



Interneto technologijos

XML Schema

Žymių ir atributų deklaracija

Paprastieji tipai



XML Schema: samprata

- Tai kalba, skirta XML dokumento (t.y., kažkieno susikurtos kalbos) struktūros apibrėžimui (t.y., autoriaus taisyklių nusakymui)
- XML Schema leidžia apibrėžti:
 - Kokios žymės ir kokie atributai leistini XML dokumente
 - Koks turi būti žymių eiliškumas
 - Kokio **tipo** duomenys turi būti žymės/atributo viduje



XML Schema: literatūra

- W3C: <http://www.w3.org/XML/Schema#dev>
- Specifikacija susideda iš trijų dalių:
 - XML Schema Part 0: Primer
 - Neformalus dokumentas, apžvelgia XML Schema
 - XML Schema Part 1: Structures
 - Formalus. Apibrėžia struktūrinius XML Schema elementus (atributų, elementų, sudėtingų duomenų tipų apibrėžimai)
 - XML Schema Part 2: Datatypes
 - Formalus. Aprašo duomenų tipus, apibrėžtus pačioje XML Schema specifikacijoje
- Structures ir Datatypes yra sudėtingi dokumentai; Pradedančiajam reiktų skaityti Primer ir trečių šalių įvadinę medžiagą:
 - <http://www.w3schools.com/schema/>



Pavyzdys: XML dokumentas

```
<?xml version="1.0" encoding="UTF-8"?>
<užsakymas data="2006-02-28">
  <pristatyti šalis="LT">
    <vardas>Jonas Jonaitis</vardas>
    <gatvė>Kalvarijų 5</gatvė>
    <miestas>Vilnius</miestas>
    <indeksas>LT-13952</indeksas>
  </pristatyti>
  <apmoka šalis="LT">
    <vardas>Petras Petraitis</vardas>
    <gatvė>Pylimo 2</gatvė>
    <miestas>Vilnius</miestas>
    <indeksas>LT-13434</indeksas>
  </apmoka>
  <pastaba>Dūžtanti prekė</pastaba>
  <prekiųSarašas>...</prekiųSarašas>
</užsakymas>
```



Pavyzdys: XML Schema

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="užsakymas" type="UžsakymoTipas" />

  <xs:complexType name="UžsakymoTipas">
    <xs:sequence>
      <xs:element name="pristatyti" type="AdresoTipas" />
      <xs:element name="apmoka" type="AdresoTipas" />
      <xs:element name="pastaba" type="xs:string"
                    minOccurs="0" />

      <xs:element name="prekiųSąrašas" type="..." />
      ...
    </xs:sequence>
    <xs:attribute name="data" type="xs:date" />
  </xs:complexType>
```



Pavyzdys: XML Schema (2)

```
<xs:complexType name="AdresoTipas">
  <xs:sequence>
    <xs:element name="vardas" type="xs:string"/>
    <xs:element name="gatvė" type="xs:string"/>
    <xs:element name="miestas" type="MiestoTipas"/>
    <xs:element name="indeksas" type="xs:string"/>
  </xs:sequence>
  <xs:attribute name="šalis" type="xs:string" fixed="LT"/>
</xs:complexType>
```

```
<xs:simpleType name="MiestoTipas">
  <xs:restriction base="xs:string">
    <xs:maxLength value="20"/>
  </xs:restriction>
</xs:simpleType>
...
</xs:schema>
```



XML Schema dokumento antraštė

- XML Schema yra W3C sukurta XML *kalba*, t.y., turi savo žymes ir atributus, bei savo vardų sritį:
 - <http://www.w3.org/2001/XMLSchema>
 - Šią vardų sritį būtina naudoti visuose XML Schema dokumentuose

- XML Schema dokumento antraštė:

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<xs:schema
```

```
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

```
...
```

```
</xs:schema>
```

- Žymės, prasidedančios prefiksu "**xs**", yra XML Schema kalbos žymės, kitos žymės yra jūsų aprašomos kalbos žymės
- Vietoj "**xs**" galima naudoti bet kokią kitą patinkantį prefiksą



Kurso turinys

- Žymių paskelbimas
- Atributų paskelbimas
- Paprasti tipai
- Sudėtingi tipai
- Integralumo apribojimai
 - Raktai, unikalumo reikalavimai
 - Nuorodos į raktus



Žymių deklaracija

- Žymės (sinonimas *elementas*) deklaracija yra:

1. žymės vardo susiejimas su žymės tipu,
2. žymės pasikartojimo skaičiaus nurodymas (nebūtina)
3. reikšmės pagal nutylėjimą nurodymas (nebūtina)

