="publisher" type="xsd:anyType"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Primer sheme (2)

XML koji je saglasan sa prethodnom shemom:

```
<?xml version=„1.0“>
<book>
   <title>Die Wilde Wutz</title>
   <author><first>D.</first>
                  <last>K.</last></author>
   <publisher>
       Addison Wesley, <state>CA</state>, USA
   </publisher>
</book>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Primer sheme (3)

Pridružuje "xsd" prostor imena sa XML shemom

ovo je element koren, sa tipom čija specifikacija sledi

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="mastersthesis" type="ThesisType"/>

  <xsd:complexType name="ThesisType">
     <xsd:attribute name="mdate" type="xsd:date"/>
     <xsd:attribute name="key" type="xsd:string"/>
     <xsd:attribute name="advisor" type="xsd:string"/>
     <xsd:sequence>
        <xsd:element name="author" type="xsd:string"/>
        <xsd:element name="title" type="xsd:string"/>
        <xsd:element name="year" type="xsd:integer"/>
        <xsd:element name="school" type="xsd:string"/>
        <xsd:element name="committeemember" type="CommitteeType"
          minOccurs="0"/>
     </xsd:sequence>
  </xsd:complexType>
</xsd:schema>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Deklaracije atributa

- Attributi mogu biti samo prostog tipa (npr. string)
- Deklaracije atributa mogu biti globalne
  - Takve deklaracije se mogu ponovo iskoristiti pomoću ref
- Deklaracije atributa su kompatibilne sa listom atributa u DTD
  - Moguće je korišćenje podrazumevanih vrednosti
  - Moguće je atribute proglasiti zahtevanim ili opcionalnim
  - Postojanje fiksnih atributa
  - Kao dodatna osobina, postoje i „zabranjeni“ atributi

Matematički fakultet

vladaf@matf.bg.ac.rs

Deklaracije atributa (2)

Primer dela sheme koji opisuje strukturu knjige i indeksa:

```
<xsd:complexType name="BookType">
  <xsd:sequence> ... </xsd:sequence>

  <xsd:attribute name="isbn" type="xsd:string" use="required"/>
  <xsd:attribute name="price" type="xsd:decimal"
                 use="optional" />
  <xsd:attribute name="curr" type="xsd:string"
                 fixed="EUR" />
  <xsd:attribute name="index" type="xsd:idrefs"
                 default="" />

</xsd:complexType>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Anonimni tipovi

Primer dela sheme koji opisuje strukturu knjige, gde se tip koji opisuje osobu ne imenuje:

```
<xsd:complexType name="BookType">
    ...
  <xsd:element name="author">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="first" type="xsd:string"/>
        <xsd:element name="last" type="xsd:string"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
    ...
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Elementi i atributi

Primer dela sheme koji opisuje strukturu cene:

```
<xsd:element name="price">
   <xsd:complexType>
      <xsd:simpleContent>
         <xsd:extension base= "xsd:decimal" >
            <xsd:attribute name="curr" type="xsd:string"/>
         </xsd:extension>
      </xsd:simpleContent>
   </xsd:complexType>
</xsd:element>
```

Primer za validan primerak cene:

```
<price curr="USD" >69.95</price>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Elementi i atributi (2)

Primer dela sheme koji opisuje strukturu cene:

```
<xsd:element name="price">
   <xsd:complexType>
     <xsd:attribute name="curr" type="xsd:string"/>
     <xsd:attribute name="val" type="xsd:decimal"/>
   </xsd:complexType>
</xsd:element>
```

Primer za validan primerak cene:

```
<price curr="USD" val="69.95" />
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Predefinisani prosti tipovi

- Numerički tipovi

  Integer, Short, Decimal, Float, Double, HexBinary, ...
- Tipovi za datume i periode

  Duration, DateTime, Time, Date, ...
- Tipovi za niske

  String, NMTOKEN, NMTOKENS, NormalizedString
- Ostali tipovi

  Qname, AnyURI, ID, IDREFS, Language, Entity, ...
- Kao zaključak, postoje 44 predefinisana prosta tipa

Matematički fakultet

vladaf@matf.bg.ac.rs

Izvedeni prosti tipovi

Primer dela sheme sa tipom gde je izvršeno ograničavanje domena:

```
<xsd:simpleType name="MyInteger">
      <xsd:restriction base="xsd:integer">
          <xsd:minInclusive value="10000"/>
          <xsd:maxInclusive value="99999"/>
      </xsd:restriction>
    </xsd:simpleType>
```

Primer definicije tipa gde je izvršeno ograničavanje domena preko regularnih izraza, tako da valute može biti zapisana samo pomoću tri velika slova:

```
    <xsd:simpleType name="Currency">
       <xsd:restriction base="xsd:string" >
          <xsd:pattern value="[A-Z]{3}"/>
       </xsd:restriction>
    </xsd:simpleType>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Izvedeni prosti tipovi (2)

