

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ**

Кафедра комп'ютерних наук



«ЗАТВЕРДЖУЮ»
декан факультету інформаційних
технологій

Глазунова О.Г.
2022 р.

«СХВАЛЕНО»

на засіданні кафедри комп'ютерних наук

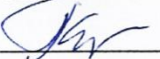
Протокол № _____ від «___» _____ 20__ р.

Завідувач кафедри

 Б. Л. Голуб

«РОЗГЛЯНУТО»

Гарант ОП «Інформаційні управляючі
системи і технології»

 Густера О.М.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ТЕХНОЛОГІЇ BIG DATA

ОСВІТНЬО-ПРОФЕСІЙНА ПРОГРАМА

«Інформаційні управляючі системи і технології»

за спеціальністю 122 «Комп'ютерні науки»

галузі знань 12 «Інформаційні технології»

Факультет інформаційних технологій

Розробник: професор кафедри комп'ютерних наук, д.т.н., проф. Хиленко В. В.

Київ 2022

1. Опис навчальної дисципліни

Технології Big Data

Галузь знань, спеціальність, освітній ступінь	
Галузь знань	12 «Інформаційні технології»
Спеціальність	122 «Комп'ютерні науки»
Освітній ступінь	Магістр
Характеристика навчальної дисципліни	
Вид	Обов'язкова
Загальна кількість годин	120
Кількість кредитів ECTS	3
Кількість змістових модулів	2
Курсовий проект (робота) (за наявності)	-
Форма контролю	Іспит
Показники навчальної дисципліни для денної та заочної форми навчання	
	денна форма навчання
Рік підготовки (курс)	1
Семестр	1
Лекційні заняття	30
Практичні, семінарські заняття	
Лабораторні заняття	30
Самостійна робота	60
Індивідуальні завдання	
Кількість тижневих аудиторних годин для денної форми навчання	12

2. Мета та завдання навчальної дисципліни

Мета: 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. Навчити студента формулювати задачі створення баз даних, баз знань, сховищ даних для розв'язання аналітичних задач (системного аналізу / Data Analysis, імітаційного моделювання) з урахуванням особливостей предметної області.

У **результаті** вивчення навчальної дисципліни студент повинен

Знати:

- принципи та методи побудови інформаційних моделей;
- принципи та методи побудови сховищ даних та баз знань;
- принципи і підходи до організації процедур імітаційного моделювання,
- особливості математичних методів, математичного і алгоритмічного забезпечення яке може використовуватись для роботи з великими об'ємами даних;
- методології аналізу результатів імітаційного моделювання;
- методологію та методи оцінки достовірності отриманих результатів;
- методи побудови та дослідження математичних моделей складних динамічних систем.

Уміти:

- формулювати задачу обробки великих інформаційних масивів відповідно до задачі, що вирішується;
- формувати бази знань (сховища даних) з урахування специфіки предметної галузі досліджень;
- розробляти структуру інформаційної моделі об'єкта дослідження;
- будувати інформаційну модель (модель інформаційних потоків) об'єкту або процесу, який вивчається;
- виконати реструктуризацію моделі або, при необхідності, її декомпозицію;
- оцінити рівень достовірності отриманих результатів;
- модифікувати побудовані інформаційні моделі у відповідності до вимог і цілей замовника досліджень з метою створення спектру моделей;
- використовувати існуючі на ринку інструментальні та програмно-алгоритмічні засоби, що можуть бути використані для вирішення задач обробки великих інформаційних масивів.

3. Програма навчальної дисципліни

ЗМІСТОВНИЙ МОДУЛЬ 1. ФОРМУЛЮВАННЯ І ФОРМАЛІЗАЦІЯ ЗАДАЧ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ

ТЕМА 1 ВСТУП ДО ДИСЦИПЛІНИ. СУЧАСНІ ТЕНДЕНЦІЇ В РОЗВИТКУ СИСТЕМ ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ.

