

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

Затверджую

В. о. декана факультету комп'ютерних
інформаційних технологій



Ігор Якименко

" _____ 2023р.

Затверджую

В.о. проєктора з науково-педагогічної
роботи



Віктор ОСТРОВЕРХОВ

" _____ 2023 р..

Затверджую

Директор ІНІІОТ



Святослав Питель

" _____ 2023р.

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	Тре- нінг, КПІЗ	СРС	Разом	Екзамен,
Денна	I	I	30	15	5	4	96	150	I
Заочна	I	II	8	4	-	-	138	150	II

31.08.2023 р.

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 №10 від 23.06.2023 р.).

Робочу програму склав: професор кафедри економічної кібернетики та інформатики
ПАСІЧНИК Роман Мирославович

Робоча програма затверджена на засіданні кафедри економічної кібернетики та інформатики, протокол № 1 від 28.08.2023 р.

Завідувач кафедри



проф. БУЯК Леся Михайлівна

Розглянуто та схвалено групою забезпечення спеціальності системний аналіз, протокол №1 від 30.08.2023 р.

Голова ГЗС



проф. ПАСІЧНИК Роман Мирославович

Гарант ОПП



доц. БАБАЛА Людмила Василівна

СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Збір та обробка даних»

1. Опис дисципліни «Збір та обробка даних»

Дисципліна – Збір та обробка даних	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни
Кількість кредитів ECTS 5	Галузь знань – 12 «Інформаційні технології»	Нормативна дисципліна, мова навчання - <i>українська</i>
Кількість залікових модулів - 4	Спеціальність – 124 «Системний аналіз»,	<i>Денна:</i> Рік підготовки:1 Семестр – 1 <i>Заочна:</i> Рік підготовки:1 Семестр – 2
Кількість змістових модулів - 2	Ступінь вищої освіти – магістр	<i>Денна:</i> лекції – 30.; практ - 15 <i>Заочна:</i> лекції – 8 год.; лабор.- 4 год
Загальна кількість годин - - 150		Самостійна робота: 92 год., тренінг(КПЗ) – 8 год. Індивідуальна робота : 5 год.
Тижневих годин: 10 год., з них аудиторних – 3 год		Вид підсумкового контролю – <i>екзамен</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 методологією розроблення програмного забезпечення, організації процесів такого розроблення, загальної архітектури програмного забезпечення, архітектурним особливостям сучасних середовищ розроблення Веб-інформаційних систем.

Передбачено вивчення базових підходів до організації роботи із фреймворками NODE, Express, NoSQL, базою даних MongoDB, методів побудови CRUD систем, основами реалізації скрепінгу у глобальній мережі. Вивчаються також методи побудови прогнозних моделей, методи кластеризації для підвищення точності моделей а також методи організації програмного інтерфейсу за допомогою мови програмування javascript.

На практичних заняттях за допомогою цих засобів створюються шаблони веб-проектів та апробуються методи автоматизованого пошуку інформації у глобальній мережі.

Метою викладання даної навчальної дисципліни є формування системи теоретичних знань про архітектуру спеціалізованих програмних систем та методи автоматизованого видобування інформації із глобальної мережі, комплекс умінь із формування програмних систем та автоматизації їх наповнення актуальною інформацією із Вебу.

Для досягнення мети поставлені такі основні **завдання**:

сформувати понятійний апарат і розуміння взаємозв'язку між основними засобами реалізації програмної структури інформаційних систем;

отримати базові знання щодо основних методів побудови архітектури програмних систем та методів скрепінгу;

отримати практичні навички щодо побудови інформаційних систем та автоматизації пошуку їх наповнення.

Об'єктом навчальної дисципліни є процеси структурування інформаційних систем.

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

2.2. Завдання вивчення дисципліни

В результаті вивчення курсу " Збір та обробка даних " студенти повинні:

- знати основні методики розроблення програмного забезпечення та методи їх практичного застосування, характеристики зрілості процесів виробництва програмного забезпечення, основні поняття архітектури сучасних програмних систем, концепції модель-подання-контроллер, SQL та NoSQL баз. даних, програмних засобів тестування Веб-сервісів, розроблення структури моделі проекту та типових маршрутизаторів;

- вміти використовувати методики розроблення програмного забезпечення, моделі зрілості процесів виробництва програмного забезпечення, об'єктно-реляційні подання SQL та NoSQL баз даних, генерувати динамічний контент веб-сторінок.

