

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ  
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ВНУТРІШНІХ СПРАВ**

**Кафедра кібербезпеки та DATA-технологій, факультет № 6**

**МЕТОДИЧНІ МАТЕРІАЛИ**

**ДО ПРАКТИЧНИХ ЗАНЯТЬ**

**з навчальної дисципліни «Моделі, методи та засоби аналітичної обробки  
великих масивів даних»**

**вибіркових компонент освітньої програми  
другого (магістерського) рівня вищої освіти**

**Спеціальність: 072 Фінанси, банківська справа, страхування та фондовий ринок  
(фінансова безпека та фінансові розслідування)**

**Харків 2023**

**ЗАТВЕРДЖЕНО**

Науково-методичною радою  
Харківського національного  
університету внутрішніх справ  
Протокол від 30.08.2023 № 7

**СХВАЛЕНО**

Вченою радою факультету № 6  
Протокол від 25.08.2023 № 7

**ПОГОДЖЕНО**

Секцією Науково-методичної ради  
ХНУВС з технічних дисциплін  
Протокол від 29.08.2023 № 7

Розглянуто на засіданні кафедри кібербезпеки та DATA-технологій  
факультету № 6 (*протокол від 15.08.2023 № 8*)

**Розробник:**

1. Професор. кафедри, к.т.н., доцент Струков В.М.

**Рецензенти:**

1. Провідний науковий співробітник Науково-дослідної лабораторії з проблем розвитку інформаційних технологій ХНУВС, канд.техн.наук, доцент Мордвинцев М.В.;
2. Канд.техн.наук, доцент, доцент кафедри програмної інженерії ХНУРЕ Лановий О.Ф.

## 1. Розподіл часу навчальної дисципліни за темами (денна форма навчання)

Номер та назва навчальної теми	Кількість годин, відведених на вивчення навчальної дисципліни						Вид контролю
	Всього	з них:					
		лекції	Семінарські заняття	Практичні заняття	Лабораторні заняття	Самостійна робота	
<b>Семестр № 2</b>							
Тема № 1. Технологічні тренди використання Великих Даних у кримінальній фінансовій сфері.	6	2				4	к/р
Тема № 2: Задачі і етапи опрацювання великих даних.	6	2				4	
Тема № 3: Елементи теорії множин і теорії графів.	20	4		4		12	к/р
Тема № 4. Підходи, методи і алгоритми кластеризації текстових документів.	10	2		2		6	
Тема № 5: Основи нейронних мереж і машинного навчання.	12	4				8	
Тема № 6: Інструментальні засоби аналітичної обробки Великих Даних: мова R.	60	8		20		32	
Тема № 7: Система Європолу ePOOLICE протидії організованій злочинності і тероризму.	6	2				4	
<b>Всього за семестр № 1:</b>	<b>120</b>	<b>24</b>		<b>26</b>		<b>70</b>	<b>залік</b>
<b>Всього по дисципліні</b>	<b>120</b>	<b>24</b>		<b>26</b>		<b>70</b>	

## 2. Розподіл часу навчальної дисципліни за темами (заочна форма навчання)

Номер та назва навчальної теми	Кількість годин, відведених на вивчення навчальної дисципліни					Вид контролю
	Всього	з них:				
		лекції	Семінарські заняття	Практичні заняття	Лабораторні заняття	
Семестр № 2						

Тема № 1. Технологічні тренди використання Великих Даних у кримінальній фінансовій сфері.	6	1			5	к/р
Тема № 2: Задачі і етапи опрацювання великих даних.	6	1			5	
Тема № 3: Елементи теорії множин і теорії графів.	20			2	18	к/р
Тема № 4. Підходи, методи і алгоритми кластеризації текстових документів.	10	1			9	
Тема № 5: Основи нейронних мереж і машинного навчання.	12	1			11	
Тема № 6: Інструментальні засоби аналітичної обробки Великих Даних: мова R.	60	1		4	55	
Тема № 7: Система Європолу ePOOLICE протидії організованих злочинності і тероризму.	6	1			5	
<b>Всього за семестр № 1:</b>	<b>120</b>	<b>6</b>		<b>6</b>	<b>108</b>	<b>залік</b>
<b>Всього по дисципліні</b>	<b>120</b>	<b>6</b>		<b>6</b>	<b>108</b>	

### 3. Методичні вказівки до практичних занять

#### Тема № 3. Елементи теорії множин і теорії графів.

##### Практичне заняття 1. Елементи теорії множин.

**Навчальна мета заняття:** сформуувати у студентів основи теоретичних знань і навичок розв'язання задач з теорії множин.

