


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій

Затверджую
Декан факультету комп'ютерних
інформаційних технологій


Ігор Якименко
2024р.

Затверджую
Директор Навчально-наукового інституту
новітніх освітніх технологій


Святослав Питель
2024р.

Затверджую
Проректор з науково-педагогічної роботи


Віктор ОСТРОВЕРХОВ
2024 р..

РОБОЧА ПРОГРАМА
з дисципліни

«Збір та обробка даних»

Ступінь вищої освіти – другий (магістерський)

Галузь знань: 12 Інформаційні технології

Спеціальність: 124 Системний аналіз

Освітньо-професійна програма «Системний аналіз»

Кафедра економічної кібернетики та інформатики

| Форма навчання | Курс | Семестр | Лекції | Практ. | ІРС | Тренінг, КПЗ | СРС | Разом | Екзамен, залік (семестр) |
|----------------|------|---------|--------|--------|-----|--------------|-----|-------|--------------------------|
| Денна | 1 | 1 | 32 | 14 | 5 | 6 | 93 | 150 | 1 |
| Заочна | 1 | 1 | 8 | 4 | | | 138 | 150 | 1 |

Тернопіль

2024

Робоча програма складена на основі освітньо-професійної програми підготовки магістра галузі знань 12 Інформаційні технології спеціальності 124 Системний аналіз, затвердженої на засіданні вченої ради ЗУНУ (протокол № 11 від 26.06. 2024 р.)

Робочу програму склав: д.т.н., професор, професор кафедри економічної кібернетики та інформатики Роман ПАСІЧНИК

Робоча програма затверджена на засіданні кафедри економічної кібернетики та інформатики, протокол № 1 від _____ 2024 р.

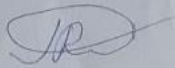
Завідувач кафедри



проф. Леся БУЯК

Розглянуто та схвалено групою забезпечення спеціальності системний аналіз, протокол №1 від _____ 2024 р.

Голова ГЗС



проф. Роман ПАСІЧНИК

Гарант ОПП



проф. Дмитро БОДНАР

СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Збір та обробка даних»

1. Опис дисципліни «Збір та обробка даних»

| Дисципліна – Збір та обробка даних | Галузь знань, спеціальність, ступінь вищої освіти | Характеристика навчальної дисципліни |
|---|---|--|
| Кількість кредитів ECTS 5 | Галузь знань – 12 «Інформаційні технології» | Статус дисципліни – обов'язкова мова навчання - <i>українська</i> |
| Кількість залікових модулів - 3 | Спеціальність – 124 «Системний аналіз», | <i>Денна:</i> Рік підготовки: 1 Семестр – 1 |
| Кількість змістових модулів - 2 | Ступінь вищої освіти – магістр | <i>Денна:</i> лекції – 32 год.; практ – 14 год. <i>Заочна:</i> лекції – 8 год.; практ – 4 год |
| Загальна кількість годин - - 150 | | Самостійна робота: <i>денна: 93 год.,</i> <i>заочна: 138</i> Тренінг – 6 год.. |
| Тижневих годин: 10 год., з них аудиторних – 3 год | | Вид підсумкового контролю – <i>залік</i> |

2. Мета й завдання вивчення дисципліни " Збір та обробка даних "

2.1. Мета вивчення дисципліни

Метою викладання дисципліни " Збір та обробка даних " є ознайомлення студентів з методологіями розроблення Веб-проектів, автоматизованого збору проектної інформації, організації оброблення проектної інформації із використанням засобів штучного інтелекту.

Предметом навчальної дисципліни є методи та засоби реалізації архітектури програмних систем та методів автоматизованого їх поповнення.

2.2. Завдання вивчення дисципліни " Збір та обробка даних " - поглиблення знань студентів щодо принципів і методології автоматизованого збору інформації для її оброблення у проекті а також методів оброблення зібраної інформації за допомогою спеціалізованих засобів, в тому числі зсобів із використанням технологій штучного інтелекту.

2.3. Найменування та опис компетентностей, формування котрих забезпечує вивчення дисциплін:

СК2. Здатність проектувати архітектуру інформаційних систем.

