



Силабус дисципліни «Технології обробки великих даних»



Викладач: *Калініна Ірина Олександрівна*

д-р техн. наук, доцент, в. о. професора кафедри інтелектуальних інформаційних систем

Очікувані результати навчання

В результаті вивчення дисципліни студент

має знати:

- методи та засоби агрегації та інтегрування великих даних різних типів;
- методи та засоби проектування сховищ для зберігання великих даних;
- методи, засоби та інструменти інтелектуальної обробки великих даних;
- методи та засоби візуалізації великих даних;
- методи та засоби отримання знання з великих даних.

має вміти:

- використовувати інструментальні засоби інтеграції великих даних різних типів;
- правильно обирати та використовувати методи та алгоритми для фільтрації, валідації та зберігання великих даних;
- правильно обирати найбільш інформативні способи візуалізації великих даних;
- отримувати знання шляхом аналізу великих даних;
- проектувати та розробляти прикладні інформаційні програми для аналізу великих даних та прийняття рішень на основі отриманої інформації.

Пререквізити

Обсяг дисципліни: 4 кредити ECTS (16 годин лекцій, 32 годин практичних робіт).

Мета: теоретична та практична підготовка майбутніх магістрів в області інформаційних систем і технологій до роботи з великими даними. Знання, отримані в результаті освоєння дисципліни, допоможуть при зборі та аналізі величезних обсягів структурованої або неструктурованої інформації, при розробці моделей даних і отриманні нових знань. Все це необхідно випускнику, що освоїв програму магістра, для вирішення різних завдань практичної і науково-дослідної діяльності.

Оригінальність навчальної дисципліни:

Авторський курс

Зміст дисципліни

Тема 1. Введення в аналіз великих даних. Основні поняття та визначення. Характеристики великих даних. Типи великих даних. Застосування в економіці, бізнесі, сільському господарстві, промисловості. Приклади використання. Великі дані в наукових сферах. Особливості застосування. Вимоги до професії аналітика великих даних

Тема 2. Бізнес мотивація та стимули для переходу до обробки великих даних. Динаміка ринку. Бізнес-архітектури. Управління бізнес-процесами. Інформаційно-комунікаційні технології.

Тема 3. Корпоративні технології та Business Intelligence для великих даних.

Тема 4. Концепції зберігання великих даних. Кластери. Файлові системи та розподілені файлові системи. NoSQL, шардінг, реплікація, теорема CAP, ACID, BASE.

Тема 5. Концепції обробки великих даних. Паралельна обробка даних. Розподілена обробка даних. MapReduce. Обробка робочих завдань. Кластер. Обробка у пакетному

«Вища математика», «Основи програмування», «Інтелектуальний аналіз даних», «Теорія ймовірностей та математична статистика», «Теорія алгоритмів та структури даних», «Теорія прийняття рішень», «Методи і системи машинного навчання».

Пореквізити

Знання, які студенти набудуть при вивченні курсу «Технології обробки великих даних» будуть необхідними при подальшому навчанні, при проходженні переддипломної практики та підготовці кваліфікаційної роботи, а також у виробничій діяльності з фахової спеціальності.

Семестровий контроль: залік

Оцінювання:

За семестр: 70 балів

За іспит: 30 балів

Види робіт:

Практичні роботи

Технічне забезпечення

Заняття з дисципліни проводяться у дистанційному форматі (*GoogleMeet*, *Zoom* тощо) або у комп'ютерних класах з використанням мов програмування та програмних середовищ: RStudio, R, Python.

Політика щодо дедлайнів

Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку.

Політика щодо академічної доброчесності

Передбачає самостійне виконання лабораторних робіт. Списування під час іспиту (в т. ч. із використанням мобільних пристроїв) заборонено. У разі виявлення плагіату або списування роботи не зараховуються.

режимі. Обробка у режимі реального часу.

Тема 6. Пакетна обробка великих даних за допомогою MapReduce. Приклади використання.

Тема 7. Технології зберігання великих даних. Дискові пристрої для зберігання. Системи зберігання оперативної пам'яті. Приклади використання.

Тема 8. Основні методи аналізу великих даних. Кількісний аналіз. Якісний аналіз. Data Mining. Статистичний аналіз. Машинне навчання. Семантичний аналіз. Візуальний аналіз.

Тема 9. Машинне навчання на великих даних.

Критерії оцінювання лабораторних/практичних/індивідуальних/робіт/ доповідей/проєктів

Максимальна кількість балів – студент з високою якістю самостійно виконав весь обсяг робіт, відповідає на всі питання, пов'язані з виконаними роботами, та робить додаткові розрахунки, які йому пропонує викладач. У викладача немає претензій щодо реалізації та вимог до виконання роботи.

70%-99% від максимальної кількості балів – студент з достатньою якістю виконав всі завдання, але в процесі роботи він робив деякі помилки, які, після вказування на них викладачем, самостійно виправляв. На деякі питання він відповідає з похибкою. Запропоновані викладачем додаткові розрахунки робить з деякою потугою. Не всі вимоги до виконання роботи дотримані.

40%-69% від максимальної кількості балів – студент самостійно виконав всі роботи, але якість реалізації недостатня (помилки при розрахунках, не всі вимоги до роботи дотримані). На питання щодо виконання робіт відповідає не зовсім чітко. Є помилки при відповідях.

1%-39% від максимальної кількості балів – студент самостійно виконав не всі роботи, при цьому якість реалізації недостатня (помилки при розрахунках, не дотримується вимог до оформлення роботи). На питання щодо виконання робіт відповідає не чітко. Є грубі помилки при відповідях.

0 балів – студент не виконав весь обсяг робіт, або виконав з грубими помилками. Він має проблеми з розрахунками, не знає теоретичного матеріалу, програмна реалізація не відповідає поставленим вимогам.

Для зарахування практичної роботи здобувач має змогу надати сертифікат (або

інший документ) про проходження курсів на платформах *Prometheus*, *Coursera* тощо та за умови співпадіння направленості курсів блоку освітньої компоненти.