

SEMESTRÁLNÍ PRÁCE

předmět AD7B36XML

Ozdravné pobyty dětí

ČVUT FEL

obor STM - Softwarové inženýrství, kombinované studium
4. semestr

Zpracovala:

 Radoslava Jandová (jandora1)

V Praze dne 10. května 2011

Obsah:

Úvod	3 str.
1. Úkol - Základní pravidla tvorby XML dokumentů	3 str.
1.1. Zadání	
1.2. Zpracování - <i>Obr. 1: Well-Formed Document</i>	
2. Úkol - Kontrola validity XML dokumentu	5 str.
2.1. Zadání	
2.2. Zpracování - <i>Obr. 2: Validace XML</i>	
3. Úkol - Kontrola validity pomocí jazyka XML Schema	6 str.
3.1. Zadání	
3.2. Zpracování - <i>Obr. 3: Validace XML dokumentu pomocí nástroje NetBeans IDE</i> <i>Obr. 4: Validace XML dokumentu pomocí validátoru W3C</i> <i>Obr. 5: Warnings označené validací</i>	
4. Úkol - Dotaz v jazyce XPath	8 str.
4.1. Zadání	
4.2. Zpracování - <i>Obr. 6 - 15: Výsledky dotazů 1 - 10</i>	
5. Úkol - Transformace pomocí XSLT	12 str.
5.1. Zadání	
5.2. Zpracování - <i>Obr. 16: Výstup transformace do HTML</i> <i>Obr. 17: Výstup transformace do XHTML</i>	
6. Úkol - XQuery, případně DOM/SAX	14 str.
6.1. Zadání	
6.2. Zpracování - <i>Obr. 17 - 20: Výstupy dotazů XQuery</i>	
Závěr	17 str.
Zdroje	17 str.

Úvod

Téma referátu: Ozdravné pobyty dětí

Popis: Organizátor zajišťuje ozdravné pobyty pro děti. Tyto pobyty se pořádají v několika lokalitách. Pobyty jsou rozděleny do turnusů. Každý turnus se koná v určené lokalitě a má přesně daný termín. Turnusy jsou rozděleny do skupin. Každá skupina je určena unikátním číslem a je vedena jedním vedoucím. Vedoucí je určen svým jménem a příjmením. V každé skupině je maximálně 10 dětí. Do skupin jsou rozděleny děti podle věku a pohlaví. U každého dítěte je uvedeno jméno, příjmení, pohlaví a rok narození.

Použitý SW: NetBeans IDE, verze 6.9.1. – program byl využitý k tvorbě základních dokumentů referátu a k validaci.

1. Úkol - Základní pravidla tvorby XML dokumentů

Odevzdáno: 24. února 2011

Obsah ZIP souboru: referat.xml, wellFormed.jpg

1.1. Zadání: viz <http://ocw.cvut.cz/moodle/mod/assignment/view.php?id=8744>

Zde odevzdáte 1. kapitulu Vašeho referátu jako textový soubor.

Smyslem této kapitoly je procvičit základní pravidla tvorby dobře vytvořených (well-formed) XML dokumentů, testování správnosti (well-formedness), procvičení základních pojmů XML - element, atribut, character data, entita apod.

Do 1. kapitoly svého referátu umístíte řešení pro Vaše téma (které si zvolíte).

Pokyny (Šablona pro 1. kapitulu):

=====

1. Vytvořte XML dokument popisující základní objekty Vaší aplikace.
2. Validujte tento dokument pomocí některého validátoru (řádkové nástroje xmllint, xmlwf - obojí je v Solariu nainstalováno).
3. Umístěte element dokumentu z úkolu 1 do nějakého jmenného prostoru.
4. Rozšiřte Vaši definici o text pomocí značek HTML -> dokument mixující elementy ze dvou jmenných prostorů.

Příklad na jmenné prostory

=====

```
<root>
<a xmlns="urn:ns:nsA" />
<b:b xmlns:b="urn:ns:nsB" />
<c xmlns:d="urn:ns:nsD" d:d="attvalue" />
</root>
```

1.2. Zpracování:

krok 1. Vytvořte XML dokument popisující základní objekty Vaší aplikace.

V programu NetBeans IDE jsem vytvořila základní schéma dokumentu XML „referat.xml“ takto:

```
<pobyty>
  <turnus id="">                                = turnus pobytu (může jich být 1 - n)
    <lokalita></lokalita>                        = lokalita, kde se turnus koná
    <termin></termin>                            = termín konání turnusu

    <skupina id="">                               = skupiny v rámci turnusu (může jich být 1 - n)
      <cisloSkupiny>< cisloSkupiny/>              = unikátní číslo skupiny

      <vedouci>                                   = vedoucí skupiny, jeden vede jednu skupinu
        <jmenoVedouciho>< jmenoVedouciho/>
        <jmenoVedouciho>< jmenoVedouciho/>
      </vedouci/>

      <deti>                                       = děti zařazené do dané skupiny
```

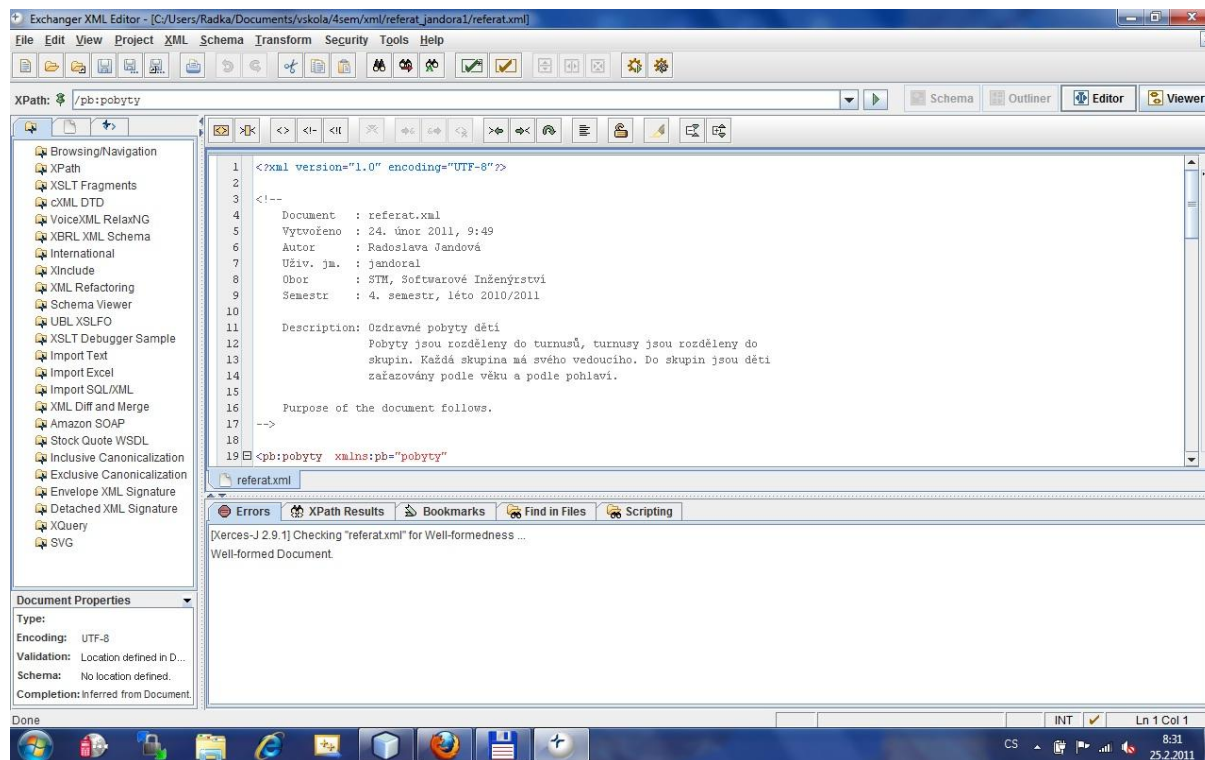
```

        <dite id="">      = identifikace dítěte (může jich být 0 - n)
        <jmenoD><jmenoD/>
        <prijmeniD><prijmeniD/>
        <pohlaviD><pohlaviD/>
        <rokNarozeni><rokNarozeni/>
    </dite/>
</deti/>
<skupina/>
</turnus/>
</pobyty>

```

krok 2. Validujte tento dokument pomocí některého validátoru

Dokument XML jsem validovala v programu Exchanger XML Editor, verze 3.3.1.



Obr. 1: Well-Formed Document

krok 3. Umístěte element dokumentu z úkolu 1 do nějakého jmenného prostoru.

