

Циклова комісія економічних, математичних дисциплін, менеджменту та туризму

Введення в інженерію великих даних

Викладач: Борисовська Юлія Олександрівна

Опис дисципліни

Дисципліна “Введення в інженерію великих даних” спрямована на оволодіння основними поняттями в галузі великих даних, ознайомлення з принципами організації сховищ великих даних, із базовими алгоритмами збереження та пошуку у сховищі великих даних, ознайомлення із переліком задач, що відносяться до організації роботи з великими даними, та способами рішення таких задач, формування навичок з реалізації сховища великих даних, організації процесів роботи зі сховищем, типового аналізу великих даних.

Мета курсу: є ознайомлення з галуззю великих даних (Big Data), пов'язаних даних (Linked Data) та з інструментарієм, що використовується для створення, зберігання, публікації, пошуку, аналізу великих даних.

Предмет вивчення: теоретичні та практичні основи оброблення великих даних, загальні методи оброблення великих даних, можливості мов програмування Python та R для аналізу та візуалізації даних

У результаті вивчення дисципліни “Введення до теорії хмарних обчислень” студент повинен оволодіти наступними **КОМПЕТЕНТНОСТЯМИ**:

- здатність застосовувати фундаментальні та міждисциплінарні знання для успішного розв’язання завдань інженерії програмного забезпечення;
- здатність брати участь у проєктуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування;
- здатність розробляти архітектури, модулі та компоненти програмного забезпечення.

Результати навчання:

- вміти застосовувати методи обчислення та структури даних для вирішення задач аналізу та синтезу алгоритмів;
- знати основні стандарти, принципи та нотації проектування програмного забезпечення та вміти застосовувати їх на практиці для моделювання структури, поведінки та внутрішніх процесів програмної системи.
- вміти застосовувати знання технічних характеристик, конструктивних особливостей, призначення і правил експлуатації програмно-технічних засобів комп'ютерних систем та мереж для вирішення технічних задач спеціальності.

Програма дисципліни:

Змістовий модуль 1 Джерела відкритих даних

1. Дослідження джерел відкритих даних. Завантаження датасету та збереження даних в форматі csv
2. Аналіз та візуалізація даних у Python
3. Кореляційний аналіз у Python

Змістовий модуль 2 Аналіз та візуалізація даних

1. Побудова лінійної регресії в Python
2. Аналіз та візуалізація даних в R
3. Розподілені обчислення даних з використанням Spark-кластера та мови R