**Місце проведення** – навчальна аудиторія.

**Час проведення:** 2 год.

##### **Навчальні питання:**

1. Поняття множини.
2. Способи задання множин.
3. Властивості множин.
4. Операції над множинами.

##### **Література:**

1. Кристофідес М. “Теорія графів”. – М.: Мир, 1978. – 432 с.
2. Капітонова Ю.В., Кривий С.Л., Летичевський О.А., Луцький Г.М., Печорін М.К. „Основи дискретної математики”, -Київ.: Наук.думка, 2012. – 579 с.
3. Конспект лекцій.
4. Р.М.Трохимчук. Теорія графів. Навчальний посібник для студентів факультету кібернетики - К.: РВЦ “Київський університет”, 1998. - 43 с.
5. Основні поняття теорії графів: сайт. URL: [https://evgavrilenko.ucoz.ru/DS/LEKCIYA\\_1.pdf](https://evgavrilenko.ucoz.ru/DS/LEKCIYA_1.pdf)

#### 1. ПІДГОТОВКА ДО ЗАНЯТТЯ.

##### 1.1. Вивчити основні поняття і терміни теорії множин.

- 1.2. Вивчити способи задання множин.
- 1.3. Вивчити властивості множин.
- 1.4. Вивчити операції над множинами.

## 2. ПЛАН ПРОВЕДЕННЯ ЗАНЯТТЯ.

### 2.1. Контроль знань теоретичного матеріалу.

#### 2.2. Розв'язання задач:

Задача 1. Для заданих множин  $A$  і  $B$  обчислити:

- (a)  $A \cup B$ ;
- (b)  $A \cap B$ ;
- (c)  $(A \cup C) \cap B$ ;
- (d)  $A \cap B \cap C$ ;

якщо  $A = \{1, 3, 5, 6\}$ ,  $B = \{1, 2, 3, 5, 7\}$  і  $C = \{2, 4, 7\}$ ;

Задача 2. Нехай  $A = \{2, 4, 5, 7, 8\}$ ,  $B = \{1, 2, 4, 7\}$  і  $C = \{2, 4, 6, 7\}$ .

Перевірити, що

- (a)  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ ;
- (b)  $(A \setminus B) \setminus C = (A \setminus C) \setminus (B \setminus C)$ ;

#### 2.3. Домашнє завдання:

Задача 1. Для заданих множин  $A$  і  $B$  обчислити:

- (e)  $A \setminus B$ ;
- (f)  $A \cap B$ ;
- (g)  $(A \setminus C) \cup (A \setminus B)$ ;
- (h)  $(A \setminus C) \cap (A \setminus B)$ ,

якщо  $A = \{x \in \mathbb{R} \mid x^2 + x - 2 \leq 0\}$ ,  $B = \{y \in \mathbb{Q} \mid y^2 + y - 2 = 0\}$ ,  $C = \{x \in \mathbb{R} \mid x^2 > 0\}$ .

Задача 2. Нехай  $A = \{2, 4, 5, 7, 8\}$ ,  $B = \{1, 2, 4, 7\}$  і  $C = \{2, 4, 6, 7\}$ .

Перевірити, що

- (d)  $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$ .

### Тема № 3. Елементи теорії множин і теорії графів.

#### Практичне заняття 2. Елементи теорії графів.

**Навчальна мета заняття:** сформулювати у студентів основи теоретичних знань і навичок розв'язання задач з теорії графів.

**Місце проведення** – навчальна аудиторія.

**Час проведення:** 2 год.

#### *Навчальні питання:*

1. Поняття графа. Види графів.
2. Способи опису графів.
3. Планарність графів.
4. Алгоритми на графах. Найкоротший шлях, найкоротший гамільтоновий цикл, найкоротше остовне дерево.