Do dokumentu byl přidán následující jmenný prostor:

```
<pb:pobyty xmlns:pb="pobyty"> ... </pobyty>
```

krok 4. Rozšiřte Vaši definici o text pomocí značek HTML -> dokument mixující elementy ze dvou jmenných prostorů.

Do dokumentu XML byla vložena následující značka HTML:

```

<pb:pobyty xmlns:pb="pobyty"
  xmlns:xhtml="http://www.w3.org/1999/xhtml"> ... </pobyty>

```

2. Úkol - Kontrola validity XML dokumentu

Odevzdáno: 1. března 2011

Obsah ZIP souboru: referat.pdf, referat.xml, referat.dtd, validaceXML.jpg

2.1. Zadání: viz <http://ocw.cvut.cz/moodle/mod/assignment/view.php?id=8746>

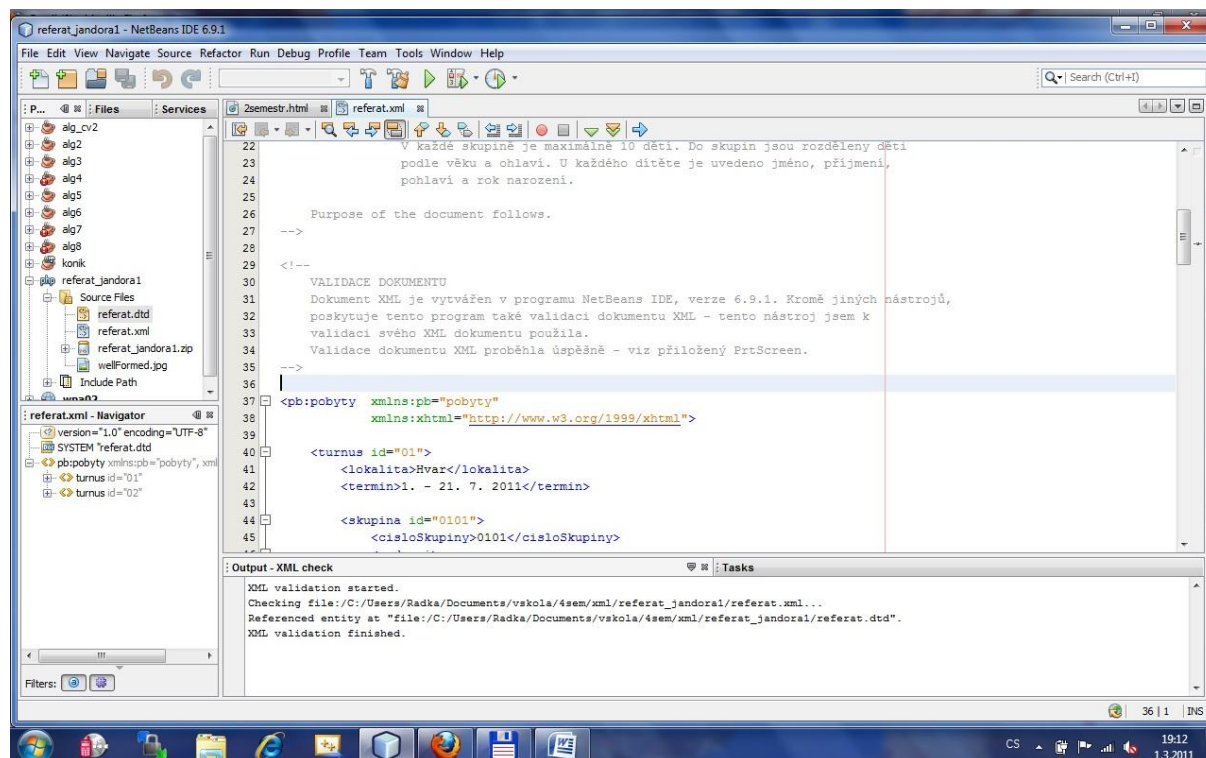
Vytvořte 2. kapitulu Vašeho referátu dokumentující definici DTD pro dokumenty z Vašeho tématu. Vytvořte definici DTD pro Vaše téma, dále vytvořte příklad dokumentu z Vašeho tématu. Zkontrolujte pomocí validátoru, zda je XML dokument validní vzhledem k DTD. Opravujte XML dokument, dokud není validní. Do 2.kapitoly vložte XML document, DTD a popis způsobu, jak jste jej validovali. Pokud si chcete téma podrobněji vyzkoušet, prostudujte si doplňující materiály.

2.2. Zpracování:

Pro dokument „referat.xml“ jsem vytvořila DTD – soubor s názvem „referat.dtd“ a provedla jsem propojení se souborem „referat.xml“:

```
<!DOCTYPE pb:pobyty SYSTEM "referat.dtd">
```

Následně jsem soubor „referat.xml“ validovala, a to pomocí nástroje „Validate XML“, který poskytuje přímo pro program NetBeans IDE. Na základě výsledku validace jsem opravila chyby, dokud validace dokumentu nebyla úspěšná.



Obr. 2: Validace XML dokumentu

3. Úkol - Kontrola validity pomocí jazyka XML Schema

Odevzdáno: 11. 3. 2011

Obsah ZIP souboru: referat.pdf, referat.xml, referat.xsd, uk3validaceNb.jpg, uk3validaceW3C.jpg, warnings.jpg

3.1. Zadání: viz <http://ocw.cvut.cz/moodle/mod/assignment/view.php?id=8748>

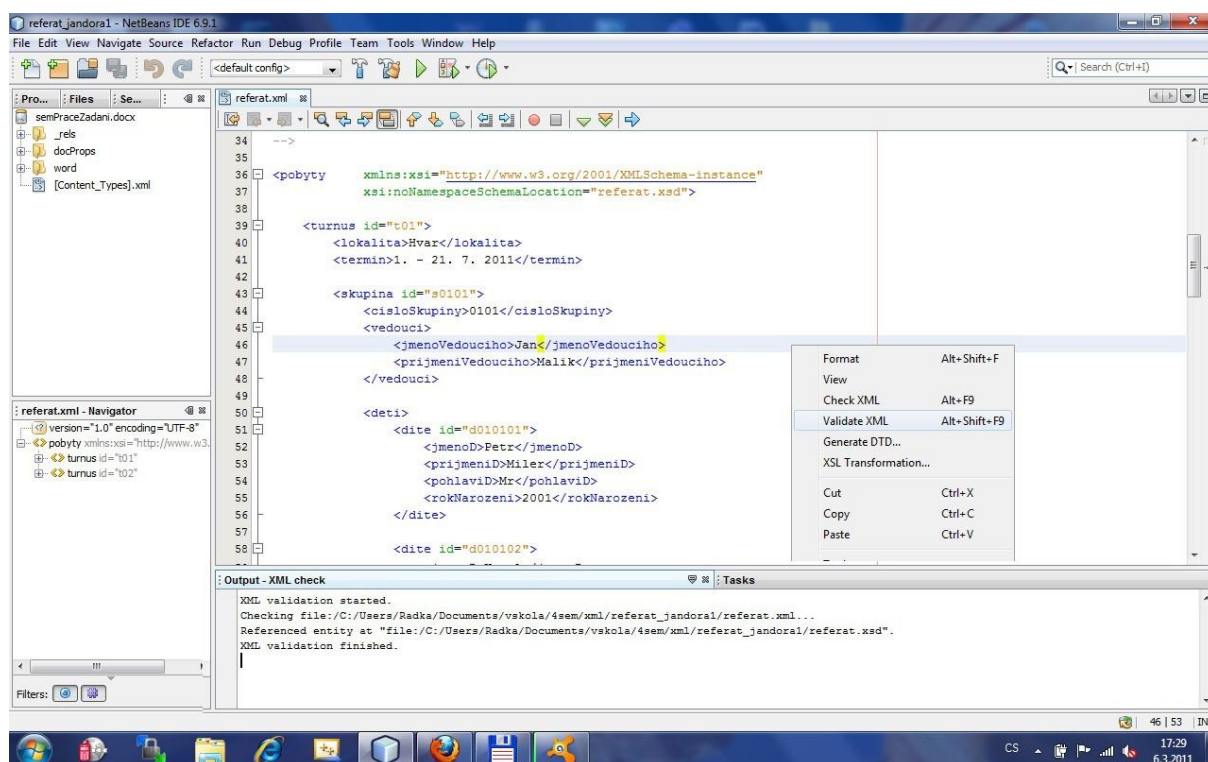
Vytvořte XML schéma v jazyce XML Schema pro XML dokumenty se strukturou danou vaším XML schématem a zkuste ho validovat. Opravujte ho, dokud není validní.

