

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КИЇВСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Кафедра технологій управління

ХЛЕВНИЙ А.О., ЗАРІЦЬКИЙ О.В.

МЕТОДИ ПОШУКУ, ОБРОБКИ ТА ВІЗУАЛІЗАЦІЇ ДАНИХ

Методичні вказівки

до виконання практичних завдань, лабораторних робіт та самостійної роботи з навчальної дисципліни галузь знань 12 «Інформаційні технології», спеціальність 122 «Комп'ютерні науки», освітній рівень – магістр, освітньо-наукова програма «Інформаційна аналітика та впливи».

Київ-2025

Рецензенти:

Доцент кафедри інформаційних систем та технологій факультету інформаційних технологій, КНУ імені Тараса Шевченка, к.т.н., доцент,
Палій Сергій Володимирович.

Декан факультету комп'ютерних наук та технологій, к.т.н., доцент.,
Фесенко Андрій Олексійович.

*Рекомендовано до публікації кафедрою технологій управління факультету інформаційних технологій Київського національного університету імені Тараса Шевченка,
протокол № 8 від «21» січня 2025 р.*

Хлевний А.О., Заріцький О.В.

Методи пошуку, обробки та візуалізації даних [Текст]: Методичні вказівки для практичних, лабораторних робіт та самостійної роботи. Кафедра технологій управління факультету інформаційних технологій Київського національного університету імені Тараса Шевченка/ А.О. Хлевний, О.В. Заріцький– К. : ВИДАВНИЦТВО, 2025. – 39с.

Методичні вказівки для практичних та лабораторних робіт, містять теоретичний та довідковий матеріал, приклади питань на модульну контрольну роботу та обсягові питання на іспит. У кожній роботі виконується індивідуальне завдання, наведено приклади їх виконання та пояснення. У методичних вказівках розкриті теми: сучасних програмних продуктів для роботи з даними R та R Studio. Завантаження, обробка та візуалізація інформації у Microsoft Power BI. Призначено для студентів другого курсу магістерської освітньої програми «Інформаційна аналітика та впливи», а також для всіх кому цікава робота з методами та засобами обробки даних.

Публікується в авторській редакції.

© А.О.Хлевний, О.В. Заріцький 2025р.

ЗМІСТ

ВСТУП	4
ЛАБОРАТОРНІ РОБОТИ	6
Лабораторна робота №1	6
Лабораторна робота №2	8
Лабораторна робота №3	13
Лабораторна робота №4	17
Лабораторна робота №5	25
ПРАКТИЧНІ РОБОТИ	29
Практична робота №1	29
Практична робота №2.....	32
Самостійна робота	33
Обсягові питання на іспит	34
Список рекомендованої літератури	36
ДОДАТКИ	38
<i>Додаток А</i>	38
<i>Теми для виконання практичних:</i>	38

ВСТУП

Дисципліна складається з одного змістовного модуля. Представлені методичні вказівки відповідають робочій навчальній програмі. Відповідно до них основна увага приділяється теоретичним засадам, методам пошуку інформації, алгоритмам роботи з різними джерелами даних, та сучасними програмними засобами, що використовують для обробки та візуалізації інформації. В рамках дисципліни вивчають основні методи роботи з інформацією: пошук, завантаження, попередня обробка даних із виключенням аномалій і заповнення пропусків даних, перетворення інформації за допомогою сучасних програмних засобів, а також візуалізація, та створення звітів.

Метою дисципліни формування у студентів системного уявлення про методи пошуку та обробки інформації, як складової інформаційної аналітики. А саме, про теорію інформації, методи пошуку інформації, сучасні програмні продукти для роботи з інформацією, процедури пошуку та обробки інформації, представлення та маніпулювання даними при аналітичній діяльності.

Попередні вимоги до опанування навчальної дисципліни:

1). розуміння теоретичних засад та методології застосування апарату сучасного статистичного аналізу процесів управління проектами та методів статистичного аналізу інформаційних проектів, знання мови створення SQL запитів до баз даних.

2) вміння працювати з базами даних та знань.

Завдання (навчальні цілі):

Формування знань, умінь та навичок для ефективного застосування методів побудови пошукових алгоритмів з використанням сучасних моделей, методів і програмних засобів для пошуку необхідної аналітичної інформації. Навчити використовувати пошук інформації, проводити систематизацію та узагальнення аналітичних даних. Моделювати та аналізувати розвиток соціальних та виробничих систем, виявляти основні закономірності, тенденції, взаємозв'язки між основними компонентами з використання інформаційних технологій прогнозно-аналітичної діяльності.

Загальні рекомендації до виконання лабораторних/практичних, особливості, зв'язок з лекційним курсом тощо.

При виконанні лабораторних та практичних робіт студенти самостійно обирають інформаційну область (свій варіант) для розв'язування практичних робіт. У завданнях лабораторних робіт 1-3 запропоновано використовувати мову програмування та обробки даних R та R Studio, а у лабораторній роботі №5 для візуалізації MS Power BI. Студенти можуть виконувати представлене завдання для самостійного виконання, а також запропонувати та погодити із викладачем аналогічне дослідження за іншою темою та з іншими даними. Це сприятиме формуванню цілісного дослідження студентом продовж навчання за обраним напрямком у розрізі освітньої програми «Інформаційна аналітика і впливи».

Звіт за виконаними завданнями створюється в електронній формі і завантажується у визначені системи електронного навчання у визначені викладачем терміни. Про це викладач безпосередньо сповіщає студентів при

видачі завдання. Виконане завдання студент завантажує в систему дистанційного навчання (система) самостійно, у будь-який час до крайнього терміну подання. За межами встановлених термінів система не зможе прийняти виконане завдання. Тому студенти мають працювати за складеними ними індивідуальними графіками виконання робіт. Після того, як усі роботи студентами будуть завантажені в систему, викладач матиме змогу перевірити такі роботи та виставити оцінки, згідно методик, прописаних у робочих навчальних програмах з даної дисципліни. Такі програми розміщені в межах даної дисципліни в системі в електронному вигляді.

Студенти мають завантажити виконані роботи згідно наведених нижче правил та демонстраційних прикладів в межах встановлених граничних термінів. Система буде зберігати останню версію завантаженої роботи і надавати її викладачу для оцінки. Отримані оцінки студентом за кожне завдання будуть складати його рейтинг та допуск при підготовці до екзамену.

ЛАБОРАТОРНІ РОБОТИ

Лабораторна робота №1

Назва: Завантаження та зчитування даних за допомогою R

Мета заняття: навчитись використовувати функцій завантаження даних з мережі до R.

Зв'язок з результатами навчання: Проводити систематизацію та узагальнення даних, виявляти ключові моменти.

Час проведення: 2 год – 1 год. виконання, 1 год. – захист .

Завдання для обов'язкового виконання:

Завдання 1. За допомогою `download.file()` завантажте файл `getdata_data_ss06hid.csv` за посиланням

<https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv> та завантажте дані в R.

Необхідно знайти, скільки `property` мають `value $1000000+`

Приклад виконання :

Приклад виконання завдання 1:

Для того, щоб завантажити файл за посиланням, необхідно мати саме посилання та файл який будемо завантажувати із сайту. Використавуємо функцію `download.file()` і як параметр задаємо посилання на файл, завантажуюмо файл.

Щоб його використати, необхідно виконати функцію `read.csv()` для зчитування даних з файлу. Після цього виконуємо фільтрацію і визначаємо довжину запису, так дізнаємось кількість випадків, що відповідає поставленій умові.

```
1 url = "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2Fss06hid.csv"
2 file = "getdata_data_ss06hid.csv"
3
4 download.file(url, file)
5 data_csv = read.csv(file)
6 str(data_csv)
7 View(data_csv)
8 ?data_csv
9
10 n_val = length(which(data_csv$VAL == 24))
11 cat("Number with value $1000000+: ", n_val)
```

Рис 1 – код виконання

	RT	SERIALNO	DIVISION	PUMA	REGION	ST	ADJUST	WGTP	NP	TYPE	ACR	AGS
1	H	186	8	700	4	16	1015675	89	4	1	1	NA
2	H	306	8	700	4	16	1015675	310	1	1	NA	NA
3	H	395	8	100	4	16	1015675	106	2	1	1	NA
4	H	506	8	700	4	16	1015675	240	4	1	1	NA
5	H	835	8	800	4	16	1015675	118	4	1	2	1
6	H	989	8	700	4	16	1015675	115	4	1	1	NA
7	H	1861	8	700	4	16	1015675	0	1	2	NA	NA
8	H	2120	8	200	4	16	1015675	35	1	1	1	NA
9	H	2278	8	400	4	16	1015675	47	2	1	1	NA
10	H	2428	8	500	4	16	1015675	51	2	1	1	NA
11	H	2677	8	400	4	16	1015675	114	2	1	1	NA
12	H	2928	8	800	4	16	1015675	51	2	1	1	NA
13	H	3283	8	400	4	16	1015675	67	3	1	1	NA
14	H	3331	8	600	4	16	1015675	0	1	2	NA	NA
15	H	3341	8	500	4	16	1015675	92	1	1	1	NA

Showing 1 to 16 of 6,496 entries, 188 total columns

```

R 4.4.1 · ~/
[LIST output truncated]

> View(data_csv)

> ?data_csv
No documentation for 'data_csv' in specified packages and libraries:
you could try '??data_csv'

> n_val = length(which(data_csv$VAL == 24))

> cat("Number with value $1000000+: ", n_val)
Number with value $1000000+: 53
> |

```

Рис 2 – результат виконання

Висновки: Отже, в цій лабораторній роботі було, вивчено як використовувати функцію завантаження даних до R (`download.file()`) із мережі Інтернет, та виконати їх зчитування і класифікацію за допомогою функції `read.csv()`.

Лабораторна робота №2

Назва: Функції HDF5 низького рівня у R.

Мета заняття: навчитись використовувати функцій низького рівня у R.

Зв'язок з результатами навчання: Проводити систематизацію та узагальнення даних, виявляти ключові моменти.

Час проведення: 4 год – 2 год. виконання, 2 год. – захист .

Завдання для обов'язкового виконання:

Завдання 1. Створити файл HDF5 та ієрархії груп у R.

