

# XML Schema

# Šta nije dobro kod DTD-a?

---

- neuobičajena, ne-XML sintaksa
- slaba podrška za XML prostore imena
- nema tipizacije podataka, naročito za sadržaj elementa
  - nije moguće definisati datumska ili numerička polja
- ograničena proširivost
- ograničene mogućnosti za opisivanje strukture podataka
  - ne može se nametnuti broj podelemenata bez nametanja redosleda
  - ne može se nametnuti redosled i broj podelemenata kada se koristi mešani sadržaj

# Šta je šema?

---

- šema je dokument koji opisuje strukturu drugih dokumenata
- prevazilazi mane DTD-a
  - šeme se pišu kao XML dokumenti, a ne korišćenjem sintakse nasleđene od SGML-a
  - potpuna podrška za rad sa XML namespaces
  - validacija tekstualnog sadržaja na bazi ugrađenih ili novodefinisanih tipova podataka (datum, ceo broj, itd)
  - kreiranje i jednostavno ponovno korišćenje modela sadržaja

# Da li je DTD i dalje potreban?

---

- šeme nemaju funkcionalnost rada sa entitetima
- DTD može biti ugrađen direktno u dokument koji opisuje, dok
- šeme moraju uvek biti smeštene u posebne fajlove
- softver i dalje potpuno podržava rad sa DTD-ima

# Šema standardi

---

- postoji više standarda za šema jezike
  - W3C XML Schema
  - RELAX NG
  - Schematron
- XML Schema je standard koga propisuje W3C
  - najrašireniji
  - najmoćniji
  - najkomplikovaniji

# W3C XML Schema

---

- kreirana od strane W3C XML Schema Group, na osnovu mnogih predloga
- nema patenata ili drugih restrikcija nad intelektualnom svojinom
- definisana kroz sledeće dokumente
  - XML Schema Part 1: Structures  
<http://www.w3.org/TR/xmlschema-1/>
  - XML Schema Part 2: Datatypes  
<http://www.w3.org/TR/xmlschema-2/>

# Primer šeme

---

- primer XML dokumenta

greeting.xml

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

```
<GREETING>
```

```
  Hello XML!
```

```
</GREETING>
```

# Primer šeme

---

- šema za prethodni dokument

greeting.xsd

```
<?xml version="1.0"?>  
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">  
  <xsd:element name="GREETING" type="xsd:string"/>  
</xsd:schema>
```

korenski element W3C XML Schema dokumenta je xsd:schema



# Povezivanje šeme i dokumenta

---

- standardni XML Instance namespace
  - identifikator je  
`http://www.w3.org/2001/XMLSchema-instance`
  - uobičajeni prefiks je `xsi`
  - u njemu su definisana četiri atributa
    - `xsi:type`
    - `xsi:nil`
    - `xsi:schemaLocation`
    - `xsi:noNamespaceSchemaLocation`

# Povezivanje šeme i dokumenta

---

- atribut `xsi:schemaLocation`
  - označava lokaciju šeme koja definiše strukturu datog dokumenta
  - vrednost atributa čine dva URI-ja
    - identifikator namespace-a
    - adresa fajla
- atribut `xsi:noNamespaceSchemaLocation`
  - lokacija šeme za elemente koji pripadaju default namespace-u

# Povezivanje šeme i dokumenta

---

- primer 1 - elementi su u default namespace-u

greeting.xml

```
<?xml version="1.0"?>
<GREETING
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="greeting.xsd">
  Hello XML!
</GREETING>
```

greeting.xsd

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="GREETING" type="xsd:string"/>
</xsd:schema>
```

# Povezivanje šeme i dokumenta

---

- primer 2 - elementi nisu u default namespace-u

greeting.xml

```
<?xml version="1.0"?>
<GREETING
  xmlns="http://www.ftn.ns.ac.yu/greetings"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="greeting.xsd">
  Hello XML!
</GREETING>
```

greeting.xsd