**Література:**

1. Кристофідес М. “Теорія графів”. – М.: Мир, 1978. – 432 с.
2. Капітонова Ю.В., Кривий С.Л., Летичевський О.А., Луцький Г.М., Печорін М.К. „Основи дискретної математики”, -Київ.: Наук.думка, 2002. – 579 с.
3. Конспект лекцій.
4. Р.М.Трохимчук. Теорія графів. Навчальний посібник для студентів факультету кібернетики - К.: РВЦ “Київський університет”, 1998. - 43 с.
5. Основні поняття теорії графів: сайт. URL:  
[https://evgavrilenko.ucoz.ru/DS/LEKCIYA\\_1.pdf](https://evgavrilenko.ucoz.ru/DS/LEKCIYA_1.pdf)

**1. ПІДГОТОВКА ДО ЗАНЯТТЯ.**

- 1.5. Вивчити основні поняття і терміни теорії графів.
- 1.6. Вивчити способи опису графів.
- 1.7. Вивчити властивість планарності графів.
- 1.8. Вивчити алгоритми на графах: найкоротший шлях, найкоротший гамільтоновий цикл, найкоротше остовне дерево.

**2. ПЛАН ПРОВЕДЕННЯ ЗАНЯТТЯ.****2.1. Контроль знань теоретичного матеріалу.****2.2. Розв’язання задач:**

Задача 1. Побудувати діаграму, матриці суміжності та інцидентності для графа  $G = (V, E)$ ;  $V = \{1, 2, 3, 4\}$ ,  $E = \{(1, 3), (2, 3), (3, 4), (4, 1), (4, 2)\}$ ;

Задача 2. Побудувати діаграму, матриці суміжності та інцидентності для графа  $G = (V, E)$ ;  $V = \{1, 2, 3, 4, 5\}$ ,  $E = \{(1, 5), (1, 4), (2, 4), (3, 5), (5, 2)\}$ .

Задача 3. Для заданого графа знайти найкоротший шлях, найкоротший гамільтоновий цикл.

**2.3. Домашнє завдання:**

Задача 1. Для заданого графа знайти найкоротший шлях, найкоротший гамільтоновий цикл, найкоротше остовне дерево.

**Тема № 4. Підходи, методи і алгоритми кластеризації текстових документів.**

**Практичне заняття 3. Обробка текстових даних.**

**Навчальна мета заняття:** сформулювати у студентів основи теоретичних знань і навичок розв’язання задач обробки текстових даних.

**Місце проведення** – навчальна аудиторія.

**Час проведення:** 2 год.

**Навчальні питання:**

1. Форми представлення текстових документів.
2. Оцінки близькості текстових документів. Метрики відстані між текстовими документами.

### 3. Методи і алгоритми кластеризації текстових документів.

#### **Література:**

1. Aggarwal C.C. Data Mining. – Cham: Springer Ltd. Publ. Switzerland, 2015. – 734p.
2. Конспект лекцій.
3. Р.М.Трохимчук. Теорія графів. Навчальний посібник для студентів факультету кібернетики - К.: РВЦ “Київський університет”, 1998. - 43 с.
4. Методи і засоби аналізу даних [Електронний ресурс]. – Режим доступу: <https://bourabai.ru/tpoi/analysis6.htm>.

## 1. ПІДГОТОВКА ДО ЗАНЯТТЯ.

- 1.1. Вивчити форми представлення текстових документів.
- 1.2. Вивчити оцінки близькості текстових документів.
- 1.3. Вивчити метрики відстані між текстовими документами.
- 1.4. Вивчити методи і алгоритми кластеризації текстових документів.

## 2. ПЛАН ПРОВЕДЕННЯ ЗАНЯТТЯ.

### **2.1. Контроль знань теоретичного матеріалу.**

#### **2.2. Розв’язання задач:**

Задача 1. Розрахувати відстань між двома текстовими документами з використанням метрики Жаккарда.