```
<xs:element name="žymėsVardas" type="tipoVardas">  
  ...  
</xs:element>
```

- Tipas `anyType` yra bazinis tipas (iš jo paveldi visi kiti tipai)
 - Jis niekaip nesuvaržo (t.y., netipizuoja) žymės turinio – žymės viduje gali būti tiek skaičiai, tiek bet koks tekstas, tiek ir kitos žymės

```
<xs:element name="KnygųSąrašas" type="xs:anyType" />
```

- Vardų srities prefiksą nurodyti būtina!

- Jei žymės tipas nenurodytas, tai tipas yra `anyType`:

```
<xs:element name="KnygųSąrašas" />
```



Žymės deklaracijos atributai

- `minOccurs` – kiek kartų mažiausiai žymė gali pasikartoti (pagal nutylėjimą: 1)
- `maxOccurs` – kiek kartų daugiausiai žymė gali pasikartoti (pagal nutylėjimą: 1)



Atributų deklaracija

- Atributo deklaracija yra:
 - atributo vardo susiejimas su atributo tipu,
 - atributo pasirodymo apribojimas (nebūtina),
 - atributo reikšmė pagal nutylėjimą

```
<xs:attribute name="atributoVardas"
               type="atributoTipas" />
```

- Deklaracijoje galima naudoti šiuos atributus:
 - use – nusako pasirodymo apribojimą,
 - Reikšmės: (optional | prohibited | required),
pagal nutylėjimą: optional



Globalios žymės ir atributai

- Globalios žymės/atributai yra sukuriami deklaracijomis, kurios yra žymės `xs:schema` vaikai
- Po sukūrimo, į juos gali būti daroma nuoroda iš kitų deklaracijų pasinaudojant `ref` atributu
- Globalios žymės deklaracija leidžia naudoti šią žymę kaip **šakninę** viso dokumento žymę
- Globalios deklaracijos negali turėti atributo `ref`, jos privalo turėti atributą `type`
- Globalios deklaracijos negali turėti atributų `minOccurs`, `maxOccurs`, ir `use`



XML Schema turinio rūšys

	Žymių/atributų turinio rūšys			
	Mišrus (mixed)	Sudėtingas (complex)	Paprastas (simple)	Tuščias (empty)
Gali turėti vaikinių žymių	Taip	Taip	Ne	Ne
Gali turėti tekstą	Taip	Ne	Taip	Ne

- Žymių turinys gali būti visų keturių rūšių
- Atributų turinys gali būti tik paprastas arba tuščias



Pavyzdžiai

- Paprastas turinys:

- Žymė: `<pavadinimas>Algebra</pavadinimas>`
- Atributas: `kalba="lt"`

- Tuščias turinys:

- Žymė: `<pavadinimas></pavadinimas>`, arba tiesiog `<pavadinimas/>`
- Atributas: `kalba=""`

- Mišrus turinys:

`<pastaba>`

Šita prekė yra

`dūžtanti`

Ryte `<i>nepristatinėti</i>!`

`</pastaba>`

- Jei pastabos vaikai būtų tik žymės, tada turinys būtų sudėtingas

Mišraus turinio pavyzdys

- Čia ne mišrus, o sudėtingas turinys:

```
<kažkas>
```

```
  <aaa>kva</aaa>
```

```
  <bbb>miau</bbb>
```

```
</kažkas>
```

Žymė "kažkas"
tekstinių **vaikų**
neturi! Turi tik
vaikines žymes.

- O va čia – mišrus:

