



Софийски университет „Св. Кл. Охридски“

Факултет по математика и информатика

*Бакалавърска програма
„Софтуерно инженерство“*



Предмет: XML технологии за семантичен Уеб

Зимен семестър, 2019/2020 год.

Тема №07: „Каталог на университети“

Курсов проект

Автори:

Даниел Димитров, фак. номер 62121

Иван Чучулски, фак. номер 62167

януари, 2020

София

Съдържание

| | | |
|-----|--|----|
| 1 | Въведение | 3 |
| 2 | Анализ на решението | 3 |
| 2.1 | Работен процес | 3 |
| 2.2 | Структура на съдържанието | 3 |
| 2.3 | Тип и представяне на съдържанието | 6 |
| 3 | Дизайн | 6 |
| 4 | Тестване | 8 |
| 5 | Заклучение и възможно бъдещо развитие | 11 |
| 6 | Разпределение на работата | 11 |
| 7 | Използвани литературни източници и Уеб сайтове | 11 |

1 Въведение

Проектът има за цел да се създаде каталог на университетите в България, който съдържа основна информация за всеки университет, оценки за професионалните направления, по които се извършва обучение в него и също така да бъде включено графично съдържание. Разгледани са дванадесет университети, намиращи се в различни региони на България, както и от двата основни тип финансиране – държавно и частно. Този каталог може да послужи на кандидат-студентите да получат нагледна представа за местата, където могат да кандидатстват и така те да направят своя информиран избор.

Каталогът е реализиран чрез XML документ, който описва университетите и географските региони в България. Връзката между регион и университет са описани чрез атрибути ID/IDREF. Графичното съдържание на каталога е представено чрез XML единици и съответните нотации за формата на единиците. XML документът е валидиран от DTD документ, който описва структурата на XML документа и декларира нотациите и единиците. Самото XML съдържание е представено в PDF документ, който е генериран с помощта на XML документ описващ XSL трансформациите.

2 Анализ на решението

2.1 Работен процес

Съдържанието на каталога е представено с XML документ, където е описана основната информация за университетите, посочени са техните региони и е указана връзка към съответното графично съдържание. След завършването на цялостния XML документ неговото съдържание бива валидирано от DTD документ. След успешното валидиране следва генерирането на PDF документа, което става с изготвения за това XSLT документ и чрез XSL-FO процесора и форматиращия процесор Apache FOP, които са част от средата Altova XMLSPY 2020.

2.2 Структура на съдържанието

Описание на XML документа, представящ информацията на университетите :

1. universitiesInBulgaria – кореновия елемент на XML документа. В DTD документа е указано, че той има два преки наследника : **universities** и **regions** и те се срещат задължително точно веднъж в документа.

1.1. universities - елемент, който съдържа информацията на университетите. В DTD документа е указано, че той съдържа един или повече елементи **university**.

1.1.1. university – елемент, който съдържа информацията на даден университет, разделена в няколко основни категории. В DTD документа е указано, че неговите преки наследници, които се срещат точно веднъж са **univDetails**, **univContacts**, **programsAccreditation** и **listOfFaculties**, а такива, които се срещат опционално са **listOfBranches** и

listOfDepartments. Елементът има и атрибут **univId**, който е от тип ID и е задължителен. Този атрибут оказва уникален идентификационен номер на всеки университет.

1.1.1.1. univDetails – елемент, който съдържа най-основните характеристики на университета. Негови преки наследници, който се срещат точно веднъж са **name**, **yearFounded**, **univImage**, **rectorName**, **numberOfStudents**, а такива, които се срещат или веднъж или не се срещат са **numberOfBachPrograms**, **numberOfMstPrograms**, **numberOfPhDPrograms**.

1.1.1.1.1. name – елемент, който съдържа името на университета. В DTD документа е указано, че той е от тип PCDATA.

1.1.1.1.2. yearFounded – елемент, който представлява годината на основаване на университета. В DTD документа е указано, че той е от тип PCDATA.

