

# XSchema

## validácia XML dokumentov

## XSchema - rozšírenie validácie

XML jazyk - definovaná syntax a sémantika

podobne ako DTD:

- definícia elementov, atribútov
- definícia štruktúry dokumentu
- definícia obsahu elementov
- definícia default hodnôt pre atribúty
- atď.

## XSchema - rozšírenie validácie (pokr.)

výhody (voči DTD):

- podpora namespace
- podpora dátových typov (kontrola platnosti, rozsahu)
- XML syntax
- rozšíriteľné

preddefinovaná namespace:

**<http://www.w3.org/2001/XMLSchema>**

## XSchema - príklad

XML  
objednávka tovaru

inštancia dokumentu,  
triedu definuje schéma

```
<?xml version="1.0"?>
<purchaseOrder orderDate="1999-10-20">
  <shipTo country="US">
    <name>Alice Smith</name>
    <street>123 Maple Street</street>
    <city>Mill Valley</city>
    <state>CA</state>
    <zip>90952</zip>
  </shipTo>
  <billTo country="US">
    <name>Robert Smith</name>
    <street>8 Oak Avenue</street>
    <city>Old Town</city>
    <state>PA</state>
    <zip>95819</zip>
  </billTo>
  <comment>Hurry, my lawn is going wild!</comment>
  <items>
    <item partNum="872-AA">
      <productName>Lawnmower</productName>
      <quantity>1</quantity>
      <USPrice>148.95</USPrice>
      <comment>Confirm this is electric</comment>
    </item>
    <item partNum="926-AA">
      <productName>Baby Monitor</productName>
      <quantity>1</quantity>
      <USPrice>39.98</USPrice>
      <shipDate>1999-05-21</shipDate>
    </item>
  </items>
</purchaseOrder>
```

## XSchemata - príklad

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
```

```
<xsd:element name="purchaseOrder" type="PurchaseOrderType"/>
<xsd:element name="comment" type="xsd:string"/>
```

```
<xsd:complexType name="PurchaseOrderType">
  <xsd:sequence>
    <xsd:element name="shipTo" type="USAddress"/>
    <xsd:element name="billTo" type="USAddress"/>
    <xsd:element ref="comment" minOccurs="0"/>
    <xsd:element name="items" type="Items"/>
  </xsd:sequence>
  <xsd:attribute name="orderDate" type="xsd:date"/>
</xsd:complexType>
```

```
<xsd:simpleType name="SKU">
  <xsd:restriction base="xsd:string">
    <xsd:pattern value="\d{3}-[A-Z]{2}" />
  </xsd:restriction>
</xsd:simpleType>
```

```
<xsd:complexType name="Items">
  <xsd:sequence>
    <xsd:element name="item" minOccurs="0" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:sequence>
          <xsd:element name="productName" type="xsd:string"/>
          <xsd:element name="quantity">
            <xsd:simpleType>
              <xsd:restriction base="xsd:positiveInteger">
                <xsd:maxExclusive value="100"/>
              </xsd:restriction>
            </xsd:simpleType>
          </xsd:element>
          <xsd:element name="USPrice" type="xsd:decimal"/>
          <xsd:element ref="comment" minOccurs="0"/>
          <xsd:element name="shipDate" type="xsd:date" minOccurs="0"/>
        </xsd:sequence>
        <xsd:attribute name="partNum" type="SKU" use="required"/>
      </xsd:complexType>
    </xsd:element>
  </xsd:sequence>
</xsd:complexType>

</xsd:schema>
```

## Deklarácie základných štruktúr

### typy elementov

- jednoduchý (**simple**) - obsahujú len hodnotu
- zložený (**complex**) - obsahujú iné elementy, môžu mať atribúty

schéma obsahuje definície zložených a jednoduchých typov pre danú triedu dokumentov...

- niektoré jednoduché typy sú preddefinované (napr. xsd:string)

... a pravidlá ich umiestnenia v dokumente

## Deklarácie základných štruktúr - complex type

### element complexType

- definuje deti (meno, typ, kardinalitu, ...) - **element**
- definuje atribúty (meno, typ, význam, ...) - **attribute**

```
<xsd:complexType name="USAddress">
  <xsd:sequence>
    <xsd:element name="name" type="xsd:string"/>
    <xsd:element name="street" type="xsd:string"/>
    <xsd:element name="city" type="xsd:string"/>
    <xsd:element name="state" type="xsd:string"/>
    <xsd:element name="zip" type="xsd:decimal"/>
  </xsd:sequence>

  <xsd:attribute name="country" type="xsd:NMTOKEN" fixed="US"/>
</xsd:complexType>
```

## Deklarácie základných štruktúr - complex type (pokr.)

### deklarácia complex type dieťaťa