Завдання 2. Запис даних у файл HDF5.

Завдання 3. Запис даних у файл HDF5.

Пояснення, аналіз, приклад.

Hierarchical Data Format, HDF (Ієрархічний формат даних) - назва формату файлів, розробленого для зберігання великого обсягу цифрової інформації. Спочатку був розроблений Національним центром суперкомп'ютерних додатків, зараз підтримується некомерційною організацією HDF Group.

Бібліотеки для роботи з форматом і пов'язані з ним утиліти доступні для використання під вільною ліцензією, схожою з ліцензією BSD. Формат HDF підтримується багатьма комерційними та некомерційними програмами, в тому числі є бібліотеки для роботи з ним в Java, Matlab, Scilab, Octave, Mathematica, IDL, Python, R і Julia. Вільно розповсюджуваний пакет HDF складається з бібліотеки, утиліти командного рядка, вихідних текстів для тестування, інтерфейсу для Java і Java-програми для перегляду HDF-файлів. HDF5 - сучасна версія формату. Отримав премію R & D100 від журналу "R & D Magazine" в 2002 році. [2]

Містить ієрархію з двох основних типів об'єктів:

Datasets - набори даних, багатовимірні масиви об'єктів одного типу

Groups - групи, є контейнерами для наборів даних і інших груп

Вміст файлів HDF5 організовано подібно ієрархічної файлової системи, і для доступу до даних застосовуються шляхи, подібні до POSIX-синтаксисом, наприклад, / path / to / resource. Метадані зберігаються у вигляді набору іменованих атрибутів об'єктів.

Формат NetCDF 4 заснований на HDF5.

Приклад виконання :

Приклад виконання завдання 1.

Створення файлу HDF5 та ієрархій груп:

Щоб використовувати можливості HDF5 спочатку треба інсталивати необхідні пакети та бібліотеки. Функція H5Fcreate() створює файл формату h5. При повторному використанні коду не можна ще раз створити той самий файл, тому завдяки H5Fopen ми відкриваємо існуючий файл.

Для створення і відкритті групи використовуються ті самі функції, але параметри вже інші. Треба в якості параметрів задати змінну, яка відповідає за файл і вказати назву групи, яку хочемо створити. Також можна створювати групи в групах, як папки в папках, принцип той самий, але замість файлу ми задаємо змінну групи.

```

1 #install.packages("BiocManager")
2 #BiocManager::install("rhdf5")
3 #library(rhdf5)
4 #h5file = H5Fcreate("newfile.h5")
5 h5file <- H5Fopen("newfile.h5")
6 h5file
7
8 #h5group1 <- H5Gcreate(h5file, "foo")
9 #h5group2 <- H5Gcreate(h5file, "baa")
10 #h5group3 <- H5Gcreate(h5group1, "foobaa")
11 h5group1 <- H5Gopen(h5file, "foo")
12 h5group2 <- H5Gopen(h5file, "baa")
13 h5group3 <- H5Gopen(h5group1, "foobaa")
14 h5group1
15 h5group2
16 h5group3
17

```

Рис 3 – код створення файлу і ієрархії файлу

```

< h5file
HDF5 FILE
  name /
  filename
  name      otype dclass dim
0 baa      H5I_GROUP
1 dataset1 H5I_DATASET FLOAT 5 x 7
2 foo      H5I_GROUP

> #h5group1 <- H5Gcreate(h5file, "foo")
> #h5group2 <- H5Gcreate(h5file, "baa")
> #h5group3 <- H5Gcreate(h5group1, "foobaa")
> h5group1 <- H5Gopen(h5f ... [TRUNCATED]

> h5group2 <- H5Gopen(h5file, "baa")

> h5group3 <- H5Gopen(h5group1, "foobaa")

> h5group1
HDF5 GROUP
  name /foo
  filename
  name      otype dclass dim
0 foobaa H5I_GROUP

> h5group2
HDF5 GROUP
  name /baa
  filename
  name      otype dclass dim
0 dataset2 H5I_DATASET INTEGER 5 x 7

> h5group3
HDF5 GROUP
  name /foo/foobaa
  filename

[1] name      otype dclass dim
0 dataset2 H5I_DATASET INTEGER 5 x 7

```

Рис 4 – результат виконання створення файлу та груп

Приклад виконання завдання 2.

Запис даних у файл HDF5:

Створіть 4 різні прості та скалярні простори даних. Простір даних встановлює розміри для наборів даних.

Тепер заповнемо файл даними. Для цього створимо простори даних, які використовуються для задання розміру наборів даних. Створюємо простір даних завдяки функції `H5Screate_simple` і задаємо розмір і максимальний розмір простору. Також можна використовувати функцію `H5Scopy` для копіювання елемента.

Далі створюється датасет вже знайомим методом, але знову вже по іншому параметри задаються. Вказується місце де створюється датасет, назва датасету, тип даних і простір даних.

Після створення датасету, потрібно вписати дані в ці датасети. Спочатку задаємо в змінні значення векторів, а потім вписуємо їх в ці датасети, завдяки функції `H5Dwrite`, а щоб вивести їх необхідно спочатку прочитати, а потім вивести. Прочитати дані необхідно завдяки функції `H5Dread`, а вивід функцією `print`.

```
18 d = c(5,7)
19 #h5space1 = H5Screate_simple(d,d)
20 #h5space2 = H5Screate_simple(d,NULL)
21 #h5space3 = H5Scopy(h5space1)
22 #h5space4 = H5Screate("H5S_SCALAR")
23 h5space1
24 h5space2
25 h5space3
26 h5space4
27 #H5sis_simple(h5space1)
28
29 #h5dataset1 = H5Dcreate( h5file, "dataset1", "H5T_IEEE_F32LE", h5space1 )
30 #h5dataset2 = H5Dcreate( h5group2, "dataset2", "H5T_STD_I32LE", h5space1 )
31 h5dataset1 <- H5Dopen(h5file, "dataset1")
32 h5dataset2 <- H5Dopen(h5group2, "dataset2")
33 h5dataset1
34 h5dataset2
35
36 A = seq(0.1,3.5,length.out=5*7)
37 #H5Dwrite(h5dataset1, A)
38 B = 1:35
39 #H5Dwrite(h5dataset2, B)
40
41 data_from_dataset1 <- H5Dread(h5dataset1)
42 data_from_dataset2 <- H5Dread(h5dataset2)
43 print(data_from_dataset1)
44 print(data_from_dataset2)
45
```

Рис 5 – запис і вивід даних у файл

```

filename
  type H5T_IEEE_F32LE
  rank 2
  size 5 x 7
  maxsize 5 x 7

> h5dataset2
HDF5 DATASET
  name /baa/dataset2
  filename
    type H5T_STD_I32LE
    rank 2
    size 5 x 7
    maxsize 5 x 7

> A = seq(0.1,3.5,length.out=5*7)

> #H5Dwrite(h5dataset1, A)
> B = 1:35

> #H5Dwrite(h5dataset2, B)
>
> data_from_dataset1 <- H5Dread(h5dataset1)

> data_from_dataset2 <- H5Dread(h5dataset2)

> print(data_from_dataset1)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,] 0.1 0.6 1.1 1.6 2.1 2.6 3.1
[2,] 0.2 0.7 1.2 1.7 2.2 2.7 3.2
[3,] 0.3 0.8 1.3 1.8 2.3 2.8 3.3
[4,] 0.4 0.9 1.4 1.9 2.4 2.9 3.4
[5,] 0.5 1.0 1.5 2.0 2.5 3.0 3.5

> print(data_from_dataset2)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]  1    6   11   16   21   26   31
[2,]  2    7   12   17   22   27   32
[3,]  3    8   13   18   23   28   33
[4,]  4    9   14   19   24   29   34
[5,]  5   10   15   20   25   30   35

```

Рис 6 – результат виконання запису і виводу даних

Приклад виконання завдання 3.

Виведення інформації про сесію:

Після завершення маніпуляцій з файлом необхідно закрити все, датасети, простори даних, групи та сам файл, це робиться завдяки функції `H5Gclose` і вказанням що закриваємо. Також можна переглянути інформацію про сесію в якій ми перебуваємо.

```

46 H5Dclose(h5dataset1)
47 H5Dclose(h5dataset2)
48
49 #H5Sclose(h5space1)
50 #H5Sclose(h5space2)
51 #H5Sclose(h5space3)
52 #H5Sclose(h5space4)
53
54 H5Gclose(h5group1)
55 H5Gclose(h5group2)
56 H5Gclose(h5group3)
57
58 H5Fclose(h5file)
59
60 sessionInfo()

```

Рис 7 – код для закриття файлу та перегляду сесії

В результаті було отримана інформація про сесію, де виводиться, версія мови програмування та пристрій на якому воно виконується, місцезнаходження, активні пакети різних типів.

```
> H5Gclose(h5group1)
> H5Gclose(h5group2)
> H5Gclose(h5group3)
> H5Fclose(h5file)

> sessionInfo()
R version 4.4.1 (2024-06-14 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 11 x64 (build 22631)

Matrix products: default

locale:
 [1] LC_COLLATE=Russian_Ukraine.utf8  LC_CTYPE=Russian_Ukraine.utf8
 [3] LC_MONETARY=Russian_Ukraine.utf8 LC_NUMERIC=C
 [5] LC_TIME=Russian_Ukraine.utf8

time zone: Europe/Kiev
tzcode source: internal

attached base packages:
 [1] stats      graphics  grDevices  utils      datasets  methods   base

other attached packages:
 [1] rhdf5_2.50.0

Loaded via a namespace (and not attached):
 [1] utf8_1.2.4          R6_2.5.1           tidyselect_1.2.1   farver_2.1.2
 [5] magrittr_2.0.3     gtable_0.3.6       glue_1.7.0         tibble_3.2.1
 [9] rhdf5filters_1.18.0 pkgconfig_2.0.3    generics_0.1.3     dplyr_1.1.4
[13] Rhdf5lib_1.28.0    lifecycle_1.0.4    ggplot2_3.5.1      cli_3.6.3
[17] fansi_1.0.6        scales_1.3.0       patchwork_1.3.0    grid_4.4.1
[21] vctr_0.6.5         compiler_4.4.1     rstudioapi_0.17.1  tools_4.4.1
[25] munsell_0.5.1      pillar_1.9.0       colorspace_2.1-1   rlang_1.1.4
> |
```

Рис 8 – інформація про сесію

Висновок: Отже, в цій лабораторній роботі було, вивчено, як використовувати функцій низького рівня у R. Завдяки пакету ВіосManager та rhdf5 використовуючи файл розширенням h5.