СК5. Здатність моделювати, прогнозувати та проектувати складні системи і процеси на основі методів та інструментальних засобів системного аналізу.

СК8. Здатність розробляти і реалізовувати наукові та прикладні проекти в галузі інформаційних технологій та дотичні до неї міждисциплінарні проекти.

СК11. Здатність розробляти і прототипувати спеціалізовані веб-системи

2.4. Передумови для вивчення дисципліни.

«Веб технології», «Основи програмування»,

2.5. Результати навчання:

РН12. Застосовувати методи машинного навчання та скрапінгу для видобування та аналізу структурованих даних із глобальних мереж.

РН13. Розробляти та застосовувати методи, алгоритми та інструменти класифікації користувачів соціальних мереж.

3. Програма дисципліни " Збір та обробка даних "

Змістовий модуль 1 – Наповнення даними проекту

Тема 1. Розроблення структури Веб-проекту

Постановка задачі проекту. Послідовність етапів реалізації проекту. Засоби реалізації проекту. CMS. Мови програмування. Фреймворки. Програмне середовище PyCharm.

Тема 2. Наповнення даними проекту

Збір даних. Основи реалізації скрапінгу на Python. Requests. BeautifulSoup 4. Lxml. Selenium. Початок роботи із BS4. Встановлення фреймворку. Перевірка працездатності фреймворку на тестовій сторінці. Обробка помилок підключення.

Тема 3. Реалізація скрапінгу за допомогою BS4

Структура бібліотеки BeautifulSoup. Основні переваги BeautifulSoup. Типові застосування BeautifulSoup

Тема 4. Програмний аналіз HTML сторінок

Фільтрація компонентів сторінок за тегами, атрибутами та текстовими фрагментами вмістимого тегів. Задання обсягу пошуку. Коректне використання атрибуту «class». Пошук за ключовими словами із можливістю відсіву надлишкової інформації. Обробка батьківських та дочірніх елементів. Робота з однорівневими елементами. Регулярні вирази.

Тема 5. Збір даних на множинах HTML сторінок та Веб-сайтів

Загальна структура скрепування всього Веб-сайту. Уникнення дублюючих посилань. Побудова внутрішніх та зовнішніх посилань Веб-сторінки. Скрепінг за допомогою бібліотеки scrapy та python3. Зберігання сканованої інформації в файлах заданих форматів.

Змістовий модуль 2 – Змістовне наповнення проекту

Тема 6. Наповнення загальної структури проекту

Вибір напрямку, об'єкта та предмета дослідження. Збір інформації про базові

напрацювання щодо предмету дослідження та його мало досліджені аспекти. Постановка задачі дослідження. Система базових відношень проекту. DFD пропонованої інформаційної системи. Ієрархічна ER діаграма бази даних інформаційної системи.

Тема 7. Моделі комп'ютерного зору

Метод навчання з учителем у побудові моделей комп'ютерного зору. Згорткові нейронні мережі. Трансформація низькорівневих ознак у моделі реальних об'єктів. Етапи маркування зображень. Підвищення якості маркування. Уточнення візуальних тегів, об'єднання міток, трансформація зображень. Ручна, напівавтоматизована, автоматизована та інтерактивна розмітки.

Тема 8. Семантична розмітка зображень

Глибокі нейронні мережі. Згорткові мережі. Основні кроки у класифікації зображень. Спеціалізована геоінформаційна система моделей класифікації зображень. Інформаційні моделі ГІС. Типи бізнес-процесів, що маркуються. Локації та методи маркування. Набори зображень для навчання та тестування. Некласифіковані зображення.

Тема 9. Структура фреймворку TensorFlow

Еволюція TensorFlow. Концепція обчислювального графа. Багатовимірні масиви як тензори. Операції над тензорами. Послідовності операцій та їх графи. Оптимізатори. Моделі. Галузі використання TensorFlow. Комп'ютерний зір, обробка природної мови, рекомендаційні системи, прогнозування часових рядів.