1.1.1.1.3. univImage - елемент, в който е поместено графичното съдържание на даден университет. В DTD документа е указано, че той има един пряк наследник, който се среща точно веднъж **src**.

1.1.1.1.3.1. src - елемент, който представлява графичното съдържание на даден университет. В DTD документа е указано, че елементът е празен и има един задължителен атрибут **href**, който е от тип ENTITY, което има зададена нотация jpg. Този атрибут указва името на файла със снимката.

1.1.1.1.4. rectorName - елемент, който съдържа името на настоящия ректор на университета. В DTD документа е указано, че той е от тип PCDATA.

1.1.1.1.5. numberOfStudents - елемент, който указва броя на учещите студенти в университета. В DTD документа е указано, че той е от тип PCDATA.

1.1.1.1.6. numberOfBachPrograms - елемент, който указва броя на предлаганите бакалавърски програми в университета. В DTD документа е указано, че той е от тип PCDATA.

1.1.1.1.7. numberOfMstPrograms - елемент, който указва броя на предлаганите магистърски програми в университета. Според DTD документа той е от тип PCDATA.

1.1.1.1.8. numberOfPhDPrograms - елемент, който указва броя на предлаганите докторантски програми в университета. Според DTD документа е той е от тип PCDATA.

1.1.1.2. univContacts - елемент, който съдържа информация за контакти и начин за свързване на университета. Негови преки наследници, който се срещат точно веднъж са **location**, **postalCode**, **phone**, **webSite**, **email**.

1.1.1.2.1. location - елемент, който описва местоположението на университета. В DTD документа е указано, че той има двама преки наследници, които се срещат точно веднъж **city**, **street**. Също така той има един задължителен атрибут **regionIDRef** от

тип IDREF, чиято стойност е уникален идентификатор на региона, в който се намира университета.

1.1.1.2.1.1. city - елемент, който описва града от адреса на университета. В DTD документа е указано, че елементът е от тип PCDATA.

1.1.1.2.1.2. street - елемент, който описва улицата от адреса на университета. В DTD документа е указано, че елементът е от тип PCDATA.

1.1.1.2.2. postalCode - елемент, който описва пощенския код на университета. В DTD документа е указано, че елементът е от тип PCDATA.

1.1.1.2.3. phone - елемент, който описва телефон за контакти на университета. В DTD документа е указано, че елементът е от тип PCDATA.

1.1.1.2.4. webSite - елемент, който указва адреса на уеб сайта на университета. В DTD документа е указано, че елементът е от тип PCDATA и има един атрибут *href*, който е задължителен и от тип CDATA, като неговата стойност е адреса на уеб сайта.

1.1.1.2.5. email - елемент, който указва официалния имейл на университета. В DTD документа е указано, че елементът е от тип PCDATA и има един атрибут *href*, който е задължителен и от тип CDATA, който също описва имейла.

1.1.1.3. programsAccreditation - елемент, който съдържа информация за направленията, по които се извършва обучение и оценки от рейтинговата система на висшите училища на МОН за даден университет. Негов пряк наследник, който се среща един или повече пъти е **program**.

1.1.1.3.1. program - елемент, който дава информация за конкретно професионално направление на университета. В DTD документа е указано, че този елемент има преки наследници, които се срещат точно веднъж **programName**, **accreditationScore**.

1.1.1.3.1.1. programName - елемент, който съдържа името на професионалното направление от университета. В DTD документа е указано, че елементът е от тип PCDATA.

1.1.1.3.1.2. accreditationScore - елемент, който съдържа оценка на професионалното направление според рейтинговата система на висшите училища на МОН. В DTD документа е указано, че елементът е от тип PCDATA.

1.1.1.4. listOfFaculties - елемент, който дава информация за факултетите на университета. Негов пряк наследник, който се среща един или повече пъти е **faculty**.

1.1.1.4.1. faculty - елемент, който името на факултет на университета. В DTD документа е указано, че елементът е от тип PCDATA.

1.1.1.5. listOfBranches - елемент, който съдържа информация за филиалите на университета. Негов пряк наследник, който се среща един или повече пъти е **branch**.

