

Exercise sheet 1

Recuperación de Información

XML, DTDs, XML Schema, DOM/SAX

1) *Play around: Check the **examples.zip** file attached to this exercise sheet and understand the example DTDs and XML Schemas therein. You should be able to explain what's going on and what the grammars define in the next lesson!*

2) *Encode the following tabular information as a well-formed XML document in two different ways.*

Employee	ID	Department	Email	Tel.
Gustav Sielmann	2028	Development	gsielmann@Dot.com	6045
Arnold Rummer	2037	Development	arummer@Dot.com	6085
Johann Neumeier	5042	Marketing	jneumeier@Dot.com	6093
Max Müller	5048	Marketing	mmueller@Dot.com	
Karl Thaler	7123	IT	kthaler@Dot.com	6152

Use one of the online XML validators to check if XML files created by you at this and previous tutorials are well-formed. Use one more online validator to check if the XML file from the exercise 2 is valid against the XML schema you created. See References below for online XML validators. Report on the validation results.

- 3) a) *Write a DTD for each of the XML files from exercise 2)*
b) *Write an XML Schema for one of the XML files from exercise 2)*
c) *Use one of the online XML validators to check if XML files created by you at this and previous tutorials are well-formed. Use one more online validator to check if the XML file from the exercise 2 is valid against the XML schema you created.*

Online XML validators:

<http://www.w3.org/2001/03/webdata/xsv>
<http://www.stg.brown.edu/service/xmlvalid/>
<http://www.hcrc.ed.ac.uk/~richard/xml-check.html>
http://www.w3schools.com/dom/dom_validate.asp
<http://apps.gotdotnet.com/xmltools/xsdvalidator/>
<http://tools.decisionsoft.com/schemaValidate.html>

Alternatively, you can use an offline tool such as e.g. XMLSpy.

5) Create an XML-Schema for the following XML document, using:

- *Global types (type referencing)*
- *Restricted types*
- *Extended types*
- *Documentation-Annotations*

```
<?xml version="1.0" encoding="iso-8859-1"?>

<Staff>
  <Marketing>
    <Employee ID="232">
      <Name>Anton Kiefler</Name>
      <Salary ContractType="Assistant Sales Director">2430,00</Salary>
      <Email>akiefler@Dot.com</Email>
      <Adress>
        <Street>Schiesstattstrasse 21</Street>
        <City>Salzburg</City>
        <Zip>5020</Zip>
        <Country>Austria</Country>
      </Adress>
    </Employee>
    <Employee ID="251">
      <Name>Michaela Stümper</Name>
      <Salary ContractType="Secretary">1380,50</Salary>
      <Email>mstuemper@Dot.com</Email>
      <Adress>
        <Street>Schiesstattstrasse 21</Street>
        <City>Salzburg</City>
        <Zip>5020</Zip>
        <Country>Austria</Country>
      </Adress>
    </Employee>
  </Marketing>
  <Development>
    <Employee ID="472">
      <Name>Fritz the Cat</Name>
      <Salary ContractType="External Researcher">4250,00</Salary>
      <Email>fcats@Dot.com</Email>
      <Adress>
        <Street>Kittygasse 5</Street>
        <City>Katzen</City>
        <Zip>5450</Zip>
        <Country>Austria</Country>
      </Adress>
    </Employee>
  </Development>
</Staff>
```

6) Create your own validator using the Xerces library which supports both DOM and SAX.

<http://xerces.apache.org/>

<http://xml.apache.org/xerces-c/>

Validate with your own validator your own XML files from exercise 3)