Національний технічний університет України
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Кафедра системного проєктування

ОБРОБКА НАДВЕЛИКИХ МАСИВІВ ДАНИХ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	***Другий (магістерський)***
Галузь знань	12 Інформаційні технології
Спеціальність	122 Комп'ютерні науки
Освітня програма	Комп'ютерні науки
Статус дисципліни	Нормативна
Форма навчання	Дистанційна/очна
Рік підготовки, семестр	1 курс, осінній
Обсяг дисципліни	Загальний обсяг 4 кредити ECTS (120 годин) Аудиторних занять 54 години: лекції – 36 годин, лабораторні – 18 годин Самостійна робота студентів – 66 годин
Семестровий контроль/ контрольні заходи	Екзамен / модульна контрольна робота/розрахунково-графічна робота
Розклад занять	Лекції: один раз на тиждень відповідно до розкладу занять, Лабораторні заняття: один раз на два тижні відповідно до розкладу занять
Мова викладання	Українська
Інформація про керівника курсу / викладачів	**Лектор**: д.т.н., проф. Рогоза Валерій Станіславович, rosvetnik@gmail.com Практичні / Семінарські: програмою дисципліни не передбачені **Лабораторні**: асистент, Яременко Вадим Сергійович, yaremenko.v.s@gmail.com
Розміщення курсу	*Лекційні матеріали*: конспекти лекцій в системі *PowerPoint* розміщені на Google-диску (адреси вказані в списку лекційних тем в розд.3) та відеолекції в системі *Camtasia* https://do.ipo.kpi.ua/course/view.php?id=2395 *Методичні матеріали до лабораторних робіт* : записані в *Google-диску* , адреси яких повідомляються студентам на першому лабораторному занятті.

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчання та результати навчання

Метою вивчення дисципліни є формування у студентів знань, умінь та навичок у розвитку та використанні **аналітичних методів обробки великих масивів даних**, а саме:

- аналізу характеристик даних, які надходять до комп'ютерної системи, усвідомлення задач, які треба вирішувати з цими даними, та вибору відповідних методів, комп'ютерного середовища та програмного забезпечення з метою ефективного вирішення поставлених задач (тобто діяльність спрямована на вивчення структури та обсягу даних і вибору відповідних засобів розв'язування задачі, а об'єктом діяльності є потоки даних великого обсягу);
- правильної оцінки та аналізу результатів, отриманих внаслідок розв'язування поставлених задач обробки даних, класифікації та відбору найбільш ефективних методів для розв'язання подібних задач у майбутньому, і модифікації існуючих підходів та розробки нових методів і програм на їх основі на базі отриманого досвіду з метою вдосконалення комп'ютерної платформи, яка є у розпорядженні дослідника (тобто діяльність спрямована на аналіз та розробку методів дослідження, а об'єктом діяльності є апаратне та програмне забезпечення).

Рівень компетентності студентів визначається тим, наскільки ефективно вони здатні використовувати набуті теоретичні знання та вміння до розв'язання описаних вище задач в своїй практичній діяльності.

В результаті засвоєння дисципліни студенти мають оволодіти знаннями та вміннями:

- побудови алгоритмів та їх програмної реалізації, призначених до оброблення надвеликих масивів даних;
- використання та вдосконалення існуючих методів, алгоритмів та програм для розв'язання задач аналізу надвеликих масивів даних ;
- вирішення задач аналізу даних в розподілених комп'ютерних середовищах;
- розв'язання задач інтелектуальної обробки надвеликих масивів даних;
- побудови алгоритмів та програм обробки надвеликих масивів даних в середовищі розподілених файлових систем.

При оцінці професіонального рівня студентів, які опанували в повному обсязі дану дисципліну, головна увага приділяється вмінню студентів поєднати теоретичні знання з їх практичним використанням в побудові розподілених інтелектуальних комп'ютерних систем обробки надвеликих масивів даних.

Вивчення даного предмету забезпечує оволодіння студентами наступних **загальних та фахових компетентостей**, а також **програмних результатів навчання:**

Загальні компетентності (ЗК):

- здатність до абстрактного мислення, аналізу та синтезу (ЗК 1);
- здатність до застосування знання у практичних ситуаціях (ЗК 2);
- здатність вчитися й оволодівати сучасними знаннями (ЗК 5);
- здатність генерувати нові ідеї (креативність) (ЗК 7).

Фахові компетентності спеціальності (ФК):

- усвідомлення теоретичних засад комп'ютерних наук (ФК 1);
- здатність використовувати математичних методи для аналізу формалізованих моделей предметної області (ФК 3);
- здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень (ФК 4);
- здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення (ФК 5);
- здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням неявних ресурсів та обмежень (ФК 7);
- здатність розробляти і реалізовувати проекти зі створення програмного забезпечення, у тому числі в непередбачуваних умовах, за нечітких вимог та необхідності застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проектом (ФК 8);
- здатність розробляти та адмініструвати бази даних та знань (ФК 9);
- здатність до проектування та програмної реалізації методів комп'ютерної обробки надвеликих за обсягом даних в інформаційних середовищах різноманітного призначення, систем управління бізнес-процесами, вбудованих систем та мереж Інтернету речей, сервіс-орієнтованих середовищ та систем високопродуктивних обчислень (ФК 15).

Програмні результати навчання (ПРН):

- мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань (ПРН 1);
- мати спеціалізовані уміння та навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур (ПРН 2);
- розробляти концептуальну модель інформаційної або комп'ютерної системи (ПРН 6);
- розробляти та застосовувати математичні методи для аналізу інформаційних моделей (ПРН 7);
- розробляти моделі та методи аналізу даних (включно з великими) (ПРН 8);
- розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими) (ПРН 9);
- створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх

ефективність та обмеження на їх застосування (ПРН 11);
- проєктувати та супроводжувати бази даних та знань (ПРН 12);
- тестувати програмне забезпечення (ПРН 14);
- збирати, формалізувати, систематизувати і аналізувати потреби та вимоги до інформаційної або комп'ютерної системи, що розробляється, експлуатується чи супроводжується (ПРН 18);
- аналізувати сучасний стан і світові тенденції розвитку комп'ютерних наук та інформаційних технологій (ПРН 19).

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Дана дисципліна вивчається в рамках циклу дисциплін професійної підготовки магістрів (з переліку нормативних освітніх компонентів) і спирається на знання, отримані студентами з предметів, які передують даній дисципліні, а саме: 1. «Алгоритмізація та програмування», 2. «Об'єктно-орієнтоване програмування», 3. «Системи баз даних», 4. «Вступ до інтелектуального аналізу даних».

Вивчення даної дисципліни служить базою для професійної роботи після отримання диплому про освіту в таких напрямках сучасних інформаційних технологій, як Інтернет речей та вбудовані системи, грід-технології для розподілених обчислень та обробки даних, технології побудови розподілених баз даних та знань, мультиагентні системи.

