



Силабус освітнього компонента

Програма навчальної дисципліни



Обробка даних Python

Шифр та назва спеціальності

121 – Інженерія програмного забезпечення
122 – Комп'ютерні науки

Інститут

ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма

Інженерія програмного забезпечення
Комп'ютерні науки та інтелектуальні системи

Кафедра

Програмна інженерія та інтелектуальні технології управління (321)

Рівень освіти

Бакалавр

Тип дисципліни

Спеціальна (фахова), Вибіркова

Семестр

7

Мова викладання

Українська, англійська

Викладачі, розробники



Прізвище Ім'я По батькові

Svitlana.Kovalenko@khp.edu.ua

канд.техн.наук, доцент, доцент кафедри програмної інженерії та інтелектуальних технологій управління

Google Scholar:

<https://scholar.google.com/citations?user=jeD1w74AAAAJ&hl>

ORCID: <https://orcid.org/0000-0001-6770-6778>

Scopus: <https://www.scopus.com/authid/detail.uri?authorId=57212035934>

Web of Science: <https://www.webofscience.com/wos/author/record/F-8252-2017>

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Завданням дисципліни є набуття знань, умінь та навичок, необхідних для збору, попередньої обробки, аналізу, візуалізації, валідації та очищення даних різних форматів та типів, роботи з API за допомогою вбудованих та сторонніх бібліотек мови програмування Python, а також в використанні даних для прийняття обґрунтованих рішень.

Мета та цілі дисципліни

Формування у студентів теоретичних та практичних знань щодо ефективної роботи з даними з використанням мови програмування Python, розуміння методів обробки, аналізу та візуалізації даних, а також використання даних для прийняття обґрунтованих рішень при розробці програмного забезпечення.

Формат занять

Лекції, лабораторні роботи, самостійна робота, консультації. Підсумковий контроль – залік.

Компетентності

- 121-K01. Здатність до абстрактного мислення, аналізу та синтезу.
- 121-K02. Здатність застосовувати знання у практичних ситуаціях.
- 121-K05. Здатність вчитися і оволодівати сучасними знаннями.
- 121-K06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- 121-K13. Здатність ідентифікувати, класифікувати та формулювати вимоги до програмного забезпечення.
- 121-K15. Здатність розробляти архітектури, модулі та компоненти програмних систем.
- 121-K19. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.
- 121-K22. Здатність накопичувати, обробляти та систематизувати професійні знання щодо створення і супроводження програмного забезпечення та визнання важливості навчання протягом всього життя.
- 121-K23. Здатність реалізовувати фази та ітерації життєвого циклу програмних систем та інформаційних технологій на основі відповідних моделей і підходів розробки програмного забезпечення.
- 121-K25. Здатність обґрунтовано обирати та освоювати інструментарій з розробки та супроводження програмного забезпечення.
- 121-K26. Здатність до алгоритмічного та логічного мислення.
- 122-ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.
- 122-ЗК2. Здатність застосовувати знання у практичних ситуаціях.
- 122-ЗК6. Здатність вчитися й оволодівати сучасними знаннями.
- 122-ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- 122-ЗК8. Здатність генерувати нові ідеї (креативність).
- 122-ЗК11. Здатність приймати обґрунтовані рішення.
- 122-СК8. Здатність проектувати та розробляти програмне забезпечення із застосуванням різних парадигм програмування: узагальненого, об'єктно-орієнтованого, функціонального, логічного, з відповідними моделями, методами й алгоритмами обчислень, структурами даних і механізмами управління.
- 122-СК9. Здатність реалізувати багаторівневу обчислювальну модель на основі архітектури клієнт-сервер, включаючи бази даних, знань і сховища даних, виконувати розподілену обробку великих наборів даних на кластерах стандартних серверів для забезпечення обчислювальних потреб користувачів, у тому числі на хмарних сервісах.
- 122-СК10. Здатність застосовувати методології, технології та інструментальні засоби для управління процесами життєвого циклу інформаційних і програмних систем, продуктів і сервісів інформаційних технологій відповідно до вимог замовника.
- 122-СК12. Здатність забезпечити організацію обчислювальних процесів в інформаційних системах різного призначення з урахуванням архітектури, конфігурування, показників результативності функціонування операційних систем і системного програмного забезпечення.