1.1.1.5.1. branch - елемент, който името на филиал на университета. В DTD документа е указано, че елементът е от тип PCDATA

1.1.1.6. listOfDepartments - елемент, който съдържа информация за департаментите на университета. Негов пряк наследник, който се среща един или повече пъти е **department**.

1.1.1.6.1. department - елемент, който името на департамент на университета. В DTD документа е указано, че елементът е от тип PCDATA

1.2. regions – елемент, който съдържа описанието на регионите. В DTD документа е указано, че в него се съдържат един или повече елементи **region**.

1.2.1. region – този елемент представя информацията за даден географски регион. В DTD документа е указано, че той има два преки наследника, които трябва да се срещат точно веднъж **regionName**, **regionDetails**.

1.2.1.1. regionName – елемент, който съдържа името на региона

1.2.1.2. regionDetails – празен елемент, който има един атрибут **regionID**, който е от тип ID и е задължителен. Този атрибут е уникален идентификационен номер на региона.

2.3 Тип и представяне на съдържанието

Тук опишете типа на съдържанието - текстово, графично и евентуално аудио/видео съдържание), което ще представите в проекта си. Укажете колко текстови и мултимедийни ресурса ще използвате (напр. 4 текстови документа, 12 графични изображения, 2 видео-материала), както и техните източници. Опишете размера и типа на файловете и начина на кодиране за всеки от ресурсите.

Информацията за дванадесетте университета е представена чрез текст, списъци и графичното съдържание чрез PDF документ. За всеки университет са описани неговите основни характеристики, представена е структура му и е дадена информация за професионалните направления, по които той извършва обучение. Графичното съдържание се състои от дванадесет на брой снимки в JPG формат, които са поместени в папка university-pictures.

3 Дизайн

За създаването на каталога са използвани технологиите XML version 1.0, DTD, XSLT, XPath, Altova XMLSPY 2020, в която са използвани XSL-FO, Apache-FOP. Текстовото съдържание е структурирано оформено в XML документ. В този документ са реферирани изображенията, чрез външни единици, които са декларирани в DTD документа, като в DTD схемата е описана и нотация, която задава формата на графичното съдържание.

Всеки университет и географски регион притежават уникален идентификационен номер, зададен, чрез задължителен атрибут от ID, съответно за университетите това е **univId**, а за регионите е **regionID**. В данните за локацията университетите е поместен и атрибут **regionIDRef**, който служи за рефериране към съответния регион, в който се намира университета. Този атрибут е от тип IDREF.