Лабораторна робота №3

Назва: Зчитування даних з WEB.

Мета заняття: навчитись завантаження WEB-даних та обробляти їх у R.

Зв'язок з результатами навчання: Проводити систематизацію та узагальнення даних, виявляти ключові моменти.

Час проведення: 4 год – 3 год. виконання, 1 год. – захист .

Завдання для обов'язкового виконання:

Завдання 1. Виведіть перші 6 назв фільмів дата фрейму.

Завдання 2. Виведіть всі назви фільмів с тривалістю більше 120 хв.

Завдання 3. Скільки фільмів мають тривалість менше 100 хв.

Пояснення, аналіз, приклад.

В цій лабораторній роботі необхідно зчитати WEB сторінку з сайту IMDb.com зі 100 фільмами починаючи з 2000 року (1 варіант – 2000 рік, 2 варіант – 2001 рік і так далі, варіанти по номеру у списку групи) виходу за посиланням

«http://www.imdb.com/search/title?count=100&release_date=2019,2019&title_type=feature».

Необхідно створити data.frame «**movies**» з наступними даними: номер фільму (**rank_data**), назва фільму (**title_data**), тривалість (**runtime_data**). Для виконання лабораторної рекомендується використати бібліотеку «**rvest**». CSS селектори для зчитування необхідних даних: **rank_data**: «**.text-primary**», **title_data**: «**.lister-item-header a**», **runtime_data**: «**.text-muted .runtime**». Для зчитування **url** використовується функція **read_html**, для зчитування даних по CSS селектору – **html_nodes** і для перетворення зчитаних html даних в текст - **html_text**. Рекомендується перетворити **rank_data** та **runtime_data** з тексту в числові значення. При формуванні дата фрейму функцією data.frame рекомендується використати параметр «**stringsAsFactors = FALSE**».

Приклад виконання :

Завдання 1. Виведіть перші 6 назв фільмів дата фрейму.

Щоб запарсити сайт по фільмам, необхідно виконати функцію **read_html** вказавши адресу сайту. Для завдання необхідно витягнути назви фільмів та їх тривалість. Знаходимо на сайті який тег і атрибути має місце де зберігається ця інформація і вказуємо у функції **html_nodes** та виконуємо після цього функцію **html_text** для витягування необхідної інформації. Далі прибираємо зайву

інформацію про рік та вікове обмеження. Інформація про тривалість зберігається у форматі години та хвилини з позначкою, що години, що хвилини. Треба конвертувати у хвилини, для цього Витягуємо тільки години тільки числа і так само хвилини після цього години множимо на 60 і додаємо хвилин, так ми конвертуємо у хвилини тривалість фільму. Реалізація продемонстрована на рис 9.

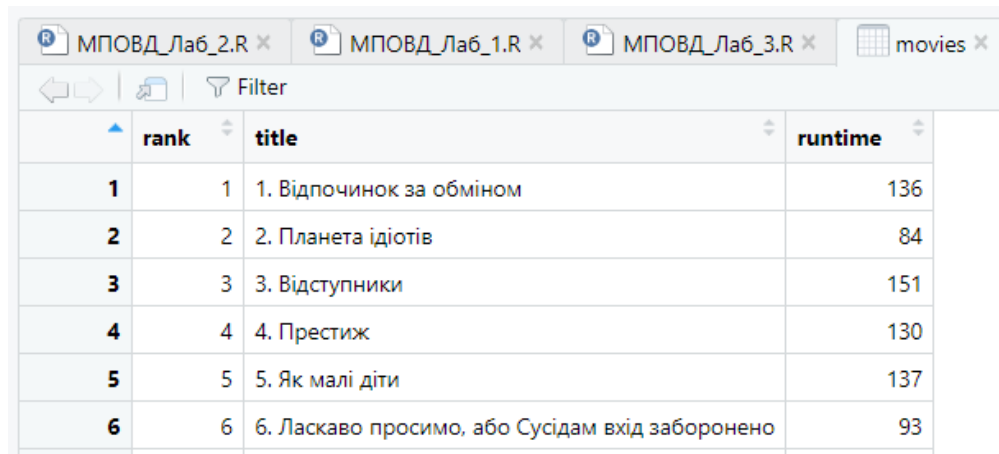
Далі заповнимо відсутніми значеннями зайві назви, щоб не було помилки, після цього обробка даних завершена і додаємо дані до датафрейму. Вивід можна побачити на рис 11, а як було реалізовано на рис 10.

```
6 webpage <- read_html(url)
7 print(webpage)
8 rank_data = 1:26
9
10 title_data <- webpage %>%
11   html_nodes("h3.ipc-title__text") %>%
12   html_text()
13
14 metadata <- webpage %>%
15   html_nodes(".sc-300a8231-7.eaXxft.dli-title-metadata-item") %>%
16   html_text()
17
18 runtime_data <- metadata[grep("^([0-9]+h|([0-9]+m)$", metadata)]
19
20 print(rank_data)
21 print(title_data)
22 print(runtime_data)
23 cat("Довжина rank_data:", length(rank_data), "\n")
24 cat("Довжина title_data:", length(title_data), "\n")
25 cat("Довжина runtime_data:", length(runtime_data), "\n")
26
27 convert_runtime <- function(runtime) {
28   hours <- 0
29   minutes <- 0
30
31   if (grep("([0-9]+)h", runtime)) {
32     hours <- as.numeric(gsub("h.*", "", runtime))
33   }
34
35   if (grep("([0-9]+)m", runtime)) {
36     minutes <- as.numeric(gsub(".*?([0-9]+)m.*", "\\1", runtime))
37   }
38
39   return(hours * 60 + minutes)
40 }
41
```

Рис 9 – обробка даних з веб-сторінки

```
44 if (length(runtime_data) < length(rank_data)) {
45   runtime_data <- c(runtime_data, rep(NA, length(rank_data) - length(runtime_data)))
46 }
47
48 movies <- data.frame(
49   rank = rank_data,
50   title = title_data,
51   runtime = runtime_data,
52   stringsAsFactors = FALSE
53 )
54
```

Рис 10 – завершення обробки та заповнення датафрейму



	rank	title	runtime
1	1	1. Відпочинок за обміном	136
2	2	2. Планета ідіотів	84
3	3	3. Відступники	151
4	4	4. Престиж	130
5	5	5. Як малі діти	137
6	6	6. Ласкаво просимо, або Сусідам вхід заборонено	93

Рис 11 – створений датафрейм

Далі виконуємо встановлені завдання. Щоб вивести перші 6 фільмів, необхідно виконати команду `head`, яка виводить кількість перших значень вказаних у параметрах, також треба вказати датафрейм, який треба вивести, через символ `$` вказуємо поле, яке треба вивести.

Завдання 2. Виведіть всі назви фільмів с тривалістю більше 120 хв.

Друге завдання потрібно вивести всі фільми яку тривали більше двох годин. Це виконується завдяки функції `subset`, яка робить вибірку за заданих умовах. Треба вказати датафрейм, умову, а також вказуємо назву поля, яке потрібно вивести.

Завдання 3. Скільки фільмів мають тривалість менше 100 хв.

Трете завдання полягає у виведенні кількості фільмів тривалістю менше 100 хвилин. Для виведення кількості необхідно виконати функцію `length`. Функція `which` повертає індекси елементів, де умова є правда за певних умов, тому будемо рахувати кількість істин і це буде кількістю фільмів менше 100 хвилин.

```
58 head(movies, 6)$title
59 subset(movies, runtime > 120)$title
60 length(which(movies$runtime < 100))
```

Рис 12 – код для виконання вибірок

```
> head(movies, 6)$title
[1] "1. Відпочинок за обміном"
[2] "2. Планета ідіотів"
[3] "3. Відступники"
[4] "4. Престиж"
[5] "5. Як малі діти"
[6] "6. Ласкаво просимо, або Сусідам вхід заборонено"

> subset(movies, runtime > 120)$title
[1] "1. Відпочинок за обміном"
[2] "3. Відступники"
[3] "4. Престиж"
[4] "5. Як малі діти"
[5] "8. 007: Казино Рояль"
[6] "10. Апокаліпсис"
[7] "15. Пірати Карибського моря: Скриня мерця"
[8] "18. Життя інших"
[9] "23. Сайлент Гілл"
[10] "24. Місія неможлива 3"

> length(which(movies$runtime < 100))
[1] 4
> |
```

Рис 13 – результат виконання вибірок

Висновки: Отже, в цій лабораторній роботі було, вивчено, як завантажити WEB-дані та перетворювати і виводити їх у R.

Лабораторна робота №4

Назва: Навчитись використовувати можливості штучного інтелекту (ШІ) (GPT-4, COPILOT, GEMINI) створених на основі нейромереж (НМ).

Мета заняття: Ознайомлення із можливостями штучного інтелекту.

Зв'язок з результатами навчання: Проводити систематизацію та узагальнення даних, виявляти ключові моменти за допомогою сучасних інструментів.

Час проведення: 4 год – 3 год. виконання, 1 год. – захист .

Завдання для обов'язкового виконання:

1. Налаштування середовища: Провести підготовку роботи та покращити робоче середовище для використання ШІ.
2. Запити до ШІ (GPT-4, COPILOT, GEMINI). Виконати запити до моделі ШІ і дослідити, які відповіді генеруються на різні запити.
3. Генерація тексту: Провести деякі бесіди по темі практичної або ключові слова по ній, задавши ролі у чаті і використати обрану модель ШІ для генерації тексту на обрану для практичних 1, 2 тему.
4. Аналіз відповідей: Проаналізувати та порівняти відповіді двох ШІ, та оцінити їх релевантність та якість.
5. Застосування в різних галузях: Розглянути можливість використання ШІ в різних сферах, таких як медицина, освіта, бізнес, наукові дослідження та інше.
6. Звіт і презентація: Підготувати звіт, який включає опис проведених досліджень та демонстрацію роботи чат-бота. Також підготувати коротку презентацію для представлення результатів лабораторної роботи.