3. Зміст навчальної дисципліни

Розподіл занять по лекціях, лабораторних заняттях та самостійної роботи студентів

Назви розділів і тем	Кількість годин				
	Всього	у тому числі			
		Лекції	Практичні	Лабораторні	СРС
Розділ 1. *Парадигма та модель великих масивів даних*					
Тема 1.1. Визначення проблематики науки про дані та аналізу надвеликих масивів даних як складової частини науки про дані https://docs.google.com/presentation/d/1kC3wmgVaJnTfMICfLAReMrc0o6ofv2He/edit?usp=share_link&ouid=111376673486030597944&rtpof=true&sd=true	2,5	2			0,5
Тема 1.2. Парадигма великих даних, лямбда-архітектура систем обробки великих масивів даних https://docs.google.com/presentation/d/1k7-lXKUaPWYYO9JY-l-jX1X-VyZ4LOxy/edit?usp=share_link&ouid=111376673486030597944&rtpof=true&sd=true	2,5	2			0,5
Тема 1.3. Модель великих обсягів даних https://docs.google.com/presentation/d/1UpSRuFQomYqOMTfrEG3gBFWbP8FrdlkK/edit?usp=share_link&ouid=111376673486030597944&rtpof=true&sd=true	2,5	2			0,5
Лабораторна робота №1	5,5			4	1,5
Разом за розділом 1	**13**	**6**		**4**	**3**

Розділ 2. ***Обчислювальна модель та фреймворк в системах оброблення великих масивів даних***					
Тема 2.1. Технологія «розподілення-редукції» обробки надвеликих обсягів даних в пакетному режимі, обчислювальна парадигма Map-Reduce https://docs.google.com/presentation/d/1rNv53bxxSo94Hzy3xCt6JMcu5mKXOwWu/edit?usp=share_link&ouid=111376673486030597944&rtpof=true&sd=true	2,5	2			0,5
Тема 2.2. Алгоритми, в яких використовується парадигма розподілених обчислень MapReduce https://drive.google.com/file/d/1VM4Q98y46xC_tsrgcQBvbQfre0mBQFvA/view?usp=share_link	2,5	2			0,5
Тема 2.3. Інструменти управління великими даними https://drive.google.com/file/d/11Yv2-ji97PkXYmY6xuiDYPRcZyU5nFKJ/view?usp=share_link https://drive.google.com/drive/folders/1_NEPLb8QpWTUI7DSSlEHDi-7lvtDjrtl?usp=share_link	2,5	2			0,5
Лабораторна робота №2	5,5			4	1,5
Разом за розділом 2	**13**	**6**		**4**	**3**
Розділ 3. ***Розв'язання задач інтелектуальної обробки великих обсягів даних (Big Data Mining)***					
Тема 3.1. Пошук подібних об'єктів у великих масивах даних https://drive.google.com/file/d/1UHwXtxw42kDLuOgvj30K_Ak7f7qnLUnB/view?usp=share_link https://drive.google.com/drive/folders/1eaMKhNCuHuYL2mOXKshVRPQ3v6z5tEiA?usp=share_link	2,5	2			0,5
Тема 3.2. Виділення частих предметних наборів даних, побудова асоціативних правил https://docs.google.com/presentation/d/1gus5U-vFe57fZw1rsiyMuDpYnMqASwSy/edit?usp=share_link&ouid=111376673486030597944&rtpof=true&sd=true	2,5	2			0,5
Тема 3.3. Кластеризація великих даних	2,5	2			0,5
Лабораторна робота №3	5,5			4	1,5
Разом за розділом 3	**13**	**6**		**4**	**3**

Розділ 4. ***Технології ефективного пошуку та аналізу документів серед великої кількості даних в мережі Web***					
Тема 4.1. Аналіз великих потоків даних. Блокчейни https://drive.google.com/drive/folders/1H0IH_CC4vptN5veJTRAMaq2Y2Olp5zMv?usp=share_link https://drive.google.com/drive/folders/136hmre681W1fxrNGjDjGWdv7zzGYSJlb?usp=share_link	2,5	2			0,5
Тема 4.2. Оцінка релевантних сторінок у ВЕБ-мережах https://drive.google.com/drive/folders/1GfSDOolJBzCb1Yx6VSffRSjLYXrNsMzk?usp=share_link https://drive.google.com/drive/folders/17PJKrT2Ql5PzSD2WOOipU3-Z9XrAxfJ8?usp=share_link	2,5	2			0,5
Тема 4.3. Тематичний і посилочний PageRank, TrustRank та WebSpam https://drive.google.com/drive/folders/1XeoBHUruJOf3pwQ6hrRC3P17oE7_CPxx?usp=share_link https://drive.google.com/drive/folders/1gjzSROswS3nM_b1QbGetrVUWc1SKolPB?usp=share_link	2,5	2			0,5
Лабораторна робота № 4	5,5			4	1,5
Разом за розділом 4	**13**	**6**		**4**	**3**
Розділ 5. ***Методи оброблення великих масивів даних в спеціалізованих системах***					
Тема 5.1. Рекомендаційні системи https://docs.google.com/presentation/d/1Ll8iE4MFlXfj4gQXp1Y0zP6wNS6DKS1Z/edit?usp=share_link&ouid=111376673486030597944&rtpof=true&sd=true	2,5	2			0,5
Тема 5.2. Соціальні мережі (1): різновиди, графове представлення, методика пошуку товариств	2,5	2			0,5
Тема 5.3. Соціальні мережі (2): пошук товариств, які пересікаються, пошук оточень товариств на графах	2,5	2			0,5
Разом за розділом 5	**7,5**	**6**			**1,5**
Розділ 6. ***Методи редукції великих даних та машинне навчання на великих даних***					
Тема 6.1. Редукція масивів даних в системі розподілених обчислень	2,5	2			0,5
Тема 6.2. Загальні принципи та моделі машинного навчання, використання перцептрону для побудови класифікатора https://docs.google.com/presentation/d/1ngGRL9C6YjOS5NUtVihzPaBpQMcnuLSM/edit?usp=share_link&ouid=11137667348	2,5	2			0,5

6030597944&rtpof=true&sd=true					
Тема 6.3. Багатокласові перцептрони, метод опорних векторів, навчання по найближчих сусідах	2,5	2			0,5
Разом за розділом 6	**7,5**	**6**			**1,5**
Захист лабораторних робіт, які залишилися незахищеними	**2**			2	
Виконання розрахунково-графічної роботи (РГР)	**15**				15
Літературний пошук для виконання модульної контрольної роботи (МКР)	**2**				2
Виконання досліджень з МКР	**4**				4
Підготовка до екзамену	**30**				30
Всього годин	**120**	**36**		**18**	**66**

4. Навчальні матеріали та ресурси

Базові

1. *Конспект лекцій*, написаний викладачем цього предмету – проф., д.т.н. Рогозою В.С. та розміщений на Google-диску в формі голосових лекцій з демонстрацією слайдів в системі Power Point, посилання на які представлені в електронній пошті групових розсилок.
2. *Методичні вказівки до лабораторного практикуму*, розроблені викладачами кафедри Системного проєктування, які проводять лабораторні роботи з цього предмету на кафедрі асс. Яременко В., та асс. Письменним В. – посилання представлені в Google classroom
3. Олещенко, Л. М. Технології оброблення великих даних. Конспект лекцій [Електронний ресурс] : навчальний посібник для студентів спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем») / Любов Михайлівна Олещенко. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 227 с.
4. Литвин В. В. Аналіз даних та знань: навч. посібник [для студ. вищ. навч. закл.] / В. В. Литвин, В. В. Пасічник, Ю. В. Нікольский. – Львів: Магнолія, 2023. – 276 с.

Додаткові

(окремі розділи з цих джерел можуть бути рекомендовані викладачами для отримання додаткових відомостей до тем, які вивчаються на заняттях)

1. Fatos Xhafa, Leonard Barolli, Admir Barolli, Petraq Papajorgji (Editors). Modeling and Processing for Next-Generation Big-Data Technologies with Applications and Case Studies. – Springer, 2015. - 516 p.
2. Wesley W. Chu (Editor). Data Mining and Knowledge Discovery for Big Data. Methodologies, Challenge and Opportunities. – Springer, 2014. - 306 p.
3. Leskovec J., Rajaraman A., Ullman J. Mining of Massive Datasets. – Cambridge University Press, 2014.
4. Zheng A., Casari A. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. – O'Reilly, 2018.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Нижче надається інформація (за розділами, темами) про зміст лекцій у формі деталізованого опису кожної лекції, яку можна розглядати як рекомендації для студентів щодо найважливіших питань, які розглядатимуться на кожній лекції. Інформація про зміст лекції представляється студентам перед початком кожної лекції.