```
<?xml version="1.0"?>
<xsd:schema
  targetNamespace="http://www.ftn.ns.ac.yu/greetings"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="GREETING" type="xsd:string"/>
</xsd:schema>
```

# Validacija dokumenata

---

- primer neispravnog dokumenta

```
<?xml version="1.0"?>
<GREETING
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="greeting.xsd">
  <P>Hello XML!</P>
</GREETING>
```

# Tipizacija podataka u W3C XML Schema

---

- prosti tipovi podataka
  - ne mogu imati podelemente ili attribute
  - primeri
    - `<datum_racuna>2007-06-01</datum_racuna>`
    - `<kolicina>17</kolicina>`
- složeni tipovi podataka
  - mogu imati podelemente ili attribute
  - primeri
    - `<name><first>John</first><last>Doe</last></name>`
    - `<person id="i73">...</person>`

# Četiri osnovna elementa šeme

---

- `xsd:element` deklarira element i dodeljuje mu tip
- `xsd:attribute` deklarira atribut i dodeljuje mu tip
- `xsd:simpleType` definiše novi prosti tip
- `xsd:complexType` definiše novi složeni tip

# Deklaracija elementa

---

- tip elementa može da se navede u okviru njegove deklaracije
  - „anonimni tip“
- kao tip elementa može da se navede neki postojeći tip
  - „imenovani tip“



# Deklaracija elementa

---

- primer XML dokumenta

```
<?xml version="1.0"?>
<SONG
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="simple_song.xsd">
  <TITLE>Hot Cop</TITLE>
  <COMPOSER>Jacques Morali</COMPOSER>
  <COMPOSER>Henri Belolo</COMPOSER>
  <COMPOSER>Victor Willis</COMPOSER>
  <PRODUCER>Jacques Morali</PRODUCER>
  <PUBLISHER>PolyGram Records</PUBLISHER>
  <LENGTH>6:20</LENGTH>
  <YEAR>1978</YEAR>
  <ARTIST>Village People</ARTIST>
</SONG>
```

# Deklaracija elementa

- šema za prethodni dokument sa anonimnim tipom

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="SONG">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element name="TITLE" type="xsd:string"/>
        <xsd:element name="COMPOSER" type="xsd:string"
                      maxOccurs="unbounded"/>
        <xsd:element name="PRODUCER" type="xsd:string"
                      minOccurs="0" maxOccurs="unbounded"/>
        <xsd:element name="PUBLISHER" type="xsd:string"
                      minOccurs="0"/>
        <xsd:element name="LENGTH" type="xsd:duration"/>
        <xsd:element name="YEAR" type="xsd:gYear"/>
        <xsd:element name="ARTIST" type="xsd:string"
                      maxOccurs="unbounded"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

# Deklaracija elementa

- šema za prethodni dokument sa imenovanim tipom

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="SONG" type="SongType"/>
  <xsd:complexType name="SongType">
    <xsd:sequence>
      <xsd:element name="TITLE" type="xsd:string"/>
      <xsd:element name="COMPOSER" type="xsd:string"
                    maxOccurs="unbounded"/>
      <xsd:element name="PRODUCER" type="xsd:string"
                    minOccurs="0" maxOccurs="unbounded"/>
      <xsd:element name="PUBLISHER" type="xsd:string"
                    minOccurs="0"/>
      <xsd:element name="LENGTH" type="xsd:duration"/>
      <xsd:element name="YEAR" type="xsd:gYear"/>
      <xsd:element name="ARTIST" type="xsd:string"
                    maxOccurs="unbounded"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
</xsd:schema>
```

# Deklaracija elementa

---

- anonimni vs imenovani tip
  - ako očekujemo da će se isti tip pojavljivati na više mesta, isplati se definisati ga na jednom mestu - koristićemo imenovani tip
  - ako ne želimo da se definicija tipa vidi kao nešto što je dostupno korisniku šeme - koristićemo anonimni tip