Pozn.: Využijte validátor <http://www.w3.org/2001/03/webdata/xsv> k testu, zda Vámi vytvořené XML schéma je validní XML Schema.

3.2. Zpracování:

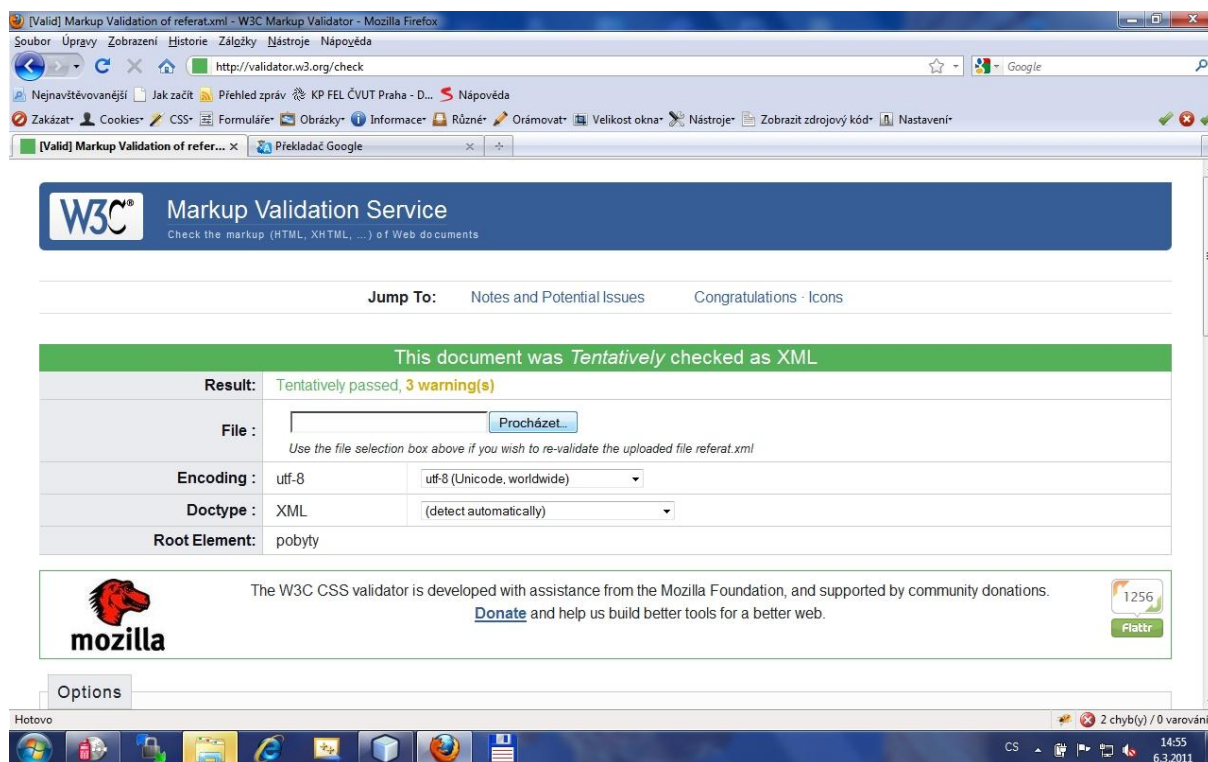
K semestrální práci jsem přidala další dokument - „referat.xsd“, ve kterém jsem vytvořila XML schéma odpovídající původně vytvořenému XML dokumentu (viz Úkol 1.).

K první validaci jsem použila validátor v rámci programu NetBeans IDE. První validace odhalila chybné propojení dokumentů „referat.xml“ a „referat.xsd“. Po opravě proběhla validace úspěšně.



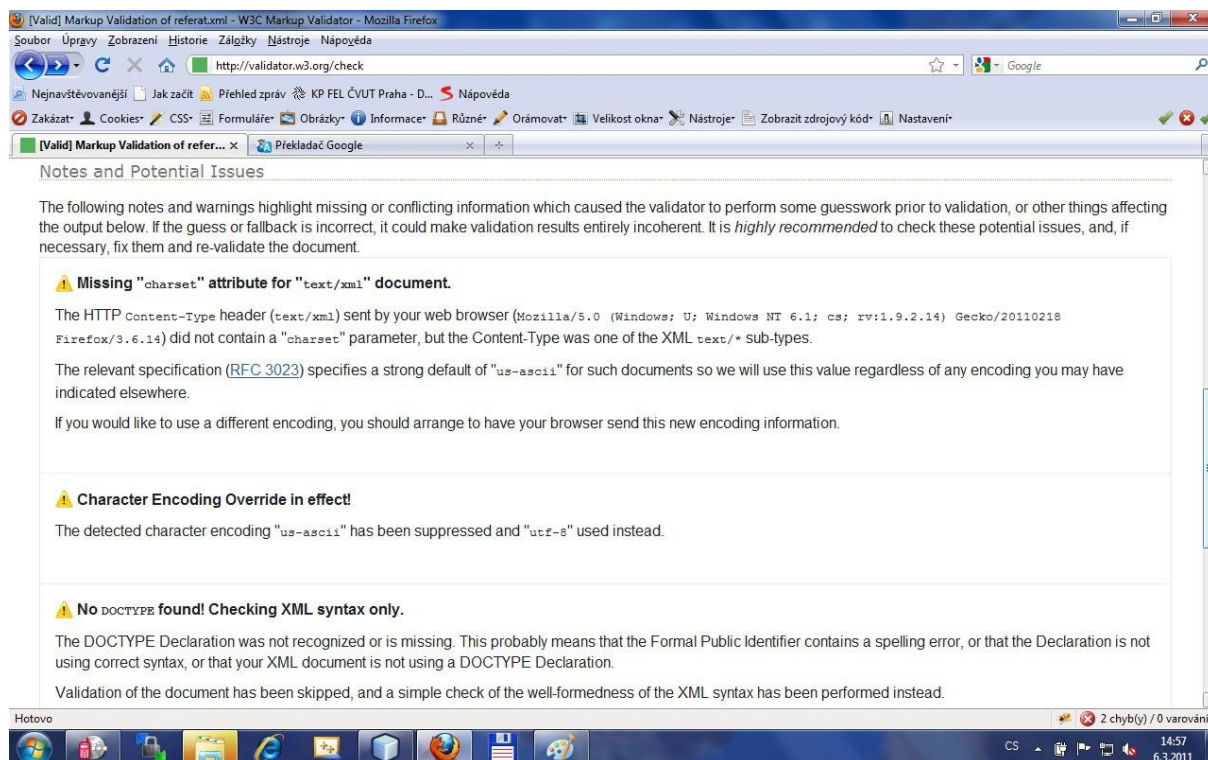
Obr. 3: Validace XML pomocí nástroje NetBeans IDE

Následně jsem dokument validovala ještě pomocí validátoru W3C, který předepisuje zadání úkolu č. 3. Validace proběhla s výsledkem „Tentatively passed“ – viz PrintScreen.



Obr. 4: Validace XML dokumentu pomocí validátoru W3C

Validace označila 3 „warnings“, které jsem ale již nedokázala odstranit.



Obr. 5: „Warnings“ označené validací.

4. Úkol - Dotaz v jazyce XPath

Odevzdáno: 10. 4. 2011

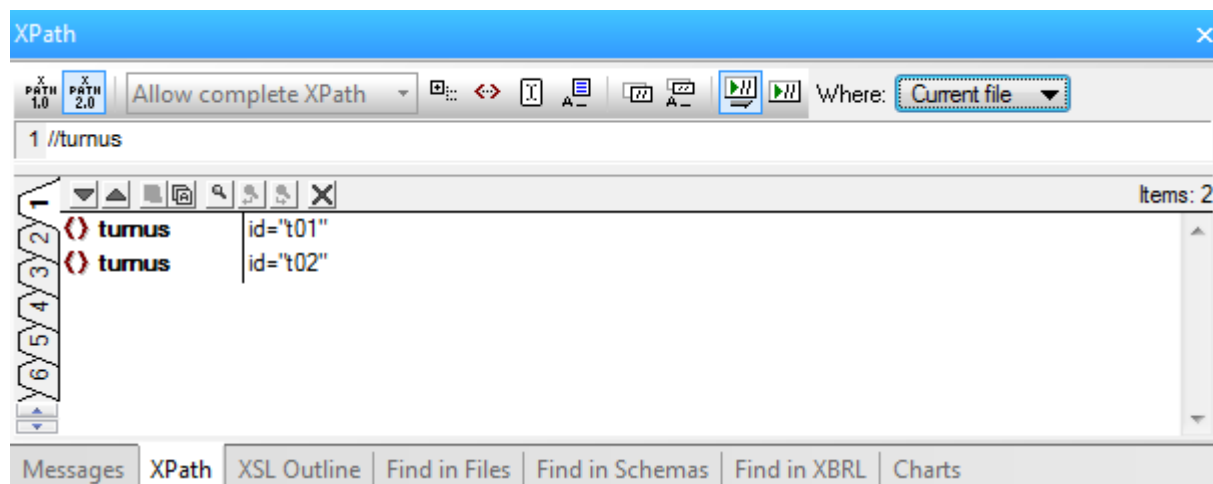
Obsah ZIP souboru: referat.pdf, referat.xml, xxx

4.1. Zadání: viz <http://ocw.cvut.cz/moodle/mod/assignment/view.php?id=8749>

Vytvořte dotazy v jazyce XPath, které se týkají obsahu Vašeho dokumentu. Demonstrujte správný výsledek pomocí nějakého nástroje.