Лекції

№ тижня	Деталізований опис лекції та заплановані питання, які вивчатимуться на кожній лекції
1	**Розділ 1. *Парадигма та модель великих обсягів даних*** **Тема 1.1. Визначення проблематики науки про дані та аналізу надвеликих масивів даних як складової частини науки про дані** Зміст теми 1.1 • Зміст науки про дані. Визначення цілей науки про дані та обробки надвеликих масивів даних. Опис головних задач, які вивчаються в даній дисципліні, визначення місця даної дисципліни серед інших предметів з інформаційних технологій, мета вивчення дисципліни та методичні рекомендації до її вивчення. • Варіанти типів даних: (а) структуровані дані, (б) неструктуровані дані, (в) дані, представлені на натуральних мовах, (г) комп'ютерні дані, (д) дані, представлені в формах графів, (е) мережеві дані, (є) аудіо- та відеодані, (ж) графічні дані, (з) потокові дані. • Визначальні ознаки великих масивів даних • Методи та техніки оброблення великих обсягів даних • Базові програмні платформи та технології обчислень: NoSQL, MapReduce, Hadoop, Python, R, Business Intelligence, реляційні системи управління базами даних з підтримкою мови SQL.
2	**Тема 1.2. Особливості оброблення великих масивів даних, лямбда-архітектура** Зміст теми 1.2 • Приклади, де ми маємо справу з великими даними, та де шукати нові вирішення проблем • Труднощі масштабування в традиційних базах даних • Чому бази даних типу MySQL не вирішують всіх проблем оброблення надвеликих масивів даних • Принципи, на яких будуються системи оброблення надвеликих масивів даних • Інструментальні засоби оброблення надвеликих масивів даних
3	**Тема 1.3. Модель великих обсягів даних** Зміст теми 1.3 • Аналіз структурної схеми обробки даних в системі з лямбда-архітектурою, інформаційна залежність даних • Характеристики даних, які доцільно підтримувати в системах великих даних: 1) необробленість даних, 2) незмінність даних, 3) безстроковість даних. • Модель даних, яка спирається на факти • Деталі реалізації моделі, яка спирається на факти
4	**Розділ 2. *Обчислювальна парадигма та фреймворки в системах оброблення великих масивів даних*** **Тема 2.1. Технологія «розподілення-редукції» обробки надвеликих обсягів даних в пакетному режимі, обчислювальна парадигма Map-Reduce** Зміст теми 2.1 • Архітектура системи кластерних обчислень • Розподілена файлова система HDFS (Hadoop Distributed File System) та її застосування в складі платформи MapReduce. • Функції складових компонентів та загальна схема обчислювальних процесів в моделі MapReduce.

5	**Тема 2.2. Алгоритми, в яких використовується парадигма розподілених обчислень MapReduce** Зміст теми 2.2 • Множення матриці на вектор та множення двох матриць із застосуванням MapReduce • Операції реляційної алгебри в MapReduce (побудова відношення, вибірки, проекції, об'єднання, перетину, різниці, натурального об'єднання відношень, угрупування та агрегування) • Системи потоків робіт, рекурсивні узагальнення обчислювальних процесів MapReduce.
6	**Тема 2.3. Інструменти управління великими даними** Зміст теми 2.3 • Фреймворк Hadoop: призначення, властивості та складові компоненти Hadoop • Використання у фреймворку, призначеному для обробки великих даних, інструментів, орієнтованих на реляційні та нереляційні бази даних (на прикладі Hadoop) • Технології імпорту даних з використанням технології Kafka • Фреймворк Spark, порівняння характеристик Hadoop та Spark • Приклади інших інструментів, призначених для фреймворків систем обробки великих даних: Hive, Impala, Cassandra • Бібліотека JCascalog для застосування в технології обчислень MapReduce
7	**Розділ 3**. ***Розв'язання задач інтелектуальної обробки великих обсягів даних (Big Data Mining)*** **Тема 3.1. Пошук подібних об'єктів у великих масивах даних, редукція масивів даних** Зміст теми 3.1 • Формулювання задачі пошуку подібних об'єктів • Коефіцієнт Жаккара та його використання в процедурі розбиття документів на k-шинглі • Обчислення мінхеш-функції, методи побудови мінхеш-сигнатур • Алгоритми хешування сигнатур з врахуванням близькості документів • Метрики та простори, які використовуються в оцінках документів • Узагальнене хешування з врахуванням близькості об'єктів • Колаборативна фільтрація • Розбиття документів на шинглі, компресія шинглів за допомогою хешування • Мінхеш • Матричне представлення множин документів • Знаходження сигнатур, матриця сигнатур • Хешування документів з врахуванням їх близькості (LSH, Locality-Sensitive Hashing), розділ елементів матриці сигнатур на смуги: хешування смуг
8	**Тема 3.2. Виділення частих предметних наборів даних, побудова асоціативних правил** Зміст теми 3.2 • Модель кошика покупок • Визначення частого предметного набору, послідовність обчислення частих предметних наборів • Асоціативні правила, достовірність та цікавість правил • Алгоритми пошуку частих предметних наборів • Пошук частих предметних пар, наївний алгоритм пошуку • Підходи до зберігання лічильників пар предметів в пам'яті: їх властивості, порівняння • Алгоритм Апріорі

	• Модифіковані версії алгоритма Apriori • Алгоритм PCY (Park-Chen-Yu) • Багатоетапний та багатохешовий алгоритми
9	**Тема 3.3. Кластеризація великих даних** Зміст теми 3.3 • Стратегії кластеризації великих обсягів даних • Ієрархічна кластеризація в евклідовому просторі • Алгоритм k середніх • Алгоритм Бреддлі-Файяда-Рейна (БФР) • Алгоритм CURE • Кластеризація в неевклідових просторах • Кластеризація для потоків даних
10	**Розділ 4.** ***Технології ефективного пошуку та аналізу документів серед великої кількості даних в мережі Web*** **Тема 4.1. Аналіз великих потоків даних. Блокчейни** Зміст теми 4.1 • Потокова модель даних • Використання блокчейнів в пересиланнях потоків даних між об'єктами розподіленої системи • Алгоритм SGD як приклад потокового алгоритму, загальна модель управління потоками даних, типи запитів до потоків даних • Вибірка даних з потоку даних: вибірка з фіксованим розміром, як оцінювати дані, які приховані в потоці, вибірка з фіксованою пропорцією, вибірки користувачів • Узагальнене вирішення задачі управління даними з потоку • Фільтрація потоків, фільтр Блума, імовірність фальшиво-позитивного результату в фільтрі Блума • Приклади деяких інших задач, пов'язаних з виділенням певних елементів з потоку даних • Запити, які здійснюються за допомогою довгого ковзного вікна: як утворюються ковзні вікна, для чого підраховуються біти в ковзних вікнах • Агоритм Датара-Гіоніса-Індика-Мотвані (DGIM), експоненційно зростаючі інтервали (вікна), правила побудови інтервалів в DGIM, дії, які треба виконати, коли приходить новий біт, покращення властивостей розподілу вікон на інтервали
11	**Тема 4.2. Оцінка релевантних сторінок у ВЕБ-мережах** Зміст теми 4.2 • Ідеї, покладені в основу алгоритму PageRank та приклади мереж, де особливо принципово важливе значення набуває боротьба зі спамом • Зв'язки як міра оцінки важливості вузлів, ранжування вузлів на графі • Що таке PageRank? Представлення ВЕБу в формі графа та матриці переходів • Яким чином описується поведінка випадкового користувача у ВЕБ-мережі • Як представляється модель зв'язків в алгоритмі PageRank • У чому полягає проблема тупиків на графі ВЕБу та яким чином можна вирішувати цю проблему • На чому полягає техніка телепортації та коли вона застосовується до ВЕБ-графу • Ітеративне обчислення PageRank за допомогою MapReduce, чому в ітеративних обчисленнях PageRank за допомогою MapReduce доцільно використовувати комбінатори

