Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

Кафедра Штучного інтелекту

Дисципліна: «Інтелектуальні технології в Internet та Semantic Web»

Лабораторна робота № 1

на тему: «РОЗРОБКА XML-ДОКУМЕНТІВ ТА ВІДПОВІДНИХ XML-СХЕМ У РЕДАКТОРІ XML»

|  |  |  |
| --- | --- | --- |
| Виконали:  ст. гр. ІТКН-17-7  Савенков Д.В. |  | Перевірив:  Бібічков І. Є. |
|  |  |  |

Харків 2020

1 РОЗРОБКА XML-ДОКУМЕНТІВ ТА ВІДПОВІДНИХ XML-СХЕМ У РЕДАКТОРІ XML

* 1. **Мета роботи**

Отримання даних із віддаленого WEB-ресурсу та їх обробка. Ознайомлення із роботою у редакторі XMLSpy або в аналогу – середовищі для роботи з кількома XML-технологіми. Розробка XML-схеми, XML-документа, його валідація.

**1.2 Хід роботи**

**Варіант -** 'https://djinni.co/set\_lang?code=en&next=/developers/'

**Посилання на гіт-репозиторій:** https://github.com/denys-savenkov/Savenkov-Semantic-Web

**1.2.1 Аргументація вибору мови програмування та засобів розробки**

Для виконання поставленого практичного завдання була обрана мова програмування Python за її легкість, зручність, наявність великої кількості бібліотек та добре складеної документації. Серед засобів розробки була обрана IDE PyCharm за її функціонал, зокрема, інтеграції механізмів та автоматизації роботи із XML та XSD файлами.

Для добування даних з веб ресурсу була обрана бібліотека Scrapy v2.0 (детальна інформація та документація: https://scrapy.org/). Однією з переваг саме цієї бібліотеки є розпаралелювання процесу «скрапінгу» та можливість зберігати дані у форматі XML.

**1.2.2 Створення XSD-схеми**

Після аналізу веб-ресурсу була створена наступна XSD-схема для подальшої валідації отриманих парсером даних:

<?xml version="1.0" encoding="UTF-8"?>

<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified" xmlns:xs="http://www.w3.org/2001/XMLSchema">

<xs:element name="candidates" type="candidatesType"/>

<xs:complexType name="locationsType">

<xs:sequence>

<xs:element type="xs:string" name="value" maxOccurs="unbounded" minOccurs="0"/>

</xs:sequence>

</xs:complexType>

<xs:complexType name="candidateType">

<xs:sequence>

<xs:element type="xs:string" name="url"/>

<xs:element type="xs:string" name="position"/>

<xs:element type="locationsType" name="locations"/>

<xs:element type="xs:string" name="salary"/>

<xs:element type="xs:float" name="years\_of\_experience"/>

<xs:element name="english\_level">

<xs:simpleType>

<xs:restriction base="xs:string">

<xs:enumeration value="Beginner/Elementary"/>

<xs:enumeration value="Advanced/Fluent"/>

<xs:enumeration value="Intermediate"/>

<xs:enumeration value="Pre-Intermediate"/>

<xs:enumeration value="Upper Intermediate"/>

<xs:enumeration value=""/>

</xs:restriction>

</xs:simpleType>

</xs:element>

<xs:element type="xs:string" name="experience"/>

<xs:element type="xs:string" name="skills"/>

<xs:element type="xs:string" name="highlights"/>

<xs:element type="xs:string" name="expectations"/>

</xs:sequence>

</xs:complexType>

<xs:complexType name="candidatesType">

<xs:sequence>

<xs:element type="candidateType" name="candidate" maxOccurs="unbounded" minOccurs="0"/>

</xs:sequence>

</xs:complexType>

</xs:schema>