# Deklaracija tipa

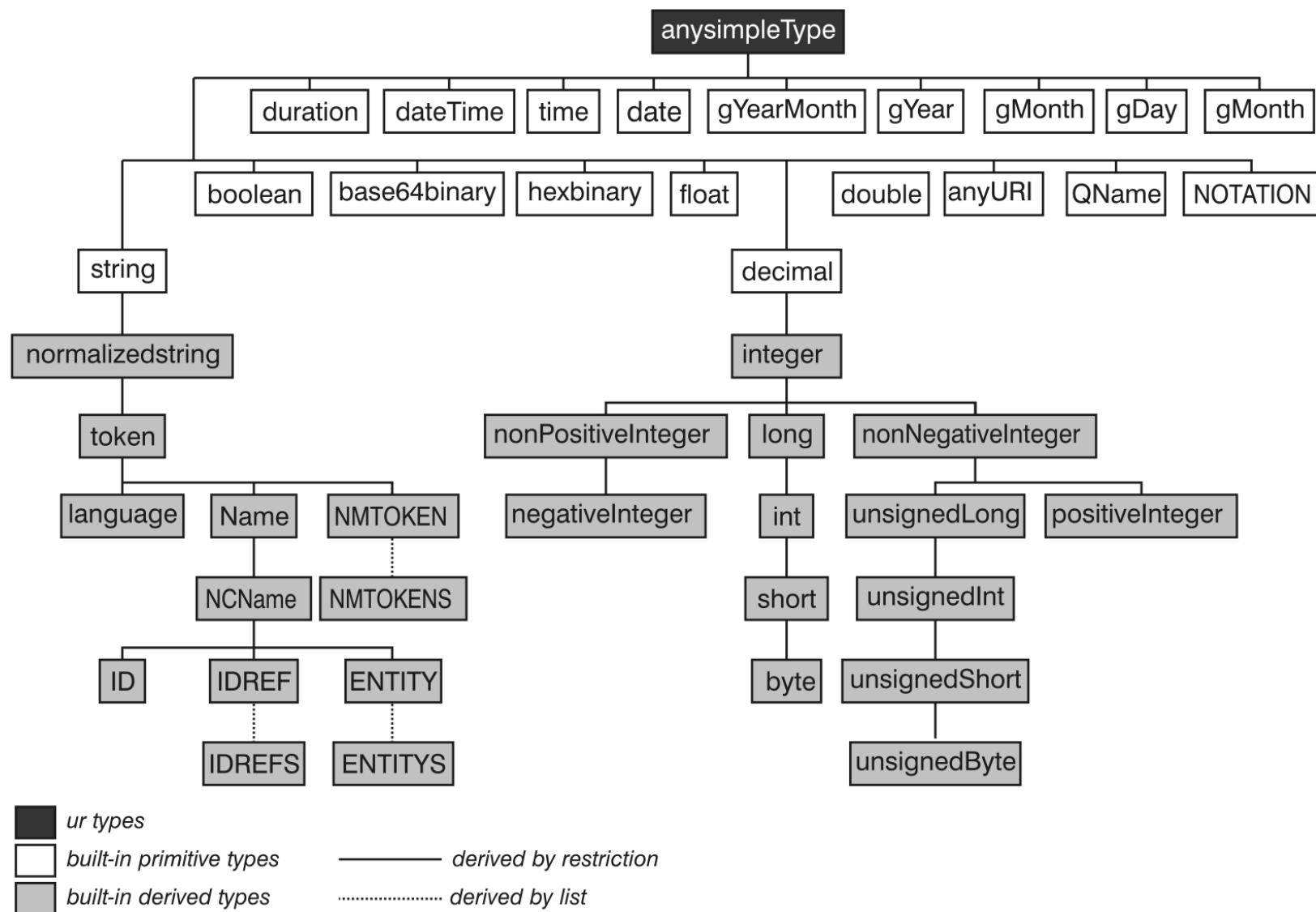
---

- svaka deklaracija novog tipa se oslanja na neki od postojećih (već definisanih) tipova
- kada definišemo prvi tip, koji tipovi mogu da se koriste?
- tipovi korišćeni u prethodnim primerima
  - xsd:string
  - xsd:year
  - xsd:duration