Результати навчання

- 121-ПР01. Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки.
- 121-ПР02. Знати кодекс професійної етики, розуміти соціальну значимість та культурні аспекти інженерії програмного забезпечення і дотримуватись їх в професійній діяльності.
- 121-ПР08. Вміти розробляти людино-машинний інтерфейс.
- 121-ПР12. Застосовувати на практиці ефективні підходи щодо проектування програмного забезпечення.
- 121-ПР13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.
- 121-ПР15. Мотивовано обирати мови програмування та технології розробки для розв'язання завдань створення і супроводження програмного забезпечення.
- 121-ПР18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних.
- 121-ПР23. Вміти документувати та презентувати результати розробки програмного забезпечення.

122-ПР1. Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.

122-ПР4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.

122-ПР10. Використовувати інструментальні засоби розробки клієнт-серверних застосувань, проектувати концептуальні, логічні та фізичні моделі баз даних, розробляти та оптимізувати запити до них, створювати розподілені бази даних, сховища та вітрини даних, бази знань, у тому числі на хмарних сервісах, із застосуванням мов веб-програмування.

122-ПР11. Володіти навичками управління життєвим циклом програмного забезпечення, продуктів і сервісів інформаційних технологій відповідно до вимог і обмежень замовника, вміти розробляти проектну документацію (техніко-економічне обґрунтування, технічне завдання, бізнес-план, угоду, договір, контракт).

122-ПР20. Розробляти архітектуру програмних систем та їх окремих компонент при побудові інтелектуальних систем управління у різних галузях, а також управляти процесами життєвого циклу програмного забезпечення інтелектуальних систем управління.

Обсяг дисципліни

Загальний обсяг дисципліни 120 год. (4 кредитів ECTS): лекції – 16 год., лабораторні роботи – 32 год., самостійна робота – 72 год.

Передумови вивчення дисципліни (пререквізити)

Вивчення цієї дисципліни безпосередньо спирається на: "Об'єктно-орієнтоване програмування", "Основи програмування", "Алгоритмізація та програмування".

Особливості дисципліни, методи та технології навчання

Методи викладання та навчання:

інтерактивні лекції з презентаціями, дискусії, лабораторні заняття, командна робота, кейс-метод, метод зворотного зв'язку з боку студентів, проблемне навчання.

Форми оцінювання:

письмові індивідуальні завдання до лабораторних робіт (CAS), оцінювання знань на лабораторних заняттях (CAS), експрес-опитування (CAS), онлайн-тести (CAS), підсумковий/семестровий контроль у формі заліку, відповідно до графіку навчального процесу (FAS).

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Основні поняття обробки даних. Основи синтаксису Python. Середовища розробки.

Поняття даних, основні завдання обробки даних. Класифікація та загальний огляд етапів та методів обробки даних. Встановлення Python. Середовища розробки. Введення та виведення даних. Змінні та типи даних : огляд типів даних в Python, операції та функції, що виконуються над різними типами даних. Умовні конструкції та цикли. Функції: визначення та виклик функцій, параметри та аргументи функцій, повернення значень з функцій. Робота зі списками та рядками. Обробка виключень. Класи. Використання вбудованих та сторонніх модулів.

Тема 2. Jupyter notebook та бібліотека NumPy

Встановлення та основи роботи з Jupyter notebook. Google Colaboratory. Історія створення NumPy. Масиви ndarray. Порівняння масивів NumPy та списків Python. Типи даних NumPy. Створення масивів. Операції над масивами. Універсальні функції. Конкатенація та розділення масивів NumPy. Broadcasting. Порівняння, маски та булева логіка. Fancy Indexing.

Тема 3. Візуалізація даних.

Бібліотеки для візуалізації даних: Matplotlib, Seaborn, Plotly. Інтерфейси Matplotlib. Встановлення стилів. Прості лінійні графіки. Налаштування графіків. Типи графіків. Візуалізація похибок. Створення декількох графіків на одному рисунку.