Цілі і задачі курсу «Біг Дейта»; зв'язок з іншими навчальними курсами та базовою системою знань, яка необхідна для вирішення задач, що потребують використання великих обсягів даних (задачі: системного аналізу / «Data analysis, імітаційного моделювання, машинного навчання, штучного інтелекту, аналітичних задач). Бази знань та сховища даних. Загальні поняття та відомості.

Сучасні тенденції в розвитку систем обробки великих обсягів даних. Системи розподіленої обробки інформації. Технології блокчейн.

ТЕМА 2. ВИКОРИСТАННЯ СУЧАСНИХ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ ДЛЯ ФОРМУВАННЯ ІНФОРМАЦІЙНОЇ БАЗИ (СПЕКТРУ) МОДЕЛЕЙ. МАСШТАБУВАННЯ БД

Апаратні засоби для обробки великих обсягів даних, суперкомп'ютери та багатопроцесорні обчислювальні комплекси. Використання хмарних технологій, використання існуючих і формування індивідуальних вузькопрофільних баз знань, використання існуючих і формування сховищ даних для створення спектру моделей імітаційного моделювання екологічних систем, перспективні програмні продукти провідних софтверних компаній та їх використання для моделювання предметної області.

ТЕМА 3. ПАРСІНГ. ІМІТАЦІЙНИЙ ЕКСПЕРИМЕНТ І ПОБУДОВА ІМІТАЦІЙНИХ СИСТЕМ

Формування вимог до системи, розробка структури системи, врахування факторів невизначеності, формування вимог до результатів імітаційних експериментів; класифікація вимог; стратегії вибору рішення; аналіз очікуваних результатів; аналіз невідповідності; аналіз невизначеності цілей; оптимізація процесу вибору рішення при побудові імітаційної системи.

ТЕМА 4. СТИСНЕННЯ ДАНИХ. МАТЕМАТИЧНА ФОРМАЛІЗАЦІЯ ЗАДАЧ В ІМІТАЦІЙНОМУ МОДЕЛЮВАННІ

Алгоритми стиснення даних. Повнота моделі; ясність (недвозначність, можливість використання бібліотечних програмних продуктів); коректність і несуперечність; масштабованість; економічна доцільність обраного типу моделі відповідно до вимог замовника; відповідність до вагових коефіцієнтів замовника.

ТЕМА 5. АНАЛІЗ ДОСТОВІРНОСТІ ОТРИМАНИХ ДАНИХ

Аналіз завдань замовника до можливої похибки імітаційного моделювання, аналіз бази даних щодо об'єкту дослідження, аналіз відповідності обраного класу математичних моделей з точки зору забезпечення ефективності імітаційного

моделювання, аналіз структурних похибок, аналіз похибок обчислень, аналіз невідповідності.

ТЕМА 6. РЕСТРУКТУРІЗАЦІЯ МОДЕЛЕЙ. ДЕКОМПОЗИЦІЯ

Методи корегування моделей у відповідності до цілі досліджень, існуюче математичне забезпечення для розв'язання окресленого кола задач, методи декомпозиції, реструктуризація моделей відповідно до вимог та цілей замовника, аналіз проблемної області; аналіз ступеня деталізації математичних моделей, що використовуються.

ЗМІСТОВНИЙ МОДУЛЬ 2. ПРОГРАМНО-АЛГОРИТМІЧНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ

ТЕМА 7. МАТЕМАТИЧНИЙ АПАРАТ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ. ОБГРУНТУВАННЯ ДОЦІЛЬНОСТІ ВИКОРИСТАННЯ АЛГОРИТМІВ СТИСНЕННЯ ДАНИХ

Методи системного аналізу, типові задачі системного аналізу; невизначеність цілей, математичне моделювання, випадкові процеси, теорія хаосу, теорія катастроф, математичні методи дослідження операцій.

ТЕМА 8. ПОБУДОВА СИСТЕМИ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ. ВИБІР СТРУКТУРИ

Оптимізація вибору структури системи відповідно до цілей модельного експерименту, визначення критеріїв відповідності і прийнятності, визначення «суттєвих» параметрів, функціональність системи, діаграма потоків даних, діаграма станів, аналіз впливу факторів невизначеності.

