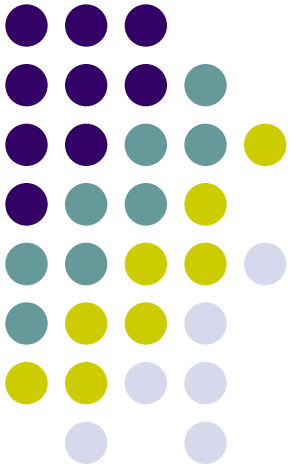
NPRG039

Advanced Aspects and New Trends in XML (and Related) Technologies

RNDr. Irena Holubová, Ph.D.

holubova@ksi.mff.cuni.cz

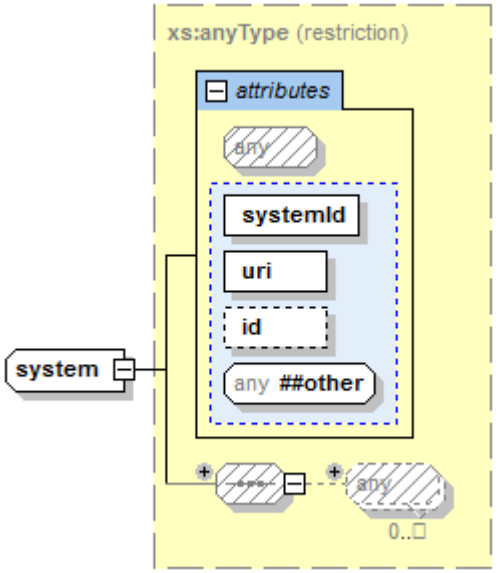


Labs 1. Modelling and generating of XML data

http://www.ksi.mff.cuni.cz/~holubova/NPRG039/indexCV.html

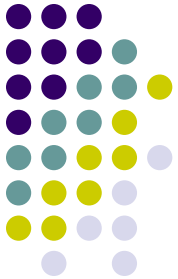
Altova XML Spy

- Installed in labs
- Does not support modelling of XML data, but supports data visualization
- Tab: Schema
  - XML schema visualization
- Menu DTD/Schema/Generate sample XML file…
  - XML data generating



http://www.altova.com/xml-editor/

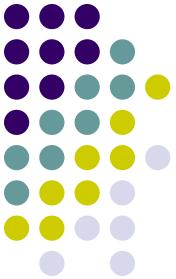
ToXgene



- Toronto XML Server Data Generator
  - http://www.cs.toronto.edu/tox/toxgene/
  - http://www.cs.toronto.edu/tox/toxgene/docs/ToXgene_manual.pdf
    - Installation, all constructs
- Annotated "XML schema"
  - Not true XSD
  - `tsl` file
- Root element tox-template:

```
<!ELEMENT tox-template
 (tox-distribution|
  simpleType|complexType|
  tox-list|
  tox-document)*>
```

tox-distribution

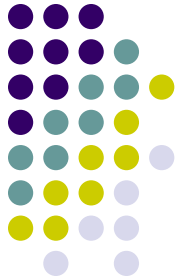


```
<tox-distribution name="age" type="normal"
  minInclusive="18" maxInclusive="127"
  mean="30" variance="15">
</tox-distribution>
```

```
<tox-distribution name="watches" type="exponential"
  minInclusive="0" maxInclusive="10" mean="4">
</tox-distribution>
```

```
<tox-distribution name="discount" type="user-defined"
min="0" max="30">
  <enumeration value="0" tox-percent="50"/>
  <enumeration value="5" tox-percent="25"/>
  <enumeration value="10" tox-percent="15"/>
  <enumeration value="30" tox-percent="10"/>
</tox-distribution>
```

simpleType



```
<simpleType name="pick_category">
  <restriction base="nonNegativeInteger">
    <tox-number tox-distribution="c1"/>
  </restriction>
</simpleType>
```

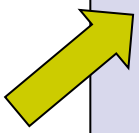
```
<simpleType name="lname">
  <restriction base="string">
    <tox-string type="lname"/>
  </restriction>
</simpleType>
```

```
<simpleType name="year">
  <restriction base="string">
    <tox-value>1942</tox-value>
  </restriction>
</simpleType>
```

tox-string types

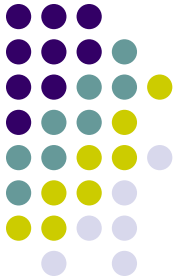
Type	Description
`text`	A textual string generated according to the TPC-H benchmark [5] rules.
`word`	A word, extracted from the list of words used in the XMark benchmark [4].
`xmrk_text`	A textual string generated according to the XMark benchmark rules.
`fname`	A first name, extracted from the list of first names in the XMark benchmark.
`lname`	A last name, extracted from the list of last names in the XMark benchmark.
`city`	A city name, extracted from the list of cities in the XMark benchmark.
`province`	A province name, extracted from the list of provinces in the XMark benchmark.
`country`	A country name, extracted from the list of countries in the XMark benchmark.
`domain`	An internet domain name, extracted from the list of domains in the XMark benchmark.
`email`	An email address of the form x`.`y`@`z, where x is an instance of `fname`, y is an instance of `lname` and z is an instance of `domain`.
`gibberish`	A random string in {a,...,z,...,A,...,Z,0,...,9} ∪ {#,^,*,(,),_,-,=,$,+,{,},[,],?,.,/,~, }.

complexType



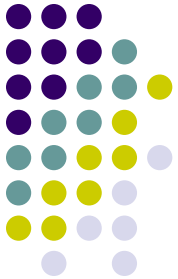
```
<complexType name="PurchaseOrderType">
  <element name="shipTo" type="USAddress"/>
  <element name="billTo" type="USAddress"/>
  <element name="comment" type="string"/>
  <element name="items"  type="Items"/>
  <attribute name="orderDate">
    <simpleType>
      <restriction base="string">
        <tox-string type="city"/>
      </restriction>
    </simpleType>
  </attribute>
</complexType>
```

tox-document



```
<tox-document name="output/review" copies="1000"
              starting-number="0">
  <element name="purchaseOrder"
           type="PurchaseOrderType"
           minOccurs="1" maxOccurs="1"/>
</tox-document>
```

tox-list



- to generate correlated content
  - e.g., a catalog of books (identified by ISBNs) and a collection of reader reviews about the books

```
<tox-list name="book_list" unique="book/isbn">
  <element name="book_rec" minOccurs="100"
                           maxOccurs="100">
    <complexType>
      <element name="isbn" type="isbn type"/>
      <element name="title">
        <simpleType>
          <restriction base="string">
            <tox-string type="text"/>
          </restriction>
        </simpleType>
      </element>
    </complexType>
  </element>
</tox-list>
```

tox-list

```
<tox-document name="books">
  ...
  <element name="book" minOccurs="100">
    <complexType>
      <tox-scan path="[book_list/book_rec]" name="a">
        <attribute name="isbn" type="isbn type">
          <tox-expr value="[$a/isbn/!]"/>
        </attribute>
        <element name="title" type="string">
          <tox-expr value="[$a/title/!]"/>
        </element>
        <element name="author" maxOccurs="5" type="string">
          <tox-string type="gibberish" maxLength="30"/>
        </element>
      </tox-scan>
    </complexType>
  </element>
  ...
</tox-document>
```

a cursor for a sequential scan over a list of elements

query relative to cursor $a

! = shorthand for CDATA

Task

- Create a model and an XML Schema expression of a (non-trivial) XSD
  - Using eXolutio (or a similar modeling tool)
- Generate respective XML documents as realistic as possible
  - Using ToXgene (or a similar generating tool)