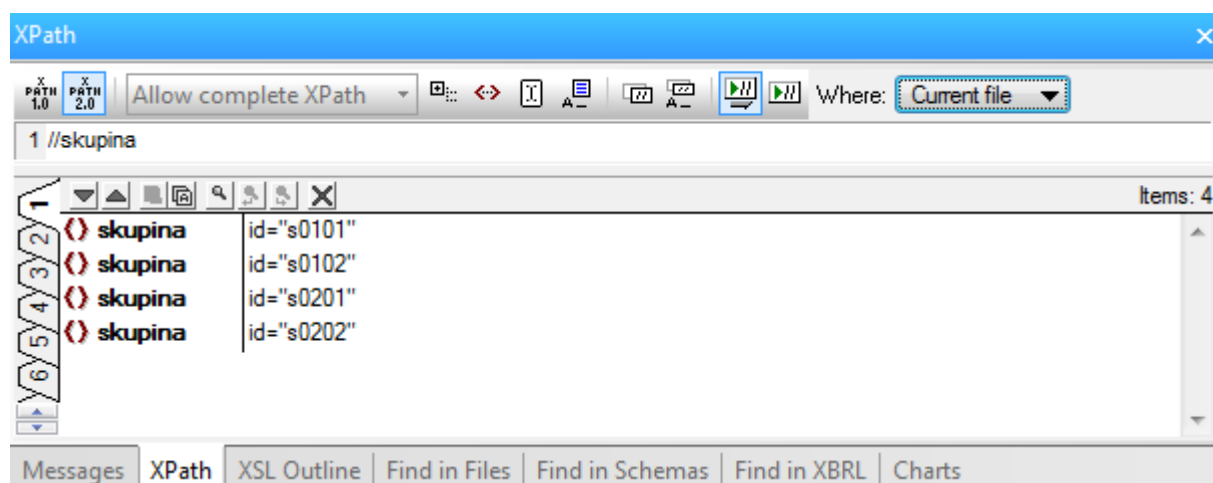
4.2. Zpracování:

Dotaz 1: Vypiš všechny turnusy.
 /pobyty/turnus
 nebo
 //turnus



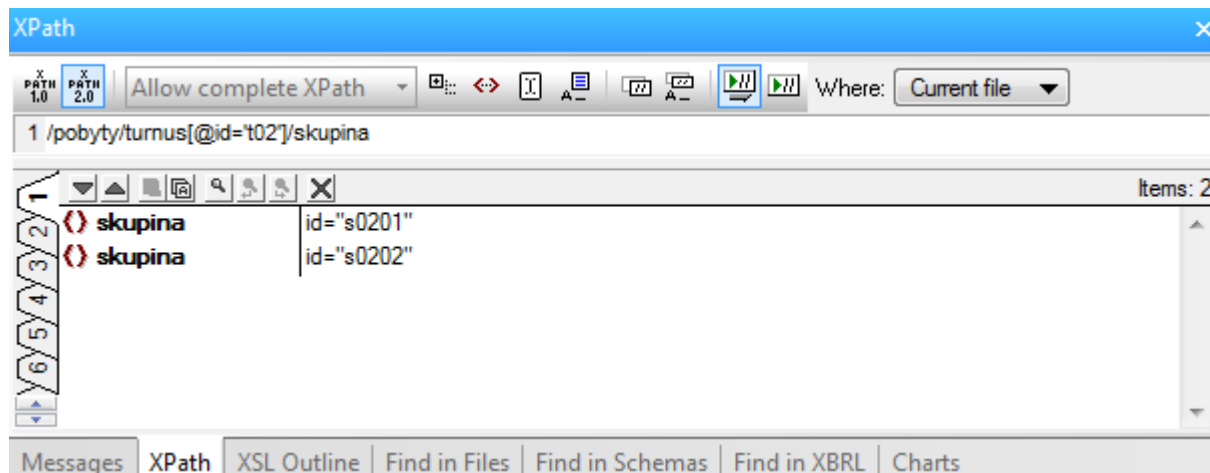
Obr. 6: Výsledek dotazu 1

Dotaz 2: Vypiš všechny skupiny bez ohledu na turnus.
 /pobyty//skupina
 nebo
 //skupina



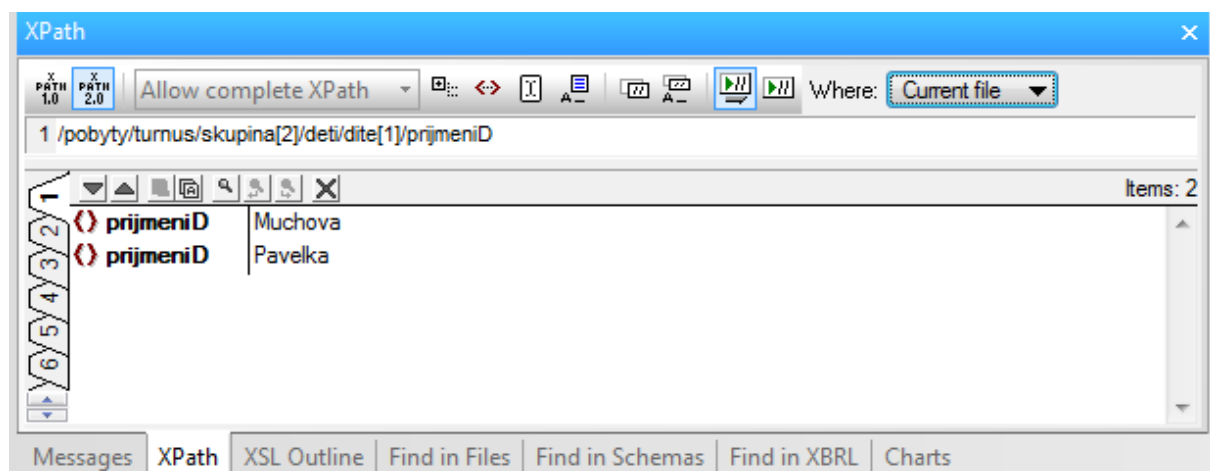
Obr. 7: Výsledek dotazu 2

Dotaz 3: Vypiš všechny skupiny z turnusu t02.
/pobyty/turnus[@id='t02']/skupina
nebo
//turnus[@id='t02']/skupina



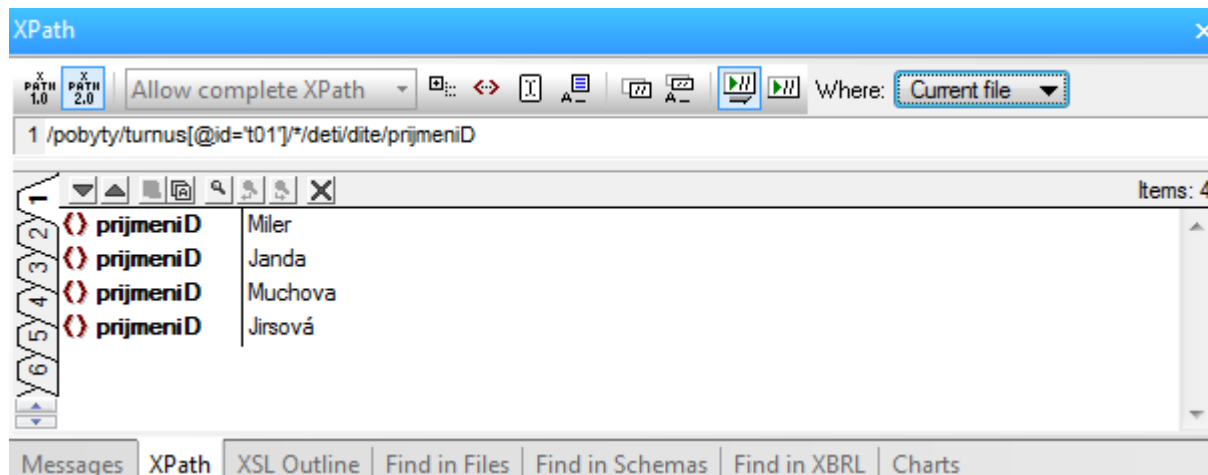
Obr. 8: Výsledek dotazu 3

Dotaz 4: Vypiš z pobytů příjmení všech dětí, které jsou zařazeny do 2 skupiny a mají pořadové číslo 1.
/pobyty/turnus/skupina[2]/deti/dite[1]/prijmeniD
nebo
// skupina[2]//dite[1]/prijmeniD



