ХЕРСОНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
(повне найменування вищого навчального закладу)
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ДИЗАЙНУ
(повне найменування інституту, назва факультету (відділення))
КАФЕДРА ПРОГРАМНИХ ЗАСОБІВ І ТЕХНОЛОГІЙ
(повна назва кафедри (предметної, циклової комісії))

Пояснювальна записка

до атестаційної роботи
бакалавра
(освітньо-кваліфікаційний рівень)

на тему: «Розробка веб-додатку для парсингу інтернет-магазинів»

Виконав: студент 4 курсу, групи 4ПР2
напряму підготовки (спеціальності)
121 -«Інженерія програмного забезпечення»
(шифр і назва напряму підготовки, спеціальності)

Дембовський А.В.
(прізвище та ініціали)
Керівник _д.т.н., професор Жарікова М.В.
(прізвище та ініціали)
Рецензент _____
(прізвище та ініціали)

Херсон – 2021

Херсонський національний технічний університет_____
(повне найменування вищого навчального закладу)

Факультет Інформаційних технологій та дизайну
Кафедра _Програмних засобів і технологій_____
_____
Освітньо-кваліфікаційний рівень___бакалавр_____
_____
Галузь знань _____12 – Інформаційні технології_____
_____
Спеціальність _____121 «Інженерія програмного забезпечення»

ЗАТВЕРДЖУЮ
Завідувач кафедри ПЗ і Т
д.т.н. проф. В.Г. Шерстюк _
“____” _____ 2021 р.

З А В Д А Н Н Я

НА ДИПЛОМНИЙ ПРОЕКТ (РОБОТУ) СТУДЕНТУ

_____Дембовському А.В._____

1. Тема проекту (роботи)__ Розробка веб-додатку для парсингу інтернет-магазинів_____
керівник проекту (роботи)____д.т.н. професор Жарікова М.В__._____,
(прізвище, ім’я, по батькові, науковий ступінь, вчене звання)
затверджена наказом вищого навчального закладу від “__” _____ №_____
2. Строк подання студентом проекту (роботи)_25.05.2021_____
3. Вихідні дані до проекту (роботи)_ ДСТУ з обробки інформації, літературні та періодичні джерела, матеріали практики _____
4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)__1. Аналіз предметної області та постановка завдання; 2. Алгоритми машинного навчання та методи дослідження; 3. Проектування та розробка програмного застосунку; 4. Хід дослідження ефективності алгоритмів_____
_____
5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)
1. Мета, об’єкт, предмет та задачі дослідження;
_____

6. Консультанти розділів проекту (роботи)

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання_____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів проекту (роботи)	Примітка
1	Відбір та вивчення літературних джерел	25.09.2021	
2	Аналіз стану вирішення завдання на сучасному етапі	30.09.2021	
3	Побудова концептуальної моделі	3.10.2021	
4	Розробка математичної моделі	10.10.2021	
5	Побудова алгоритму функціонування програмного продукту	15.10.2021	
6	Написання вихідного коду програми	01.11.2021	
7	Налагодження програмного коду	15.11.2021	
8	Оформлення пояснювальної записки	02.12.2021	

Студент _____ __Дембовський А.В.
(підпис) (прізвище та ініціали)

Керівник проекту (роботи) _____ Жарікова М.В.___
(підпис) (прізвище та ініціали)

РЕФЕРАТ

Об’єкт дослідження – веб-додаток для парсингу інтернет-магазинів.

В першому розділі проаналізовано предметну область за поставлено задачу дослідження.

В другому розділі описано використані алгоритми і технології парсингу, такі як SQL, CSV, RSS, XLS, JSON, Document Object Model.

В третьому розділі описано проектування та розробку веб-додатку.

В червертому розділі розділі описано практичну реалізацію веб-додатку.

Повний обсяг роботи 72 сторінки. Містить 21 рисунок, 1 таблицю, 2 додатока, 12 літературних джерел.

Ключові слова: парсинг, веб-сайт, посилання, контент, інформація, алгоритм, структура, веб-ресурс.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

Web – система доступу до пов'язаних між собою документів на різних комп'ютерах, підключених до Інтернету

HTML – HyperText Markup Language

URL – Uniform Resource Locator

HTTP – HyperText Transfer Protocol

XML – eXtensible Markup Language (розширювана мова розмітки)

БД – база даних

NoSQL – not only Structured query language

CSV - Comma-Separated Values

JSON - JavaScript Object Notation

DOM - Document Object Model

URL - Uniform Resource Locator

CAPTCHA - Completely Automated Public Turing test to tell Computers and Humans Apart

DFD - Data Flow Diagrams

ВСТУП

Парсинг даних – це автоматичний збір інформації розміщеної у відкритому доступі в Інтернет та її систематизація за допомогою спеціальних програм. В цілому, парсер виконує ті ж функції, що й ручний збір інформації, тільки в рази швидше та зводячи до 0 можливість помилки.

Парсинг – це лінійне зіставлення послідовності слів з правилами мови. Поняття «мова» розглядається в найширшому контексті. Це може бути звичайна людська мова (наприклад, українська), яка використовується для комунікації людей. А може бути і формалізована мова, зокрема, будь-яка мова програмування.

Парсинг сайтів – послідовний синтаксичний аналіз інформації, розміщеної на Інтернет-сторінках. Текст Інтернет-сторінок – це ієрархічний набір даних, структурований за допомогою людських і комп'ютерних мов. Людською мовою надана інформація, знання, заради яких, власне, люди і користуються мережею Інтернет. Комп'ютерні мови (HTML, JavaScript, CSS) визначають, як інформація виглядає на моніторі, розмітку сторінки, її стиль.