```
<kažkas>
```

```
  <aaa>kva</aaa>
```

```
  <bbb>miau</bbb>
```

```
  kuku
```

```
</kažkas>
```

O čia turi ir
vaikines žymes, ir
vaikinį tekstinį
mazgą "kuku".



XML Schema tipų sistema

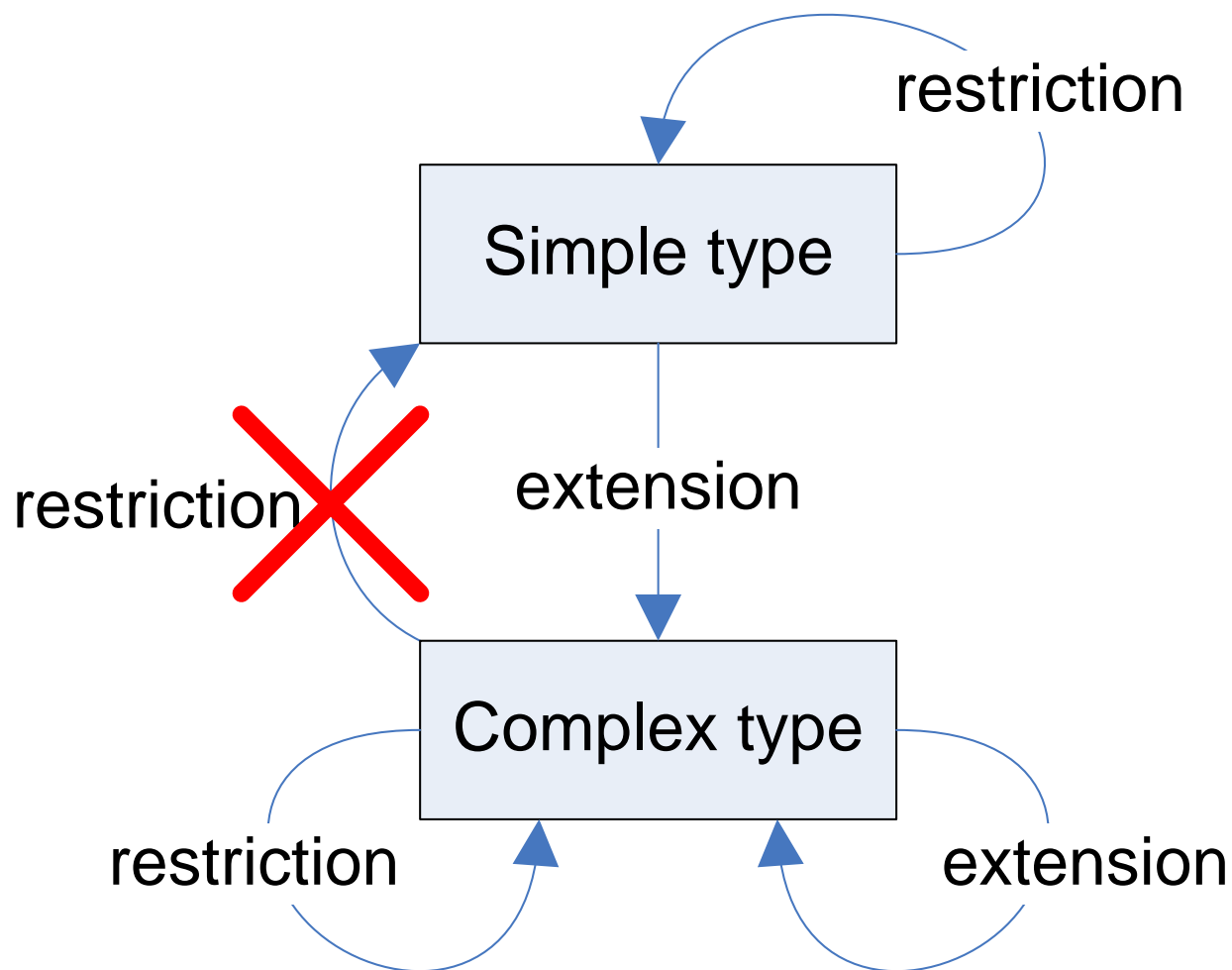
- Paprastieji tipai (`simpleType`)
 - Standartiniai duomenų tipai
 - `xs:string`, `xs:integer`, `xs:date`, ...
 - Išvestiniai duomenų tipai
 - Susikuriate savo paprastus duomenų tipus, *apribodami* standartinius duomenų tipus
 - Skirti paprastam ir tuščiam turiniui aprašyti
 - **Atributai žymės viduje negalimi!!!**
- Sudėtingi tipai (`complexType`)
 - Standartinių nėra, visus turite susikurti (*išvesti*) patys
 - Skirti visų rūšių turiniams aprašyti
 - **Žymių viduje galima naudoti atributus**



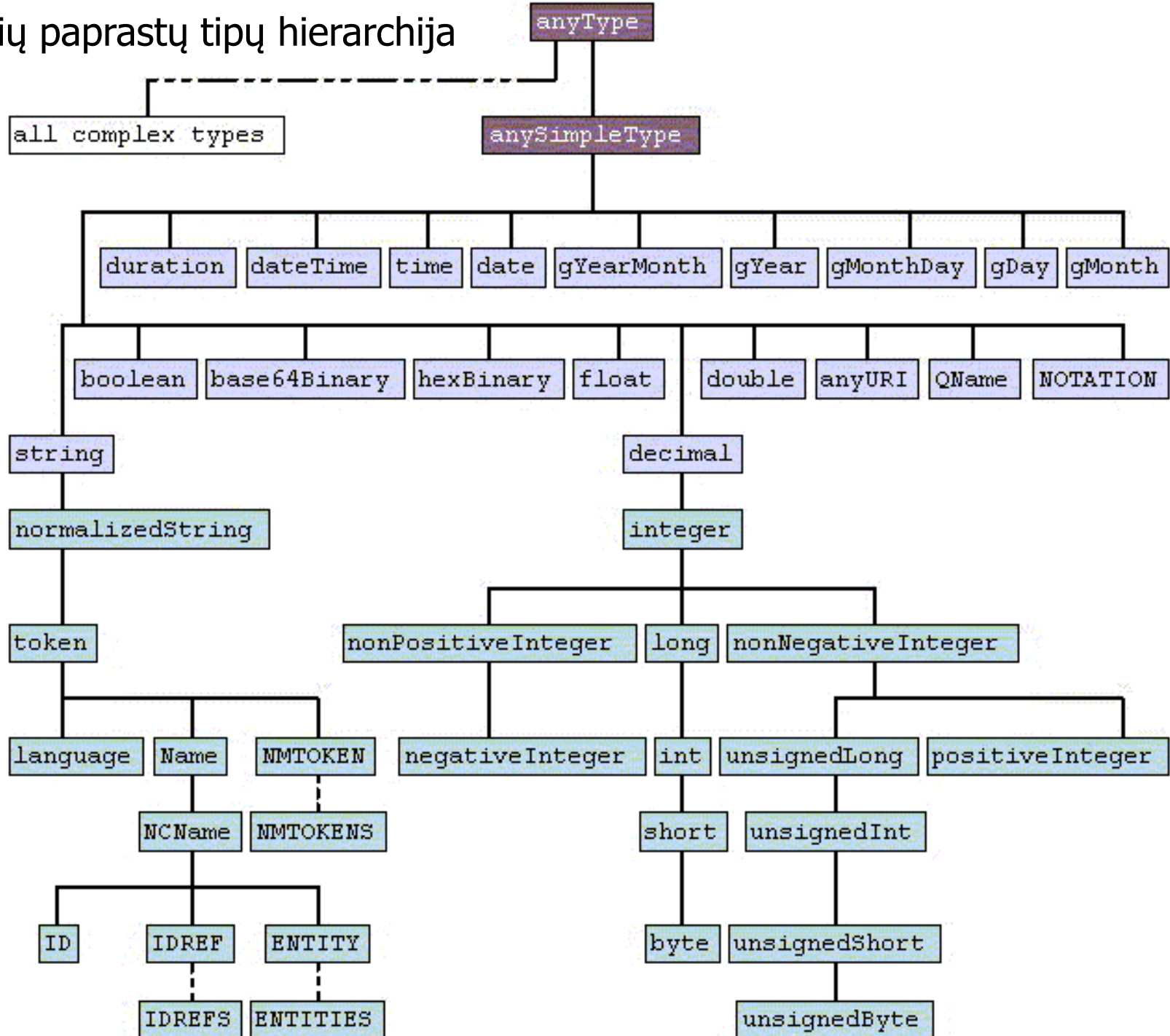
Nauju tipų išvedimo būdai

- Yra du tipų išvedimo būdai:
 - Apribojimas (restriction) – naujas tipas būna "siauresnis"/"mažesnis" nei senas
 - Išplėtimas (extension) – naujas tipas būna "platesnis"/"didesnis" už seną

Tipu su paprastu turiniu išvedimo būdai



Standartinių paprastų tipų hierarchija



string	Confirm this is electric	
normalizedString	Confirm this is electric	Tabuliacija ir eilutės pabaigos simboliai pakeisti į tarpą.