Задача 2. Розрахувати відстань між двома текстовими документами з використанням метрики Левенштейна.

Задача 3. Розрахувати відстань між двома текстовими документами з використанням метрики  $Tf\_idf$ .

## **Тема № 6. Інструментальні засоби аналітичної обробки Великих Даних: мова R.**

### **Практичне заняття № 4 на тему «Типи даних і операції в R».**

**Навчальні цілі:** формування практичних навичок використання типів даних і основних операцій з ними в середовищі інструментальної системи RStudio.

Кількість годин - 4 години.

Місце проведення – комп’ютерний клас.

#### **Навчальні питання:**

1. Загальні відомості про мову R.
2. Типи даних і об’єктів мови R.
3. Операції мови R.
4. Основні операції мови R.

#### **Література.**

1. Конспект лекцій.

2. Комп'ютерна статистика : підручник / Р. Є. Майборода. – К. : ВПЦ "Київський університет", 2019. – 589 с.

## 1. ПІДГОТОВКА ДО ЗАНЯТТЯ.

- 1.5. Вивчити загальні відомості про мову R.
- 1.6. Вивчити типи даних і об'єктів мови R.
- 1.7. Вивчити формати і порядок застосування основних операцій мови R.
- 1.8. Ознайомитися з інструментальними середовищами для мови R.
- 1.9. Підготувати звіт до практичного заняття.

## 2. ПЛАН ПРОВЕДЕННЯ ЗАНЯТТЯ.

- 2.1. Контроль засвоєння теоретичних відомостей про типи даних в R.
- 2.2. Контроль засвоєння теоретичних відомостей про формати і порядок застосування основних операцій мови R.
- 2.3. Відпрацювати операції створення векторів різних типів в середовищі R+RStudio.
- 2.4. Відпрацювати операції створення іменованих векторів різних типів в середовищі R+RStudio.
- 2.5. Відпрацювати операції обробки векторів різних типів в середовищі R+RStudio: відображення на екрані (в консолі), звернення до елементів векторів за номером, за іменем, пошук суми, добутку, сортування у порядку зростання(спадання) елементів, пошук номеру максимального (мінімального) елементу.
- 2.6. Розв'язання задач:  
Задача 1. Створити цілочисельний вектор з 5 елементів і визначити: суму, добуток і кількість від'ємних елементів цього вектору.
- 2.7. Захистити роботу.

**Тема № 6. Інструментальні засоби аналітичної обробки Великих Даних: мова R.**

**Практичне заняття № 5 на тему «Операції з матрицями в RStudio».**

**Навчальні цілі:** формування практичних навичок виконання операцій з матрицями в середовищі інструментальної системи RStudio.

Кількість годин - 4 години.

Місце проведення – комп'ютерний клас.

**Навчальні питання:**

1. Матриці в R і операції з ними.
2. Функції роботи з матрицями.
3. Впорядкування та пошук у масивах.

**Література.**

1. Конспект лекцій.
2. Комп'ютерна статистика : підручник / Р. Є. Майборода. – К. : ВПЦ "Київський університет", 2019. – 589 с.



## 1. ПІДГОТОВКА ДО ЗАНЯТТЯ.

- 1.1. Вивчити матричний тип даних і операції з ним.
- 1.2. Вивчити функції роботи з матрицями.
- 1.3. Вивчити засоби впорядкування та пошук у масивах.

## 2. ПЛАН ПРОВЕДЕННЯ ЗАНЯТТЯ.

2.1. Контроль засвоєння теоретичних відомостей про типи даних «матриці» в R.

2.2. Контроль засвоєння теоретичних відомостей про формати і порядок застосування основних операцій з матрицями у мові R.

2.3. Відпрацювати операції створення матриць в середовищі RStudio.

2.4. Відпрацювати операції з номерами і іменами рядків і стовпчиків матриць.

2.5. Відпрацювати операції обробки матриць в середовищі R+RStudio: відображення на екрані (в консолі), звернення до елементів матриць за номером, за іменем, пошук суми, добутку.

2.6. Відпрацювати операції сортування у порядку зростання(спадання) елементів, пошук номеру максимального (мінімального) елементу.

