



# **МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**Національний технічний університет України**

**«Київський політехнічний інститут»**

**Факультет прикладної математики**

**Кафедра програмного забезпечення комп'ютерних систем**

## **ЛАБОРАТОРНА РОБОТА №1**

**з дисципліни**

**«Бази даних-2»**

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

**Виконав: студент групи КП-42  
Сущик Андрій Миколайович**

**Київ 2016**

## Варіант 19

*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. Формат XML-документу:

```
<data>
  <page url="www.server.com/index.html">
    <fragment type="text">
... знайдений текст
    </fragment>
    <fragment type="image">
... url зображення
    </fragment>
  </page>
  <page url="www.server.com/index1.html">
    <fragment type="text">
... знайдений текст
    </fragment>
    <fragment type="image">
... url зображення
    </fragment>
  </page>
...
</data>
```

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.

#### *Вимоги до інтерфейсу користувача*

Використовувати консольний (текстовий) інтерфейс користувача.

#### *Вимоги до інструментарію*

1. Мова програмування – Python 2.7
2. Бібліотека для обробки XML – libxml2 або (xml.dom, xml.xpath.. - вбудовані пакети) або lxml – на вибір студента.
3. Середовище розробки програмного забезпечення – PyCharm Community Edition 2016 (опціонально)

#### *Варіант*

19.	<a href="http://www.shkola.ua">www.shkola.ua</a>	Кількість текстових фрагментів по кожному фрагменту	<a href="http://www.fishing-mart.com.ua">www.fishing-mart.com.ua</a>
-----	--	---	--

#### *Вміст програми*

[www.github.com/Sushchyk/DBLabs/Lab1](https://www.github.com/Sushchyk/DBLabs/Lab1)