Тема 10. Розпізнавання зображень за допомогою фреймворку TensorFlow

Навчальний та тестовий набори зображень. Формування наборів за допомогою фільтрованого скрапінгу. Кількість епох. Частка валідаційного набору. Dataset - набір зображень. Конвеєри обробки даних із аугментацією. Оцінка точності моделі.

4. Структура залікового кредиту дисципліни " Збір та обробка даних "

денна

| | Лекції | Практичні заняття | Самостійна робота | Інд. роб. ст.. | Тренінг | Контроль заходи |
|---|--------|-------------------|-------------------|----------------|---------|--------------------|
| Змістовий модуль 1 – Наповнення даними проекту | | | | | | |
| Тема 1. Розроблення структури Веб-проекту | 3 | 1 | 9 | 2 | 3 | поточне опитування |
| Тема 2. Наповнення даними проекту | 3 | 1 | 9 | | | |
| Тема 3. Реалізація скрапінгу за допомогою BS4 | 3 | 1 | 9 | | | |
| Тема 4. Програмний аналіз HTML сторінок | 3 | 1 | 9 | | | |
| Тема 5. Збір даних на множинах HTML сторінок та Веб-сайтів | 3 | 1 | 9 | | | |
| Змістовий модуль 2 – Змістовне наповнення проекту | | | | | | |
| Тема 6. Наповнення загальної структури проекту | 3 | 1 | 9 | 3 | 3 | поточне опитування |
| Тема 7. Моделі комп'ютерного зору | 3 | 2 | 9 | | | |
| Тема 8. Семантична розмітка зображень | 3 | 2 | 10 | | | |
| Тема 9. Структура фреймворку TensorFlow | 4 | 2 | 10 | | | |
| Тема 10. Розпізнавання зображень за допомогою фреймворку TensorFlow | 4 | 2 | 10 | | | |
| Разом | 32 | 14 | 93 | 5 | 6 | |

| | Лекції | Практичні заняття | Самостійна робота | Тренінг | Контрольні заходи |
|---|----------|-------------------|-------------------|---------|--------------------|
| Змістовий модуль 1 – Наповнення даними проекту | | | | | |
| Тема 1. Розроблення структури Веб-проекту | 0.5 | | 13 | | поточне опитування |
| Тема 2. Наповнення даними проекту | 0.5 | | 13 | | |
| Тема 3. Реалізація скрапінгу за допомогою BS4 | 0.5 | 1 | 14 | | |
| Тема 4. Програмний аналіз HTML сторінок | 0.5 | 1 | 14 | | |
| Тема 5. Збір даних на множинах HTML сторінок та Веб-сайтів | 1 | | 14 | | |
| Змістовий модуль 2 – Змістовне наповнення проекту | | | | | |
| Тема 6. Наповнення загальної структури проекту | 1 | | 14 | | поточне опитування |
| Тема 7. Моделі комп'ютерного зору | 1 | | 14 | | |
| Тема 8. Семантична розмітка зображень | 1 | 1 | 14 | | |
| Тема 9. Структура фреймворку TensorFlow | 1 | | 14 | | |
| Тема 10. Розпізнавання зображень за допомогою фреймворку TensorFlow | 1 | 1 | 14 | | |
| Разом | 8 | 4 | 138 | | |

5. Тематика практичних занять

Практичне заняття 1. Розроблення структури Веб-проекту.

1. Постановка задачі проекту.
2. Послідовність етапів реалізації проекту.
3. Засоби реалізації проекту.
4. Програмне середовище PyCharm.

Практичне заняття 2. Наповнення даними проекту

1. Збір даних. Основи реалізації скрапінгу на Python.
2. Requests. BeautifulSoup 4. Lxml. Selenium.
3. Початок роботи із BS4. Встановлення фреймворку.
4. Перевірка працездатності фреймворку на тестовій сторінці.

Практичне заняття 3. Реалізація скрапінгу за допомогою BS4

1. Структура бібліотеки BeautifulSoup.
2. Основні переваги BeautifulSoup.
3. Типові застосування BeautifulSoup

Практичне заняття 4. Програмний аналіз HTML сторінок

1. Фільтрація компонентів сторінок за тегами, атрибутами та текстовими фрагментами вмістимого тегів.
2. Задання обсягу пошуку.
3. Пошук за ключовими словами із можливістю відсіву надлишкової інформації.
4. Обробка батьківських та дочірніх елементів.