2.7. Розв'язання задач:

Задача 1. Створити 2 цілочисельних вектора з 3 елементів і створити матрицю  $3 \times 2$  і  $2 \times 3$ . Роздрукувати створені матриці.

Задача 2. Задати імена стовпчиків матриць.

Задача 3. Задати імена рядків матриць.

Задача 4. Знайти мінімальний і максимальний елемент матриць.

Задача 5. Відсортувати рядки створеної матриці за зростанням.

Задача 6. Відсортувати рядки створеної матриці за спаданням.

Задача 7. Визначити: суму і добуток елементів матриць.

2.8. Захистити роботу.

**Тема № 6. Інструментальні засоби аналітичної обробки Великих Даних: мова R.**

**Практичне заняття № 6 на тему «Операції з фреймами і списки в RStudio».**

**Навчальні цілі:** формування практичних навичок виконання операцій з структурованими даними: фреймами і списками в середовищі інструментальної системи RStudio.

Кількість годин - 4 години.

Місце проведення – комп'ютерний клас.

**Навчальні питання:**

1. Фрейми в R і операції з ними.
2. Списки в R і операції з ними.
3. Функції роботи з фреймами.
4. Впорядкування та пошук у фреймах.
5. Злиття наборів даних — merge.

### ***Література.***

1. Конспект лекцій.
2. Комп'ютерна статистика : підручник / Р. Є. Майборода. – К. : ВПЦ "Київський університет", 2019. – 589 с.

## **1. ПІДГОТОВКА ДО ЗАНЯТТЯ.**

- 1.1. Вивчити тип даних «фрейми» і операції з ними.
- 1.2. Вивчити списки і операції з ними.
- 1.3. Вивчити функції роботи з фреймами.

## **2. ПЛАН ПРОВЕДЕННЯ ЗАНЯТТЯ.**

- 2.1. Контроль засвоєння теоретичних відомостей про типи даних «фрейми» в R.
- 2.2. Контроль засвоєння теоретичних відомостей про формати і порядок застосування основних операцій з фреймами у мові R.
- 2.3. Відпрацювати операції створення фреймів в середовищі RStudio.
- 2.4. Відпрацювати операції перегляду і редагування (зміни) імен рядків і стовпчиків фреймів.
- 2.5. Відпрацювати операції обробки фреймів в середовищі R+RStudio: відображення на екрані (в консолі), звернення до елементів фреймів за номером, за іменем, пошук суми, добутку, відображення на екрані заданого діапазону рядків фрейму.
- 2.6. Відпрацювати операції пошуку максимального (мінімального) елементу заданого стовпчика фрейму.
- 2.7. Відпрацювати операції пошуку номеру максимального (мінімального) елементу заданого стовпчика фрейму.
- 2.8. Відпрацювати операції сортування у порядку зростання(спадання) елементів стовпчиків , пошук номеру максимального (мінімального) елементу.
- 2.9. Розв'язання задач:  
Задача 1. Створити фрейм з 5 записів, який містить наступну інформацію про успішність студентів групи: 1) прізвище, 2) ініціали, 3) оцінка з дисципліни 1, 4) оцінка з дисципліни 2, 5) вік, використовуючи функцію `data.frame()`. Визначити студента з найбільшим сумарним балом.
- 2.10. Захистити роботу.

## **Тема № 6. Інструментальні засоби аналітичної обробки Великих Даних: мова R.**

### **Практичне заняття № 7 на тему «Базова графіка. Побудова діаграм.».**

**Навчальні цілі:** формування практичних навичок застосування засобів візуалізації великих даних в середовищі інструментальної системи RStudio.

Кількість годин - 4 години.

Місце проведення – комп'ютерний клас.

**Навчальні питання:**

1. Стовпцеві діаграми.
2. Методика застосування функції `barplot()` для малювання стовпцевих діаграм.
3. Кругові діаграми.
4. Методика застосування функції `pie()` для малювання кругових діаграм.
5. Експорт даних з Microsoft Excel.
6. Збереження табличних даних (імпорт).

