

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЦЕНТРАЛЬНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

Кафедра кібербезпеки та програмного забезпечення

**СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Big Data**

другого (магістерського) рівня вищої освіти

м. Кропивницький

1. Загальна інформація

Назва дисципліни	Big Data
Викладач	Дресва Ганна Миколаївна, доктор філософії (PhD) з Комп'ютерної інженерії, старший викладач кафедри кібербезпеки та програмного забезпечення, https://kbpz.kntu.kr.ua/kafedra/drieieva-hanna-mykolaivna?view
Контактний телефон	+380958323438 (лектор)
E-mail:	gannadreeva@gmail.com
Консультації	<i>Очні консультації</i> згідно розкладу консультацій <i>Онлайн консультації</i> : Telegram, Viber - в робочі дні з 8 ³⁰ до 14 ²⁰

2. Анотація до дисципліни

Дисципліна «Big Data» належить до переліку вибірових навчальних дисциплін та є складовою частиною нормативно-методичного забезпечення навчального процесу за освітнім ступенем «Магістр».

Основна мета дисципліни – Навчити студента використовувати відомі методи роботи з великими об'ємами даних при вирішенні професійних (виробничих) задач. Ознайомити студента з відповідними існуючими програмними продуктами. Надати базові знання щодо математичного апарату (математичних методів та алгоритмів) які використовуються для обробки великих об'ємів даних, методів аналітичної обробки великих даних та імітаційного моделювання.

3. Мета і завдання дисципліни

Метою викладання дисципліни є ознайомлення з технологіями аналізу великих даних як інструменту для розв'язання прикладних проблем. Цілі курсу передбачають формування здатностей до пошуку, оброблення та аналізу інформації з різних джерел, формування моделей представлення великих даних, визначення основних методів отримання та аналізу даних.

Основними **завданнями** вивчення дисципліни є формування наступних компетенцій: здатність до пошуку, оброблення та аналізу інформації з різних джерел; здатність приймати обґрунтовані рішення; здатність досліджувати, розробляти та обирати технології створення великих і надвеликих систем; здатність ідентифікувати, класифікувати та описувати роботу програмно-технічних засобів, комп'ютерних систем, мереж та їхніх компонентів.

4. Формат дисципліни

Для денної форми навчання:

Викладання курсу передбачає для засвоєння дисципліни традиційні лекційні заняття із застосуванням мультимедійних презентацій, поєднуючи із лабораторними роботами та виконанням курсової роботи.

Формат очний (offline / Face to face)

Для заочної форми

Під час сесії формат очний (Face to face), у міжсесійний період – дистанційний(online).

5. Результати навчання

У результаті вивчення дисципліни здобувач вищої освіти буде вміти: знаходити необхідні дані, аналізувати та оцінювати їх; аналізувати проблематику, ідентифікувати та формулювати конкретні проблеми, що потребують вирішення, обирати ефективні методи їх вирішення; приймати ефективні рішення з питань розроблення, впровадження та експлуатації комп'ютерних систем і мереж, аналізувати альтернативи, оцінювати ризики та імовірні наслідки рішень.

6. Обсяг дисципліни

Ознака дисципліни, вид заняття	Кількість годин
Кількість кредитів/годин	4/120
Кількість змістовних модулів	2
Залік	
Всього	120

7. Технічне й програмне забезпечення/обладнання

Лекційні заняття проводяться в аудиторіях обладнаних мультимедійним проектором. Лабораторні роботи виконуються у аудиторіях кафедри кібербезпеки та програмного забезпечення, обладнаних відповідним апаратним та програмним забезпеченням (ауд 501, 507, 508, 517).

Оскільки при вивченні дисципліни використовуються інформаційні технології навчання, система дистанційної освіти Moodle, необхідно мати комп'ютерну техніку (з виходом у глобальну мережу) та оргтехніку для комунікації з викладачами, виконання тестових завдань в системі дистанційної освіти та підготовки (друку) лабораторних робіт та виконання курсової роботи.

Програмне забезпечення: Мова програмування R (платформа R - studio) та OpenServer v.5.3 та вище для роботи з базами даних.

8. Політика дисципліни

Академічна доброчесність:

Очікується, що студенти будуть дотримуватися принципів академічної доброчесності, усвідомлювати наслідки її порушення.

Детальніше за посиланням URL : <http://www.kntu.kr.ua/?view=article&id=432>

Відвідування занять

Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають лекції і лабораторні заняття курсу.

Пропущені заняття повинні бути відпрацьовані не пізніше, ніж за тиждень до залікової сесії.

Поведінка на заняттях

Недопустимість: запізнь на заняття, списування та плагіат, несвоєчасне виконання поставленого завдання.

