|  |  |  |
| --- | --- | --- |
|  | **Софийски университет „Св. Кл. Охридски”**  Факултет по математика и информатика  *Бакалавърска програма „Софтуерно инженерство”* |  |

**Предмет: XML технологии за семантичен Уеб**

***Зимен семестър, 2020/2021 год.***

**Тема №09: „Каталог на ИТ фирми“**

**Курсов проект**

*Автори:*

*Симона Любенова, фак. номер 62258*

*Ати Мускова, фак. номер 62308*

януари, 2021

София

Съдържание

Table of Contents

[Съдържание 2](#_Toc61915614)

[Table of Contents 2](#_Toc61915615)

[1 Въведение 3](#_Toc61915616)

[2 Анализ на решението 3](#_Toc61915617)

[2.1 Работен процес 3](#_Toc61915618)

[2.2 Структура на съдържанието 4](#_Toc61915619)

[2.3 Тип и представяне на съдържанието 6](#_Toc61915620)

[3 Дизайн 6](#_Toc61915621)

[4 Тестване 7](#_Toc61915622)

[5 Заключение и възможно бъдещо развитие 8](#_Toc61915623)

[6 Разпределение на работата 8](#_Toc61915624)

[7 Използвани литературни източници и Уеб сайтове 8](#_Toc61915625)

# Въведение

В интернет пространството може да се намери информация за всички ИТ фирми в страната. Това обаче отнема доста време, понеже предложените каталози с фирми не са структурирани по удобен за потребителите начин. Поради тази причина решихме да създадем единен каталог, който да класифицира фирмите по два основни за нас критерия, а именно регион и сектор. Според нас тези два критерия са основни както за хората, търсещи работа, така и за тези, търсещи помощ за своите проекти. Използвайки придобитите знания, създадохме XML документ, който да съдържа най-важната според нас информация за всяка фирма. Валидирайки го чрез DTD документ и оформяйки го чрез XSL документ, успяхме да представим съдържанието в PDF формат, използвайки JVM и библиотеките apache и saxon. Графичното съдържание в XML документа е представено чрез единици (enties), а връзките между фирми и региони и фирми и сектори се представят чрез ID/IDREF.

В останалата част от този документ е описана по-подробно работата ни върху този курсов проект – структурата на съдържанието, дизайн и тестване на курсовия проект на тема „Каталог на ИТ фирми“.

# Анализ на решението

## Работен процес

След подробното проучване на зададения ни проблем изграждането на крайния продукт протече по следния начин:

* Набавихме необходимите според нас данни (като лого, информация за контакти и други) от официалните сайтове на избраните фирми, публични държавни архиви и единни национални портали.
* Създадохме DTD документ, в който посочихме основните елементи и атрибути, дефинирайки информацията, която могат да съдържат.
* На база на получения DTD документ създадохме XML схема, която беше валидирана от дадения документ.
* На базата на валидираната XML схема създадохме XSL документа, който да оформи желаната от нас информация.
* Чрез JVM и допълнителните библиотеки apache и saxon създадохме кратка програма, която обработва XML, XSL и DTD файловете и на базата на тях създава нашия краен продукт – PDF документ.
* Описахме процеса на създаване на курсовия проект в текущия документ.

## Структура на съдържанието

* catalogue – коренов елемент на XML документа. Той съдържа следните поделемнти – региони(regions), сектори(sectors) и фирми(companies).
  + regions - съдържа поне един елемент от единствения негов поделемент – регион (region).
    - region – той съдържа един поделемент (regionName) и два атрибута – идентификационен номер на региона (regionID) и референция на компанията (companyRef).
      * regionName – елемент от тип #PCDATA. Съдържанието на този елемент е текстово съдържание – името на региона.
      * regionID – задължителен атрибут от тип ID, който се реферира от фирмите, които се намират на територията на даден регион.
      * companyRef – задължителен атрибут от тип IDREF, който свързва фирмите със съответните региони.
  + sectors – съдържа поне един елемент от единствения си поделемент – сектор (sector).
    - sector – този елемент съдържа един поделемент – име на сектора (sectorName) и два атрибута – идентификационен номер на сектора (sectorID) и референция на компаниите (companyRef).
      * sectorName – този елемент е от тип #PCDATA. Съдържа текстово съдържание – името на сектора.
      * sectorID – това е задължителен атрибут от тип ID, който се реферира от фирмите, които се намират в даден сектор.
      * companyRef – задължителен атрибут от тип IDREF, който свързва фирмите със съответните им сектори.
  + companies – съдържа поне един елемент от единствения си поделемент – фирма (company).
    - company – този елемент съдържа следните поделемeнти – име на компанията(companyName), локация(location), контакти(contacts), ЕИК(UID), както незадължителните описание(description), изображение(image); съдържа атрибутите – идентификационен номер на компанията (companyID), референция към регионите(regionRef), референция към секторите (sectorRef).
      * companyName – елемент от тип #PCDATA, който съдържа името на фирмата.
      * location – съдържа задължителен елемент град(city), както и незадължителния елемент адрес(address).
        + city – елемент от тип #PCDATA. Съдържа текстово съдържание – името на града, в който се намира фирмата.
        + address – елемент от тип #PCDATA, който съдържа текстово съдържание – точният адрес на фирмата (квартал, улица, блок, вход и други).
      * contacts – съдържа задължителния елемент – телефон(phone), както и незадължителните елементи – имейл(email), сайт(site) и фейсбук(facebook).
        + phone – задължителен елемент от тип #PCDATA. Съдържа текстово съдържание – телефон за връзка.
        + email – незадължителен елемент от тип #PCDATA. Съдържа текстово съдържание – официален имейл за контакт с фирмата.
        + site – незадължителен елемент от тип #PCDATA. Съдържа текстово съдържание – официален сайт на фирмата.
        + facebook – елемнт от тип #PCDATA. Съдържа текстово съдържание – официална фейсбук страница на фирмата.
      * UID – елемент от тип #PCDATA. Съдържа текстово съдържание – ЕИК на фирмата.
      * description – елемент от тип #PCDATA. Съдържа текстово съдържание –описание на фирмата.
      * image – елемент от тип EMPTY с един атрибут href
        + href = задължителен атрибут от тип ENTITY. Връзка с елемент, който не е в XML формат.
      * comapnyID – задължителен атрибут от тип ID, който ще се реферира от регионите и секторите, в които се намира компанията.
      * regionRef – задължителен атрибут от тип IDREF. Свързва регионите с фирмите.
      * sectorRef – задължителен атрибут от тип IDREF. Свързва секторите с фирмите.