```
<xsd:complexType name="PurchaseOrderType">
  <xsd:sequence>
    <xsd:element name="shipTo" type="USAddress"/>
    <xsd:element name="billTo" type="USAddress"/>
    <xsd:element ref="comment" minOccurs="0"/>
    <xsd:element name="items" type="Items"/>
  </xsd:sequence>
  <xsd:attribute name="orderDate" type="xsd:date"/>
</xsd:complexType>
```

### ref - referencia na existujúci element (deklaráciu)

```
<xsd:element name="billTo" type="USAddress"/>
<xsd:element ref="comment" minOccurs="0"/>
<xsd:element name="items" type="Items"/>
```

## Ohraničenia kardinality

deklarácia dovolého počtu prvkov  
(DTD: ?, +, \*)

deklarácia *element* - atribúty **minOccurs**, **maxOccurs**

default hodnota 1 (pre obe)

dovolené možnosti: < 0 , 'unbounded' >  
minOccurs <= maxOccurs

## Ohraničenia kardinality (pokr.)

deklarácia atribútu - ak výskyt, tak práve raz

element *attribute* - atribút **use**  
*required* | *optional* | *prohibited*

default hodnoty - atribút **default**

- pre atribút - hodnota ak neuvedený (**use = optional**)
- pre element - ak uvedený bez hodnoty

konštantné hodnoty - atribút **fixed**  
(nesmie byť spolu s **default**)

## Globálne elementy a atribúty

deti elementu **schema**

môžu byť:

- koreňovým elementom inštancie dokumentu
- referencované (atribút **ref**)

nedovoľujú:

- použiť referencie na iný element  
(priamo jednoduchý/zložený typ)
- ohraničovať kardinalitu  
(bez *minOccurs*, *maxOccurs*, *use*)

## Jednoduché typy

preddefinované / deklarácia vlastných

Príklady preddefinovaných:

**string**  
**byte**, **unsignedByte**  
**integer**, **positiveInteger**, **nonNegativeInteger**, ...  
**int**, **long**, **short**  
**decimal**, **float**, **double**  
**boolean**  
**time**, **dateTime**, **duration**, **date**

**Name**, **anyURI**, **language**  
**ID**, **IDREF**, ...

## Jednoduché typy (pokr.)

deklarácia nových typov deriváciou z preddefinovaných

ohraničenie (**xsd:restriction**):

- platný interval
- dĺžka
- enumerácia
- vzor

zoznam (**xsd:list**) - vektor hodnôt jedného typu

zjednotenie (**xsd:union**) - vektor hodnôt jedného z viacerých typov

## Jednoduché typy - ohraničenie intervalov

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

dve samostatne aplikovateľné ohraničenia intervalu:

dolné: **minInclusive**, **minExclusive**

horné: **maxInclusive**, **maxExclusive**

## Jednoduché typy - ohraničenia dĺžky

```
<simpleType name="string6">
  <restriction base="string">
    <maxLength value="6"/>
  </restriction>
</simpleType>
```

Znakový reťazec s max.  
dĺžkou 6 znakov

```
<simpleType name="decimal14">
  <restriction base="decimal">
    <totalDigits value="14"/>
    <fractionDigits value="2"/>
  </restriction>
</simpleType>
```

Reálne číslo, max. číslic 14,  
2 desatinné miesta

*Poznámka:* default namespace

## Jednoduché typy - ohraničenie vzorom

```
<xsd:simpleType name="SKU">
  <xsd:restriction base="xsd:string">
    <xsd:pattern value="\d{3}-[A-Z]{2}"/>
  </xsd:restriction>
</xsd:simpleType>
```

aplikovateľné na (takmer) **všetky** jednoduché typy  
(pravidlá podobné *PERL*u)

**Chapter \d** - Chapter 1, ..., Chapter 9

**Chapter\s\d** - Chapter, znak whitespace, znak čísla

## Jednoduché typy - ohraničenie vzorom (pokr.)

triedy znakov:

**\d** - znak čísla (digit)      **\D** - iný znak než číslo  
**\s** - znak whitespace      **\S**  
**\w** - znak/číslo (word)      **\W**  
**\p{IsGreek}** - znak UNICODE Greek  
**\P{IsGreek}** - znak nie z (UNICODE Greek)

automat:

**a\*x** - x, ax, aaax, ...  
**a?x** - x, ax  
**a+x** - ax, aax, ...  
**(a|b)+x** - ax, bx, aabx, babx, ...

## Jednoduché typy - ohraničenie vzorom (pokr.)

vymenované znaky:

**[abc]x** - ax, bx, cx      ~= **[a-c]x**  
**[-ac]x** - -x, ax, cx      ~= **[ac-]x**  
**[^0-9]x** - ľubovoľný znak okrem čísla, nasleduje x

automat (ohraničenia):

**ab{2}x** - abbx  
**ab{2,4}x** - abbx, abbbx, abbbb  
**ab{2,}x** - abbx, abbbx, abbbb, ...  
**(ab){2}x** - ababx

<xsd:pattern value="\d{3}-[A-Z]{2}"/>

## Jednoduché typy - vymenovanie (enumerácia)