Приклад виконання:

Для виконання лабораторної роботи було використано модель GPT-4o mini. Спочатку спробуємо зробити різні запити до чату, наприклад якщо задати питання загального характеру та якщо вузькопрофільний запит відбувся.

Судячи з результатів запитів які були отримані, то можна прийти до висновку, що при питаннях загального характеру отримаємо короткий опис

тільки головних фактів про цікаве нам питання. Якщо задати питання вузькопрофільне, то буде детальний опис, але тільки по конкретній темі.

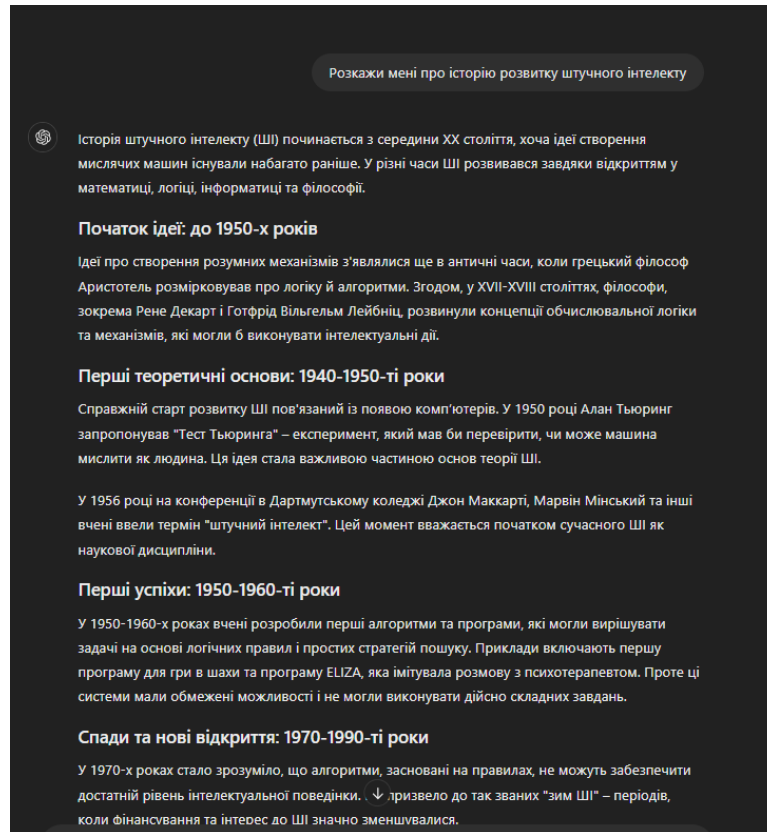


Рис 14 – результат запиту 1

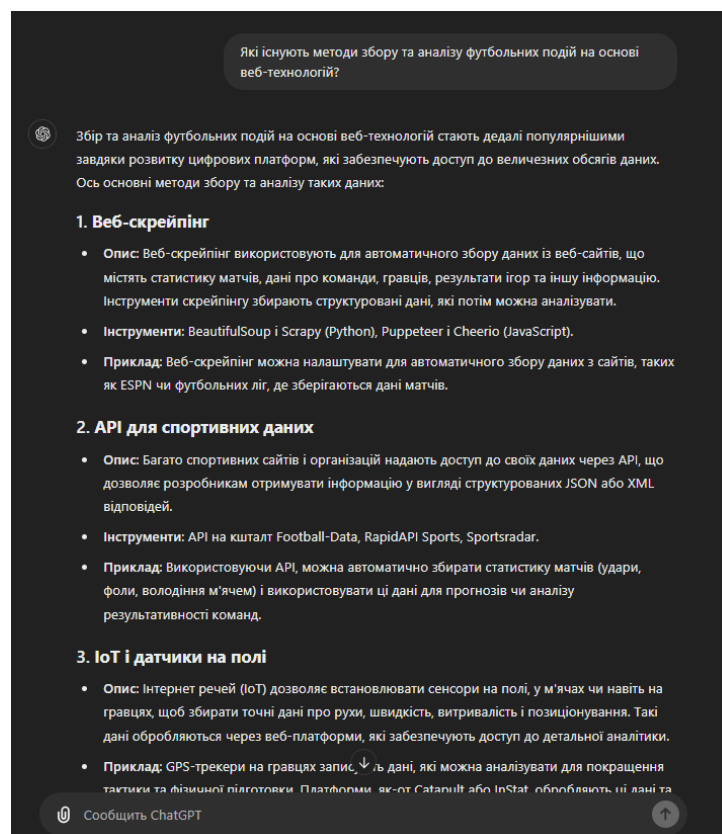


Рис 15 – результат запиту 2

Тепер спробуємо задати питання стосовно теми практичних робіт, а саме футбольний клуб «Динамо Київ», але із контекстом яку роль ми виконуємо, наприклад тренер, фанат, журналіст або аналітик.

В результаті коли давався запит про цей клуб із завданням однієї із чотирьох ролей, то при запиті в ролі тренера методів тренувань, ШІ надав базові методи і короткий їх опис. При відповіді фанату на теж запитання, ШІ розповів красиво які технології використовуються, але не надав базу.

Коли я в якості журналіста задав питання чим визначний клуб Динамо Київ, то мені ШІ почав розповідати історію клубу, а коли як аналітик, то видав чим краще цей клуб за інші команди України.