## Тип и представяне на съдържанието

Съдържанието на каталога е представено в PDF формат. Имаме изображения – логата на фирмите. Те са представени като снимков материал( JPG формат), които се намират в папка images към курсовия проект.

Броят на фирмите, присъстващи в каталога, е 13. Всички изображения, които присъстват в проекта, се намират в папката /images. Имената на използваните лого-изображения са:

* Фирма VMware – „vmware.jpg“
* Фирма Ubisoft – „ubisoft.jpg“
* Фирма Uran Company – „uran.jpg“
* Фирма IT Master Soft – „itmastersoft.jpg“
* Фирма Gameloft – „gameloft.jpg“
* Фирма Wollow – „wollow.jpg“
* Фирма Flat Rock Technologies – „flatrock.jpg“
* Фирма Code Plus LTd – „codeplusltd.jpg“
* Фирма Miracle Mill AG – „miraclemill.jpg“
* Фирма Dev Wave – „devwave.jpg“
* Фирма Areus – „areus.jpg“
* Фирма New Line Technologies – „newlinetechnologies.jpg“
* Фирма Scale Focus – „scalefocus.jpg“

# Дизайн

При изграждането на нашия курсов проект създадохме и използвахме следните типове файлове:

* XML
* DTD
* XSL
* JAVA
* PDF

Изображенията са представени чрез външни XML entities. Връзките реферират към изображенията. В каталога също сме използвали атрибути ID/IDREF. Те свързват фирмите с региона, в който се намират, както и фирмите със съответния сектор.

Документът е валидиран чрез DTD документ. В него са описани типовете елементи и атрибути, както и честотата им на срещане и дали задължително трябва да присъстват в XML документа. Например:

* <!ELEMENT regions (region+)>
* <!ELEMENT company (companyName, location, contacts, UID, description?,image?)>
* <!ELEMENT contacts (phone, email?, site?, facebook?)>
* <!ELEMENT location (city, address?)>

Единиците са валидирани по следния начин:

* <!ENTITY vmware SYSTEM "images/vmware.jpg" NDATA jpg>
* <!ENTITY ubisoft SYSTEM "images/ubisoft.jpg" NDATA jpg>
* <!ENTITY itmastersoft SYSTEM "images/itmastersoft.jpg" NDATA jpg>

XML документът е генериран в PDF документ чрез XSL документ и с помощта на създадена от нас програма на езика JAVA, използвайки допълнителните библиотеки apache и saxon. Имаме една заглавна страница, а останалите 13 страници показват данните на всяка една фирма по аналогичен начин: лого, заглавие, локация, контакти, ЕИК и описание.

# Тестване

XML документът е валидиран чрез DTD документ, а с помощта на източник №1 е реализирана валидация.

Резултатът от вече валидирания XML документ и XSL документа е следният:



# Заключение и възможно бъдещо развитие

Курсовият ни проект на тема „Каталог на ИТ фирми – 1“ показва първоначална версия на краен продукт, който би бил полезен на много хора. В бъдеще може да се увеличи броят на фирмите в него, както и да се добавят нови връзки между тях. Каталогът може да се подобри и чрез добавяне на събития и конференции, които могат да привлекат нови кадри към фирмата, както и актуални стажантски и практикантски програми.

# Разпределение на работата

За изграждане на нашия проект разделихме работата си, както следва:

* Симона Любенова
  + Създаване на XML документ
  + Създаване на XSL документ
  + Генериране на PDF документ
* Ати Мускова
  + Набавяне на данни
  + Създаване на DTD документ
  + Документация на курсовия проект.

По време на разработка на курсовия проект всички решения бяха обсъдени и одобрени и от двамата автори.

# Използвани литературни източници и Уеб сайтове

1. <https://www.xmlvalidation.com/documentation.html>
2. [http://urancompany.com](http://urancompany.com/)
3. <https://wollow-soft.com/bg-BG>
4. [https://codeplusmarketing.com](https://codeplusmarketing.com/)
5. https://itmaster-soft.com/
6. https://www.areusdev.com
7. https://www.flatrocktech.com
8. https://newline.tech
9. https://miraclemill.com
10. https://devwave.net
11. https://www.scalefocus.com
12. https://www.ubisoft.com/en-us/
13. https://www.vmware.com/
14. https://www.gameloft.com/en/