12	**Тема 4.3. Тематичний і посилочний PageRank, TrustRank та WebSpam** Зміст теми 4.3 • Тематичний PageRank • З якою метою застосовується тематичний PageRank • Матричний опис ітерацій, за допомогою яких обчислюється тематичний PageRank • Як застосовувати тематичний PageRank в пошуковій системі • Посилочний спам • Спам-ферма, аналіз спам-ферми • Боротьба з посилочним спамом • Алгоритм TrustRank, спамна маса • Хаби та авторитетні сторінки ВЕБу • Визначення хабів і авторитетних сторінок • Метод обчислення хабів і авторитетних сторінок
13	**Розділ 5**. ***Методи оброблення великих масивів даних в спеціалізованих системах*** **Тема 5.1. Рекомендаційні системи** Зміст теми 5.1 • Модель рекомендаційної системи • Представлення профілів документів • Профілі користувачів • Створення рекомендацій об'єктів користувачам на підставі вмісту об'єктів • Алгоритм класифікації • Оцінки подібності об'єктів, колаборативна фільтрація • Зменшення розмірності даних
14	**Тема 5.2. Соціальні мережі (1): різновиди, графове представлення, методика пошуку товариств** Зміст теми 5.2 • Представлення соціальних мереж в формі графів • Різновиди соціальних мереж • Метрики графів соціальних мереж • Техніки кластеризації соціальних мереж • Техніки прямого пошуку товариств
15	**Тема 5.3. Соціальні мережі (2): пошук товариств, які пересікаються, пошук оточень товариств на графах** Зміст теми 5.3 • Методи розрізання графів соціальних мереж • Оцінка максимальної подібності товариств • Пошук товариств, які пересікаються • Випадкові бликання в соціальній мережі • Побудова оточень на графах соціальних мереж
16	**Розділ 6**. ***Методи редукції великих даних, машинне навчання на великих даних*** **Тема 6.1. Редукція масивів даних в системі розподілених обчислень** Зміст теми 6.1 • Математичні операції з матрицями приналежності: обчислення властивих значень та властивих векторів • Метод головних компонентів • Матриця відстаней • Сингулярний розклад

	• Зниження розмірності за допомогою сингулярного розкладу • Запити з використанням концептів
17	**Тема 6.2. Загальні принципи та моделі машинного навчання, використання перцептрону для побудови класифікатора** Зміст теми 6.2 • Загальні принципи та моделі машинного навчання на великих наборах даних • Підходи до машинного навчання • Архітектура систем, які спираються на методах машинного навчання • Використання перцептрону для побудови класифікатора • Методика навчання перцептрону • Збіжність перцептрону, властивості перцептрону як пристрою, який здатний до навчання • Алгоритм Віннов
18	**Тема 6.3. Багатокласові перцептрони, метод опорних векторів, навчання по найближчих сусідах** Зміст теми 6.3 • Принципи побудови багатокласових перцептронів • Удосконалення алгоритму Віннов – змінний поріг • Збалансований алгоритм Віннов • Грубий сепаратор • Паралельна реалізація перцептронів на підставі моделі MapReduce • Метод опорних векторів • Узагальнена методика побудови цільової функції • Пошук рішень в методі опорних векторів за допомогою градієнтного спуску • Навчання по найближчих сусідах

Лабораторні заняття

Основними цілями виконання лабораторних занять є:

- знайомість студентів з сучасними програмно-апаратними платформами оброблення надвеликих масивів даних,
- набуття практичних навичок в обробці та дослідженнях великих масивів даних,
- вміння правильної оцінки отриманих результатів досліджень.

№№ тижнів	Теми лабораторних робіт та їх зміст	Сумарна кількість годин для проведення лабораторних досліджень та обробки результатів досліджень
1-2-3-4	**Тема 1:** Вступ до Apache Spark та огляд парадигми Map Reduce *Мета роботи* Ознайомлення з основними поняттями та можливостями Apache Spark та алгоритму MapReduce, отримання базових навичок налаштування середовища PySpark для обробки великих даних *План роботи* 1. Ознайомитися з програмними інструментами, необхідними для виконання даного завдання 2. Налаштувати середовище 3. Запустити і вивчити логіку роботи програми “word count”	4

	4. Реалізувати та проаналізувати роботу алгоритму “inverted index” 5. Розібрати теоретичні засади бібліотеки Apache Spark та файлової системи HDFS 6. Підготувати протокол з результатами виконання роботи, який має містити наступні пункти: 6.1. Кроки налаштування середовища 6.2. Запуск та пояснення роботи програми «word count» 6.3. Запуск та пояснення роботи програми «inverted index» 6.4. Відповіді на теоретичні запитання	
5-6-7-8	**Тема 2:** Дослідження модифікованого алгоритму PageRank під парадигму MapReduce *Мета роботи* Ознайомлення з алгоритмом PageRank та його модифікацією під парадигму MapReduce, здобуття навичок використання PySpark для реалізації модифікованого алгоритму PageRank та виконання експериментів з великими обсягами даних *План роботи* 1. Ознайомитися з задачами, які розв’язуються за допомогою алгоритму PageRank та його модифікацією під парадигму MapReduce та дослідити принципи роботи алгоритму PageRank, а також його модифікації під парадигму MapReduce. 2. Розробити модифікований алгоритму PageRank за допомогою PySpark та алгоритму MapReduce під конкретну запропоновану викладачем задачу. Реалізувати модифікований алгоритм PageRank за допомогою PySpark та обчислювальної моделі MapReduce, використовуючи підходи, що відповідають парадигмі MapReduce. 3. Виконати експерименти з великими даними, а саме використати розроблений модифікований алгоритм PageRank для обробки великих даних та виконати експерименти з різними параметрами алгоритму. 4. Проаналізувати результати виконаних експериментів та порівняти їх з результатами, отриманими з використанням стандартного алгоритму PageRank. 5. Підготувати протокол з результатами виконання роботи, який має містити наступні пункти: 5.1. Опис кроків налаштування середовища 5.2. Опис запуску та пояснення роботи програми «PageRank» на щонайменше 5-ти різних наборах даних 5.3. Відповіді на теоретичні запитання	4
9-10-11-12	**Тема 3:** Використання програмних інструментальних засобів для створення онтологій *Мета роботи* Здобуття навичок побудови онтології з застосуванням програмного інструментального засобу Protege та дослідження застосування онтологій до обробки надвеликих масивів даних *План роботи* 1. Проаналізувати одну з існуючих онтологій та описати її елементи у протоколі (наприклад, взяти одну зі списку: https://www.bbc.co.uk/ontologies) 2. Зареєструватись на https://webprotege.stanford.edu (також, можна використовувати офлайн-версію Protege)	4