Primer definicije tipa gde je izvršeno ograničavanje domena preko enumeracije:

```
<xsd:simpleType name="Currency">
   <xsd:restriction base="xsd:string" >
      <xsd:enumeration value="ATS"/>
      <xsd:enumeration value="EUR"/>
      <xsd:enumeration value="GBP"/>
      <xsd:enumeration value="USD"/>
   </xsd:restriction>
</xsd:simpleType>
```

- Napomene:
  - Najveći broj predefinisanih tipova je izveden restrikcijom iz drugih predefinisanih tipova, npr tip Integer je izveden iz tipa Decimal
  - Od 44 predefinisana tipa, samo njih 19 su osnovni tipovi

Matematički fakultet

vladaf@matf.bg.ac.rs

Prosti tip listi

- Postoji više vrsta prostih tipova za liste:
  - Predefinisani tipovi listi: IDREFS, NMTOKENS
  - Korisnički definisani tipovi listi
    Primer sheme za korisnički definisan tip liste

```
<xsd:simpleType name = ”intList“ >
   <xsd:list itemType = “xsd:integer“ />
</xsd:simpleType>
```

- Karakteristike:
  - Elementi u primerku takve liste su razdvojeni belinama

```
“5  -10  7       -20“
```

  - Brušenja za restrikcije kod ovih listi su: length, minLength, maxLength, enumeration

Matematički fakultet

vladaf@matf.bg.ac.rs

Prosti tip listi sa restrikcijom

Primer sheme za korisnički definisan tip liste sa restrikcijom:

```
<xsd:simpleType name = “Participants“ >
    <xsd:list itemType = “xsd:string“ />
<xsd:simpleType>

<xsd:simpleType name = “Medalists“ >
    <xsd:restriction base = “Participants“ >
        <xsd:length value = “3“ />
    </xsd:restriction>
</xsd:simpleType>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Prosti tip unije

- Odgovara znaku | kod DTD
- Ima isto značenje kao slogovi sa promenljivim delom u Pascal-u ili kao unije u C-u
- Instance su validne ako su validne za jedan od pobrojanih tipova

  Primer sheme sa prostim tipom unije:

```
<xsd:simpleType name = “Potpurri“ >
   <xsd:union memberTypes = “xsd:string intList“/>
</xsd:simpleType>
```

  Primer XML instanci validnih za gornju shemu:

  “fünfzig“ “1 3 17“ “wunderbar“ “15“
- Za prosti tip unije su podržana brušenja pattern i enumeration

Matematički fakultet

vladaf@matf.bg.ac.rs

Element choice

Primer sheme za knjigu koja ima ili element author ili element editor:

```
<xsd:complexType name = “Book“ > <xsd:sequence>
   <xsd:choice>
      <xsd:element name = “author“ type = “Person“
                                 maxOccurs = “unbounded“ />
      <xsd:element name = “editor“ type = “Person“ />
   </xsd:choice>
</xsd:sequence> </xsd:complexType>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Grupe elemenata

Opis sheme kada se treba postići da ako element book sadrži element editor, tada book takođe sadrži i element sponsor:

```
<xsd:complexType name = „Book“ > <xsd:sequence>
    <xsd:choice>
      <xsd:element name = „Author“ type = „Person“ .../>
        <xsd:group ref = „EditorSponsor“ />
    </xsd:choice> </xsd:sequence> </xsd:complexType>

<xsd:group name = „EditorSponsor“ > <xsd:sequence>
  <xsd:element name =„Editor“ type=„Person“ />
   <xsd:element name = „Sponsor“ type = „Org“ />
</xsd:sequence> </xsd:group>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Grupe atributa

Opis sheme sa grupom atributa:

```
<xsd:attributeGroup name = „PriceInfo“ >
   <xsd:attribute name = „curr“ type = „xsd:string“ />
   <xsd:attribute name = „val“ type = „xsd:decimal“ />
</xsd:attributeGroup>

<xsd:complexType name = „Book“ >
   ...
   <xsd:attributeGroup ref = „PriceInfo“ />
</xsd:complexType>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Definicija ključeva

- Ključevi jednoznačno identifikuju element i definisani su kao deo elementa
- Uveden je specijalni element koji se ugnježdava, nazvan key
  - U okviru tog novog elementa uvedeni su:
    - selector: opisuje kontekst na koji se odnosi ključ
    - field: opisuje koje je polje ključ u kontekstu opisanim selektorom
  - Ako ima više elemenata field u okviru ključa, tada se radi o tzv. kompozitnom ključu
- Vrednosti za selector i za field su XPath izrazi
- Validacija ključa u XML-u se realizje na sledeći način:
  1. Evaluira se selector i dobije sekvenca čvorova
  2. Evaliraju se vrednosti za field na sekvenci čvorova i dobije se skup uređenih n-torki vrednosti
  3. Proverava se da li ima duplikata u skupu uređenih n-torki vrednosti