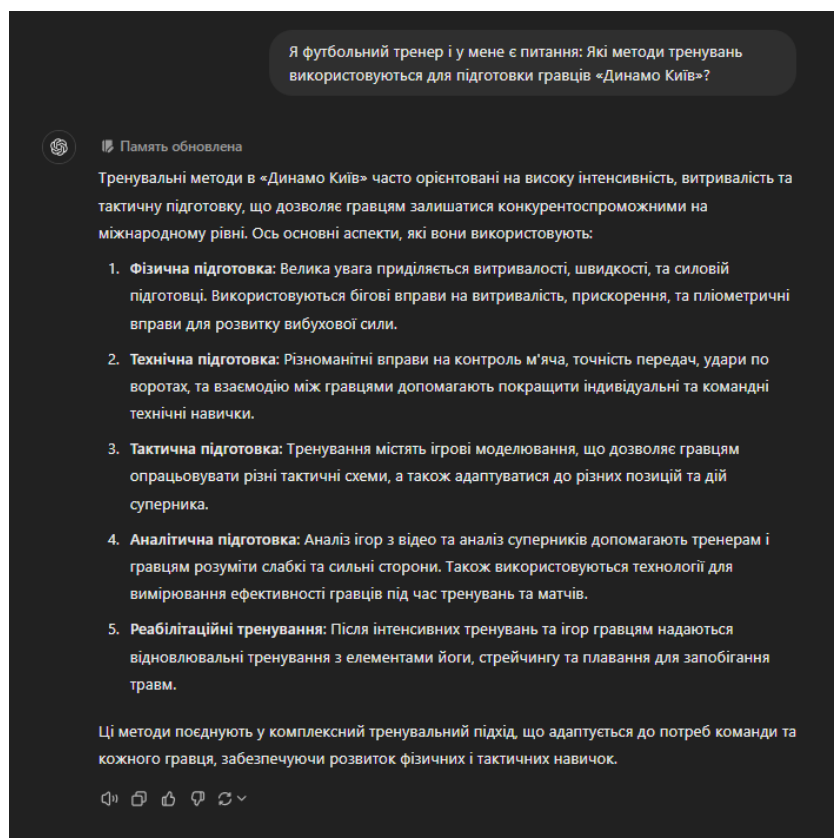


Рис 16 – результат запиту з роллю футбольний тренер

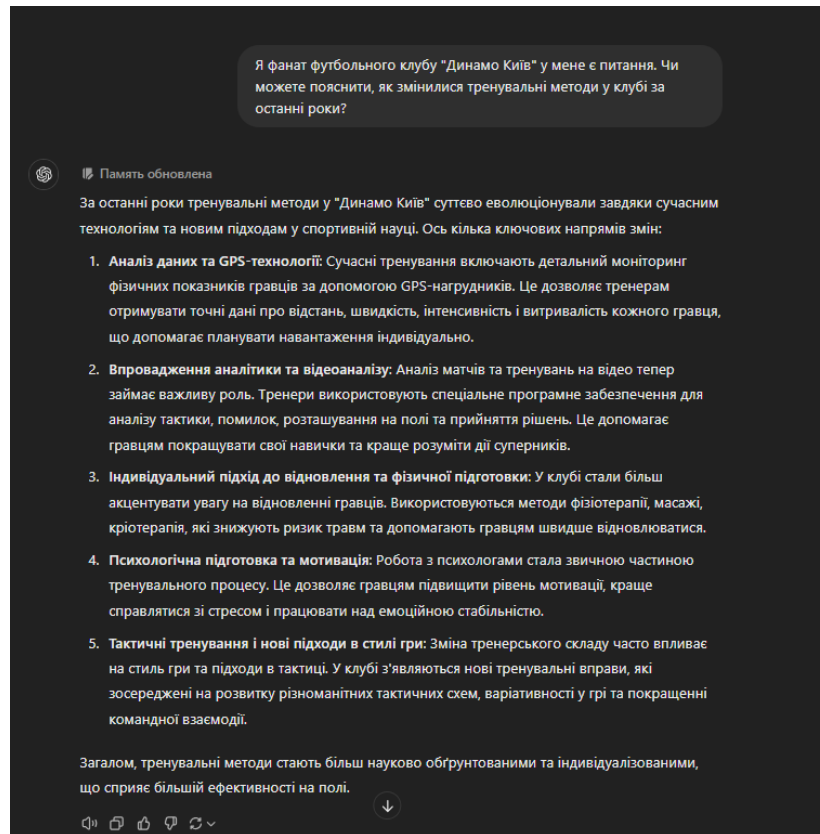


Рис 17 – результат запиту з роллю фаната

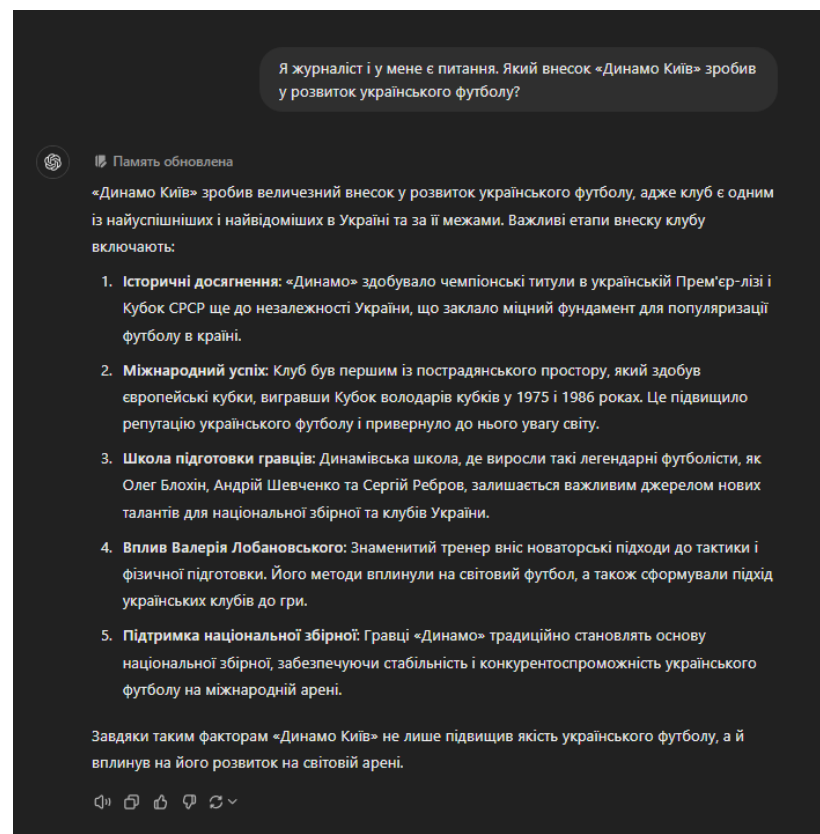


Рис 18 – результат запиту з роллю журналіст

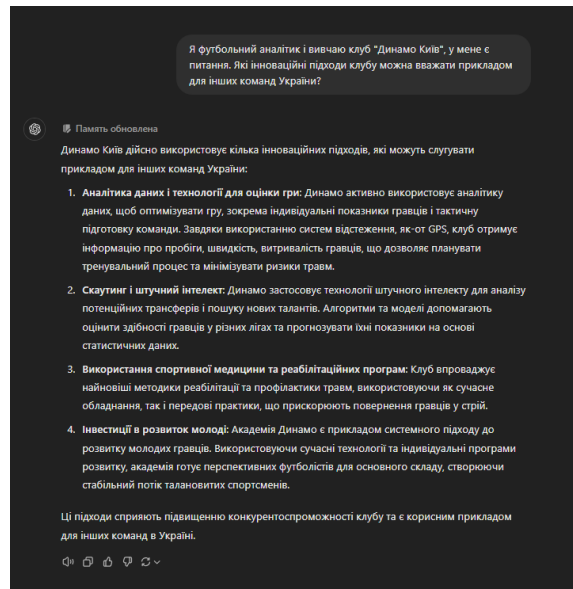


Рис 19 – результат запиту з роллю футбольного аналітика

Спробуємо для порівняння двох запитів задати по суті одне і те саме питання, але перше питання буде в простій формі, а друге буде більш в складному форматі. Було задано питання що таке машинне навчання і питання основні алгоритми в машинному навчанні і як вони працюють.

В результаті отримуємо, що при першому питанні видає короткий опис що це і деякі важливі пункти. А при другому питанні видає детальний опис конкретніше що цікавить, а саме алгоритми машинного навчання.

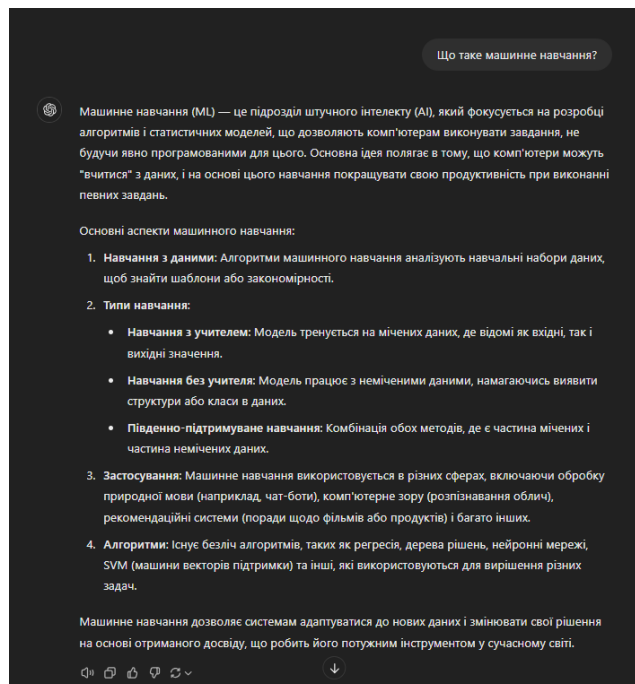


Рис 20 – результат запиту з простим питанням

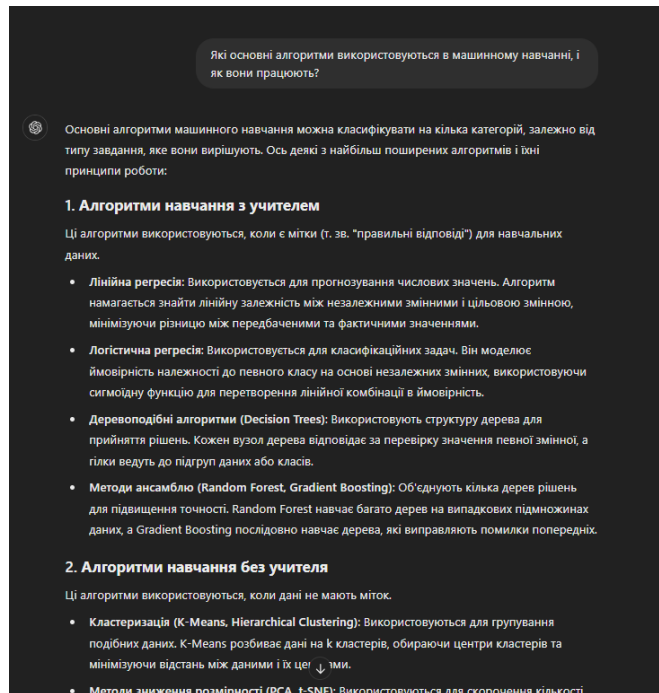


Рис 21 – результат запиту з питанням більш розширеним

В медицині є багато моментів в яких може бути корисним ШІ. На рисунку на слайді можна побачити опис як може допомогти чат.

Також треба зазначити, що він попереджає, що не треба сліпо вірити ШІ. В сфері медицини, щоб підтвердити інформацію потрібно, щоб лікар перевіряв цю інформацію, бо брехня може коштувати життям.

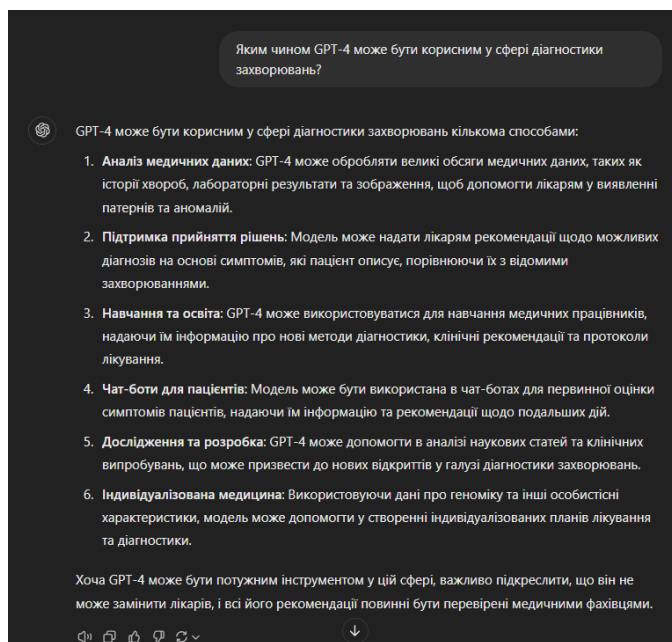


Рис 22 – результат запиту на застосування у медицині

GPT досить потужний інструмент для вивчення в різних галузях, що робить його корисним для освіти і швидкому навчанню, бо замість довгого

пошуку потрібної інформації, чат може надавати швидкі відповіді, але все ж необхідно перевіряти правдивість наданої інформації і ставитись до неї критично. На малюнку у цьому слайді продемонстровані пункти, які чат може бути корисним в освіті.