	3. Створити власну онтологію, з близько 10-ма класами у ній (наприклад, взяти предметну область з бакалаврської або магістерської роботи) 4. Дослідити засоби для експорту онтологій (додаткове завдання) 5. Дати відповіді на контрольні запитання.	
13-14-15-16	**Тема 4:** Використання Apache Kafka для високошвидкісної обробки, потокової аналітики та інтеграції даних *Мета роботи* Здобуття навичок налаштування системи потокової обробки і менеджменту даних та інтегрування її з «виробниками» та «споживачами» інформації, знайомість з теоретичними підставами роботи Apache Kafka та можливостями її застосування для обробки потоків великих даних та здобуття вмінь практичної роботи з застуванням Apache Kafka *План роботи* 1. Запустити Apache Kafka з використанням інструменту Docker: перше завдання - розгорнути Apache Kafka в середовищі контейнеризації за допомогою Docker, що дозволяє створити ізольоване середовище для роботи з Kafka, та дослідити можливості цього середовища. 2. Створити умовного «Виробника» даних, а саме, написати програму, яка виступатиме у ролі "Виробника" даних. Ця програма буде генерувати та передавати дані через Apache Kafka до теми (topic) на сервері Kafka. При цьому належить визначити формат та зміст даних, що генеруються. 3. Створити умовного «Споживача» даних, а саме, написати програму, яка виступатиме у ролі "Споживача" даних. Ця програма повинна підписатися на тему Kafka та зчитувати дані, які надходять від "Виробника". При цьому належить обробити ці дані та вивести їх на екран або зробити іншу корисну обробку даних, яка ілюструє взаємодію «Виробника» даних із «Споживачем» даних. 4. Представити відповіді на контрольні запитання, які стосуються принципів роботи Apache Kafka, та висловити свої висновки щодо особливостей роботи з цією системою та контейнеризацією.	4
17-18	**Підсумкове заняття**: Захист лабораторних робіт, які залишилися незахищеними, та отримання кінцевих оцінок за виконання циклу лабораторних робіт	2
Всього годин		18

Коментар. Теми лабораторних робіт мають узагальнюючі назви, і тому не співпадають з назвами тем лекцій.

6. Самостійна робота студентів (СРС)

Самостійна робота студентів (СРС) запланована в обсязі 66 годин (таблиця розділу 3) і передбачає: вивчення матеріалів лекцій (9 год.), підготовку до виконання лабораторних робіт та написання звітів з виконаних лабораторних робіт (6 год.), виконання розрахунково-графічної роботи (15 год.), літературний пошук для виконання модульної контрольної роботи (2 год.), виконання досліджень з модульної контрольної роботи та написання звіту (4 год.), підготовку до екзамену (30 год.).

Модульна контрольна робота (МКР)

Мета проведення виконання досліджень з модульної контрольної роботи полягає у розвинені

у студентів навичок до поглибленого аналізу задач оброблення великих даних за темою, запропонованою викладачем та узгодженою зі студентом, більш глибокого вивчення спеціальних методів та алгоритмів оброблення великих масивів даних на прикладі об'єкту досліджень по темі МКР, а також заохочування студентів до розробки власних (можливо) нестандартних методів та підходів та креативного мислення з використанням знань, умінь та навичок, здобутих на лекціях та лабораторних заняттях. МКР виконується як самостійне дослідження за пропонованою темою, яке має завершуватися звітом у формі письмового реферату. Викладач пропонує кожному студенту конкретне завдання для поглибленого аналізу.

Список завдань для виконання в рамках МКР, наведено нижче. Результати виконання досліджень, підготовка письмового звіту на захист досліджуваної теми оцінюються студенту як виконання модульної контрольної роботи.

Коментар: для зручності вибору студентом теми досліджень, теми згруповані між собою за близькістю проблематики; всього пропонується дванадцять груп тем.

Питання, які пропонуються студентам для самостійного аналізу в рамках виконання досліджень з модульної контрольної роботи (МКР)

Група 1
1. Визначення цілей науки про дані та обробки надвеликих масивів даних. Опис головних задач, які вивчаються в даній дисципліні, визначення місця даної дисципліни серед інших предметів з інформаційних технологій, мета вивчення дисципліни та методичні рекомендації до її вивчення. 2. Варіанти типів даних: (а) структуровані дані, (б) неструктуровані дані, (в) дані, представлені на натуральних мовах, (г) комп'ютерні дані, (д) дані, представлені в формах графів, (е) мережеві дані, (є) аудіо- та відеодані, (ж) графічні дані, (з) потокові дані. 3. Процеси data science: підготовка та дослідження даних, побудова моделей даних, відображення та автоматизація оброблення даних
Група 2
4. Принципи побудови традиційних інкрементних структур сховищ даних та їх недоліки для зберігання надвеликих масивів даних 5. Розподілені файлові системи 6. Інфраструктури розподіленого програмування, інтеграції даних та машинного навчання 7. Лямбда-архітектура системи, призначеної для оброблення надвеликих масивів даних 8. Бази даних NoSQL, їх переваги та недоліки 9. Повністю інкрементні архітектури систем оброблення надвеликих масивів даних та їх недоліки. 10. Сучасні тенденції в розвитку технологій (еластичні «хмари», ефективна еко-система з відкритим кодом для великих обсягів даних, приклад – система Hadoop)
Група 3
11. Загальні принципи оброблення даних в науці про дані 12. Етапи оброблення даних, визначені в науці про дані: *Етап 1* – Визначення цілей дослідження даних та створення проектного завдання *Етап 2* – Збирання даних *Етап 3* – Очищення, інтеграція та перетворення даних *Етап 4* – Дослідження даних *Етап 5* – Побудова моделей даних *Етап 6* – Представлення результатів та побудова додатків на їх підставі
Група 4
13. Задачі машинного навчання в Data Science, інструменти мови Python 14. Процес моделювання 15. Створення нових показників та вибір моделі 16. Типи машинного навчання, які використовуються в аналізі даних методами Data Science
Група 5
17. Властивості даних

18. Модель представлення даних, яка спирається на факти
19. Граф-схеми
20. Приклад побудови моделі даних для великих масивів даних (застосування каркасу серіалізації Apache Thrift, вузли, ребра, властивості, об’єкти даних)

Група 6

21. Вимоги до сховищ даних
22. Застосування сховища даних за схемою «ключ-значення»
23. Зберігання головного масиву даних в розподіленій файловій системі
24. Вертикальний розподіл даних
25. Низькорівневий характер розподільних файлових систем

Група 7

26. Приклади пакетної обробки великих масивів даних та мотивація до застосування спеціальних методів обробки даних, приклад Google
27. Парадигма розподілених обчислень для великих обсягів даних MapReduce
28. Низькорівневий характер моделі MapReduce
29. Конвеєрні схеми для пакетних обчислень

Група 8

30. Типові обмеження інструментальних засобів обробки даних
31. Модель даних JCascalog
32. Композиція запитань та динамічне створення запитань
33. Предикатні макрокоманди та їх динамічне створення

Група 9

34. Архітектура системи кластерних обчислень
35. Задачі-розподільники, групування даних по ключу
36. Задачі-редуктори
37. Приклад застосування моделі MapReduce
38. Принципи побудови розподіленої файлової системи HDFS (Hadoop Distributed File System) та її застосування в складі платформи MapReduce.