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

1. Здатність ідентифікувати, класифікувати та описувати проектні завдання, знаходити раціональні методи й підходи до їх розв'язання.
2. Здатність проектувати системи, включаючи проведення моделювання їх архітектури, поведінки та процесів функціонування окремих підсистем і модулів.

2.4. Передумови для вивчення дисципліни.

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.

2.5. Результати навчання:

РН7. Розробляти інтелектуальні системи в умовах слабо структурованих даних

2.6. Завдання лекційних занять

Мета проведення лекцій полягає у тому, щоб ознайомити студентів із головними питаннями курсу " Збір та обробка даних ".

Завдання проведення лекцій полягає у:

- викладенні студентам у відповідності з програмою та робочим планом основних питань курсу " Збір та обробка даних ";
- сформуванні у студентів цілісної системи теоретичних знань з курсу " Збір та обробка даних ".

2.6. Завдання проведення практичних занять

Мета проведення практичних занять полягає у тому, щоб виробити у студентів практичні навички використання теоретичного матеріалу. Завдання проведення практичних занять у глибокому засвоєнні та закріпленні теоретичних знань, одержаних на лекціях.

Зміст вміння, що забезпечується вищеописаними компетентностями, – освоєння суті та вироблення навиків використання основних методів реалізації архітектури інформаційних систем та засобів автоматизації їх наповнення із глобальної мережі.

3. Програма дисципліни " Збір та обробка даних "

Змістовий модуль 1 – Аналіз архітектури проекту

Тема 1. Основи роботи з NODE

Загальна характеристика програмного середовища NODE. Встановлення програмного середовища. Перевірка працездатності встановленого продукту. Взаємодія NODE із браузерами користувачів.

Тема 2. Основи роботи із фреймворком Express

Загальна характеристика середовища розробки Express. Перевірка працездатності менеджера пакетів NPM. Файл package.json. Формування залежностей проекту. Встановлення середовища Express. Імпорт системних модулів а також модулів проміжного програмного забезпечення та створення

обробників маршрутів в Express. Управління статичними файлами.

Тема 3. Основи роботи із базою даних MongoDB

Завантаження відкритого інсталлятора MongoDB (категорія community). Інсталяція MongoDB. Запуск MongoDB та підключення її робочого каталогу. Встановлення драйвера MongoDB в середовищі Express. Тестове підключення MongoDB в середовищі Express.

Тема 4. CRUD система на основі Express

Загальна структура найпростішої CRUD системи. Загальна структура домашньої сторінки. Структура базового відношення проекту. Подання базових сторінок проекту в шаблонізаторі EJS. Шаблон подань проекту. Засоби запуску CRUD системи.

Тема 5. Формування 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.

Змістовий модуль 2 – Візуалізація інформації Веб-системи

Тема 6. Основи реалізації скрепінгу на Python

Базові бібліотеки скрепінгу. Реалізація запитів до Веб-сервера. Обробка помилок при відкритті зовнішніх сторінок через Веб-сервери. Сканування веб-сторінок та аналіз вмісту тегів. Структурування процесу за допомогою підпрограм.

Тема 7. 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 8. 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.

Тема 9. Використання API

Загальна структура API. Формування роутера для запуску функцій обробки запитів користувача. Основи побудови контролерів для обробки запитів користувача. Побудова сервера для діалогу із користувачем API. Робота з Google API.

Тема 10. Збереження результатів скрепінгу

Завантаження файлів за видобутими посиланнями. Використання модуля os.

Завантаження CSV файлів. Використання драйверів баз даних. Створення баз даних із відповідним кодуванням. Типи кодування текстових файлів. Читання csv та pdf файлів. Підходи до читання docx файлів. Уникання пасток скрепінгу.