В DTD документа е указано каква трябва да е структурата, последователността и честотата на срещане на елементите. Също така какъв тип от стойности те приемат, както и вида на техните атрибути. Декларирани са нотация за JPG изображенията и единиците, които реферират самите изображения.

```
universities-in-bulgaria.dtd
1  <?xml version="1.0" encoding="UTF-8"?>
2
3  <!ELEMENT universitiesInBulgaria (universities, regions)>
4    <!ELEMENT universities (university+)>
5      <!ELEMENT university (univDetails, univContacts, programsAccreditation, listOfFaculties, listOfBranches?, listOfDepartments?)>
6        <!ELEMENT univDetails (name, yearFounded, univImage, rectorName, numberOfStudents,
7          numberOfBachPrograms?, numberOfMstPrograms?, numberOfPhDPrograms?)>
8          <!ELEMENT name (#PCDATA)>
9          <!ELEMENT yearFounded (#PCDATA)>
10         <!ELEMENT univImage (src)>
11           <!ELEMENT src EMPTY>
12         <!ELEMENT rectorName (#PCDATA)>
13         <!ELEMENT numberOfStudents (#PCDATA)>
14         <!ELEMENT numberOfBachPrograms (#PCDATA)>
15         <!ELEMENT numberOfMstPrograms (#PCDATA)>
16         <!ELEMENT numberOfPhDPrograms (#PCDATA)>
17
18        <!ELEMENT univContacts (location, postalCode, phone, webSite, email)>
19          <!ELEMENT location (city, street)>
20            <!ELEMENT city (#PCDATA)>
21            <!ELEMENT street (#PCDATA)>
22          <!ELEMENT postalCode (#PCDATA)>
23          <!ELEMENT phone (#PCDATA)>
24          <!ELEMENT webSite (#PCDATA)>
25          <!ELEMENT email (#PCDATA)>
26
```

```
universities-in-bulgaria.dtd
47  <!-- universities attribute list -->
48  <!ATTLIST university univId ID #REQUIRED>
49  <!ATTLIST src href ENTITY #REQUIRED>
50  <!ATTLIST location regionIDRef IDREF #REQUIRED>
51  <!ATTLIST webSite href CDATA #REQUIRED>
52  <!ATTLIST email mailAddress CDATA #REQUIRED>
53
54  <!-- regions attribute list -->
55  <!ATTLIST regionDetails regionID ID #REQUIRED>
56
```

```

universities-in-bulgaria.dtd
57 <!-- notations -->
58 <!NOTATION jpg PUBLIC "jpg viewer">
59
60 <!-- entities -->
61 <!ENTITY su SYSTEM "university-pictures/su.jpg" NDATA jpg>
62 <!ENTITY aubg SYSTEM "university-pictures/aubg.jpg" NDATA jpg>
63 <!ENTITY varnafree SYSTEM "university-pictures/varnafree.jpg" NDATA jpg>
64 <!ENTITY vtu SYSTEM "university-pictures/vtu.jpg" NDATA jpg>
65 <!ENTITY musofia SYSTEM "university-pictures/musofia.jpg" NDATA jpg>
66 <!ENTITY natifz SYSTEM "university-pictures/natifz.jpg" NDATA jpg>
67 <!ENTITY muzikalna SYSTEM "university-pictures/muzikalna.jpg" NDATA jpg>
68 <!ENTITY nbu SYSTEM "university-pictures/nbu.jpg" NDATA jpg>
69 <!ENTITY plovdivski SYSTEM "university-pictures/plovdivski.jpg" NDATA jpg>
70 <!ENTITY rusenski SYSTEM "university-pictures/rusenski.jpg" NDATA jpg>
71 <!ENTITY unss SYSTEM "university-pictures/unss.jpg" NDATA jpg>
72 <!ENTITY shumenski SYSTEM "university-pictures/shumenski.jpg" NDATA jpg>
73