Практичне заняття 5. Збір даних на множинах HTML сторінок та Веб-сайтів

1. Загальна структура скрепування всього Веб-сайту.
2. Уникнення дублюючих посилань.
3. Побудова внутрішніх та зовнішніх посилань Веб-сторінки.
4. Скрапінг за допомогою бібліотеки scrapy та python3.

Практичне заняття 6. Наповнення загальної структури проекту

1. Вибір напрямку, об'єкта та предмета дослідження.

2. Збір інформації про базові напрацювання щодо предмету дослідження та його мало досліджені аспекти.
3. Постановка задачі дослідження. Система базових відношень проекту.
4. DFD пропонованої інформаційної системи.
5. Ієрархічна ER діаграма бази даних інформаційної системи.

Практичне заняття 7. Моделі комп'ютерного зору

1. Метод навчання з учителем у побудові моделей комп'ютерного зору.
2. Згорткові нейронні мережі. Трансформація низькорівневих ознак у моделі реальних об'єктів.
3. Етапи маркування зображень. Підвищення якості маркування.
4. Уточнення візуальних тегів, об'єднання міток, трансформація зображень.

Практичне заняття 8. Семантична розмітка зображень

1. Глибокі нейронні мережі. Згорткові мережі.
2. Основні кроки у класифікації зображень.
3. Спеціалізована геоінформаційна система моделей класифікації зображень.
4. Типи бізнес-процесів, що маркуються.

Практичне заняття 9. Структура фреймворку TensorFlow

1. Еволюція TensorFlow. Концепція обчислювального графа.
2. Багатовимірні масиви як тензори. Операції над тензорами.
3. Послідовності операцій та їх графи. Сесії.
4. Оптимізатори. Моделі. Галузі використання TensorFlow..

Практичне заняття 10. Розпізнавання зображень за допомогою фреймворку

TensorFlow

1. Навчальний та тестовий набори зображень.
2. Формування наборів за допомогою фільтрованого скрапінгу. Кількість епох.
3. Частка валідаційного набору. Dataset - набір зображень.
4. Конвеєри обробки даних із аугментацією. Оцінка точності моделі.

6. Самостійна робота

Виконання індивідуального завдання полягає у побудові проекту інформаційної системи «Система обробки завантаженої інформації» відповідно до обраного напрямку.

1. Кожен студент вибирає напрямок за яким завантажуватиме інформацію.
2. На основі специфіки обраного напрямку, студент проводить проектування та розроблення, які включають:

- Аналіз потреб клієнтів у обраній галузі
- Розроблення проекту інформаційної системи щодо забезпечення потреб клієнтів
- Реалізація розробленого проекту засобами Python, scikit-learn, TensorFlow та аналогічним програмним забезпеченням

3. Студент повинен підготувати звіт, який містить:

- Опис проблем вибраної тематики
- Опис проекту інформаційної системи
- Документування роботи інформаційної системи за допомогою скрін-шотів

Роботу необхідно здати викладачу у вигляді письмового звіту та презентації розробленого додатку. Захист роботи відбувається у формі усної презентації з демонстрацією функціоналу додатку.

7. Організація та проведення тренінгу з дисципліни " Збір та обробка даних "

Цей тренінг охоплює базові аспекти розроблення інформаційних систем для завантаження інформації із глобальних мереж. Учасники отримають досвід роботи із прикладною бібліотекою засобів скрапінгу Beautiful Soap 4.

Мета тренінгу:

Забезпечити учасників теоретичними знаннями та практичними навичками в галузі скрапінгу інформації із Веб-сайтів.