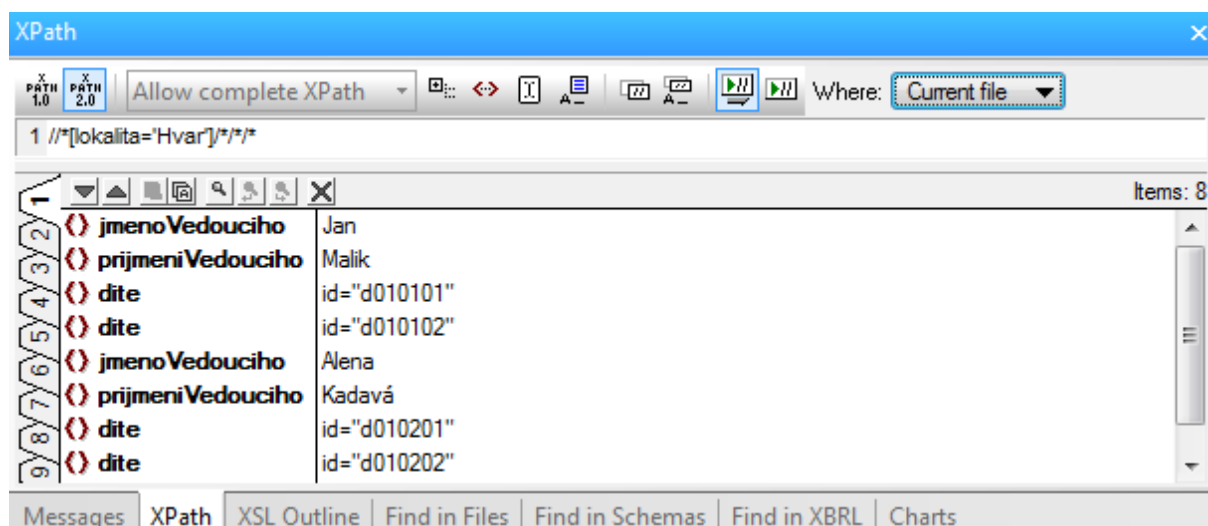
Obr. 9: Výsledek dotazu 4

Dotaz 5: Vypiš příjmení všech dětí z turnusu t01.
 /pobyty/turnus[@id='t01']/*/dite/prijmeniD
 nebo
 //turnus[@id='t01']//prijmeniD



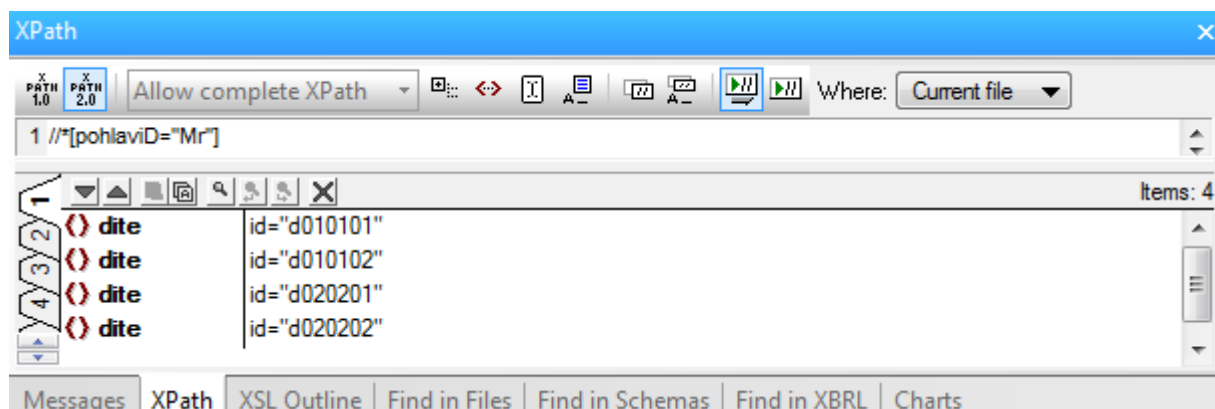
Obr. 10: Výsledek dotazu 5

Dotaz 6: Vypiš jména a příjmení všech vedoucích z lokality Hvar a id dětí v jejich skupinách.
 /*[lokalita='Hvar']/*/*/*



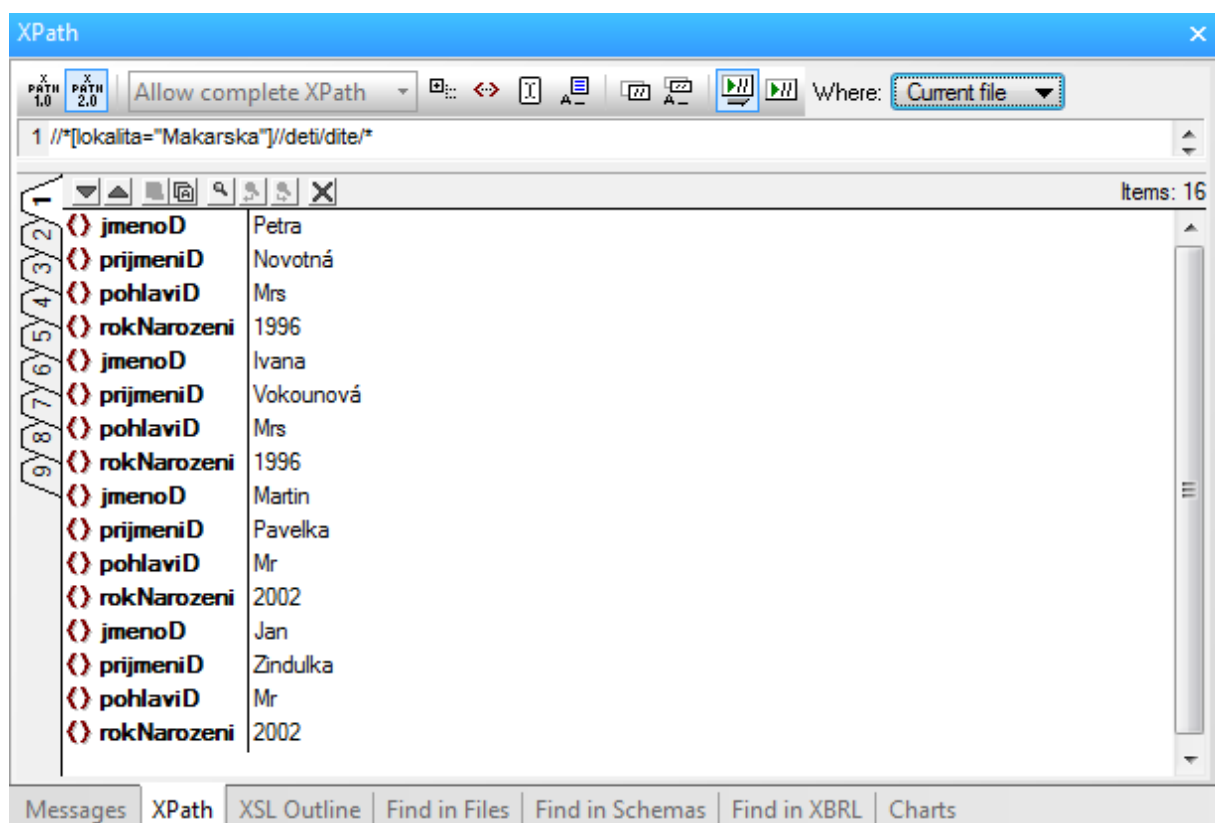
Obr. 11: Výsledek dotazu 6

Dotaz 7: Vypiš všechny chlapce.
 /*[pohlaviD="Mr"]
 nebo
 /*[pohlaviD!="Mrs"]