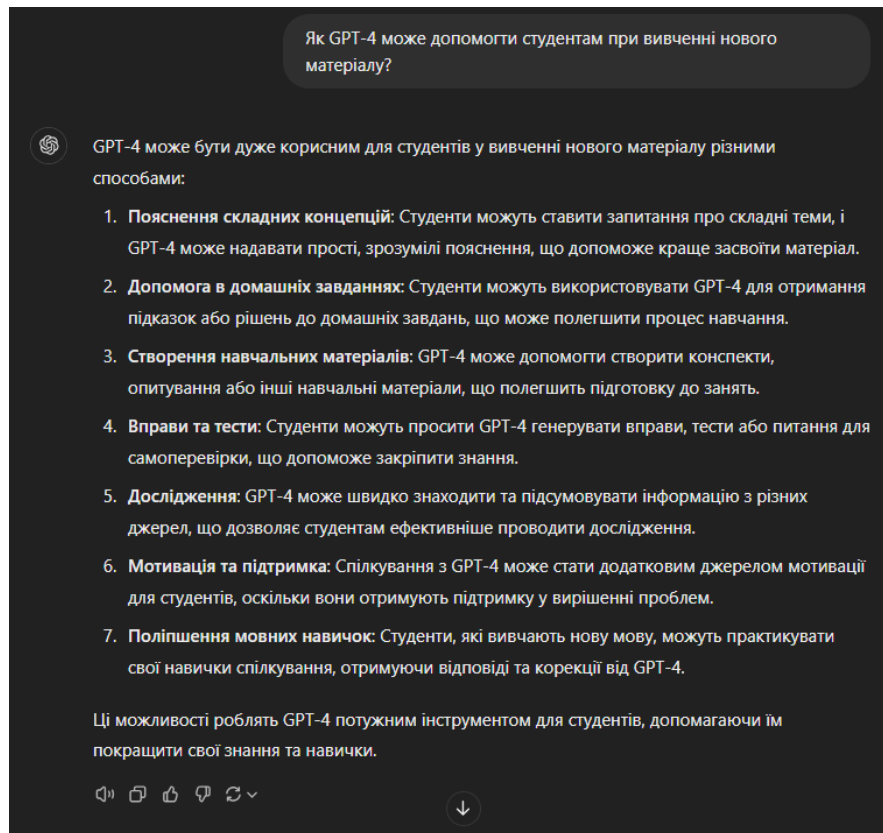


Рис 23 – результат запиту на застосування у навчанні

GPT для бізнесу може бути дуже корисним інструментом, в особливості у роботі з клієнтами. Це дозволить більш ефективно працювати з клієнтами, менше витратити коштів і звісно покращувати якість цього обслуговування.

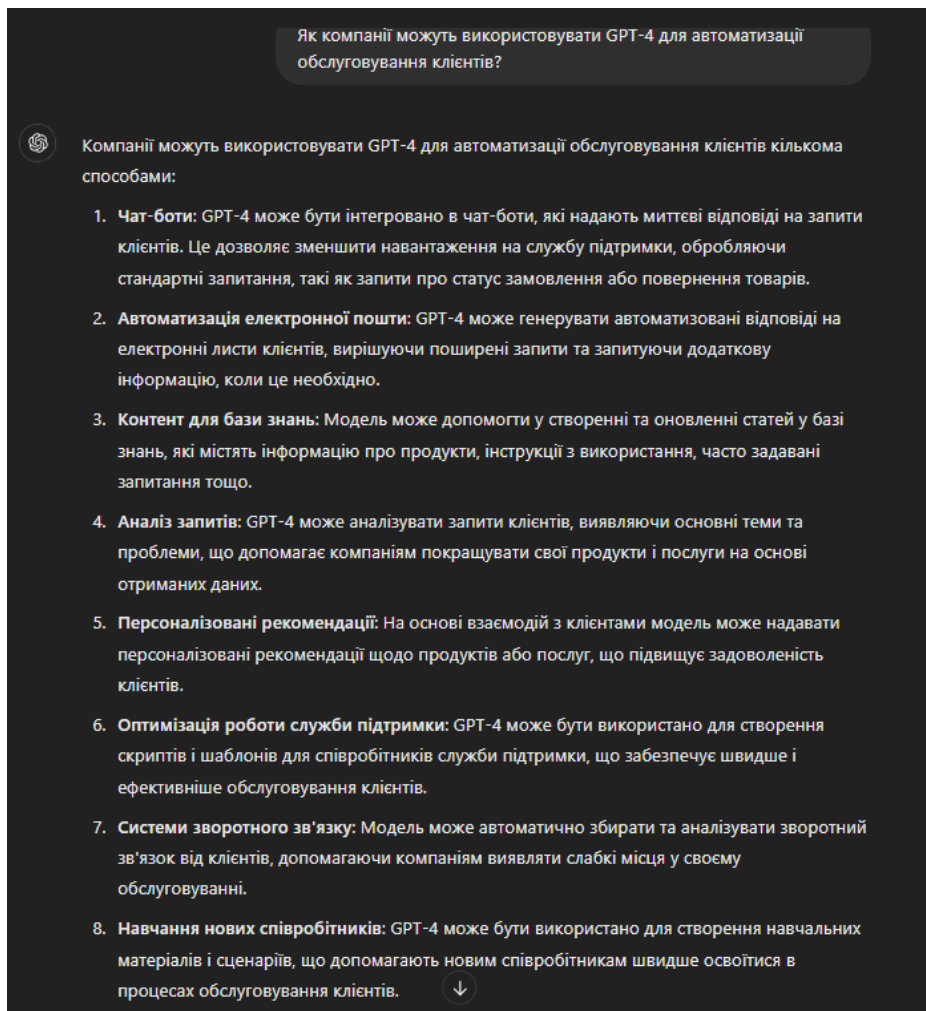


Рис 24 – результат виконання запиту на застосування у бізнесі

Висновки: Отже, в цій лабораторній роботі було, ознайомлено із можливостями штучного інтелекту на основі моделі GPT-4 та досліджено, як ця технологія може бути використана в різних сферах. Також визначено, що для коректної роботи потрібно додавати контекст для більшої ймовірності коректної відповіді на питання яке саме цікавить користувача.

Лабораторна робота №5

Назва: Обробка та візуалізація даних за допомогою MS Power BI.

Мета заняття: навчитись обробляти дані та візуалізувати їх у Power BI.

Зв'язок з результатами навчання: Проводити систематизацію та узагальнення даних, виявляти ключові моменти по візуалізації.

Час проведення: 7 год – 4 год. виконання, 3 год. – захист .

Завдання для обов'язкового виконання:

Завдання 1. Завантажте файл з набором даних за посиланням: <https://drive.google.com/file/d/1zWMFjR9hgxJ-g-TMuYvr0R-cuSffQL8X/view?usp=sharing>. Оберіть регіон України як варіант для проведення лабораторної роботи.

Завдання 2. Провести аналіз по рокам і типам підприємств обраного регіону по відношенню до всіх регіонів України в цілому, підготуйте табличне представлення та завантажте дані у MS Power BI.

Завдання 3. Візуалізуйте дані підприємств обраного регіону та всіх регіонів України, у вигляді діаграми MS Power BI із зв'язком її до таблиці по роках.

Пояснення, аналіз, приклад.

Для заповнення пропущених даних у таблицях скористайтесь двома з перелічених алгоритмів: заповнення пропущених параметрів за допомогою середнього значення; інтерполяція за сусідніми точками; середнє за n сусідніми точками; медіана N за найближчими значеннями; метод найвішого прогнозу; заповнення пропусків з використанням лінійної регресії; метод розрахунку середнього за відповідною датою; метод пошуку відсотка від максимуму або іншої знакової величини та ін. Поясніть чому використали саме ці методи заповнення пропущених даних.

Коротко про методи заповнення даними:

Метод середнього значення. У разі вибору цього методу всі відсутні дані просто замінюються середнім арифметичним значенням для всіх спостережень. Цей метод може не підходити, коли ряд не постійний або коли є великі систематичні коливання у змінних ряду. З іншого боку, повне середнє часто є кращим апіорним (неупередженим) припущенням для відсутніх даних. Цей метод має багато переваг, проте не володіє гнучкістю до несезонних змін і навіть може призвести до труднощів у визначенні лінії тренду, зробити її менш помітною для розпізнавання.

Метод наївного прогнозу. Наївний прогноз полягає в тому, що деякий основний період прогнозованого ряду найкраще описує майбутнє цього ряду. Тому моделі наївного прогнозу, як правило, є простою функцією від значень прогнозованої змінної у близькому минулому. Найпростішою моделлю наївного прогнозу є відповідність припущенню, що "завтра буде як сьогодні":

$$Y_{t+1}=Y_t ,$$

де: Y_t – відоме значення, а Y_{t+1} – прогнозоване значення. У роботі застосовано наївний прогноз з використанням найпростішої функції. Основним недоліком наївного прогнозування є дуже низька точність прогнозу.

Відновлення пропусків на основі регресійних моделей. Методи лінійної регресії дають змогу отримати правдоподібно заповнені дані. Однак реальним даним властивий деякий розкид значень, який у разі заповнення пропусків на основі лінійної регресії відсутній. Як наслідок, варіація значень характеристики стає меншою, а кореляція між двома характеристиками штучно посилюється. Тому цей метод заповнення пропусків є ефективнішим, чим нижча варіація значень характеристики, пропуски в якій потрібно заповнити, і чим нижчим є відсоток пропущених параметрів від загальної кількості масиву даних.

Надано п'ять листів Excel-файлу:

1. «Кількість підприємств» - Кількість підприємств з розподілом на великі, середні, малі та мікропідприємства по регіонах у 2010-2019 роках
2. «Обсяг-регіон» - Обсяг реалізованої продукції (товарів, послуг) підприємств з розподілом на великі, середні, малі та мікропідприємства по регіонах у 2010-2019 роках
3. «Оплата праці» - Витрати на оплату праці підприємств з розподілом на великі, середні, малі та мікропідприємства по регіонах у 2010-2019 роках
4. «Витрати на персонал» - Витрати на персонал підприємств з розподілом на великі, середні, малі та мікропідприємства по регіонах у 2010-2019 роках
5. «Кількість зайнятих працівників» - Кількість зайнятих працівників на підприємствах з розподілом на великі, середні, малі та мікропідприємства по регіонах у 2010–2019 роках

Кожен з яких має схожу структуру (колонки):

1. Роки - перераховано роки 2010-2019 для кожної області
2. Дані по великих підприємствах
3. Дані по підприємствах середнього розміру
4. Дані по підприємствах малого розміру та мікропідприємствах
5. Дані окремо по мікропідприємствах

Приклад виконання :

Приклад виконання завдання 1: Переходимо за посиланням і скачуємо файл з даними «Вхідні дані ЛР10». Обираємо регіон для подальшої обробки та візуалізації.

Приклад виконання завдання 2: Початкові дані містять пропущені значення. Для того, щоб зменшити їх вплив при візуалізації потрібно заповнити кожне відсутнє значення двома обраними методами.