Перелік задач для тренінгу:

1. Вибір напрямку пошуку.
2. Сканування глобальної мережі для виявлення тематичних Веб-сайтів.
2. Вибір головного Веб-сайту для завантаження та класу css тегу, що містить потрібні зображення.
3. Вибір програмного забезпечення завантаження зображень.
4. Завантаження 1000 зображень із обраного джерела
5. Демонстрація скріншоту каталогу із завантаженими зображеннями

Порядок проведення тренінгу:

Вступна частина проводиться з метою ознайомлення студентів із запропонованими завданнями тренінгу.

Організаційна частина полягає у створенні робочого настрою у колективі студентів.

Практична частина реалізується шляхом виконання одного вибраного завдання тренінгу.

Підведення підсумків. Обговорення результатів виконаних завдань. Обмін думками з питань, що виносились на тренінг.

8. Методи навчання.

У навчальному процесі застосовуються: лекції, в тому числі з використання мультимедійного проєктора та інших ТЗН; лабораторні роботи, індивідуальні заняття; самостійна робота студентів, робота в Інтернет.

9. Засоби оцінювання та методи демонстрування результатів навчання

- У процесі вивчення дисципліни «Збір та обробка даних» використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:
 - поточне опитування;
 - залікове модульне тестування та опитування;
 - оцінювання лабораторних робіт;
 - оцінювання тренінгів;
 - оцінювання результатів самостійної роботи.

10. Критерії, форми поточного та підсумкового контролю

Підсумковий бал (за 100-бальною шкалою) з дисципліни " Збір та обробка даних " визначається як середньозважена величина, в залежності від питомої ваги кожної складової залікового кредиту:

| Модуль 1 | | Модуль 2 | | Модуль 3 | Модуль 4 | Модуль 5 |
|--|---|--|--|---|--|--|
| 10% | 10% | 10% | 10% | 5% | 15% | 40% |
| Поточне оцінювання | Модульний контроль 1 | Поточне оцінювання | Модульний контроль 2 | Тренінги | Самостійна робота | Екзамен |
| Оцінка за даний модуль визначається як середнє арифметичне за роботу на практичних заняттях №1-5 | Підсумкова письмова робота за темами №1-5 | Оцінка за даний модуль визначається як середнє арифметичне за роботу на практичних заняттях №6-10. | Підсумкова письмова робота за темами №6-10 | Визначається як середнє арифметичне за виконання завдань за темами №1-5 тренінгу. | Оцінка за виконаний і представлений звіт із оброблених отриманих даних | 1. Теоретична частина: 2 запитання по 20 балів (40 балів) 2. Практична частина 2 задачі по 30 балів кожна (60 балів). |

11. Інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна

| № | Найменування | Номер теми |
|----|--|------------|
| 1. | Персональний комп'ютер | 1-10 |
| 2. | Електронний варіант презентацій | 1-10 |
| 4. | Використання програмного інструментарію Python 3, - пакет Beatiful Soup, - Фреймворк TensorFlow. | 1-10 |

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. WEB SCRAPING ЗА ДОПОМОГОЮ SCRAPY І PYTHON 3. <https://www.8host.com/blog/web-scraping-s-pomoshhyu-scrapy-i-python-3/>
2. Mitchell R. Website scraping with Python. 2018 / <https://edu.anarcho-copy.org/Programming%20Languages/Python/Web%20Scraping%20with%20Python,%202nd%20Edition.pdf>
3. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition. O'Reilly Media. 2019
4. Francois Chollet. Deep Learning with Python, 2nd Edition, Francois Chollet. Manning, 2021.
5. Thushan Ganegedara. TensorFlow in Action. Manning, 2022
6. Andrew Trask. Grokking Deep Learning. MANNING, 2019. // <https://edu.anarcho-copy.org/Algorithm/grokking-deep-learning.pdf>
7. Young A., Meck B., Cantelon M. Node.js in action. 2017./ <https://dokumen.pub/nodejs-in-action-2nbsped-1617292575-9781617292576.html>
8. Powers S. Learning Node.js. 2012. / <https://pepa.holla.cz/wp-content/uploads/2016/12/Learning-Node.pdf>
9. Banks A., Porcello E. Learning React: Functional Web Development with React and Redux. 2017 / <https://morioh.com/p/6a42a1b8bb58>