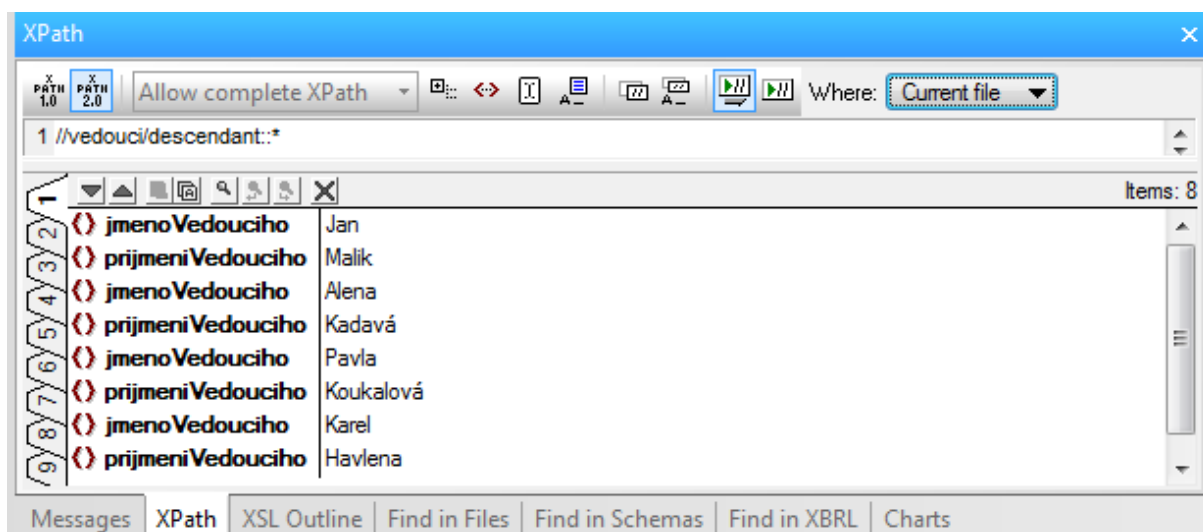
Obr. 12: Výsledek dotazu 7

Dotaz 8: Vypiš vše o dětech z lokality Makarska.
 /**[lokalita="Makarska"]//deti/dite/*



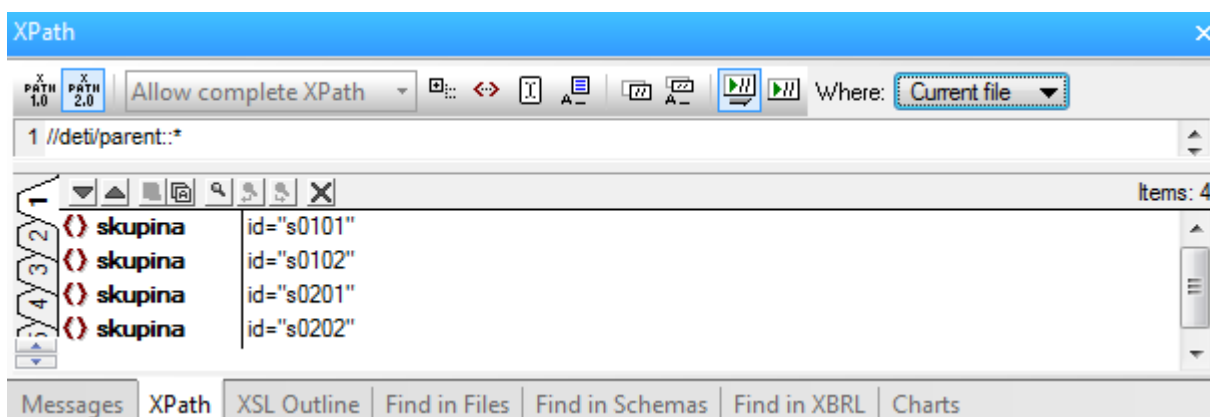
Obr. 13: Výsledek dotazu 8

Dotaz 9: Vypiš všechny vedoucí.
 //vedouci/descendant::*



Obr. 14: Výsledek dotazu 9

Dotaz 10: Vypiš všechny „rodiče“ elementu děti.
 //deti/parent::*



Obr. 15: Výsledek dotazu 10

5. Úkol – Transformace pomocí XSLT

Odevzdáno: do 28. 4. 2011

Obsah ZIP souboru: referat_jandora1.pdf
 referatHTML.xml + referatHTML.xls + referatHTML.png,
 referatXHTML.xml + referatXHTML.xls + referatXHTML.png
 referatXML.xml

5.1. Zadání: viz <http://ocw.cvut.cz/moodle/mod/assignment/view.php?id=8750>

Vytvořte a vyzkoušejte nejméně 3 transformace Vašeho dokumentu do XHTML, HTML a jiného XML. Výsledky a jejich dokumentaci vložte do kapitoly č. 5.

5.2. Zpracování:

Část A – transformace do HTML

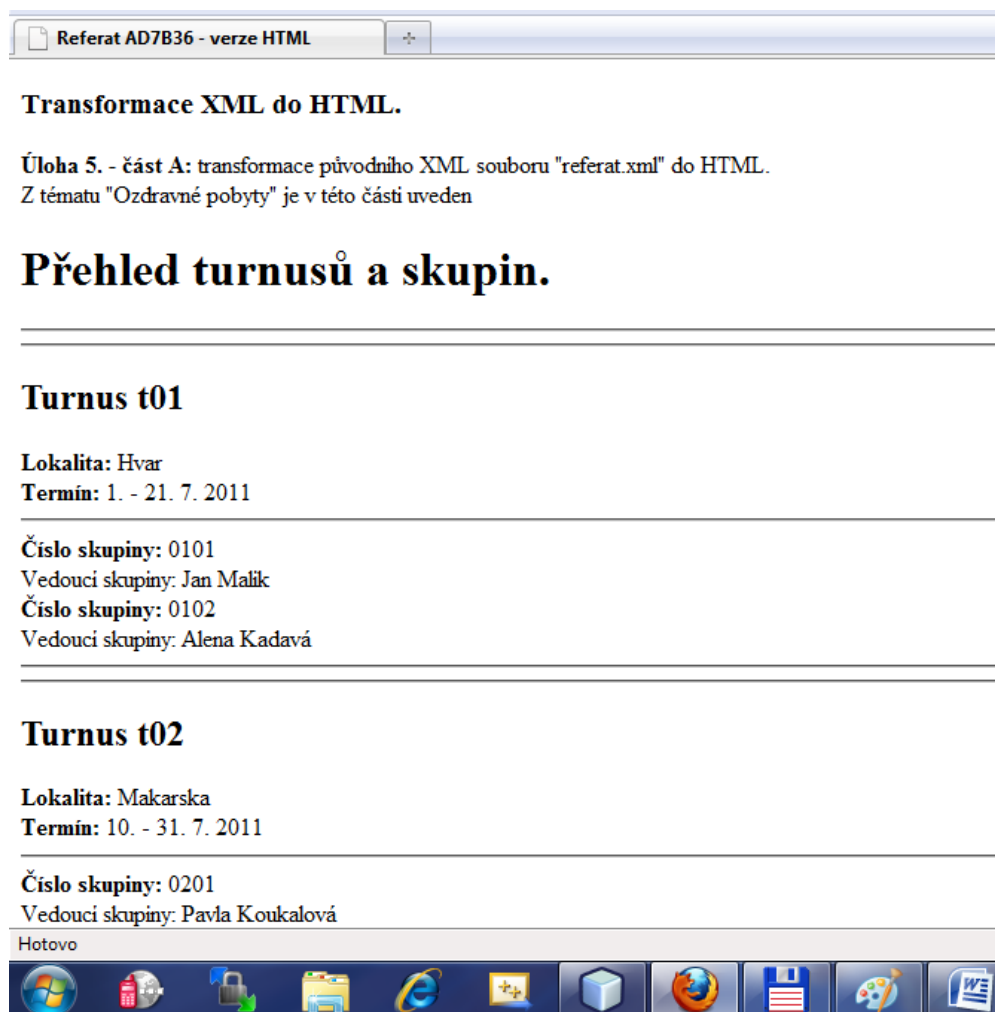
Transformace do HTML je uvedena v souboru „referatHTML.xsl“ a PrintScreen výstupu v „referatHTML.png“.

V této části jsem uvedla přehled turnusů (včetně lokality a termínu konání) a skupin v těchto turnusech (včetně čísla skupiny a vedoucího).