```

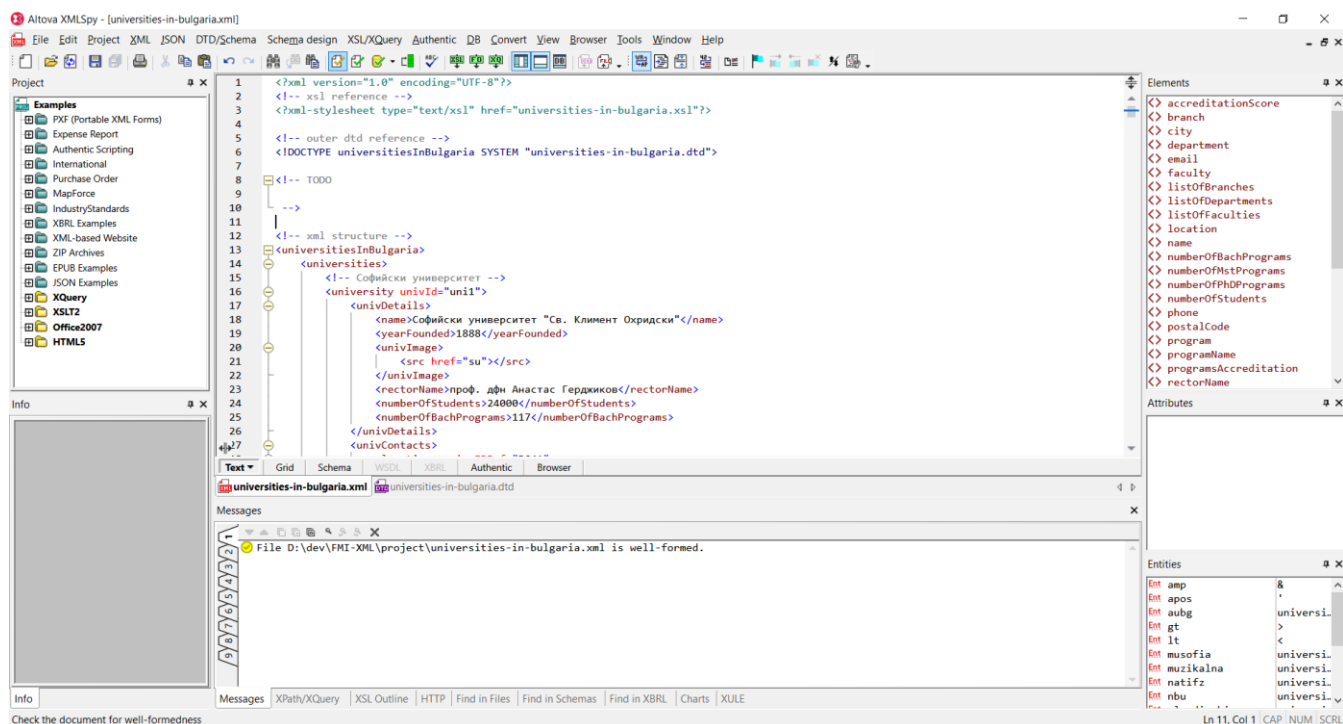
Генерирането на PDF документите се извършва с помощта на XSL документ, в който са описани трансформации, които визуализират текстовото съдържание на основния XML документ, както и задават размер и позиция на изображенията. Изпълняването на XSL трансформациите се извършва в средата Altova XMLSPY 2020. Използвана е технологията Apache FOP, която представя всяка страница като набор от кутии, на които се задават параметри за позициониране, както и каква информация от XML документа да бъде поместена в нея, като тя бива селектирана с помощта на XPath изрази. За всеки университет е представен в три страници - обща информация, структура и професионални направления, като в XSL документа са зададени основните елементи на всяка страница и има няколко шаблона за представяне на графично съдържание и представяне на множеството от елементи като факултети и филиали. Тъй като някои от университетите нямат всички структури, е използвана XSL конструкцията `xsl:if`, чрез която се извършва проверка за наличие на елемента за даден университет.

Чрез графичния интерфейс на програмата Altova XMLSPY 2020 се подават XML документа и XSL документа към процесора Apache FOP и се указва къде да бъде записан резултата от приложените трансформации, който е документ в PDF формат.