### ***Література.***

1. Конспект лекцій.
2. Комп'ютерна статистика : підручник / Р. Є. Майборода. – К. : ВПЦ "Київський університет", 2019. – 589 с.

## **1. ПІДГОТОВКА ДО ЗАНЯТТЯ.**

- 1.1. Вивчити методику будування стовпчикових діаграм.
- 1.2. Вивчити методику будування кругових діаграм.
- 1.3. Вивчити методику експорту даних з Microsoft Excel.
- 1.4. Вивчити методику збереження (імпорту) табличних даних.

## **2. ПЛАН ПРОВЕДЕННЯ ЗАНЯТТЯ.**

- 2.1. Контроль засвоєння теоретичних відомостей щодо методики будування стовпчикових діаграм в R.
- 2.2. Контроль засвоєння теоретичних відомостей щодо методики будування кругових діаграм в R.
- 2.3. Контроль засвоєння теоретичних відомостей щодо методики експорту і збереження (імпорту) табличних даних з Microsoft Excel
- 2.4. Сформувати два вектора з 5 цілих чисел.
- 2.5. Відпрацювати методику побудови стовпчикової діаграми з формуванням її параметрів.
- 2.6. Відпрацювати методику побудови кругової діаграми з формуванням її параметрів.
- 2.7. Відпрацювати методику експорту даних з Microsoft Excel.
- 2.8. Відпрацювати методику збереження (імпорту) табличних даних.
- 2.9. Захистити роботу.

## **Тема № 6. Інструментальні засоби аналітичної обробки Великих Даних: мова R.**

**Практичне заняття № 8 на тему «Візуалізація даних на географічній мапі».**

**Навчальні цілі:** формування практичних навичок застосування засобів візуалізації великих даних на географічній мапі в середовищі інструментальної системи RStudio.

Кількість годин - 4 години.

Місце проведення – комп'ютерний клас.

**Навчальні питання:**

1. Пакети для роботи з географічними картами.
2. Методика застосування пакета `maptools`.
3. Функції для роботи з географічними картами.
4. Методика застосування функції `plot()` для роботи з географічними картами.
5. Методика малювання кругових діаграм на мапі. Пакет `plotrix` . Функція `floating.pie()`.

### ***Література.***

1. Конспект лекцій.
2. Комп'ютерна статистика : підручник / Р. Є. Майборода. – К. : ВПЦ "Київський університет", 2019. – 589 с.

## **1. ПІДГОТОВКА ДО ЗАНЯТТЯ.**

- 1.1. Вивчити методику роботи з географічними картами.
- 1.2. Вивчити методику застосування пакета `maptools`.
- 1.3. Вивчити методику застосування функції `plot()` для роботи з географічними картами.
- 1.4. Вивчити методику малювання кругових діаграм на мапі з використанням пакету `plotrix`, функції `floating.pie()`.

## **2. ПЛАН ПРОВЕДЕННЯ ЗАНЯТТЯ.**

- 2.1. Контроль засвоєння теоретичних відомостей щодо методики роботи з географічними картами в R.
- 2.2. Контроль засвоєння теоретичних відомостей щодо методики застосування пакета `maptools`.
- 2.3. Контроль засвоєння теоретичних відомостей щодо методики застосування функції `plot()` для роботи з географічними картами.
- 2.4. Інсталювати на комп'ютері пакети (бібліотеки) `sp`, `maptools`.
- 2.5. Завантажити у робочу область, використовуючи функцію `library ()` пакети (бібліотеки) `sp`, `maptools`.
- 2.6. Відпрацювати методику побудови кругової діаграми з формуванням її параметрів.
- 2.7. Завантажити політичну карту земної кулі `wrld_simpl`.
- 2.8. Надрукувати політичну карту земної кулі `wrld_simpl` функцією `plot`.
- 2.9. Відпрацювати методику відображення на картах кругових діаграм.
- 2.10. Завантажити у робочу область пакет `raster`.
- 2.11. Завантажити політичну карту України.
- 2.12. Зберегти політичну карту України у вигляді змінної `Ukraine` для подальшого використання.
- 2.13. Захистити роботу.