При організації освітнього процесу в Центральнотукаїнському національному технічному університеті студенти, викладачі та адміністрація діють відповідно до: Положення про організацію освітнього процесу; Положення про організацію вивчення навчальних дисциплін вільного вибору; Положення про рубіжний контроль успішності і сесійну атестацію студентів ЦНТУ; Кодексу академічної доброчесності ЦНТУ.

9. Навчально - методична карта дисципліни

Тема 1. Вступ до дисципліни. Сучасні тенденції в розвитку систем обробки великих обсягів даних. BigData сьогодні.

Тема2. Системи управління великими даними.

Тема 3. Аналіз даних в R

Тема 4. Математичний апарат для вирішення задач обробки великих обсягів даних. Обґрунтування доцільності.

Використання алгоритмів стиснення даних.

Тема 5. BIG DATA, екосистема APACHE HADOOP.

Основні концепції Hadoop®; Встановлення та налаштування: локально і в хмарі; HDFS: Архітектура, реплікація, читання і запис даних; Команди HDFS.

Тема 6. MapReduce в Hadoop

Робочий процес MapReduce.. Hadoop MapReduce і HDFS. Виконання MapReduce. Архітектура і робота першої версії MapReduce. Hadoop API(типи, класи). WordCount

(Congigure Job, Mapper, Reducer). Reducer як Combiner. Типи даних в Hadoop. InputSplit, InputFormat, OutputFormat. Shuffle і Sort в Hadoop. Запуск і налагодження завдань. Hadoop Streaming. Streaming в MapReduce.

Лабораторна робота 1. Знайомство з мовою R.

Лабораторна робота 2. Стратегії роботи з великими масивами даних.

Лабораторна робота 3. Обробка даних. Вибір ознак (Feature Selection).

Лабораторна робота 4. Обробка даних. Вибір екземплярів (Instance Selection).

Лабораторна робота 5. Обробка даних. Дискретизація для класифікації (Discretization)

Лабораторна робота 6.. Запуск і налагодження завдань. Hadoop Streaming. Streaming в MapReduce.

10. Система оцінювання та вимоги

Види контролю: поточний, підсумковий.

Методи контролю: спостереження за навчальною діяльністю студентів, усне опитування, письмовий контроль, тестовий контроль.

Форма підсумкового контролю: залік.

Контроль знань і умінь (поточний і підсумковий) з дисципліни «Big Data» здійснюється згідно з кредитною трансфернонакопичувальною системою організації навчального процесу. Рейтинг студента із засвоєння дисципліни визначається за 100-бальною шкалою. Він складається з рейтингу навчальної роботи (засвоєння теоретичного матеріалу під час аудиторних занять та самостійної роботи, виконання практичних та індивідуальних завдань), для оцінювання якої призначається максимум 100 балів у семестрі.

Розподіл балів, які отримують студенти при вивченні дисципліни «Big Data»

Змістовний модуль 1								Змістовний модуль 2								Залік
T1		T2		T3		T4		T4		T5		T6		T7		
Л	ЛБ	Л	ЛБ	Л	ЛБ	Л	ЛБ	ЛБ	Л	ЛБ	Л	Л	ЛБ	Л	ЛБ	-
4	10	4	10	4	10	2	-	10	2	10	4	4	10	4	12	

Л – лекція. ЛБ – лабораторна робота.

Шкала оцінювання: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ЄКТС	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90-100	A	Відмінно	зараховано
82-89	B	Добре	
74-81	C		
64-73	D	Задовільно	
60-63	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
1-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

Критерії оцінювання. Еквівалент оцінки в балах для кожної окремої теми може бути різний, загальну суму балів за тему визначено в навчально-методичній карті. Розподіл балів між видами занять (лекції, лабораторні заняття, самостійна робота) можливий шляхом спільного прийняття рішення викладача і студентів на першому занятті: оцінку «відмінно» (90-100 балів, A) заслуговує студент, який:

- всебічно, систематично і глибоко володіє навчально-програмовим матеріалом;
- вміє самостійно виконувати завдання, передбачені програмою, використовує набуті знання і вміння у нестандартних ситуаціях;
- засвоїв основну і ознайомлений з додатковою літературою, яка рекомендована програмою;
- засвоїв взаємозв'язок основних понять дисципліни та усвідомлює їх значення для професії, яку він набуває;
- вільно висловлює власні думки, самостійно оцінює різноманітні життєві явища і факти, виявляючи особистісну позицію;
- самостійно визначає окремі цілі власної навчальної діяльності, виявив творчі здібності і використовує їх при вивченні навчально-програмового матеріалу, проявив нахил до наукової роботи.