token	Confirm this is electric	Ištrina tarpus pradžioje ir pabaigoje. Du (ir daugiau) tarpai keičiami vienu.
integer	...-1, 0, 1, ...	Nuo minus begalybės iki plus begalybės
positiveInteger	1, 2, ...	
negativeInteger	... -2, -1	
nonNegativeInteger	0, 1, 2, ...	
nonPositiveInteger	... -2, -1, 0	
long	-9223372036854775808, ... -1, 0, 1, ... 9223372036854775807	
unsignedLong	0, 1, ... 18446744073709551615	
decimal	-1.23, 0, 123.4, 1000.00	
float	-INF, -1E4, -0, 0, 12.78E-2, 12, INF, NaN	
double	-INF, -1E4, -0, 0, 12.78E-2, 12, INF, NaN	
boolean	true, false, 1, 0	
duration	P1Y2M3DT10H30M12.3S	1 year, 2 months, 3 days, 10 hours, 30 minutes, and 12.3 seconds
dateTime	1999-05-31T13:20:00.000-05:00	May 31st 1999 at 1.20pm Eastern Standard Time which is 5 hours behind Co-Ordinated Universal Time, see (2)
date	1999-05-31	
time	13:20:00.000	
gYear	1999	
gYearMonth	1999-02	the month of February 1999, regardless of the number of days
gMonth	--05	May
gMonthDay	--05-31	every May 31st
gDay	---31	the 31st day
language	en-GB, en-US, fr	valid values for xml:lang as defined in XML 1.0

Paprastųjų tipų išvedimas

- Pavyzdys: aprašome paprastą turinį – sąskaitos numerį, t.y., tam tikrą šabloną atitinkancią eilutę (string):