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

	Лекції	Прак- тичні занят- тя	Самост робота	Інди від робо та	Контрол заходи
Тема 1. Основи роботи з NODE	3	2	9		поточне опит.
Тема 2. Основи роботи із фреймворком Express	3	1	9	1	поточне опит.
Тема 3. Основи роботи із базою даних MongoDB	3	2	9		поточне опит.
Тема 4. CRUD система на основі Express	3	1	9	1	поточне опит.
Тема 5. Формування загальної структури проекту	3	2	9		модульн контр
Тема 6. Основи реалізації скрепінгу на Python	3	1	9		поточне опит.
Тема 7. Програмний аналіз HTML сторінок	3	2	9	1	поточне опит.
Тема 8. Збір даних на множинах HTML сторінок та Веб-сайтів	3	1	9		поточне опит.
Тема 9. Використання API	3	2	10		модульн контр
Тема 10. Збереження результатів скрепінгу	3	1	10	1	поточне опит.
Разом	30	15	92	5	ректорс. контр

заочна

	Лекції	Практичні заняття	Самост робота
Тема 1. Основи роботи з NODE	-	1	9
Тема 2. Основи роботи із фреймворком Express	1		9
Тема 3. Основи роботи із базою даних MongoDB	1		9
Тема 4. CRUD система на основі Express	1		9
Тема 5. Формування загальної структури проекту		1	9
Тема 6. Основи реалізації скрепінгу на Python	1		9
Тема 7. Програмний аналіз HTML сторінок	1	1	9
Тема 8. Збір даних на множинах HTML сторінок та Веб-сайтів	1	1	9
Тема 9. Використання API	1		10
Тема 10. Збереження результатів скрепінгу	1		10
Разом	8	4	138

5. Тематика практичних занять

Практичне заняття 1. Тема. Основи роботи з NODE

1. Загальна характеристика програмного середовища NODE.
2. Встановлення програмного середовища.
3. Перевірка працездатності встановленого продукту.
4. Взаємодія NODE із браузерами користувачів.

Практичне заняття 2. Тема: Основи роботи із фреймворком Express

1. Формування залежностей проекту.
2. Встановлення середовища Express.
3. Імпорт системних модулів
4. Управління статичними файлами

Практичне заняття 3. Основи роботи із базою даних MongoDB

1. Інсталяція MongoDB.
2. Запуск MongoDB та підключення її робочого каталогу.
3. Встановлення драйвера MongoDB в середовищі Express.
4. Тестове підключення MongoDB в середовищі Express.

Практичне заняття 4. CRUD система на основі Express

1. Загальна структура найпростішої CRUD системи.
2. Структура базового відношення проекту.
3. Подання базових сторінок проекту в шаблонізаторі EJS.
4. Шаблон подань проекту.

Практичне заняття 5. Формування загальної структури проекту

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

Практичне заняття 6. Основи реалізації скрепінгу на Python

1. Базові бібліотеки скрапінгу.
2. Реалізація запитів до Веб-сервера.
3. Обробка помилок при відкритті зовнішніх сторінок через Веб-сервери.
4. Сканування веб-сторінок та аналіз вмісту тегів.

Практичне заняття 7. Програмний аналіз HTML сторінок

1. Фільтрація компонентів сторінок за тегами.
2. Пошук за ключовими словами із відсівом надлишкової інформації.
3. Обробка батьківських та дочірніх елементів..
4. Робота з однорівневими елементами.

Практичне заняття 8. Збір даних на множинах HTML сторінок та Веб-сайтів

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

Практичне заняття 9. Використання API

1. Приклади оголошення масивів..
2. Способи вводу масиву із клавіатури..
3. Багатовимірні масиви.
4. Способи використання рядків.

Практичне заняття 10. Збереження результатів скрапінгу

1. Завантаження файлів за видобутими посиланнями.
2. Використання драйверів баз даних.
3. Створення баз даних із відповідним кодуванням.
4. Читання csv та pdf файлів.

6. Тематика лабораторних занять

Лабораторне заняття 1. Тема. Основи роботи з NODE – 3 год.

1. Загальна характеристика програмного середовища NODE.
2. Встановлення програмного середовища.
3. Перевірка працездатності встановленого продукту.
4. Взаємодія NODE із браузерами користувачів.

Практичне заняття 2. Тема: Основи роботи із фреймворком Express

1. Формування залежностей проекту.
2. Встановлення середовища Express.
3. Імпорт системних модулів
4. Управління статичними файлами

Практичне заняття 3. Основи роботи із базою даних MongoDB

1. Інсталяція MongoDB.
2. Запуск MongoDB та підключення її робочого каталогу.
3. Встановлення драйвера MongoDB в середовищі Express.
4. Тестове підключення MongoDB в середовищі Express.