Výstupní kód pro HTML transformaci:

```
<xsl:output method="html" encoding="utf-8" indent="yes" doctype-public="-//W3C//DTD HTML 4.01//EN"/>
```

PrintScreen výstupu:



Obr. 16: Výstup transformace do HTML

Část B – transformace do XHTML

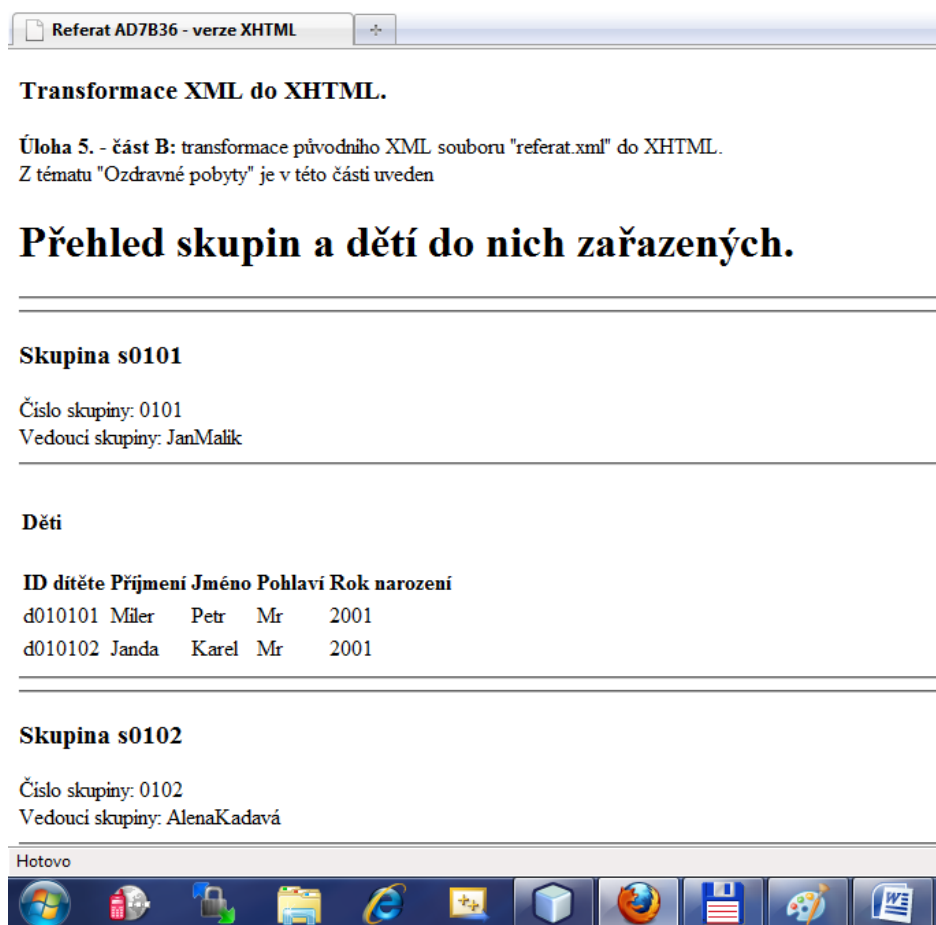
Transformace do XHTML je uvedena v souboru „referatXHTML.xsl“ a PrintScreen výstupu v „referatXHTML.png“.

V této části jsem uvedla přehled skupin (včetně čísla skupiny a vedoucího) a dětí do nich zařazených (včetně všech údajů, které jsou o nich uvedeny).

Výstupní kód pro XHTML transformaci:

```
<xsl:output method="html" encoding="utf-8" indent="yes" doctype-public="-//W3C//DTD HTML 4.01//EN"/>
```

PrintScreen výstupu:



Obr. 17: Výstup transformace do XHTML

Část C – transformace do jiného XML

Transformace do jiného XML je uvedena v souboru „referatXML.xsl“.
V této části jsem transformovala původní XML na seznam účastníků pobytů.

6. Úkol - XQuery, případně DOM/SAX

Odevzdáno: do 12. 5. 2011

Obsah ZIP souboru: referat_jandora1.pdf, referat.xml,
dotazy = q01.query, q02.query, q03.query,
výstupy = out01.xml, out02.xml, out03.xml,

6.1. Zadání: viz <http://ocw.cvut.cz/moodle/mod/assignment/view.php?id=8752>

Podklady : [SAX - ContentHandler](#), [DOM - Document](#)

Podmínky pro splnění DÚ:

- Vyzkoušejte si alespoň 3 dotazy v jazyce XQuery, které se týkají Vaší problematiky a Vašich dokumentů. V řešení uveďte zdrojové XML dokumenty, dotazy v XQuery a protokol o výsledcích dotazů získaných zvoleným nástrojem.

Pokud chcete získat body navíc (až 15 bodů), můžete vyřešit následující úkoly:

- Prostřednictvím rozhraní DOM nebo SAX implementovat vyhodnocení jednoho XQuery dotazu nad daty z předchozích úloh.
- Prostřednictvím rozhraní SAX spočítat zvolenou charakteristiku XML dokumentu (např. hloubku, počet elementů, průměrnou délku názvů elementů, počet atributů apod.) nebo XSD (např. počet složených datových typů, počet elementů se smíšeným obsahem, počet lokálně a globálně definovaných jednoduchých datových typů apod.). Každý si vymyslí vlastní téma.

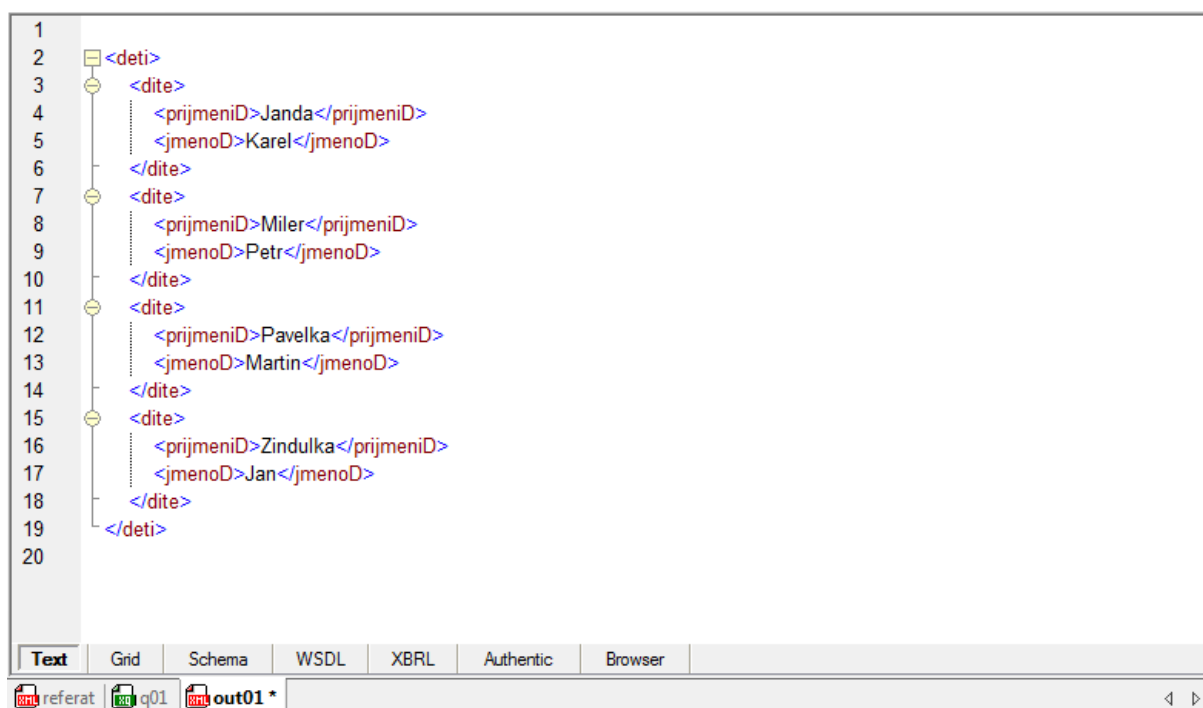
- Prostřednictvím rozhraní DOM provést vybranou transformaci XML dokumentu nebo XSD, která změní jeho strukturu (např. obrátit pořadí elementů, převést atributy na elementy, převést elementy s jednoduchým datovým typem na atributy, převést všechny lokálně definované jednoduché/složené datové typy na globálně definované nebo obráceně apod.) Každý si vymyslí vlastní téma.