Група 10

39. Множення матриці на вектор та множення двох матриць із застосуванням MapReduce
40. Операції реляційної алгебри в MapReduce (побудова відношення, вибірки, проекції, об’єднання, перетину, різниці, натурального об’єднання відношень, угрупування та агрегування)
41. Системи потоків робіт, рекурсивні узагальнення обчислювальних процесів MapReduce

Група 11

42. Коефіцієнт Жаккара та його використання в процедурі розбиття документів на k-шинглі
43. Обчислення мінхеш-функції, методи побудови мінхеш-сигнатур
44. Алгоритми хешування сигнатур з врахуванням близькості документів
45. Метрики та простори, які використовуються в оцінках документів
46. Узагальнене хешування з врахуванням близькості об’єктів
47. Індекси (по символам, по позиціям та по суфіксу) та їх практичне використання

Група 12

48. Потокові моделі даних: (а) на підставі узагальнюючих ознак, (б) на підставі вибору даних з вікон фіксованих розмірів
49. Техніки побудови та використання блокчейнів в процесах пересилання потоків даних між об’єктами розподіленої системи
50. Методи фільтрації даних: вибір ключових атрибутів в фільтрі Блума, підрахунок різних елементів
51. Моменти потоків: методика визначення моментів різних порядків, оцінювання «одиниць» у вікні, експоненційно затухаючі вікна та зберігання елементів в експоненційно затухаючих вікнах)

Розрахунково-графічна робота (РГР)

Мета виконання розрахунково-графічної роботи (РГР) полягає у розвинені у студентів навичок до практичного застосування та поглиблення знань з предмету «Обробка надвеликих масивів даних» під час розв'язання конкретної задачі по запропонованій тематиці, розвиток у студентів креативності та творчості в пошуку ефективних методів розв'язання задач.

При розробці тематики завдань з РГР та вимог до виконання цих завдань були враховані наступні особливості тої області інформаційних технологій, яку вивчають студенти в предметі «Обробка надвеликих масивів даних».

В предметі «Обробка надвеликих масивів даних» студенти вивчають аналітичні методи обробки структурованих та неструктурованих даних з особливим акцентом на обробку великих масивів даних. Дані можуть надходити до системи обробки даних як в формі потоків даних з ВЕБ-мережі, так і з баз даних, які є складовими частинами системи. В цих двох випадках вимоги до ефективності роботи системи є різними.

Для даних, які надходять в інформаційних потоках, мають застосовуватися методи прискореної обробки для того, щоб ці дані не були втрачені, причому швидкість обробки звичайно досягається за рахунок зменшення точності аналізу. Але якщо інформація зберігається в базі даних, то вимоги до швидкості обробки даних зменшуються і є можливість отримання більш точного результату аналізу.

Ще одним фактором, який треба враховувати, вибираючи методи аналізу великих масивів даних, є різноманітність задач аналізу даних. Наприклад, до типових задач обробки даних можна віднести аналіз лексичної подібності текстових документів, боротьба зі спамом, оцінка інформаційної вартості сторінок ВЕБу, побудова рекомендаційних систем, тощо. Суть кожного з цих класів задач принципово відрізняється від інших і вимагає застосування специфічних підходів дослідження.

Нарешті необхідно брати до уваги, що для розв'язання практично кожного класу задач існують не один, а кілька альтернативних методів, вибір яких може суттєво вплинути на швидкість та точність аналізу. Вміння вдалого вибору методу аналітиком даних приходить з досвідом, тим не менш оволодінню цим досвідом сприяє добра теоретична підготовка дослідника.

В програмі предмету «Обробка надвеликих масивів даних» вибір класів задач та підходів до їх розв'язання був ретельно вивірений з врахуванням важливості задач аналізу даних, з якими зустрічається аналітик даних в своїй професійній діяльності.

Викладач пропонує кожному студенту конкретну задачу до розв'язання із наведеного нижче списку тем.

Теми завдань з розрахунково-графічної роботи:

Тема 1. Оцінка приналежності документів до тематики, обраної користувачем

Тема 2. Оцінка лексичної подібності документів через розбиття документів на шинглі

Тема 3. Визначення лексичної близькості текстових документів за допомогою мінхеш-сигнатур

Тема 4. Визначення лексичної близькості текстових документів з використанням обчислювальних хеш-функцій

Тема 5. Визначення лексичної подібності текстових документів методом хешування з врахуванням близькості

Тема 6. Оцінка рангу сторінок ВЕБу методом PageRank

Тема 7. Оцінки рангу сторінок ВЕБу з використанням техніки подолання тупиків

Тема 8. Виявлення «павучих пасток» серед сторінок ВЕБу та вихід з них через телепортацію

Тема 9. Техніка зміщеного випадкового блукання в побудові тематичного Page Rank

Тема 10. Оцінка хабності та авторитетності сторінок ВЕБу

Тема 11. Виявлення частих предметних наборів

Тема 12. Побудова асоціативних правил для наборів даних

Тема 13. Ієрархічна кластеризація об'єктів в евклідовому просторі L_2

Тема 14. Ієрархічна кластеризація об'єктів в евклідовому просторі L_1

Тема 15. Ієрархічна кластеризація об'єктів в евклідовому просторі L_∞

Тема 16. Порівняльний аналіз результатів ієрархічної кластеризації об'єктів в евклідових

просторах L_1, L_2 та L_∞

Тема 17. Ієрархічна кластеризація об'єктів в неевклідовому просторі з визначенням кластроїдів та використанням міри відстані по Жаккару

Тема 18. Ієрархічна кластеризація об'єктів в неевклідовому просторі з визначенням кластроїдів та використанням редакційної міри відстані

Тема 19. Ієрархічна кластеризація об'єктів в неевклідовому просторі без застосування кластроїдів

Тема 20. Порівняльний аналіз результатів ієрархічної кластеризації об'єктів в неевклідовому просторі з використанням та без використання кластроїдів

Тема 21. Модель рекомендаційної системи, яка побудована на підставі фільтрації вмісту об'єктів

Тема 22. Модель рекомендаційної системи, яка побудована на принципах колаборативної фільтрації по схемі користувач-користувач

Тема 23. Модель рекомендаційної системи, яка побудована на принципах колаборативної фільтрації по схемі об'єкт-об'єкт

Тема 24. Зниження вимірності простору властивостей об'єктів методом головних компонент (методом *PCA*)

Тема 25. Застосування методу сингулярного розкладу матриці даних (методу *SVD*) для оцінки важливості концептів груп об'єктів

Тема 26. Застосування методу сингулярного розкладу матриці даних (методу SVD) для прогнозування властивостей нових об'єктів

Тема 27. Застосування методу *CUR*-декомпозиції для апроксимації характеристичної матриці великих розмірів матрицями менших розмірів

Студентам пропонується наступний потижневий план поетапного виконання розрахунково-графічної роботи:

Тижні учбового семестру	Етапи роботи	Навчальні години, виділені для виконання РГР
1	Отримання завдання з РГР	1
2, 3, 4	Виконання літературного пошуку по темі РГР з метою аргументованого вибору методу до розв'язання поставленої задачі	2
5 – 15	Виконання обчислень за обраним методом для розв'язання поставленої задачі	7
16	Оцінка результатів розв'язання задачі за обраним методом	2
17	Написання звіту з виконання РГР	2
18	Захист перед викладачами виконаної РГР	1
	Загальна кількість годин СРС для виконання РГР	**15**

Рекомендована література до виконання СРС

Крім інформаційних джерел, рекомендованих в розділі 4, передбачається, що кожний студент має виконати інформаційний пошук в доступних для нього джерелах (Інтернет, наукові статті, та книжки). Деякі видання можуть бути використані під час створення проєкту програмної реалізації методу або алгоритму, запропонованих студентом. Як приклад, наступні книги дають уявлення про програмну реалізаці методів класифікації великих даних та методів машинного навчання:

1. Fatos Xhafa, Leonard Barolli, Admir Barolli, Petraq Papajorgji (Editors). Modeling and Processing for Next-Generation Big-Data Technologies with Applications and Case Studies. – Springer, 2015. – 516 p.
2. Wesley W. Chu (Editor). Data Mining and Knowledge Discovery for Big Data. Methodologies,

Challenge and Opportunities. – Springer, 2014. – 306 p.
3. Leskovec J., Rajaraman A., Ullman J. Mining of Massive Datasets. – Cambridge University Press, 2014.
4. Zheng A., Casari A. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. – O'Reilly, 2018.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Система вимог, які викладач ставить перед студентом:

- відвідування занять – лекцій та лабораторних практикумів;
- проявлення активності, своєчасна підготовка звітів рефератів за обраною темою СРС та РГР, відключення телефонів на заняттях, використання засобів зв'язку для пошуку інформації на Google-диску викладача чи в Інтернеті;
- захист лабораторних робіт має відбуватися в терміни та часи, визначені викладачами, які проводять ці заняття;
- захист індивідуальних завдань з модульної контрольної роботи та розрахункової роботи відбувається в дні та години, визначені викладачами;
- до студентів застосовуються правила призначення заохочувальних та штрафних балів, які наведені нижче в розділі 8 даного силабусу;
- до студентів застосовується політика дедлайнів та перескладань, яка викладена нижче в розділі 8 даного силабусу;
- реалізується політика щодо академічної доброчесності: самостійність виконання завдань модульних контрольних робіт, лабораторних робіт та курсових робіт;
- в умовах воєнного часу допускається звітування щодо виконаних завдань з лабораторного практикуму, модульних контрольних робіт та розрахунково-графічних робіт в інші дні і часи, ніж передбачені планом занять, але лише за умов узгодженості з викладачами.

8. Види контрольних заходів та рейтингова система оцінювання результатів навчання (РСО)

Види контрольних заходів

Контроль за виконанням студентами програми даного предмету здійснюється в формах: *поточного контролю, календарного контролю* та *семестрового контролю, який завершується екзаменом.*

Поточний контроль направлений на оцінювання сприйняття студентами теоретичних засад та практичних методів розв'язання типових задач з даної дисципліни, і включає у себе: оцінку виконання циклу лабораторних робіт (ЛР), оцінку виконання досліджень в рамках модульної контрольної роботи (МКР) та оцінку виконання розрахунків по темі розрахунково-графічної роботи (РГР).

Лабораторні роботи (ЛР)

Кожна виконана та захищена лабораторна робота оцінюється за 10-бальною шкалою: максимальна оцінка за виконану лабораторну роботу складає 10 балів, мінімальна позитивна оцінка складає 6 балів. Таким чином, сумарна максимальна оцінка за виконання всього циклу лабораторних робіт складає 40 балів, а сумарна мінімальна позитивна оцінка – 24 бали.

Критерії оцінювання виконання кожної лабораторної роботи:

10 балів	Правильно виконані реалізація алгоритмів та розрахункова частина, досконало і повно проведений аналіз отриманих результатів, представлені правильні відповіді на теоретичні запитання і оформлено повний звіт за виконану роботу
8 – 9 балів	Правильно виконані реалізація алгоритмів та розрахункова частина, але аналіз отриманих результатів не є повним, не всі відповіді на теоретичні запитання є вичерпними, звіт про виконання завдання оформлено з недоліками
6 – 7 балів	Робота виконана не в повному обсязі, не представлено аналіз всіх отриманих результатів, в звіті представлені відповіді не на всі теоретичні запитання, а деякі відповіді мають помилки, звіт неповний

< 6 балів	Експериментальна частина роботи виконана з грубими помилками, немає відповідей на всі теоретичні запитання, звіт не відображає результатів досліджень, більшість висновків по роботі є неправильними або відсутні

Модульна контрольна робота (МКР)

Для виконання модульної контрольної роботи (МКР) кожному студенту пропонується одне питання, на яке треба підготувати письмову відповідь у формі реферату. Відповідь на питання оцінюється за 20-бальною шкалою: максимальна оцінка - 20 балів, мінімальна позитивна оцінка – 12 балів.

Критерії оцінювання виконання МКР:

19 -20 балів	Виконано літературний пошук відомих результатів, отриманих дослідниками по тематиці запропонованого для аналізу завдання, у звіті представлена повна і аргументована відповідь на питання, зроблені узагальнюючі висновки по темі, звіт з МКР містить всі необхідні матеріали, які свідчать про самостійність та креативність студента в дослідженні запропонованої теми
17–18 балів	Літературний пошук виконано поверхнево без відповідного аналізу інформації, представленої у наведених літературних джерелах, представлені у звіті результати досліджень по запропонованій тематиці не достатньо аргументовані, узагальнюючі висновки або надто поверхневі, або зовсім відсутні, звіт не дає достатнє уявлення про самостійність студента в представлених результатах досліджень
15 – 16 балів	Літературний пошук не виконано, представлені у звіті результати досліджень по запропонованій тематиці слабо або зовсім неаргументовані, узагальнюючі висновки відсутні, звіт не дає уявлення про самостійність студента в представлених результатах досліджень
12 – 14 балів	Літературний пошук не виконано, представлені у звіті результати досліджень по запропонованій тематиці неаргументовані, узагальнюючі висновки відсутні, звіт не дає уявлення про самостійність студента в представлених результатах досліджень
< 12 балів	Звіт не відповідає вимогам до МКР

Розрахунково-графічна робота (РГР)

Для виконання розрахунково-графічної роботи (РГР) кожному студенту пропонується одне завдання. Виконання завдання оцінюється за 40-бальною шкалою: найвища оцінка – 40 балів, мінімальна позитивна оцінка – 24 бали.

Критерії оцінювання виконання РГР:

37 – 40 балів	Здійснений вибір методу для розв’язання задачі, сформульованої студентові для аналізу та дослідження, розрахунки виконані в повному обсязі відповідно до поставленого завдання, у звіті наведені всі проміжні результати обчислень з відповідними поясненнями, отриманий кінцевий результат розрахунків відповідає очікуваному, в разі, якщо це вимагає завдання, студентом написана комп’ютерна програма, яка була використана в розрахунках, в звіті представлені висновки щодо властивостей застосованого в розрахунках методу.
33 – 36 балів	Вибір методу для розв’язання задачі, сформульованої студентові для аналізу та дослідження, не достатньо обґрунтований, розрахунки виконані в обсязі відповідно до поставленого завдання, але у звіті не наведені проміжні результати обчислень з відповідними поясненнями, отриманий кінцевий результат розрахунків відповідає очікуваному, але не представлені повні висновки щодо властивостей застосованого в розрахунках методу.
29 – 32 балів	Відсутнє обґрунтування методу для розв’язання задачі, сформульованої студентові для аналізу та дослідження, відсутні результати проміжних розрахунків, представлено хоч і правильний кінцевий результат розрахунків, але необхідні пояснення стосовно цього результату не достатні для доведення, що вони відповідають очікуваним, висновки щодо властивостей застосованого в

	розрахунках методу є неповними.
24 – 28 балів	Відсутнє обґрунтування застосованого методу для розв'язання поставленої задачі, відсутні результати проміжних розрахунків, кінцевий результат розрахунків представлено без коментарів, не наведені висновки щодо властивостей застосованого в розрахунках методу.
< 24 балів	Результати виконання РГР не відповідають вимогам до РГР

Календарний контроль здійснюється в формі проміжної атестації студентів. Метою проведення атестації є підвищення якості навчання студентів та моніторинг виконання графіка освітнього процесу студентами. Умовою першої атестації є виконання та захист всіх лабораторних робіт на час атестації. Умовою другої атестації є виконання та захист всіх лабораторних робіт на час атестації, а також отримання позитивних оцінок за виконання РГР та МКР.