ТЕМА 9. СУЧАСНІ ПРОГРАМНО-АЛГОРИТМІЧНІ ЗАСОБИ ДЛЯ РОБОТИ З ВЕЛИКИМИ ДАНИМИ

Програмні оболонки, вибір інтерфейсу та мови програмування при створенні програмних продуктів BIG DATA, Apache Hadoop, MapReduce, використання бібліотек та стандартних (ринкових) програмних продуктів, потоки даних.

ТЕМА 10. ОСОБЛИВОСТІ ЕКОЛОГІЧНИХ СИСТЕМ ЯК ОБ'ЄКТІВ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ BIG DATA

Особливості типових математичних моделей екологічних систем; типові проблемні ситуації процесу імітаційного моделювання екологічних систем; які типи моделей ефективно використовувати у відповідності до цілей імітаційного експерименту, діаграма варіантів використання; діаграма дій; альтернативні мови моделювання.

ТЕМА 11. МОДИФІКАЦІЯ ІНФОРМАЦІЙНИХ ТА МАТЕМАТИЧНИХ МОДЕЛЕЙ.

Аналіз чутливості моделей до зміни окремих параметрів та коефіцієнтів, оцінка витрат часу при створенні різних груп моделей і оцінка об'єму досліджень (інтенсивність використання) різних моделей відповідно до градації частоти вирішення практичних задач, що вирішуються для екологічних систем, побудова спектру моделей для забезпечення можливості вирішення різнопланових завдань замовника.

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин											
	денна форма						Заочна форма					
	усього	у тому числі					усього	у тому числі				
		л	п	лаб	інд	с.р.		л	п	лаб	інд	с.р.
1	2	3	4	5	6	7	8	9	10	11	12	13
Змістовий модуль 1. ФОРМУЛЮВАННЯ І ФОРМАЛІЗАЦІЯ ЗАДАЧ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ												
Тема 1. Вступ до дисципліни. Сучасні тенденції в розвитку систем обробки великих обсягів даних	2	2										
Тема 2. Використання сучасних інформаційно-комунікаційних технологій для формування інформаційної бази (спектру) моделей. Масштабування БД	6	4		2								
Тема 3. Парсинг. Імітаційний експеримент і побудова імітаційних систем	6	4		2								
Тема 4. Стиснення даних. Математична формалізація задач в імітаційному моделюванні	14	2		2		10						
Тема 5. Аналіз достовірності отриманих даних	4	2		2								
Тема 6. Реструктуризація моделей. Декомпозиція	8	4		4								
Разом за змістовим модулем 1	40	18		12		10						
Змістовий модуль 2. ПРОГРАМНО-АЛГОРИТМІЧНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ												
Тема 7. Математичний апарат для вирішення задач обробки великих обсягів даних. Обґрунтування доцільності використання алгоритмів стиснення даних	6	2		4								
Тема 8. Побудова системи імітаційного моделювання. Вибір структури	14	2		2		10						
Тема 9. Сучасні програмно-алгоритмічні засоби для роботи з великими даними	6	2		4								
Тема 10. Особливості екологічних систем як об'єктів імітаційного моделювання з використанням технологій Big Data	16	2		4		10						
Тема 11. Модифікація інформаційних та математичних моделей	8	4		4								
Разом за змістовим модулем 2	50	12		18		20						
Усього годин	90	30		30		30						

5. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1.	Вибір предмета дослідження при імітаційному моделюванні якого потрібно опрацьовувати великі обсяги даних	2
2.	Математична та програмно-алгоритмічна формалізація задачі	4
3.	Побудова інформаційної моделі, бази даних, бази знань	6
4.	Аналіз достовірності отриманих результатів, хмарні обчислення	6
5.	Методи побудови та застосування альтернативних моделей, аналіз можливості застосування штучного інтелекту	6
6.	Вибір програмних оболонок, бібліотек та обчислювальних методів для комплексного вирішення задач імітаційного моделювання	6
Разом		30