6.2. Zpracování:

Dotaz 1: Vypiš všechny chlapce zařazené do pobytů a seřaď je podle příjmení.

```
<deti>
{
for   $dite in (fn:doc("referat.xml")//pobyty/turnus/skupina/deti/dite)
where $dite/pohlaviD = "Mr"
order by $dite/prijmeniD

return
<dite>
    { $dite/prijmeniD }
    { $dite/jmenoD }
</dite>
}
</deti>
```



Obr. 18: Výstup 1. dotazu XQuery.

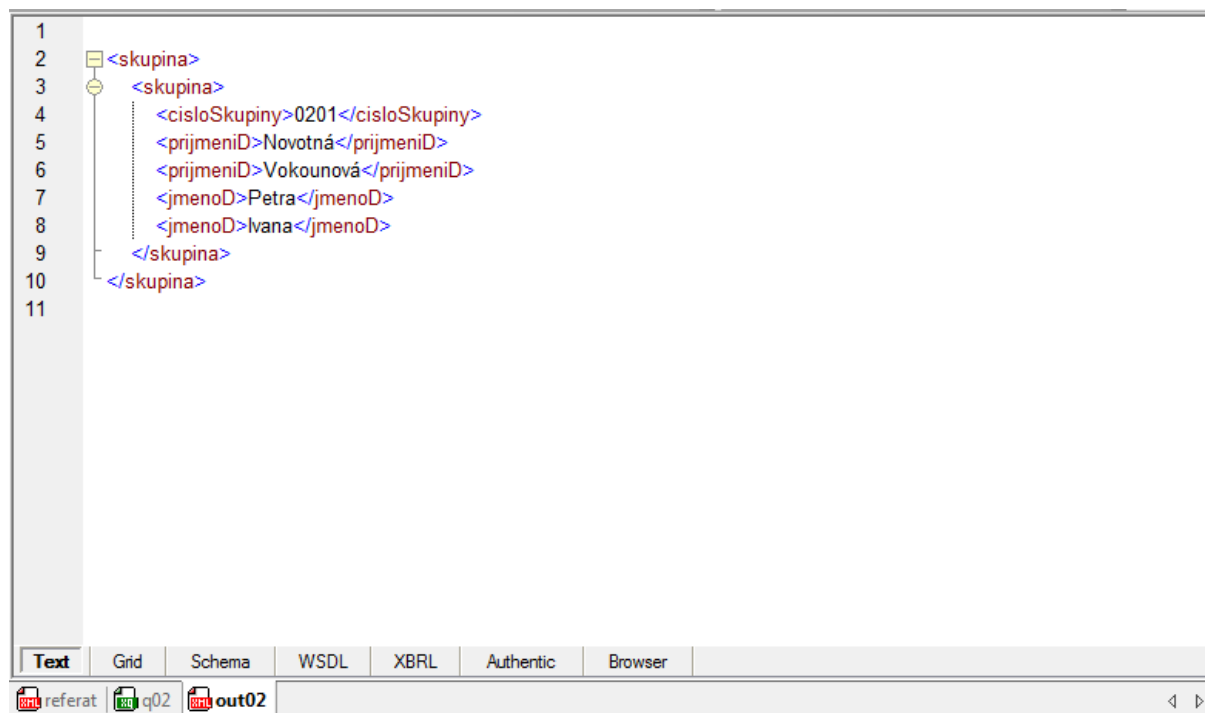
Dotaz 2: Vypiš číslo skupiny a děti zařazené do skupny, kde je vedoucí Koukalová.

```
<skupina>
{
for   $skupina in (fn:doc("referat.xml")//pobyty/turnus/skupina)
where $skupina/vedouci/prijmeniVedouciho = "Koukalová"
return
    <skupina>
        { $skupina/cisloSkupiny }
        { $skupina/deti/dite/prijmeniD }
        { $skupina/deti/dite/jmenoD }
```

```

    </skupina>
  }
</skupina>

```



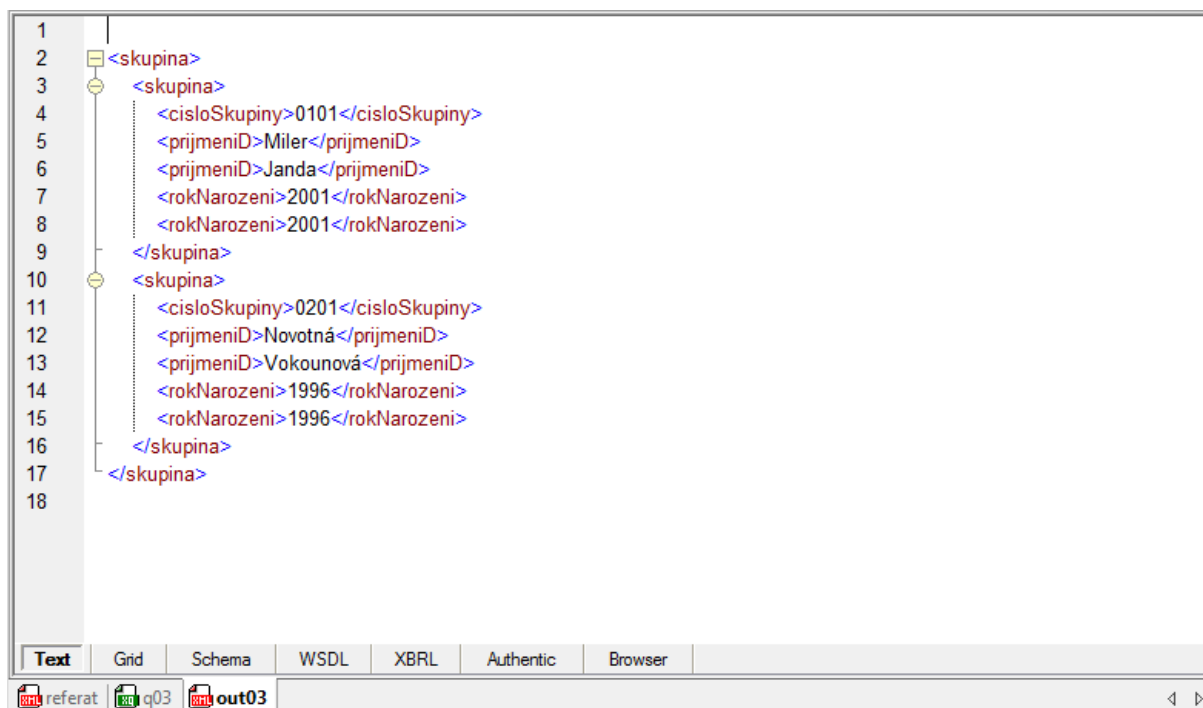
Obr. 19: Výstup 2. dotazu XQuery.

Dotaz 3: Vypiš číslo skupiny, příjmení a jméno dětí starších než děti narozené v roce 2002.

```

<skupina>
{
for   $skupina in (fn:doc("referat.xml")//pobyty/turnus/skupina)
where $skupina/deti/dite/rokNarozeni < "2002"
order by $skupina/deti/dite/prijmeniD
return
    <skupina>
        {
            {$skupina/cisloSkupiny}
            {$skupina/deti/dite/prijmeniD}
            {$skupina/deti/dite/rokNarozeni}
        }
    </skupina>
}
</skupina>

```



Obr. 20: Výstup 3. dotazu XQuery.

Závěr

V rámci této semestrální práce jsem měla možnost prakticky si vyzkoušet vše, co bylo obsahem jednotlivých přednášek k předmětu AD7B36XML. Vzhledem k tomu, že ve své praxi nemám možnost se s tvorbou dokumentů XML setkat, velmi jsem tuto možnost uvítala.

Zdroje

- [1] *Přednášky k předmětu AD7B36XML*. Dostupné z WWW: <<http://ocw.cvut.cz/moodle>>.
- [2] MLÝNKOVÁ, I, NEČASNÝ, Michal, POKORNÝ, RICHTA, Karel, KAMIL, VOJTĚCH. *Technologie XML, Principy a aplikace v praxi*, ISBN: 978-80-247-2725-7, vydáno 2008.
- [3] SW: *Exchanger XML Editor*. Verze 3.3.1. Dostupné z WWW: <<http://www.freexmleditor.com>>.
- [4] KOSEK, Jiří. *Profesionální školení a poradenství v oblasti technologií XML, RELAX NG*. Video z cyklu Programátorské večery, rok 2005. Dostupné z WWW: <<http://www.avc-cvut.cz/avc.php?id=2309>> [cit. 2011-03-04].
- [5] SW: *Validátor*. Dostupné z WWW: <<http://www.w3.org/2001/03/webdata/xsv>> [cit. 2011-03-04].
- [6] SW: *Altova XML Spy*. Dostupné z WWW: <<http://ocw.cvut.cz/moodle>>.