

## Aufgabe 2-0: BASE XML SETUP

### Steps Perfor med

Download the base XML for Team 13:

1. Original download: <http://online.univie.ac.at/vlvz?kapitel=707&semester=S2015&format=xml> is empty, since study programme changed, communication with Mag. Mangler and adaption to new URL: <http://online.univie.ac.at/vlvz?kapitel=706&semester=S2015&format=xml>
2. Cleaning of XML through search/replace for XML entities with the following results  
Replace &apos; -> &apos; 721 occurrences  
Replace &quot; -> &quot; 15 occurrences  
Replace &amp; -> &amp; 507 occurrences  
Replace &lt; -> &lt; 0 occurrences  
Replace &gt; -> &gt; 0 occurrences  
remove nbsp; in Line 466 -> This was a random space in the middle of a word  
Corrected numeric entities &# -> &# 439
3. Development of XSL based transformation in 3 steps to recognize the hierarchical structure and transform to English tagnames.

All 3 xsl transformations are performed using xsltproc XSL\_FILE INPUT\_XML > OUTPUT.XML

1 extractHierarchyAndRename.xsl: clean anchor levell at the beginning of file, rename tag names for numeric chapters to tag "chapter" and add a specific attribute to identify the level of a chapter

```
xsltproc 1\ extractHierarchyAndRename.xsl  
baseXML_706_download.xml > output_step1.xml
```

2 constructParentRelation.xsl: find the parent level ID through the order of chapters and add parentID attribute (preceding-sibling::chapter[@level = \$parentLevel][1]/@id)

```
xsltproc 2\ constructParentRelation.xsl output_step1.xml  
> output_step2.xml
```

3 constructHierarchyAndRename.xsl create the hierarchy as suggested and perform language transformation. Perform any kind of calculation and information structure needed for exercises.

```
xsltproc 3\ constructHierarchyAndRename.xsl  
output_step2.xml > output_step3.xml
```

Pretty-print result XML

```
xmllint --format output_step3.xml >  
baseXML_tranformed.xml
```

<p><b>Solution</b></p>	<pre>&lt;?xml version="1.0"?&gt; &lt;studies name="7.6. Masterstudien Geschichte" period="Sommersemester 2015" numberOfStudyPrograms="9"&gt;   &lt;study name="7.6.1 MA Geschichte 2014" id="706_2" parentID="706_1"&gt;     &lt;module name="Pflichtmodul 1: Einführung in die Schwerpunkte" id="706_3" parentID="706_2"&gt;       &lt;category name="Schwerpunkt-Einführung: Vorlesungen" id="706_4" parentID="706_3"&gt;         &lt;course id="070116" name="Introduction to Global History" title="Introduction to Global History" type="V0" groupcount=           &lt;group appointmentcount="13" signlanguage="No" livestream="No" learningplatform="MOODLE" learningplatformurl="https             &lt;appointment date="17-Mar-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="24-Mar-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="14-Apr-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="21-Apr-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="28-Apr-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="05-May-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="12-May-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="19-May-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="02-Jun-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="09-Jun-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="16-Jun-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="23-Jun-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=             &lt;appointment date="30-Jun-2015" start="1315" end="1445" duration_min="90" room="H&amp;#xF6;rsaal 34" zip="1010" city=           &lt;courselanguage&gt;Englisch&lt;/courselanguage&gt;           &lt;lecturer givenname="Martina" name="Kaller" email="martina.kaller@univie.ac.at" role="V - Verantwortlicher Leiter           &lt;lecturer givenname="Berthold" name="Molden" email="berthold.molden@univie.ac.at" role="P - Mitanbieter"/&gt;         &lt;/group&gt;       &lt;/category&gt;     &lt;/course&gt;   &lt;/study&gt;   &lt;category name="Schwerpunkt-Einführung: Kurse" id="706_5" parentID="706_3"&gt;     &lt;course id="070281" name="Lekt&amp;#xF6;rskurs" title="Lekt&amp;#xF6;rskurs - Reading Group for Digital Humanities and Digital       &lt;group id="070281" name="Lekt&amp;#xF6;rskurs" title="Lekt&amp;#xF6;rskurs - Reading Group for Digital Humanities and Digital         &lt;appointment date="16-Mar-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1         &lt;appointment date="23-Mar-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1         &lt;appointment date="13-Apr-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1         &lt;appointment date="20-Apr-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1         &lt;appointment date="27-Apr-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1         &lt;appointment date="04-May-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1         &lt;appointment date="11-May-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1         &lt;appointment date="18-May-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1         &lt;appointment date="01-Jun-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1         &lt;appointment date="08-Jun-2015" start="1600" end="1830" duration_min="150" room="Seminarraum Geschichte 1" zip="1       &lt;courselanguage&gt;Deutsch&lt;/courselanguage&gt;       &lt;lecturer givenname="Mareike" name="K&amp;#xF6;nig" email="mareike.koenig@univie.ac.at" role="V - Verantwortlicher Le     &lt;/group&gt;   &lt;/category&gt; &lt;/studies&gt;</pre>
<p><b>Source Files</b></p>	<p><b>XML DOWNLOAD:</b>  <a href="https://raw.githubusercontent.com/fridlutz/UNIVIE/master/INTERROP/Session%202/0%20XML_Clean/baseXML_706_download.xml">https://raw.githubusercontent.com/fridlutz/UNIVIE/master/INTERROP/Session%202/0%20XML_Clean/baseXML_706_download.xml</a></p> <p><b>XML RESULT:</b>  <a href="https://raw.githubusercontent.com/fridlutz/UNIVIE/master/INTERROP/Session%202/0%20XML_Clean/baseXML_706_transformed.xml">https://raw.githubusercontent.com/fridlutz/UNIVIE/master/INTERROP/Session%202/0%20XML_Clean/baseXML_706_transformed.xml</a></p> <p><b>DEV ENVIRONMENT:</b>  <a href="https://github.com/fridlutz/UNIVIE/tree/master/INTERROP/Session%202/0%20XML_Clean">https://github.com/fridlutz/UNIVIE/tree/master/INTERROP/Session%202/0%20XML_Clean</a></p>