оцінку «добре» (82-89 балів, B) - заслуговує студент, який:

- повністю опанував і вільно (самостійно) володіє навчально-програмовим матеріалом, в тому числі застосовує його на практиці, має системні знання достатньому обсязі відповідно до навчально-програмового матеріалу, аргументовано використовує їх у різних ситуаціях;
- має здатність до самостійного пошуку інформації, а також до аналізу, постановки і розв'язування проблем професійного спрямування;

- під час відповіді допустив деякі неточності, які самостійно виправляє, добирає переконливі аргументи на підтвердження вивченого матеріалу; оцінку «добре» (74-81 бал, C) заслуговує студент, який:

- в загальному роботу виконав, але відповідає на екзамені з певною кількістю помилок;
- вміє порівнювати, узагальнювати, систематизувати інформацію під керівництвом викладача, в цілому самостійно застосовувати на практиці, контролювати власну діяльність;

- опанував навчально-програмовий матеріал, успішно виконав завдання, передбачені програмою, засвоїв основну літературу, яка рекомендована програмою;

оцінку «задовільно» (64-73 бали, D) - заслуговує студент, який:

- знає основний навчально-програмовий матеріал в обсязі, необхідному для подальшого навчання і використання його у майбутній професії;

- виконує завдання, але при рішенні допускає значну кількість помилок;

- ознайомлений з основною літературою, яка рекомендована програмою;

- допускає на заняттях чи екзамені помилки при виконанні завдань, але під керівництвом викладача знаходить шляхи їх усунення.

оцінку «задовільно» (60-63 бали, E) - заслуговує студент, який:

- володіє основним навчально-програмовим матеріалом в обсязі, необхідному для подальшого навчання і використання його у майбутній професії, а виконання завдань задовольняє мінімальні критерії. Знання мають репродуктивний характер.

оцінка «незадовільно» (35-59 балів, FX) - виставляється студенту, який:

виявив суттєві прогалини в знаннях основного програмового матеріалу, допустив принципові помилки у виконанні передбачених програмою завдань.

оцінку «незадовільно» (35 балів, F) - виставляється студенту, який:

- володіє навчальним матеріалом тільки на рівні елементарного розпізнавання і відтворення окремих фактів або не володіє зовсім;

- допускає грубі помилки при виконанні завдань, передбачених програмою;

- не може продовжувати навчання і не готовий до професійної діяльності після закінчення університету без повторного вивчення даної дисципліни.

Підсумкова (загальна оцінка) курсу навчальної дисципліни. Є сумою рейтингових оцінок (балів), одержаних за окремі оцінювані форми навчальної діяльності: поточне та підсумкове тестування рівня засвоєння теоретичного матеріалу під час аудиторних занять та самостійної роботи (модульний контроль); оцінка (бали) за виконання практичних індивідуальних завдань. Підсумкова оцінка виставляється після повного вивчення навчальної дисципліни, яка виводиться як сума проміжних оцінок за змістові модулі. Остаточна оцінка рівня знань складається з рейтингу з навчальної роботи, для оцінювання якої призначається 100 балів з атестації (залік) перший семестр , та 60 балів навчальної роботи то максимум 40 балів екзамен (2 семестр).

11. Рекомендована література

Базова

1. Методичні рекомендації до виконання лабораторних робіт з навчальної дисципліни «Big Data» для студентів денної та заочної форми навчання за спеціальностями 123 «Комп'ютерна інженерія», 122 «Комп'ютерні науки» / уклад. В.В. Босько, Л.В. Константинова,

- Дресва Г.М. — Кропивницький: ЦНТУ, 2024. — 73 с. <https://dspace.kntu.kr.ua/handle/123456789/14069>
2. Andreas Weigend. Data for the People: How to Make Our Post-Privacy Economy Work for You. : Basic Books, 2017. — 272

Допоміжна

3. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018.
4. Ghavami Peter. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics. CreateSpace Independent Publishing Platform, 2016. — 204 p
5. Feeney K. et al. Engineering Agile Big-Data Systems //Kevin Feeney, Jim Davies, James Welch,
6. Sebastian Hellmann, Christian Dirschl, Andreas Koller, Pieter Francois, Arkadiusz Marciniak. — River Publishers, 2018. — 436 p.
7. Raheem N. Big Data: A Tutorial-Based Approach. Taylor & Francis Group LLC, CRC Press, 2019. — 203 p

Інформаційні ресурси

1. <http://www.nbuv.gov.ua/eb/ep.html> - Національна бібліотека України імені В.І.Вернадського
2. <http://dspace.nbuv.gov.ua/> - Наукова електронна бібліотека періодичних видань НАН України
3. <http://www.elibukr.org/> - Електронна бібліотека України
4. http://uk.wikipedia.org/wiki/Головна_сторінка - електронна енциклопедія Вікіпедія (українською мовою)
5. SQL Fiddle [Електронний ресурс]. – Режим доступу: <http://sqlfiddle.com/about.html>
6. <https://dou.ua/> - Співтовариство програмістів.
7. <https://habr.com> - веб-сайт, який поєднує ознаки соціальної мережі і колективного блогу, створений для публікації новин, аналітичних статей, думок, пов'язаних із інформаційними технологіями, бізнесом та Інтернетом.
8. <https://www.coursera.org>