Практичне заняття 4. CRUD система на основі Express

1. Загальна структура найпростішої CRUD системи.
2. Структура базового відношення проекту.

3. Подання базових сторінок проекту в шаблонізаторі EJS.
4. Шаблон подань проекту.

Практичне заняття 5. Формування загальної структури проекту

1. Збір інформації про базові напрацювання щодо предмету дослідження та його мало досліджені аспекти.
2. Постановка задачі дослідження.
3. DFD запропонованої інформаційної системи.
4. Ієрархічна ER діаграма бази даних інформаційної системи.

7. Комплексне практичне індивідуальне завдання.

Виконання індивідуального завдання полягає у побудові проекту інформаційної системи відповідно до обраної теми.

Комплексне практичне індивідуальне завдання (КПЗ) виконується самостійно при консультуванні викладачем протягом вивчення дисципліни у відповідності до графіка навчального процесу.

Із виконується з метою закріплення, поглиблення й узагальнення знань, одержаних студентами за час навчання, та придбання практичних навичок їх застосування при розробці. При розробці студент повинен використовувати знання та вміння, набуті на заняттях і в ході самостійної роботи. Прийняті технічні та технологічні рішення повинні бути обґрунтовані.

Для досягнення мети студент повинен виконати всі передбачені технологією етапи для кожного виду робіт.

8. Самостійна робота

Тематика	К-сть годин
Тема 1. Основи роботи з NODE Вивчення лекційного матеріалу. Самостійне поглиблено вивчення питання «Загальна характеристика програмного середовища NODE.»	9
Тема 2. Основи роботи із фреймворком Express Вивчення лекційного матеріалу, підготовка до практичного заняття. Вибір тематики для проекту інформаційної системи у вибраній предметній галузі.	9
Тема 3. Основи роботи із базою даних MongoDB Вивчення лекційного матеріалу, підготовка до практичного заняття. Самостійне поглиблено вивчення питань "Тестове підключення MongoDB в середовищі Express." Виконання індивідуального завдання "Запуск MongoDB та підключення її робочого каталогу "	9
Тема 4. CRUD система на основі Express Вивчення лекційного матеріалу, підготовка до практичного заняття. Виконання <i>індивідуального завдання</i> "Подання базових сторінок проекту в шаблонізаторі EJS "	9
Тема 5. Формування загальної структури проекту Вивчення лекційного матеріалу, підготовка до практичного заняття. Самостійне вивчення питань: Постановка задачі дослідження . Виконання <i>індивідуального завдання</i> " Ієрархічна ER діаграма бази даних інформаційної системи."	9
Тема 6. Основи реалізації скрепінгу на Python Вивчення лекційного матеріалу, підготовка до практичного заняття.	9

Самостійне вивчення питання "Базові бібліотеки скрепінгу. Реалізація запитів до Веб-сервера", Виконання <i>індивідуального завдання</i> "Сканування веб-сторінок та аналіз вмісту тегів."	
Тема 7. Програмний аналіз HTML сторінок Вивчення лекційного матеріалу, підготовка до практичного заняття. Самостійне вивчення питань " Пошук за ключовими словами із можливістю відсіву надлишкової інформації.". Виконання <i>індивідуального завдання</i> " Фільтрація компонентів сторінок за тегами"	9
Тема 8. Збір даних на множинах HTML сторінок та Веб-сайтів Вивчення лекційного матеріалу, підготовка до практичного заняття. Самостійне вивчення питання "Загальна структура скрепування всього Веб-сайту", Виконання <i>індивідуального завдання</i> "Скрепінг за допомогою бібліотеки scrapy та python3"	9
Тема 9. Використання API Вивчення лекційного матеріалу, підготовка до практичного заняття. Самостійне вивчення питань "Формування роутера для запуску функцій обробки запитів користувача.". Виконання <i>індивідуального завдання</i> "Побудова сервера для діалогу із користувачем API "	10
Тема 10. Збереження результатів скрепінгу Вивчення лекційного матеріалу, підготовка до практичного заняття. Самостійне вивчення питань "Уникання пасток скрепінгу". Виконання <i>індивідуального завдання</i> "Завантаження файлів за видобутими посиланнями. Завантаження CSV файлів "	10
Тренінг	4
Разом:	96