# Ugrađeni tipovi

- definisani u XML Schema Part 2: Datatypes

- logički
- tekstualni
- URI
- numerički
- vremenski
- XML tipovi



# Ugrađeni numerički tipovi

float	IEEE 754 32-bitni broj u pokretnom zarezu
double	IEEE 754 64-bitni broj u pokretnom zarezu
decimal	decimalni broj proizvoljne preciznosti
integer	proizvoljno velik ceo broj
nonPositiveInteger	ceo broj manji ili jednak nuli
negativeInteger	ceo broj manji od nule
long	64-bitni ceo broj
int	32-bitni ceo broj
short	16-bitni ceo broj
byte	8-bitni ceo broj
nonNegativeInteger	ceo broj veći ili jednak nuli
unsignedLong	neoznačeni 64-bitni ceo broj
unsignedInt	neoznačeni 32-bitni ceo broj
unsignedShort	neoznačeni 16-bitni ceo broj
unsignedByte	neoznačeni 8-bitni ceo broj
positiveInteger	ceo broj veći od nule

# Ugrađeni vremenski tipovi

dateTime	trenutak u vremenu
gMonth	mesec u godini
gYear	godina
gMonthDay	datum nevezan za godinu
gDay	dan u mesecu
duration	dužina vremenskog intervala, bez fiksnog početka ili kraja
date	konkretan dan u vremenu
time	konkretno vreme u toku dana



# Ostali ugrađeni tipovi

string	
normalizedString	string koji ne sadrži TAB, CR, LF
token	string čija normalizovana vrednost ne sadrži više uzastopnih razmaka ili razmake na krajevima stringa
boolean	logički tip
anyURI	relativna ili apsolutna internet adresa
hexBinary	binarni podaci zapisani u heksadecimalnom obliku
base64Binary	binarni podaci zapisani u Base64 obliku

# Deklaracija novog prostog tipa

---

- novi prosti tipovi se definišu na osnovu postojećih - ugrađenih
- mehanizam definisanja novog tipa
  - restrikcija: navođenje ograničenja na vrednost novog tipa
  - lista: definisanje konačne liste mogućih vrednosti
  - unija: kombinovanje vrednosti više različitih osnovnih tipova
- `<xsd:simpleType name="ImeTipa">...</simpleType>`

# Deklaracija novog prostog tipa

---

- restrikcija
  - skup mogućih vrednosti novog tipa je podskup vrednosti osnovnog tipa
  - primer: `integer` i `nonNegativeInteger`
  - ograničenja se navode pomoću 12 *facet-a*
    - `length`
    - `minLength`
    - `maxLength`
    - `pattern`
    - `enumeration`
    - `whiteSpace`
    - `maxInclusive`
    - `maxExclusive`
    - `minInclusive`
    - `minExclusive`
    - `totalDigits`
    - `fractionDigits`

# Deklaracija novog prostog tipa

---

- restrikcija
  - primer: string dužine 1-255 znakova (minLength i maxLength)

```
<xsd:simpleType name="Str255">  
  <xsd:restriction base="xsd:string">  
    <xsd:minLength value="1"/>  
    <xsd:maxLength value="255"/>  
  </xsd:restriction>  
</xsd:simpleType>
```

# Deklaracija novog prostog tipa

---

- restrikcija
  - primer: nabranje mogućih vrednosti (enumeration)

```
<xsd:simpleType name="Tseason">  
  <xsd:restriction base="xsd:string">  
    <xsd:enumeration value="spring"/>  
    <xsd:enumeration value="summer"/>  
    <xsd:enumeration value="autumn"/>  
    <xsd:enumeration value="winter"/>  
  </xsd:restriction>  
</xsd:simpleType>
```

# Deklaracija novog prostog tipa

- restrikcija
  - primer: definisanje tipa za novčane iznose (pattern)

```
<xsd:simpleType name="money">
  <xsd:restriction base="xsd:string">
    <xsd:pattern value="\p{Sc}\d+(\.\d\d)?"/>
    <!--
      Regularni izraz:
      \p{Sc}      oznaka valute $, €, £, ¥.
      \d          decimalna cifra
      \d+         jedna ili više decimalnih cifara
      \.          znak tačka
      (\.\d\d)?  nula ili jedan string oblika .35
    -->
  </xsd:restriction>
</xsd:simpleType>
```