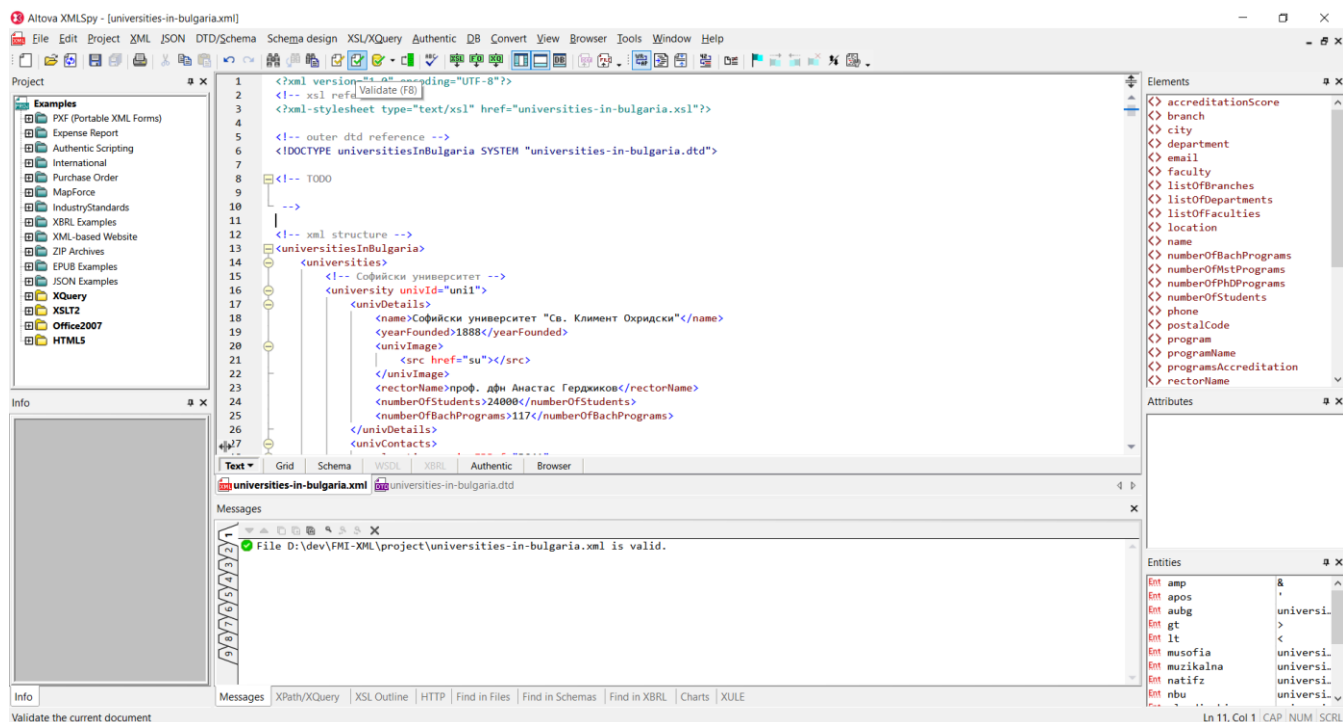
4 Тестване

Съдържанието на готовия XML документ е проверено за това, че е добре оформено и валидирано спрямо DTD документа в средата Altova XMLSPY 2020. Това става като се отворят XML файла, съдържащ информацията за университетите и DTD файла, описващ структурата, последователността и честотата на срещане на елементите, какъв тип от стойности те приемат, както и вида на техните атрибути. В DTD файла са описани и нотацията и XML единиците, нужни за включването на не-XML графичното съдържание. XML файла реферира този с DTD декларациите и след зареждането им от графичния интерфейс на средата Altova XMLSPY 2020 избираме да XML документа да бъде проверен за това, дали е добре оформен и дали в него са спазени DTD декларациите.

Проверка за това, че документа е добре оформен :