Семестровий контроль визначає умови допуску студентів до екзамену та виставлення оцінки по предмету з врахуванням оцінок поточного контролю та оцінки, отриманої студентом на екзамені. Умовами допуску до екзамену є отримання студентом оцінок, не менших ніж мінімально позитивні оцінки за кожний з трьох видів поточного контролю в семестрі, а саме: з циклу лабораторних робіт (ЛР), виконання модульної контрольної роботи (МКР) та виконання завдання з розрахунково-графічної роботи (РГР).

Рейтингова система оцінювання результатів навчання (РСО)

Рейтингова система оцінювання складається з трьох складових: *стартової системи оцінювання, екзаменаційної системи оцінювання*, та *заохочувальних* або *штрафних балів.*

Стартова система оцінювання призначена для оцінювання якості і своєчасності виконання циклу лабораторних занять (ЛР) впродовж навчального семестру, а також якості і своєчасності виконання завдань з модульної контрольної роботи (МКР) та розрахунково-графічної роботи (РГР).

Екзаменаційна система оцінювання застосовується на екзамені і призначена для оцінювання теоретичних знань студента по даному предмету. В екзаменаційній оцінці враховуються повнота відповідей студента на екзаменаційні запитання, а також точність та аргументованості відповідей на додаткові запитання екзаменатора. Під час оцінки знань студентів на екзамені також приймаються до уваги розуміння студентом запитань, поставлених екзаменатором, та повнота відповідей на них, розуміння студентом зв'язку поставлених запитань з іншими темами даної дисципліни, а також темами інших навчальних предметів, які мають безпосереднє відношення до екзаменаційних запитань. Максимальна екзаменаційна оцінка становить 100 балів.

Критерії оцінювання відповідей студента на екзаменаційні питання:

95 – 100 балів	Змістовна і повна відповідь на теоретичне питання білету, вміння студента глибоко аналізувати теоретичні та практичні аспекти підходів та методів, які запропоновані для аналізу в питаннях білету
85 – 94 бали	Добра повна відповідь на теоретичне питання білету, вміння студента аналізувати теоретичні та практичні аспекти підходів та методів, які запропоновані для аналізу в питаннях білету, але з невеликими уточненням та поправками з боку екзаменатора
75 – 84 бали	Добра відповідь на питання, але з деякими поправками з боку екзаменатора
65 – 74 бали	Задовільна відповідь на всі питання, але у відповіді є помилки та неточності
60 – 64 бали	Задовільна відповідь на більш ніж половину питань, у відповіді є помилки та неточності, які студенту було важко виправити без допомоги викладача
< 60 балів	Незадовільна відповідь на більшу частину або на всі питання

Заохочувальні бали можуть надаватися студенту викладачем в разі, якщо: 1) студент достроково і якісно виконав всі завдання з лабораторного практикуму і на часі захистив їх перед викладачем; 2) або якщо студент приймав участь у модернізації лабораторних робіт; 3) або якщо студент виконав глибокі дослідження в рамках домашньої модульної контрольної роботи (МКР), 4) або якщо студент своєчасно представив розрахунково-графічну роботу (РГР), проявивши креативність, та отримав оригінальні результати, які виходять за рамки формальних вимог; 5) або студент підготував до публікації наукову статтю в фаховому журналі України, в якій він представив результати своїх досліджень з МКР або РГР. Максимальний сумарний заохочувальний бал по всіх

трьох видах поточного навчання (ЛР, МКР та РГР) складає +6.

Штрафні бали можуть бути застосовані до студента в разі, якщо: 1) студент без поважних причин не на часі захищав виконані лабораторні роботи та/або не на часі представив результати з МКР та/або РГР; 2) студент без поважних причин не був готовий до здачі екзамену і здавав екзамен в додатковий термін, призначений викладачем. Сумарний штрафний бал по всіх трьох видах поточного навчання може досягати −6.

Заохочувальні та штрафні бали враховується під час виставлення кінцевої оцінки по предмету.

Рейтингова оцінка r_P з даного предмету формується як сума балів *стартової оцінки* r_C з ваговим коефіцієнтом 0,6 та *екзаменаційної оцінки* r_E з ваговим коефіцієнтом 0,4, а також заохочувального або штрафного балів r_0, але не може перевищувати 100 балів.

$$r_P = 0{,}6\, r_C + 0{,}4\, r_E + r_0$$

Якщо наведена формула дає суму, яка перевищує 100 балів (це можливо внаслідок врахування оцінки r_0), то приймається, що $r_P = 100$.

Стартова система оцінювання r_C, максимально можлива величина якої становить 100 балів, складається з трьох оцінок: оцінки за виконання лабораторного практикуму $r_{ЛАБ}$, оцінки з модульної контрольної роботи $r_{МКР}$, та оцінки за виконання розрахунково-графічної роботи $r_{РГР}$:

$$r_C = r_{ЛАБ} + r_{МКР} + r_{РГР}$$

В разі існування поважних причин, студенту можуть бути призначені індивідуальні терміни складання всіх форм контрольних заходів (лабораторного практикуму, модульної контрольної роботи, розрахунково-графічної роботи та екзамену з дисципліни), які мають бути узгоджені з викладачами по даній дисципліні.

В разі, якщо не існує поважних причин несвоєчасного захисту лабораторного практикуму, та/або підготовки реферату з модульної контрольної роботи, та/або підготовки звіту з розрахунково-графічної роботи, викладач може призначити додаткові терміни для звітування із зазначених вище форм звітності, але ці терміни не можуть бути пізніше, ніж за три дні до початку екзаменаційної сесії.

В цьому випадку застосовуються штрафні санкції зі знижкою до 6 балів. Якщо й ці умови не виконані, то питання про надання термінів звітування студента з зазначених форм навчання, або надання студенту академічної відпустки, або виключення з університету відносяться до компетенції деканату та ректорату університету.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Кількість балів	*Оцінка*	*Кількість балів*	*Оцінка*
95 … 100	Відмінно	60 … 64	Достатньо
85 … 94	Дуже добре	Менше 60	Незадовільно
75 … 84	Добре	Не виконані умови допуску	Не допущено
65 … 74	Задовільно		

9. Консультації і контакти із науково-педагогічними працівниками

Консультації по матеріалах лекцій проводяться в дні, виділені в розкладі занять для лекцій, в години одразу після чергової лекції у вигляді очних або дистанційних консультацій, контактний телефон ведучого викладача – проф., д.т.н. Рогози В.С. 067-467-65-53, e-mail: rosvetnik@gmail.com

Консультації з лабораторного практикуму проводяться в дні та години, виділені для лабораторного практикуму ст.викл. Яременко Вадимом Сергійовичем: контактний телефон 063-155-2206, e-mail: yaremenko.v.s@gmail.com

Робочу програму навчальної дисципліни (силабус):

Складено професором кафедри системного проєктування, д.т.н. проф. **Рогозою Валерієм Станіславовичем**

Ухвалено кафедрою системного проєктування (протокол № 13 від 17 червня 2024 р.)

Погоджено методичною комісією НН ІПСА (протокол № 10 від 24 червня 2024 р.)

Погоджено науково-методичною комісією КПІ ім. Ігоря Сікорського зі спеціальності 122 (протокол № 11 від 28 червня 2024 р.)