# Deklaracija novog prostog tipa

---

- lista
  - tip koji definišemo predstavljajući listu elemenata nekog tipa
  - elementi liste su razdvojeni razmacima
  - primer

```
<element name="ages">  
  <simpleType>  
    <list itemType="positiveInteger"/>  
  </simpleType>  
</element>
```

```
<ages>4 9 12</ages>
```

# Deklaracija novog prostog tipa

---

- unija
  - vrednost novog tipa može biti vrednost bilo kog tipa koji je član unije
  - primer

```
<element name="timePeriod">  
  <simpleType>  
    <union memberTypes="season xsd:date"/>  
  </simpleType>  
</element>
```

```
<timePeriod>summer</timePeriod>  
<timePeriod>2007-05-25</timePeriod>
```



# Deklaracija novog složenog tipa

---

- deklaracija složenog tipa koristi jedan od modela sadržaja
  - sekvenca podelemenata
  - izbor između podelemenata
  - neuređeni skup podelemenata
- deklaracija složenog tipa može da sadrži deklaracije atributa
- ako je deklaracija složenog tipa prazna, definisan je prazan tip

# Deklaracija novog složenog tipa

---

- sekvenca - sequence
  - sekvenca podelemenata ili drugih modela sadržaja
  - primer

```
<xsd:complexType name="TName">  
  <xsd:sequence minOccurs="1" maxOccurs="1">  
    <xsd:element name="first" type="xsd:string"/>  
    <xsd:element name="middle" type="xsd:string"/>  
    <xsd:element name="last" type="xsd:string"/>  
  </xsd:sequence>  
  <xsd:attribute name="title" type="xsd:string"/>  
</xsd:complexType>
```

# Deklaracija novog složenog tipa

- sekvenca - sequence
  - primer: sekvenca u sekvenci

```
<xsd:complexType name="TName">
  <xsd:sequence minOccurs="1" maxOccurs="1">
    <xsd:element name="first" type="xsd:string"/>
    <xsd:element name="middle" type="xsd:string"/>
    <xsd:element name="last" type="xsd:string"/>
    <xsd:sequence minOccurs="0" maxOccurs="unbounded">
      <xsd:element name="title" type="xsd:string"/>
      <xsd:element name="year" type="xsd:year"/>
    </xsd:sequence>
  </xsd:sequence>
  <xsd:attribute name="title" type="xsd:string"/>
</xsd:complexType>
```

# Deklaracija novog složenog tipa

---

- sekvenca - sequence
  - redosled podelemenata je bitan
  - broj pojavljivanja podelemenata je bitan
    - ako se ne navede, podrazumeva se `minOccurs="1"` i `maxOccurs="1"`
  - cela sekvenca može da se ponavlja
    - ima attribute `minOccurs` i `maxOccurs`

# Deklaracija novog složenog tipa

---

- izbor - choice
  - bira se jedan od podelemenata ili modela sadržaja
  - primer

```
<xsd:complexType name="TName">  
  <xsd:choice minOccurs="1" maxOccurs="1">  
    <xsd:element name="first" type="xsd:string"/>  
    <xsd:element name="middle" type="xsd:string"/>  
    <xsd:element name="last" type="xsd:string"/>  
  </xsd:choice>  
  <xsd:attribute name="title" type="xsd:string"/>  
</xsd:complexType>
```

# Deklaracija novog složenog tipa

---

- neuređeni skup podelemenata - `all`
  - u sadržaju se mogu pojaviti svi podelementi u bilo kom redosledu
  - podelement se može pojaviti tačno jednom
  - ne može da sadrži sekvence ili izbore
  - ne može da bude uključen u sekvence ili izbore
  - primer

```
<xsd:complexType name="TName">
  <xsd:all minOccurs="0 ili 1" maxOccurs="1">
    <xsd:element name="first" type="xsd:string"/>
    <xsd:element name="middle" type="xsd:string"/>
    <xsd:element name="last" type="xsd:string"/>
  </xsd:all>
  <xsd:attribute name="title" type="xsd:string"/>
</xsd:complexType>
```

