НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КПІ»

ФAКУЛЬТЕТ ПРИКЛАДНОЇ МАТЕМАТИКИ

Кафедра програмного забезпечення комп’ютерних систем

ЛАБОРАТОРНА РОБОТА №1

*з дисципліни*

*«Бази даних на основі XML»*

**ТЕМА: « Вивчення базових операцій обробки XML-документів »**

Підготував: студент групи КП-41

Худер Карім Нідаль

Перевірив:.

Петрашенко Андрій Васильович

Оцінка

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(дата, підпис)

Київ 2016

**Метою роботи** є здобуття практичних навичок створення програм, орієнтованих на обробку XML-документів засобами мови Python.

**Завдання роботи** полягає у наступному:

1. Виконати збір інформації зі сторінок Web-сайту за варіантом.
2. Виконати аналіз сторінок Web-сайту для подальшої обробки текстової та графічної інформації, розміщеної на ньому.
3. Реалізувати функціональні можливості згідно вимог, наведених нижче.

**Функціональні вимоги**

1. На основі базової адреси Web-сайту виконати обхід наявних сторінок сайту, відокремлюючи текстову та графічну інформацію від тегів HTML. Пошук вузлів виконувати засобами XPath. Наступну сторінку для аналізу цього ж сайту обрати як одне із гіперпосилань на даній сторінці (тег <a href=”url”/>). Обмежитись аналізом 20 сторінок сайту. Зберегти XML у вигляді файлу.
2. Виконати аналіз отриманих даних засобами XML згідно варіанту та вивести результати у консольне вікно. Відбір вузлів виконувати засобами XPath.
3. Проаналізувати вміст Web-сторінок інтернет-магазину (див. варіант). Отримати ціну, опис та зображення для 20 товарів з нього за допомогою DOM-парсеру та мови XPath для пошуку відповідних вузлів. Результат записати в XML-файл.
4. Перетворити отриманий XML-файл у XHTML-сторінку за допомогою мови XSLT. Дані подати у вигляді XHTML-таблиці та записати його у файл.

**Product.xsl**

*<?***xml version="1.0" encoding="UTF-8"***?>*<**xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"**>  
  
<**xsl:output method="xml" doctype-system  
="http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"  
 doctype-public="-//W3C//DTD XHTML 1.0 Transitional//EN" indent="yes" encoding="utf-8"** />  
  
<**xsl:template match="/"**>  
 <**html**>  
 <**head**>  
 <**meta http-equiv="Content-Type" content="text/xhtml; charset=utf-8"** />  
 <**title**>Laba 1</**title**>  
 </**head**>  
 <**body**>  
 <**table**>  
 <**tbody**>  
 <**xsl:for-each select="data/category/product"**>  
 <**tr style="margin-bottom: 25px;"**>  
 <**td style="border-bottom: 1px solid #ddd; padding: 15px;"**>  
 <**img style="max-width: 250px"**><**xsl:attribute name="src"**><**xsl:value-of select='image'** /></**xsl:attribute**></**img**>  
 </**td**>  
 <**td style="border-bottom: 1px solid #ddd; padding: 15px;"**>  
 <**h3**><**xsl:value-of select="header"**/></**h3**>  
 </**td**>  
 <**td style="border-bottom: 1px solid #ddd; padding: 15px;"**>  
 <**span**><**xsl:value-of select="price"**/></**span**>  
 </**td**>  
 <**td style="border-bottom: 1px solid #ddd; padding: 15px;"**>  
 <**p**><**xsl:value-of select="description"**/></**p**>  
 </**td**>  
 </**tr**>  
 </**xsl:for-each**>  
 </**tbody**>  
 </**table**>  
 </**body**>  
 </**html**>  
</**xsl:template**>  
  
</**xsl:stylesheet**>

**Main3**

page = html.parse(**'http://www.fishing-mart.com.ua/'**)  
data = ET.Element(**"data"**)  
categories = page.xpath(**"//div[@class='block\_content']/ul[contains(@class, 'tree')]/li/ul/li/a/@href"**)[0:17]  
**for** cat **in** categories:  
 categoriesElement = ET.SubElement(data, **'category'**, url=cat)  
 p = html.parse(cat)  
 products = p.xpath(**"//ul[@id='product\_list']/li/\*/a[@class='product\_img\_link']/@href"**)  
 **for** pr **in** products:  
 productElement = ET.SubElement(categoriesElement, **"product"**, url=pr)  
 p = html.parse(pr)  
 *#Print Header* header = p.xpath(**"//div[@id='primary\_block']/h1/text()"**)[0]  
 ET.SubElement(productElement, **'header'**, type=**"text"**).text = header  
 *#print Image* image = p.xpath(**"//div[@id='primary\_block']//img[@id='bigpic']/@src"**)[0]  
 ET.SubElement(productElement, **'image'**, type=**"image"**).text = image  
 *#print price* price = p.xpath(**"//div[@id='primary\_block']//span[@id='our\_price\_display']/text()"**)[0]  
 ET.SubElement(productElement, **'price'**, type=**"text"**).text = price  
 *#print description* desc = join(p.xpath(**"//div[@id='more\_info\_sheets']/div[@id='idTab1']//text()"**), **" "**)  
 ET.SubElement(productElement, **'description'**, type=**"text"**).text = desc

tree = ET.ElementTree(data)  
tree.write(**"data-3.xml"**, pretty\_print=True, encoding=**"UTF-8"**)

**Product.html (фрагмент)**

<!DOCTYPE **html** PUBLIC **"-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"**>  
<**html xmlns="http://www.w3.org/1999/xhtml"**>  
 <**head**>  
 <**meta http-equiv="Content-Type" content="text/xhtml; charset=utf-8"** />  
 <**title**>Laba 1</**title**>  
 </**head**>  
 <**body**>  
 <**table**>  
 <**tbody**>  
 <**tr style="margin-bottom**: 25**px**;**"**>  
 <**td style="border-bottom**: 1**px solid #ddd**; **padding**: 15**px**;**"**>  
 <**img style="max-width**: 250**px" src="http://www.fishing-mart.com.ua/1091-4537-large/bratfishing-felchen-hot-shot-2-14-imx8.jpg"** />  
 </**td**>  
 <**td style="border-bottom**: 1**px solid #ddd**; **padding**: 15**px**;**"**>  
 <**h3**>Bratfishing Felchen Hot Shot 2-14 IMX8</**h3**>  
 </**td**>  
 <**td style="border-bottom**: 1**px solid #ddd**; **padding**: 15**px**;**"**>  
 <**span**>$624.00</**span**>  
 </**td**>  
 <**td style="border-bottom**: 1**px solid #ddd**; **padding**: 15**px**;**"**>  
 <**p**>Bratfishing Felchen - карбоновый бюджетный спиннинг лайт класса. Изготовлен из карбона марки IM8. По весу очень легкий. Данный спиннинг позиционируется как спиннинг класса Hot Shot…</**p**>  
 </**td**>  
 </**tr**>