```
<xsd:simpleType name="USState">
  <xsd:restriction base="xsd:string">
    <xsd:enumeration value="AK"/>
    <xsd:enumeration value="AL"/>
    <xsd:enumeration value="AR"/>
    <!-- and so on ... -->
  </xsd:restriction>
</xsd:simpleType>
```

**všetky** jednoduché typy (okrem *boolean*)

## Jednoduché typy - vektor (list)

```
<xsd:simpleType name="listOfMyIntType">
  <xsd:list itemType="myInteger"/>
</xsd:simpleType>
```

neatomický jednoduchý typ

whitespace oddelené hodnoty jedného **atomického** typu

<listOfMyInt>20003 15037 95977 95945</listOfMyInt>

*Poznámka:* string obsahujúci whitespace

## Jednoduché typy - vektor (pokr.)

### kombinácia ohraničení

```
<xsd:simpleType name="USStateList">
  <xsd:list itemType="USState"/>
</xsd:simpleType>

<xsd:simpleType name="SixUSStates">
  <xsd:restriction base="USStateList">
    <xsd:length value="6"/>
  </xsd:restriction>
</xsd:simpleType>
```

## Jednoduché typy - zjednotenie (union)

### vektor hodnôt jedného typu zo zoznamu atomických typov a vektorov

```
<xsd:simpleType name="zipUnion">
  <xsd:union memberTypes="USState listOfMyIntType"/>
</xsd:simpleType>

<zips>CA</zips>
<zips>95630 95977 95945</zips>
<zips>AK</zips>
```

## Anonymné typy

### ak nie je treba používať v iných súvislostiach...

```
<xsd:complexType name="Items">
  <xsd:sequence>
    <xsd:element name="item" minOccurs="0" maxOccurs="unbounded">
      <xsd:complexType>
        <xsd:sequence>
          <xsd:element name="productName" type="xsd:string"/>
          <xsd:element name="quantity">
            <xsd:simpleType>
              <xsd:restriction base="xsd:positiveInteger">
                <xsd:maxExclusive value="100"/>
              </xsd:restriction>
            </xsd:simpleType>
          </xsd:element>
          <xsd:element name="USPrice" type="xsd:decimal"/>
          <xsd:element ref="comment" minOccurs="0"/>
          <xsd:element name="shipDate" type="xsd:date" minOccurs="0"/>
        </xsd:sequence>
        <xsd:attribute name="partNum" type="SKU" use="required"/>
      </xsd:complexType>
    </xsd:element>
  </xsd:sequence>
</xsd:complexType>
```

## Obsah elementov

### simple type, pridaný atribút

```
<internationalPrice currency="EUR">423.46</internationalPrice>

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

## Obsah elementov (pokr.) - zmiešaný obsah

```
<letterBody>
<salutation>Dear Mr. <name>Robert Smith</name>.</salutation>
Your order of <quantity>1</quantity> <productName>Baby
Monitor</productName> shipped from our warehouse on
<shipDate>1999-05-21</shipDate>. ....
</letterBody>

<xsd:element name="letterBody">
  <xsd:complexType mixed="true">
    <xsd:sequence>
      <xsd:element name="salutation">
        <xsd:complexType mixed="true">
          <xsd:sequence>
            <xsd:element name="name" type="xsd:string"/>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
      <xsd:element name="quantity" type="xsd:positiveInteger"/>
      <!-- etc. -->
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
```

*Poznámka: úplná validácia (oproti DTD)*

## Obsah elementov (pokr.) - prázdny element

```
<internationalPrice currency="EUR" value="423.46"/>

<xsd:element name="internationalPrice">
  <xsd:complexType>
    <xsd:complexContent>
      <xsd:restriction base="xsd:anyType">
        <xsd:attribute name="currency" type="xsd:string"/>
        <xsd:attribute name="value" type="xsd:decimal"/>
      </xsd:restriction>
    </xsd:complexContent>
  </xsd:complexType>
</xsd:element>
```

(definujeme complex type bez definovania elementov)

## Elementy bez validácie

preddefinovaný typ, ktorý môže obsahovať ľubovoľné znaky - **anyType**

```
<xsd:element name="anything" type="xsd:anyType"/>
```

default typ pre všetky deklarácie

## Deklarácia možného výskytu detí

**xsd:sequence** - presne určené poradie  
(minOccurs, maxOccurs)

**xsd:choice** - v dokumente práve jedno z detí

deti je možné aj združovať do skupín: **xsd:group**  
(sequence, choice, ...)