# Novi složeni tip sa mešanim sadržajem

---

- tekst može biti isprepletan sa podelementima
- atribut `mixed` ima vrednost `true`
- primer

```
<element name="p">  
  <complexType mixed="true">  
    <choice minOccurs="0" maxOccurs="unbounded">  
      <element name="b" type="string"/>  
      <element name="i" type="string"/>  
    </choice>  
  </complexType>  
</element>
```

```
<p>Without question, <b>Ramses II</b> is the  
<i>greatest</i> pharaoh <b>of all time!</b></p>
```

# Deklaracija atributa

---

- atributi moraju biti prostog tipa

- deklaracija

```
<attribute
```

```
  name="ime atributa"
```

```
  type="ime tipa"
```

```
  form="qualified/unqualified"
```

```
  use="optional/required/prohibited"
```

```
  default="podrazumevana vrednost"
```

```
  fixed="fiksna vrednost">
```



# Deklaracija atributa

---

- primeri

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

```
<xsd:attribute name="title" type="xsd:string"  
  default="Mr." use="optional"/>
```

```
<xsd:attribute name="season" type="Tseason"/>
```

# Primer šeme

- primer dokumenta

```
<?xml version="1.0"?>
<SONG xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns="http://www.ftn.ns.ac.yu/songs"
      xsi:schemaLocation="http://www.ftn.ns.ac.yu/songs
                           songs.xsd">
  <TITLE>Hot Cop</TITLE>
  <PHOTO ALT="Victor Willis" WIDTH="100" HEIGHT="200"/>
  <COMPOSER>
    <NAME><GIVEN>Jacques</GIVEN><FAMILY>Morali</FAMILY></NAME>
  </COMPOSER>
  <PRODUCER>
    <NAME><GIVEN>Jacques</GIVEN><FAMILY>Morali</FAMILY></NAME>
  </PRODUCER>
  <PUBLISHER>PolyGram Records</PUBLISHER>
  <YEAR>1978</YEAR>
  <ARTIST>Village People</ARTIST>
  <LENGTH>6:20</LENGTH>
</SONG>
```

# Primer šeme

---

- tip koji predstavlja osobu

```
<xsd:complexType name="PersonType">
  <xsd:sequence>
    <xsd:element name="NAME">
      <xsd:complexType>
        <xsd:sequence>
          <xsd:element name="GIVEN" type="xsd:string"/>
          <xsd:element name="FAMILY" type="xsd:string"/>
        </xsd:sequence>
      </xsd:complexType>
    </xsd:element>
  </xsd:sequence>
</xsd:complexType>
```

# Primer šeme

- tip koji predstavlja pesmu

```
<xsd:complexType name="SongType">
  <xsd:sequence>
    <xsd:element name="TITLE" type="xsd:string"/>
    <xsd:element name="COMPOSER" type="xsd:string"
                  maxOccurs="unbounded"/>
    <xsd:element name="PRODUCER" type="xsd:string"
                  minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element name="PUBLISHER" type="xsd:string"
                  minOccurs="0"/>
    <xsd:element name="LENGTH" type="xsd:duration"/>
    <xsd:element name="YEAR" type="xsd:gYear"/>
    <xsd:element name="ARTIST" type="xsd:string"
                  maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>
```

# Primer šeme

---

- definicija korenskog elementa

```
<xsd:element name="SONG" type="SongType"/>
```

# Primer šeme

---

- cela šema

```
<xsd:schema
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://www.ftn.ns.ac.yu/songs">
  <xsd:complexType name="PersonType">
    ...
  </xsd:complexType>
  <xsd:complexType name="SongType">
    ...
  </xsd:complexType>
  <xsd:element name="SONG" type="SongType"/>
</xsd:schema>
```