8. Тренінг з дисципліни

Виконання Комплексного практичного індивідуального завдання «Проектування веб додатків». Вибір тематики проєктованого додатку. Основні відношення бази даних. DFD та ER діаграми бази даних додатку. Елементи інтерфейсу додатку

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 результатів КППЗ;
- ректорська контрольна робота;
- екзамен;

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:

Заліковий модуль 1	Заліковий модуль 2	Заліковий модуль 3	Екзамен	Разом
20%	20%	20%	40%	100%
1. Усне опитування під час заняття (5 теми по 10 балів = 50 балів) 2. Письмова робота = 50 балів	1. Усне опитування під час заняття (5 тем по 10 балів = 50 балів) 2. Письмова робота = 50 балів	1. Написання та захист КПЗ = 60 балів. 3. Виконання завдань під час тренінгу = 40 балів	1. 3 запитання по 20 балів = 60 балів 2. Задача = 40 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 Beautiful Soup, Фреймворки Node, Express.	1-10
5	Індивідуальні завдання для самостійного виконання	1-10

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Larry Ullman. Modern JavaScript: Develop and Design. Berkeley, CA. 2012. 625 p. <https://pepa.holla.cz/wp-content/uploads/2016/08/Modern-JavaScript.pdf>
2. Flanagan D. JavaScript: The Definitive Guide: Master the World's Most-Used Programming Language // <https://dokumen.pub/qdownload/javascript-the-definitive-guide-master-the-worlds-most-used-programming-language-9781491951989-1491951982.html>
3. Mario Casciaro. Node.js Design Patterns Kindle Edition. 2014 // <https://github.com/mystroken/JSBooks/blob/master/%5BNode.js%20Design%20Patterns%20Kindle%20Edition%20by%20Mario%20Casciaro%20-%202014%5D.pdf>
4. Angel Stoyanov. Java Script. Templates. Use of a standard object modeling language. [/https://www.scribd.com/document/205710294/Angel-Stoyanov?language_settings_changed=English](https://www.scribd.com/document/205710294/Angel-Stoyanov?language_settings_changed=English)

5. Josip Maras. Secrets of the JavaScript Ninja. https://www.scribd.com/book/511817184/Secrets-of-the-JavaScript-Ninja?utm_medium=cpc&utm_source=google_search&utm_campaign=3Q_Google_DSA_NB_RoW&utm_term=&utm_device=c&gclid=Cj0KCQiAlKmeBhCkARIsAHy7WVtBlk5iq_4f9sHVOwh8Y_58-3S5mhr-DMR1jIWxj7LBOLxHj4gzTD8aApb4EALw_wcB
6. Elliot E. Javascript Application Programming. Resilient web architecture with Node, HTML5 and modern JS libraries. 2014/<https://www.pdfdrive.com/programming-javascript-applications-robust-web-architecture-with-node-html5-and-modern-js-libraries-e177459843.html>.
7. WEB SCRAPING C ПОМОЩЬЮ SCRAPY И PYTHON 3. <https://www.8host.com/blog/web-scraping-s-pomoshhyu-scrapy-i-python-3/>
8. Young A., Meck B., Cantelon M. Node.js in action. 2017./ <https://dokumen.pub/nodejs-in-action-2nbsped-1617292575-9781617292576.html>
9. Powers S. Learning Node.js. 2012. / <https://pepa.holla.cz/wp-content/uploads/2016/12/Learning-Node.pdf>
10. Brown E. Web Development with Node and Express. Full use of the JavaScript stack. 2014 / https://www.vanmeegern.de/fileadmin/user_upload/PDF/Web_Development_with_Node_Express.pdf
11. Learning Node.js. Free e-book. / <https://riptutorial.com/Download/node-js.pdf>
12. Banks A., Porcello E. Learning React: Functional Web Development with React and Redux. 2017 / <https://morioh.com/p/6a42a1b8bb58>
13. Mitchell R. Website scraping with Python. 2018 / <https://edu.anarcho-copy.org/Programming%20Languages/Python/Web%20Scraping%20with%20Python,%202nd%20Edition.pdf>