Matematički fakultet

vladaf@matf.bg.ac.rs

Definicija ključeva (2)

Primer sheme u kojoj je isbn definisano kao ključ za books u bib:

```
<element name = „bib“> <complexType> <sequence>
   <element book maxOccurs = „unbounded>
     <complexType> <sequence> ... </sequence>
       <attribute name = „isbn“ type = „string“ />
   </complexType> </element> </sequence>
   <key name = „constraintX“ >
       <selector xpath = „book“ />
       <field xpath = „@isbn“ />
   </key>
</complexType> </element>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

Reference (strani ključevi)

- Strani ključevi predstavljaju deo definicije elementa
- Uveden je specijalni element koji se ugnježdava, nazvan keyref (sa atributom refer) i u okviru njega elementi selector i field (sa atributom xpath)
  - selector: određuje kontekst stranih ključeva
  - field(s): specificira strani ključ
  - refer: daje opseg za reference (ograničenja za ključ)

Primer sheme za knjige koje referišu prema drugim knjigama:

```
<keyref name = “constraintY“ refer = “constraintX“ >
   <selector xpath = “book/references“ />
   <field xpath = “@isbn“ />
</keyref>
```

Matematički fakultet

vladaf@matf.bg.ac.rs

XML i programerske paradigme

XML i OO

- Encapsulacija
  - OO sakriva podatke
  - XML čini da podaci budu eksplicitni
- Hijerarhija tipova
  - OO definiše relacije podskup/nadskup
  - XML deli strukturu, pa skupovne relacije nemaju smisla
- Podaci i ponašanje
  - OO ih pakuje zajedno u jednu celinu
  - XML razdvaja podatke od njihove interpretacije

Matematički fakultet

vladaf@matf.bg.ac.rs

XML i relacione baze podataka

- Strukturne razlike
  - Drvo naspram tabele
  - Heterogene naspram homogenih
  - Opcionalni tipovi naspram striktnog tipiziranja
  - Nenormalizovani podaci naspram normalizovanih
- Neke od sličnosti
  - Logička i fizička nezavisnost podataka
  - Deklarativna semantika
  - Generički model podataka

Matematički fakultet
vladaf@matf.bg.ac.rs

Programerski modeli procesiranja XML-a

- Ogromna korist od XML-a su standardni parseri i standardni API-ji (nezavisni od jezika) za njihovo procesiranje
- DOM je objektno-orjentisana reprezentacija XML drveta parsiranja
  - **DOM objekti** sadrže
    - metode kao što su getFirstChild, getNextSibling, koje predstavljaju uobičajen način prolaska kroz drvo
    - Takođe mogu da modifikuju samo DOM drvo, tj. da izmene XML, korišćenjem metoda insertAfter, itd
- SAX se koristi u situacijama kada nisu potrebni svi podaci
  - **Interfejs za parser** je ključan u ovom pristupu:
    - On poziva funkciju svaki put kada parsira instrukciju procesiranja, element itd.
    - Razvijeni kod može odrediti šta treba raditi u datom slučaju, npr. modifikovati strukturu podataka ili ukloniti dete delove podataka

Matematički fakultet

vladaf@matf.bg.ac.rs

XML upiti

- *Upitni jezik* predstavlja alternativni pristup procesiranju XML podataka
  - Definiše se neka vrsta *šablona* koji opisuje prolaske (tj. putanje) od korenog čvora usmerenog grafa koji predstavlja XML
  - Potencijalna korist ovakvog pristupa ogleda se u eksploataciji paralelizma, pogleda, mapranja shema itd.
  - Kod jezika XML, osnova za ovakve šablone se naziva XPath
    - XPath takođe može deklarisati neka ograničenja na vrednosti koje se traže
    - XPath kao rezultat upita vraća *skup čvorova* koji predstavlja poklapanja

Matematički fakultet

vladaf@matf.bg.ac.rs

XPath

- U svom najprostijem obliku, XPath liči na opis putanje u sistemu datoteka:

```
/mypath/subpath/*/morepath
```

- Međutim, XPath vraće *skup čvorova* koji predstavljaju XML čvorove (i njihova poddrveta) koji se nalaze na kraju zadate putanje
- XPaths na samom kraju putanje može sadržavati *testove za čvorove*, i tako kreirati filter po tipu čvora metodama `text()`, `processing-instruction()`, `comment()`, `element()`, `attribute()`
- XPath vodi računa o uređenju, može se postaviti upit tako da se vodi računa o uređenju i dobiti odgovor koji poštuje dato uređenje

Matematički fakultet

vladaf@matf.bg.ac.rs

Zahvalnica

Delovi materijala ove prezentacije su preuzeti iz:

- Skripte iz predmeta Uvod u veb i internet tehnologije, na Matematičkom fakultetu Univeziteta u Beogradu, autor prof. dr Filip Marić
- Skripte iz predmeta Informatika na Univerzitetu Milano Bicocca, autor dr Mirko Cesarini