```
<xs:simpleType name="SąskaitosNr">  
  <xs:restriction base="xs:string">  
    <xs:length value="20" />  
    <xs:pattern value="LT[0-9]{18}" />  
  </xs:restriction>  
</xs:simpleType>
```

Reguliari
išraiška

- Tokį tipą galima priskirti tiek žymei, tiek ir atributui:
 - <Sąskaita>LT123456789012345678</Sąskaita>
 - <Sąskaita nr="LT123456789012345678">...</Sąskaita>

Paprastųjų tipų struktūra (1)

```
<simpleType name = NCName>
```

```
  Content: (restriction | list | union)
```

```
</simpleType>
```

```
<restriction base = QName>
```

```
  Content: (simpleType?,
```

```
    (minExclusive | minInclusive |  
     maxExclusive | maxInclusive |  
     totalDigits | fractionDigits |  
     length | minLength | maxLength |  
     enumeration | pattern)*
```

```
)
```

```
</restriction>
```

Apribojamo
tipo
pavadinimas

Apribojimai
(facets)



Paprastųjų tipų struktūra – sąrašo tipas

```
<list itemType = QName>  
  Content: (simpleType?)  
</list>
```

■ Pavyzdys:

```
<xs:simpleType name="intList">  
  <xs:list itemType="xs:integer"/>  
</xs:simpleType>
```

■ Tarkim, žymė skaičiai turi tipą intList:

```
<skaičiai>20003 15037 95977 95945</skaičiai>
```

■ Sąrašo tipą galima apriboti, nurodant jo ilgį, šabloną, ir pan.



Paprastųjų tipų struktūra – sąjungos tipas

```
<union memberTypes = List of QName>  
  Content: (simpleType*)  
</union>
```

■ Pavyzdys:

```
<xs:simpleType name="intDateUnion">  
  <xs:union memberTypes="intList xs:date"/>  
</xs:simpleType>
```

■ Tarkim, žymė sąrašas turi tipą intDateUnion:

```
<sąrašas>20003 2006-01-02 -150 555</sąrašas>
```




Paprastų tipų apribojimai

- Visiems iš `xs:string` išvestiems tipams galima taikyti šiuos apribojimus:
 - `length` - ilgis
 - `minLength` – mažiausias ilgis
 - `maxLength` – didžiausias ilgis
 - `pattern` – šablonas (reguliari išraiška)
 - `enumeration` – reikšmių išvardijimas
- Visiems skaitiniams tipams ir datoms:
 - `pattern`, `enumeration` – tas pats
 - `minInclusive` – mažiausias įtraukiant (matematiškai "[")
 - `minExclusive` – mažiausias neįtraukiant (matematiškai "(")
 - `maxInclusive` – didžiausias įtraukiant (matematiškai "]")
 - `maxExclusive` – didžiausias neįtraukiant (matematiškai ") ")
 - `totalDigits` – koks turi būti skaičiaus ilgis (skaitmenų skaičius)
 - `fractionDigits` – kiek gali būti skaitmenų po kablelio



enumeration pavyzdys

```
<xs:simpleType name="Šalys">
  <xs:restriction base="xs:string">
    <xs:enumeration value="LT" />
    <xs:enumeration value="US" />
    <xs:enumeration value="RU" />
    <!-- čia galim rašyti ir kitus apribojimus -->
  </xs:restriction>
</xs:simpleType>
```