## Deklarácia možného výskytu detí - príklad

```
<xsd:complexType name="PurchaseOrderType">
  <xsd:sequence>
    <xsd:choice>
      <xsd:group ref="shipAndBill"/>
      <xsd:element name="singleUSAddress" type="USAddress"/>
    </xsd:choice>
    <xsd:element ref="comment" minOccurs="0"/>
    <xsd:element name="items" type="Items"/>
  </xsd:sequence>
  <xsd:attribute name="orderDate" type="xsd:date"/>
</xsd:complexType>

<xsd:group name="shipAndBill">
  <xsd:sequence>
    <xsd:element name="shipTo" type="USAddress"/>
    <xsd:element name="billTo" type="USAddress"/>
  </xsd:sequence>
</xsd:group>
```

## Deklarácia možného výskytu detí (pokr.)

### xsd:all

- každé dieťa sa môže objaviť najviac raz
- v ľubovoľnom poradí

ale:

- iba ako jediný top-level model ľubovoľnej deklarácie
- deti musia byť individuálne elementy

```
<xsd:complexType name="PurchaseOrderType">
  <xsd:all>
    <xsd:element name="shipTo" type="USAddress"/>
    <xsd:element name="billTo" type="USAddress"/>
    <xsd:element ref="comment" minOccurs="0"/>
    <xsd:element name="items" type="Items"/>
  </xsd:all>
  <xsd:attribute name="orderDate" type="xsd:date"/>
</xsd:complexType>
```

## Prázdne hodnoty

chceme signalizovať existenciu elementu, ale bez udanej hodnoty - **xsi:nil**

```
<xsd:element name="shipDate" type="xsd:date" nillable="true"/>

<shipDate xsi:nil="true"></shipDate>
```

- aplikovateľné iba na elementy
- nemôže mať obsah, iba atribúty

**xsi** namespace:

**<http://www.w3.org/2001/XMLSchema-instance>**

## Schémy - čo sme vynechali

rozšírený kurz o schémach:

- anotácie (dokumentácia) schémy
- zoskupovanie atribútov
- prepájanie viacerých schém (target namespace)
- derivácia nových typov (dedenie, rozširovanie, obmedzovanie, final deklarácie)
- abstraktné elementy a typy
- slovníky typov
- kontrola „nekontrolovaných“ oblastí
- a mnohé ďalšie detaily

## Kontrola unikátnosti

(pomenovaná) požiadavka unikátnosti hodnoty  
(kombinácie hodnôt) v nejakej množine prvkov -  
**xsd:unique**

**xsd:selector xpath = expression**  
- výber množiny hodnôt

**xsd:field xpath = expression**  
- lokálne určenie unikátneho parametra

*Poznámka:* obmedzený XPath

## Kontrola unikátnosti - príklady

```
<element name="Auftrag" type="elmAuftrag">
  <unique name="dummy1">
    <selector xpath="Teilauftrag"/>
    <field xpath="@Teilauftrags_Nummer"/>
  </unique>
</element>
```

```
<unique name="dummy2">
  <selector xpath="regions/zip"/>
  <field xpath="@code"/>
  <field xpath="part/@number"/>
</unique>
```

## Definovanie použitej schémy - najjednoduchší prípad

schéma bez deklarovania vlastnej namespace

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema">
```

```
  <!-- declarations -->
```

```
</schema>
```

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Auftrag
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="d:\xml\_XSD\Auftrag.xsd"

  Auftrags_Nummer = "ZP999549"
>
<!-- and so on... -->
```

## XSchem - zhrnutie

**nástroj na (relatívne rozsiahlu) validáciu  
XML dokumentov**

základné deklarácie štruktúr (podobne ako DTD)  
(definície elementov, atribútov)

kontrola dátových typov  
(široká preddefinovaná skupina jednoduchých typov)  
(kontrola platnosti, rozsahu, vzor, ...)

## XSchema - zhrnutie

definícia štruktúry

(poradie elementov, povolený počet, content model)

kontrola unikátnosti

široké a univerzálne praktické použitie

(“dedenie”, slovníky typov, kombinácie viacerých schém)

## Ďakujem za pozornosť



## Použitá (a študijná) literatúra

[www.w3.org](http://www.w3.org)

oficiálne stránky WWW Consortia

(odporúčania sú na adresách namespace technológií)

[www.w3schools.com](http://www.w3schools.com)

tutoriály k väčšine web-technológií

[www.zvon.org](http://www.zvon.org)

český server - veľa materiálov, tutoriály