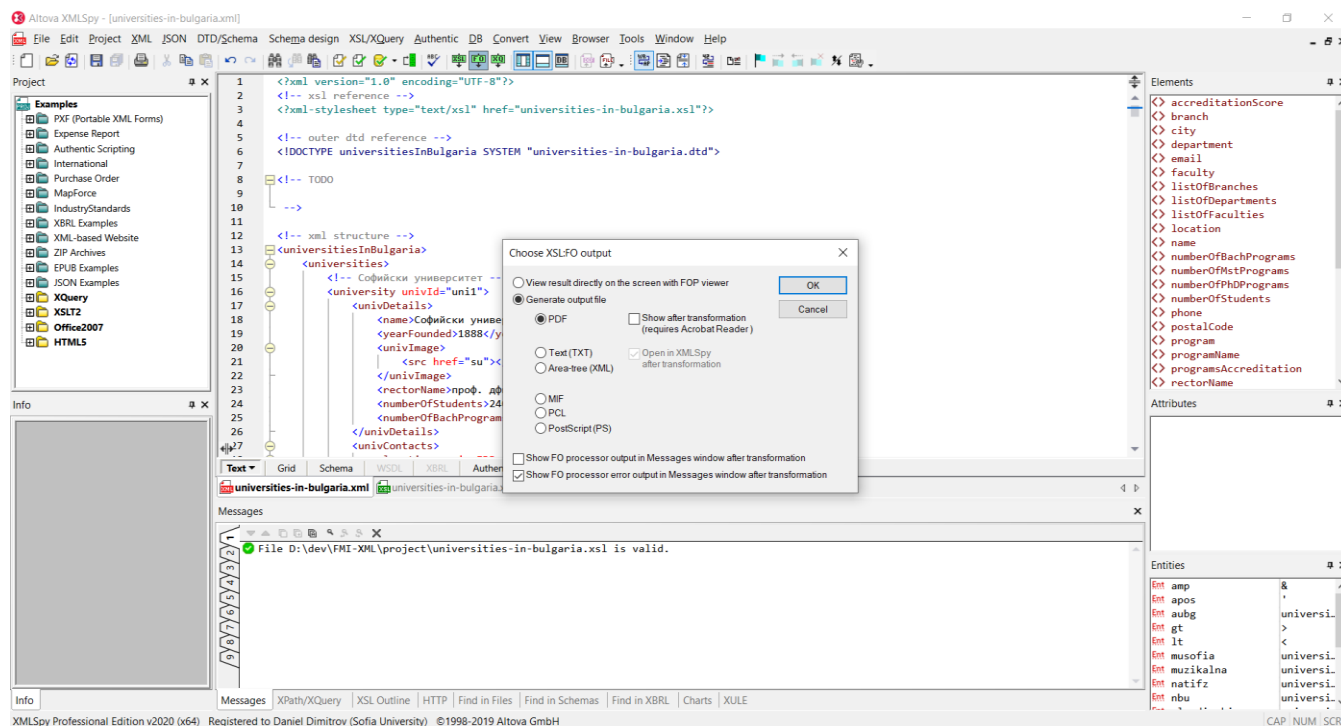
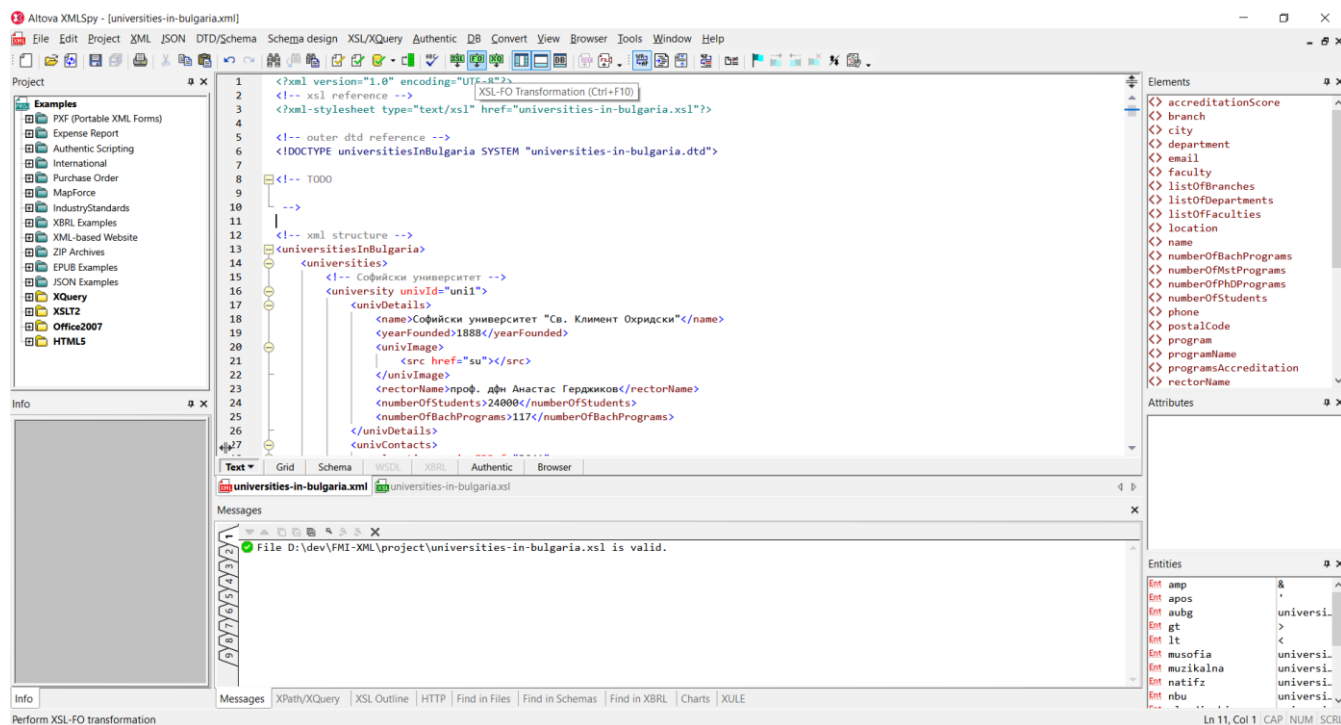