Було:

Тернопільська	2010	к	к	440719,0	120374,7
	2011	547815,4	1163574,8	618399,8	180817,1
	2012	553186,1	1497862,9	669606,5	202893,4
	2013	691835,0	1430111,1	695088,3	201125,6
	2014	586203,6	1695690,1	741568,6	217425,7
	2015	к	к	862419,0	249246,1
	2016	к	к	999483,7	293706,0
	2017	к	к	1479566,2	379262,2
	2018	1480819,4	4155422,0	1877693,3	571608,2
	2019	к	к	2217834,6	652164,3

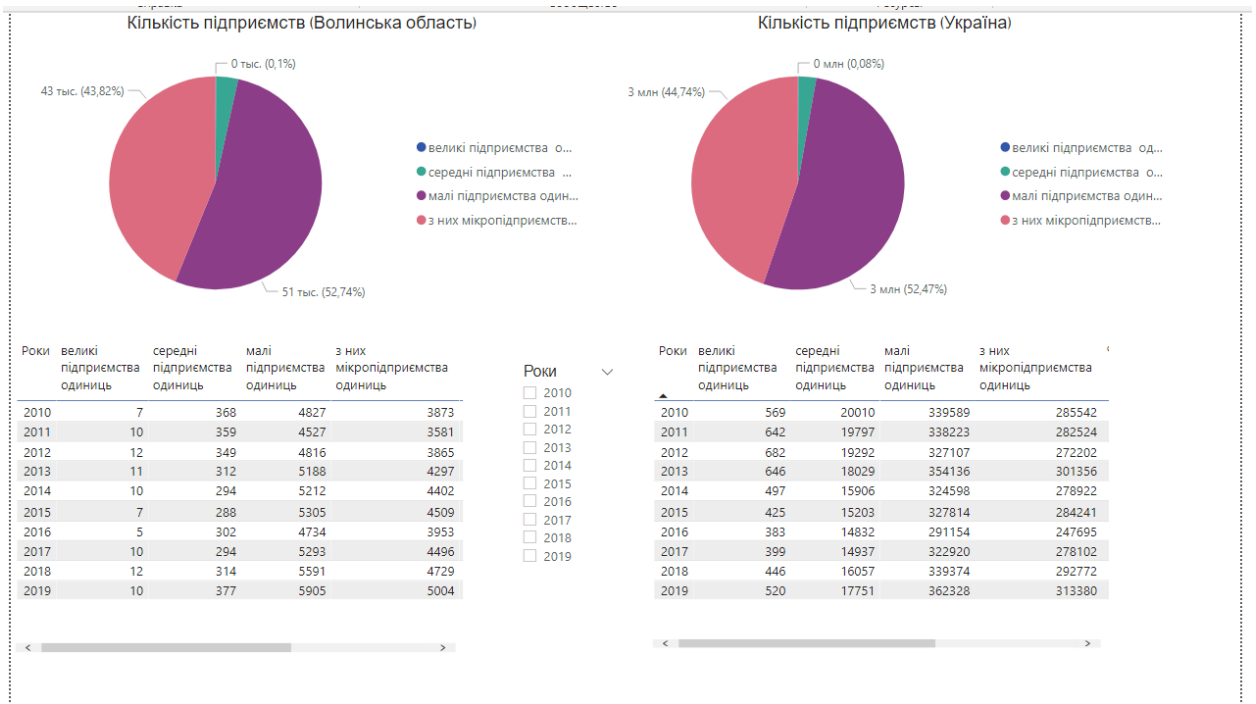
Після заповнення:

Тернопільська	2010	771971,9	1988532,2	440719,0	120374,7
	2011	547815,4	1163574,8	618399,8	180817,1
	2012	553186,1	1497862,9	669606,5	202893,4
	2013	691835,0	1430111,1	695088,3	201125,6
	2014	586203,6	1695690,1	741568,6	217425,7
	2015	771971,9	1988532,2	862419,0	249246,1
	2016	771971,9	1988532,2	999483,7	293706,0
	2017	771971,9	1988532,2	1479566,2	379262,2
	2018	1480819,4	4155422,0	1877693,3	571608,2
	2019	771971,9	1988532,2	2217834,6	652164,3

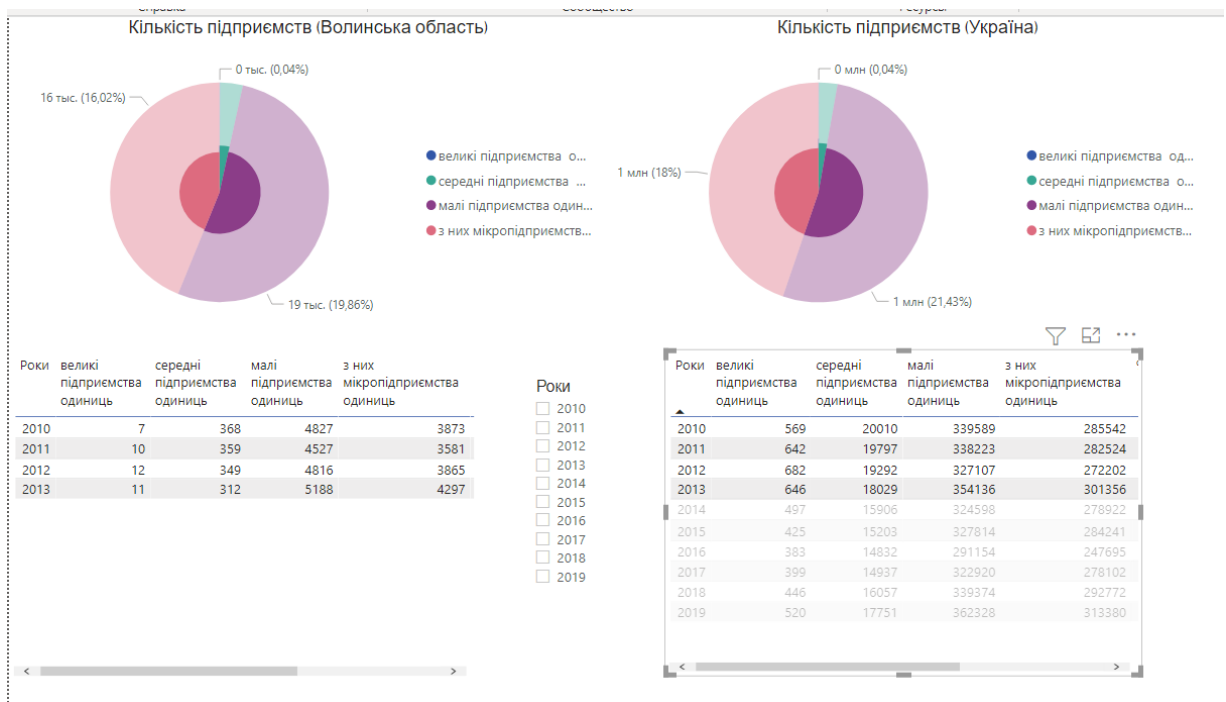
У даному випадку скористались методом середнього значення.

Приклад виконання завдання 3:

Для візуалізації використали кругову діаграму у кожній діаграмі власна таблиця знизу.



При виборі певних років, відображення обраного з'являється на усіх діаграмах одразу.



Поясніть свій вибір візуального супроводу у MS Power BI.

ПРАКТИЧНІ РОБОТИ

Практична робота №1

Назва: Пошук інформації для її обробки по обраних індивідуальних варіантах завданнях.

Мета: Набути практичних навичок роботи з методами пошуку інформації.

Зв'язок з результатами навчання: Виявляти особливості застосування методів пошуку даних у мережі Інтернет, та створення запитів.

Час проведення: 3 год. – 2 год виконання, 1 год. – захист.

Завдання для обов'язкового виконання:

1. Оберіть тему для практичної роботи у Додатку А, чи запропонуйте свою викладачеві.
2. Проведіть інформаційний пошук по обраній темі в Інтернеті, результати збережіть.

Пояснення, аналіз, приклад

Інформаційний пошук (англ. Information retrieval) — наука про пошук неструктурованої інформації.

У Всесвітній Мережі пошук інформації здійснюється двома способами:

- За допомогою URL адреси;
- За допомогою пошукових служб;
- За допомогою ШІ (GPT-4, COPILOT, GEMINI);
- За допомогою парсингу сайтів.

Перший спосіб найшвидший і простий, якщо знаєте де шукати (URL адреса). Для виклику документа, що містить потрібну вам інформацію, необхідно в адресному рядку браузера набрати адрес цієї сторінки або сайту: (ввести URL адресу в полі Адреса).

Другий спосіб з використання пошукових серверів. Він універсальний, дозволяє "знайти" будь-яку потрібну вам інформацію. Саме тут користувач повинен продемонструвати свій рівень розуміння задачі і формулювання мети.

За принципом дії розрізняють кілька видів пошукових служб. Найбільш поширеними є веб-каталоги та пошукові системи.

Веб-каталоги. Веб-каталоги (інша назва тематичні каталоги) містять базу даних посилань на веб-сайти, які розподілені на окремі тематичні рубрики. Наприклад, до однієї рубрики віднесені усі посилання, що стосуються подорожей, до іншої – комп'ютерів, до третьої – бізнесу тощо. У свою чергу, кожна рубрика поділяється на підрубрики. Так, у рубриці Бізнес можна виділити підрубрики Промисловість, Будівництво, Транспорт тощо.

Окремі пошукові каталоги поруч із посиланнями на ресурс надають оцінку його популярності у користувачів. Такі каталоги називають рейтинговими системами.

Заповнюють базу даних каталогу, як правило, вручну працівники служби. Вони переглядають веб-сторінки, обирають ті, що на їх погляд можуть представляти інтерес для користувачів, заносять URL-адреси таких веб-сторінок з коротким описом до відповідної рубрики бази даних.

Один і той самий сайт може бути віднесений до кількох рубрик. Число сайтів в Інтернеті величезне, тому при ручному опрацюванні веб-сторінок база даних інформаційнопошукової системи веб-каталогів наповнюється й оновлюється досить повільно. Власники сайтів для популяризації своїх ресурсів можуть самостійно зареєструвати свій сайт у веб-каталогах.