Тема 4. Обробка даних за допомогою Pandas

Об'єкти Pandas. Властивості Series та DataFrame. Побудова об'єктів Pandas. Об'єкт Index. Індикація та вибір даних. Операції з даними. Збереження індексу, Універсальні функції та вирівнювання індексу. Операції між DataFrame та Series. Маніпуляції з даними. Групування даних. Обробка відсутніх даних. Ієрархічна індикація. Поєднання наборів даних: concat(), append(), merge(), join(). Зведені таблиці. Візуалізація в Pandas.

Тема 5. Робота з текстом.

Обробка тексту методами стандартної бібліотеки Python. Бібліотека re та регулярні вирази. Шаблони і класи символів. Квантифікатори. Групи. Обробка рядків в Pandas. Векторизація рядкових операторів. Методи Pandas, що використовують регулярні вирази

Тема 6. Робота з графікою. PIL та opencv-python

Встановлення бібліотек. Представлення зображень в бібліотеках. Колірні моделі. Формати графічних даних. Основні маніпуляції з графічними даними: зміна розмірів, обрізання, транспонування, поворот, вставка одного зображення в інше, створення «водяних знаків». Визначення об'єктів на зображенні. Pytesseract та розпізнавання тексту.

Тема 7. Збір даних. Бібліотеки requests та BeautifulSoup

Бібліотека requests. GET та POST -запити. Передача параметрів, обробка відповіді сервера. JSON. Встановлення та імпорт BeautifulSoup. Завантаження веб-сторінки. Розбір HTML-сторінки. Вибір елементів та отримання даних. Навігація по сторінці. Обробка динамічного контенту. Selenium. Інтеграція з Pandas. Збереження даних.

Теми практичних занять

Практичні заняття в рамках дисципліни не передбачені.

Теми лабораторних робіт

Тема 1. Робота в середовищі Jupyter Notebook з обробки та візуалізації даних.

Тема 2. Робота з даними засобами Pandas.

Тема 3. Об'єднання та обробка даних, що знаходяться в різних джерелах.

Тема 4. Обробка текстових даних засобами Pandas.

Тема 5. Обробка цифрових зображень.

Тема 6. Створення веб-застосунку для збору та збереження даних з використанням сторонніх API.

Самостійна робота

Студентам рекомендуються додаткові матеріали (відео, статті) для самостійного вивчення та опрацювання.

Література та навчальні матеріали

Основна література

1. Wes McKinney. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, 3rd edition, 2022, 540 p.
2. Jake VanderPlas. Python Data Science Handbook, 2d edition, O'Reilly Media, 2022, 550 p.
3. Robert Johansson. Numerical Python. Scientific Computing and Data. Science Applications with Numpy, SciPy and Matplotlib. Second Edition, 2019, 685 p.
4. Ravishankar Chityala, Sridevi Pudipeddi. Image Processing and Acquisition using Python. CRC Press, 2021, 452 p.
5. Al Sweigart. Automate the Boring Stuff with Python, 2nd edition: Practical Programming for Total Beginners, No Starch Press, 2020, 901 p.

Додаткова література

1. Sandipan Dey. Hands-On Image Processing with Python: Expert techniques for advanced image analysis and effective interpretation of image data. Packt, 2018, 492 p.
2. Richard L. Halterman. Fundamentals of Python Programming, 2019, 669 p.
3. John Canning, Alan Broder, Robert Lafore. Data Structures & Algorithms in Python, Pearson Education, Inc., 2023, 1050 p.
4. Fabio Nelli. Python data analytics : with Pandas, NumPy, and Matplotlib, Apress, New York, NY, 2018, 555 p.
5. Takatomo Honda Flask Web Development from Scratch: Introduction to Developing Web Applications with Python, 2019, 137 p.

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

100% підсумкової оцінки складаються з результатів оцінювання у вигляді заліку (40%) та поточного оцінювання (60%):

- 6 лабораторних робіт (по 7%);
- 2 контрольні роботи (по 9%).

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

Погодження

Силабус погоджено

08.06.2023

Завідувач кафедри
Ігор ГАМАЮН

08.06.2023

Гаранти ОП
Андрій КОПП
Юлія ЛІТВІНОВА