Валидиране на документа в средата Altova XMLSPY 2020 :



Виждаме, че XML документа е добре оформен и е валиден спрямо DTD файла.

Генерирането на PDF документите извършваме като първо заредим XML документа и документа с XSL трансформациите и след това от графичния интерфейс на средата Altova XMLSPY 2020 избираме опцията за прилагане на XSL-FO трансформации и запазване на резултата в PDF формат.



Резултатът е PDF файл, който съдържа готовия каталог с приложеното графично съдържание в него.

5 Заключение и възможно бъдещо развитие

В заключение, можем да кажем, че изградихме един информативен и семпъл каталог на най-добрите университети в България, който може да послужи на кандидат-студенти да направят своя информиран избор. Като насоки за подобряване можем да посочим обогатяване на информацията, като възможни данни са по-подробно излагане на структурата на университетите, на техните учебни програми и отзиви от бивши възпитаници. За съжаление за тези неща не успяхме намерим достатъчно ресурси. Спрямо технологиите не срещнахме никакви пречки и затруднения, но като недостатък може да споменем липсата на възможност на XSL-FO за направа на различни визуализации на данните в графичен формат като хистограми или непрекъснати линии например. С технологии, които поддържат такива графики и наличие на повече данни за определени категории, можем да например да визуализираме как се променят във времето показатели като оценките по направления и очаквани заплати от завършилите.

6 Разпределение на работата

Двамата автори съвместно измислихме общата структура на представяне на информацията за университетите и изготвиха DTD файла. Даниел намери и изготви информация за десет от дванадесетте университети и съответните им изображения. Иван намери информация и изображения за останалите 2 университета и въведе събраните данни във валиден XML документ. След това двамата заедно измислихме дизайна на PDF документа, който представлява каталога и Иван реализира файла с XSL трансформациите.

7 Използвани литературни източници и Уеб сайтове

1. Рейтинговата система на висшите училища на МОН : www.rsvu.mon.bg/rsvu4/#/universities
2. www.universiteti.start.bg/
3. www.myeducation.bg/bg/pages/nai-dobriyat-onlain-katalog-v-bylgariya/
4. www.moodle.org/mod/glossary/view.php?id=7607&mode=letter&hook=%D0%92&sortkey=CREATION&sortorder=as
5. Софийски университет "Св. Климент Охридски" : www.uni-sofia.bg
6. Американски университет в България – Благоевград : www.aubg.edu
7. Варненски свободен университет "Черноризец Храбър" : www.vfu.bg
8. Великотърновски университет "Св. св. Кирил и Методий" : www.uni-vt.bg
9. Медицински университет - София : www.mu-sofia.bg
10. НАТФИЗ "Кръстьо Сарафов" - София : www.natfiz.bg
11. Национална музикална академия "Проф. Панчо Владигеров" – София : www.nma.bg
12. Нов български университет - София : www.nbu.bg
13. Пловдивски университет "Паисий Хилендарски" : www.uni-plovdiv.bg

14. Русенски университет "Ангел Кънчев" : www.uni-ruse.bg
15. Университет за национално и световно стопанство - София : www.unwe.bg
16. Шуменски университет "Епископ Константин Преславски" : www.shu.bg
17. http://www.datypic.com/sc/fo11/e-fo_block-container.html
18. <https://w3schools.sinsixx.com/xslfo/default.asp.htm>
19. <http://videlibri.sourceforge.net/cgi-bin/xidelcgi>
20. <https://stackoverflow.com/questions/767851/xpath-find-if-node-exists>