6. Самостійна робота

№ з/п	Назва теми	Кількість годин
1.	Вибір екологічних систем як об'єкта імітаційного моделювання і обґрунтування вибраної структури бази даних (сховища даних)	10
2.	Математична формалізація задач імітаційного моделювання і побудова інформаційних моделей для конкретних класів екологічних систем. обґрунтування доцільності використання методів і розподілених технологій обробки даних та паралельних обчислень на прикладах конкретних задач.	10
3.	Реструктуризація задач імітаційного моделювання. Аналіз вибору математичного і програмно-алгоритмічного забезпечення для розв'язання сформованих задач	10
Разом		30

7. Методи навчання.

- М1. Лекція (проблемна, інтерактивна).
- М2. Лабораторна робота.
- М3. Проблемне навчання.
- М8. Дослідницький метод.

8. Форми контролю.

- МК1. Тестування.
- МК4. Методи усного контролю.
- МК5. Екзамен.

9. Розподіл балів, які отримують студенти.

Оцінювання студента відбувається згідно положенням «Про екзамени та заліки у НУБіП України» від 27.02.2019 р. протокол №7.

Рейтинг здобувача вищої освіти, бали	Оцінка національна за результати складання екзаменів заліків	
	екзаменів	заліків
90-100	відмінно	зараховано
74-89	добре	
60-73	задовільно	
0-59	незадовільно	не зараховано

Для визначення рейтингу студента (слухача) із засвоєння дисципліни **R**_{дис} (до 100 балів) одержаний рейтинг з атестації (до 30 балів) додається до рейтингу студента (слухача) з навчальної роботи **R**_{нр} (до 70 балів):

$$R_{\text{дис}} = R_{\text{нр}} + R_{\text{ат.}}$$

10. Методичне забезпечення

Опорні конспекти з курсу. Тестові завдання для поточного та підсумкового контролю. Електронний навчальний курс. (знаходиться в процесі розробки).

11. Рекомендована література

– основна;

1. Вигерс Карл. Разработка требований к программному обеспечению. – Пер., с англ. - М.:Издательско-торговый дом "Русская Редакция", 2004. -576с.: ил
2. Леффингуелл Д., Уидриг Д. Принципы работы с требованиями к программному обеспечению. М.: ИД "Вильямс", 2002.
3. Алистер Коберн Современные методы описания функциональных требований к системам. - М.: издательство "Лори", 2002. - 263 с
4. Мацяшек Лешек. Анализ требований и проектирование систем. Разработка информационных. - Пер. с англ. - М.: Издательский дом "Вильямс", 2002. - 432 с.: ил. - Парал. тит. Англ.
5. IEEE Guide to the Software Engineering Body of Knowledge. (1) - SWEBOOK®, 2004

– допоміжна.

6. Громов, Каменова Моделирование бизнеса. Методология ARIS. - М.: Вестъ- МетаТехнология, 2001
7. Коберн А. Быстрая разработка программного обеспечения. - М.: Лори, 2002. 314 с.
8. Брауде Э. Технологии разработки программного обеспечения.- СПб: Питер, 2004. - 655 с.: ил.
9. А. Якобсон, Г. Буч, Дж. Рамбо. Унифицированный процесс разработки программного обеспечения. - СПб.: Питер , 2002. - 496 с

12. Інформаційні ресурси

1. Сбор и анализ требований к программному продукту (Версия 1.03)
pmi.ru/profes/Software_Requirements_Khimonin.pdf
2. АНАЛИЗ ТРЕБОВАНИЙ К ИНФОРМАЦИОННЫМ СИСТЕМАМ
<http://ivan-shamaev.ru/wp-content/uploads/2013/06/Information-systems-analysis-and-requirements-analysis.pdf>
3. Microsoft Solutions Framework. Модель процессов MSF, версия 3.1
<https://www.microsoft.com/Rus/Download.aspx>