У різних каталогах рубрики можуть називатися по різному, бази даних можуть суттєво відрізнятися за обсягом, працівниками пошукових служб можуть опрацьовуватися різні ресурси Всесвітньої павутини. Тому результати пошуку на різних веб-каталогах будуть відрізнятися один від одного. Якщо потрібна інформація не знайдена в одному тематичному каталозі, то можна спробувати її знайти в іншому.

Пошукові системи, або іншими словами індексні бази даних або пошукові сервери. Бази даних пошукових систем містять значно більше посилань на веб-сторінки, ніж бази даних тематичних каталогів. Це досягається завдяки тому, що опрацювання веб-сторінок виконують не люди, а спеціальні програми, які називають пошуковими роботами. Ці програми переглядають вміст веб-сторінок, визначають слова, які найчастіше зустрічаються у тексті та зберігають результати опрацювання вмісту сторінок у базі даних пошукової системи.

Пошукова система (ПС) – спеціалізований ресурс, який використовується для пошуку необхідної користувачам інформації в мережі Інтернет. Основою будь-якої пошукової системи є складний програмний комплекс, що працює на базі пошукових алгоритмів, що постійно оновлюються і модифікуються.

У змаганні за популярність пошукові системи намагаються надати якомога більше послуг своїм користувачам. Серед них:

- пошук ресурсів не лише за точним формулюванням пошукового запиту, а й з урахуванням різних форм слів, їх синонімів;
- складний (розширений) пошук: за точною фразою, з обов'язковими чи забороненими словами, за датою публікації сторінок тощо;
- пошук різних видів повідомлень та об'єктів – зображень, новин, блогів, мап, музики, відео тощо;
- пошук сторінок за регіональними ознаками – певною мовою або зареєстрованих у визначених територіальних доменах;

- створення власних сторінок користувачів і користувацької пошукової системи;
- перевірка правопису та переклад веб-сторінок;
- пошук означень в енциклопедіях;
- пошук інших сторінок, які посилаються на задану сторінку та інші.

Практична робота №2

Назва: Представлення результатів інформаційного пошуку.

Мета: Навчитися інформативно представляти результати інформаційного пошуку, обробляти дані для кращого візуального сприйняття.

Зв'язок з результатами навчання: Проводити систематизацію та узагальнення даних, виявляти зв'язки між ними. Виявляти основні поняття, відношення і характеристики даних по темі дослідження.

Час проведення: 4 год. – 3 год виконання, 1 год. – захист.

Завдання для обов'язкового виконання:

Завдання 1. Обробити відомими методами зібрані у практичній роботі №1 дані для створення презентації результату інформаційного пошуку.

Завдання 2. Використовуючи власний досвід представлення інформації, донести основні повідомлення результату інформаційного пошуку.

Завдання 3. Презентувати і захистити результат.

Самостійна робота

Мета та завдання самостійної роботи студентів. Студенти здійснюють підготовку лабораторних та практичних робіт до захисту, оформлюють роботи у вигляді звітів. Основний зміст самостійної роботи студентів з дисципліни “Методи пошуку обробки та візуалізації даних” полягає у вивченні та застосуванні системи знань у галузі теорії та практики пошуку та обробки даних, які застосовуються при виконанні лабораторних і практичних робіт.

Форми організації самостійної роботи. Індивідуальна

Зміст самостійної роботи. Самостійна робота передбачає також вивчення методичних вказівок до лабораторних робіт і ознайомлення із додатковою літературою, пов’язаною з виконанням цих робіт.

Для більшого розуміння перетворення даних на самостійну роботу виноситься такі теми:

1. Алгоритми роботи пошукової системи Google.
2. Установка та основи роботи з ПЗ MS Power BI.
3. Сучасні методи обробки великих масивів даних.
4. Бази знань та їх використання у експертних системах.
5. Робота із запитамі до сучасних чатботів, таких як GPT-4, COPILOT, GEMINI.

Тільки постійне самостійне навчання дає можливість якомога ближче наблизитися до вершини знань певної галузі, оволодіти такою сумою знань і вмінь, які б дали змогу заявити про себе як про професіонала. Студент, який хоче оволодіти дисципліною, має добре розуміти: на занятті викладач подає основи знань, навчає, як учити, підкреслює ті ключові істини дисципліни, які пробуджують у молодій людини потяг до поглиблення й удосконалення усіх знань. Збагачення загальною сумою знань, накопичених людством, розширення загального світогляду, усвідомлення наявної перспективи щодо реалізації певних знань є основним мотивом до сумлінного навчання.

Обсягові питання на іспит

1. Теорія інформації. Історія. Визначення.
2. Етапи сприйняття інформації.
3. Аксиоми інформації.
4. Інформаційне забруднення.
5. Базові закони інформації.
6. Властивості та види інформації.
7. Підходи теорії інформації.
8. Інформаційний пошук. Визначення.
9. Види та методи пошуку інформації.
10. Інформаційно-пошукові системи. Визначення.
11. Організація та технологія пошуку інформації. Критерії.
12. Переробка інформації – це
13. Представлення інформації - це
14. Стратегії інформаційного пошуку.
15. Назвіть відомі Вам методи обробки інформації.
16. Бази знань. Визначення. Класифікація.
17. Класифікація знань по категоріях.
18. Процедурні і декларативні знання.
19. Стратегії здобуття знань.
20. Опишіть роботу інтерпретатора у базі знань.
21. Експертна система – це
22. Опишіть елементи експертної системи.
23. Алгоритми пошуку Google.
24. Пошукова оптимізація.
25. Data Mining. Визначення.
26. Етапи Data Mining.
27. Компоненти Data Mining.
28. Класифікація Data Mining.
29. Представлення інформації – це
30. Обробка інформації- це
31. Гейміфікація. Підходи та переваги.
32. Інформаційний пошук. Визначення.
33. Класифікація даних за способом їх отримання, опишіть.
34. Представлення інформації – це
35. Обробка інформації- це
36. По яких основних критеріях класифікують дані?
37. Класифікація по числу змінних, опишіть.

- 38.Класифікація даних за наявністю або відсутністю упорядкування в часі, опишіть.
- 39.Класифікація даних по типу шкали вимірювання ознаки, опишіть.
- 40.Назвіть і коротко опишіть методи обробки інформації.

Список рекомендованої літератури

Основна:

1. Положення про організацію освітнього процесу у Київському національному університеті імені Тараса Шевченка- К.: КНУ імені Тараса Шевченка, 2018. – 112 с.
2. Teslia, Iurii & Yehorchenkova, Nataliia & Yehorchenkov, Oleksii & Khlevna, Iulia & Kataieva, Yevheniia & Klievanna, Ganna & Khlevnyi, Andrii & Ivanov, Ivan & Sazonov, Anton. (2023). Development of a multilingual intelligent project planning and monitoring system. Eastern-European Journal of Enterprise Technologies. 2. 82-94. 10.15587/1729-4061.2023.277618.
3. Khlevna, I., Teslia, I., Yehorchenkova, N., Yehorchenkov, O., Kataieva, Y., Jamečný, Lubomír, Khlevnyi, A., Latysheva, T., Veretelnyk, V., & Ohirko, I. (2023). Development of the concept of building a corporate standard of portfolio management in the course of territory restoration planning in the context of Making-City project . Eastern-European Journal of Enterprise Technologies, 4(3 (124), 6–18. <https://doi.org/10.15587/1729-4061.2023.285799> (Scopus).
4. І.В. Заріцький І.В., Заріцький О.В.. Методологія інтелектуального аналізу та вибору типу електричної машини для систем запуску авіаційних газотурбінних двигунів перспективних літаків військово-транспортної авіації. – Збірник наукових праць. Державний науково-дослідний інститут авіації. – вип.№20(27).- 2024. - ст. 71-82. DOI: <https://doi.org/10.54858/dndia.2024-20-9>
5. Viktor Morozov, Vladyslav Deineha and Andrii Khlevnyi, «Research on the Use of Machine Learning Methods for Forecasting Time Series when Making Management Decisions in IT Projects Under Martial Law», CEUR Workshop Proceedings (CEUR-WS.org) (Scopus).
6. The application of practical management projects for the implementation of an automated parking system in the context of the "Smart City" initiative, Andrii Khlevny, Anatolii Vintsyk and Iulia Khlevna. Рукопис, доповідь на міжнародній конференції у Братиславі (Словаччина), (CEUR-WS.org) (Scopus).
7. Mike Lukides. What is data science? The future belongs to the companies and people that turn data into products (англ.). Radar. O'Reilly Media (2 June 2010).
8. Data Science and Predictive Analytics, Biomedical and Health Applications using R, Authors: Ivo D. Dinov, Textbook 2023.
9. Information Retrieval: Advanced Topics and Techniques, Association for Computing Machinery, New York, NY, United States, (December 2024).

Додаткова:

1. Oleh Zaritskyi, Andrii Miroshnyk. Machine learning in Counter-Unmanned Aircraft Systems. Information Technology and Implementation (Satellite) (IT&I-2023), 2024. Ukraine. – pp. 100-103. ISBN 978-960-801-870-9.
2. Oleh Zaritskyi, Andrii Miroshnyk. A combined method for recognizing and intercepting unmanned aerial vehicles using machine learning methods. XII Міжнародна науково-практична конференція «Сучасні проблеми і досягнення в галузі радіотехніки, телекомунікацій та інформаційних технологій», 10-12

грудня 2024 р., м. Запоріжжя – 127-131 pp. ISBN 978-617-529-487-1, <https://eir.zp.edu.ua/handle/123456789/19725>.

3. Kirill Abramov, Andrii Khlevnyi. Rationale for the approach to the development of an analytical text message processing system. // Information Technology and Implementation (Satellite): Conference Proceedings, November 20-21, 2023, Kyiv, Ukraine / Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Publisher Individual entrepreneur Picha Y.V., 2023.

4. Maksym Hudymenko, Andrii Khlevnyi. Using the GAN generative model and deep learning methods for image formation. // Information Technology and Implementation (Satellite): Conference Proceedings, November 20-21, 2023, Kyiv, Ukraine / Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Publisher Individual entrepreneur Picha Y.V., 2023.

5. Nikita Podolskyi, Andrii Khlevnyi. Development of Data Science technology in customer segmentation using support vector machine classification method. // Information Technology and Implementation (Satellite): Conference Proceedings, November 20-21, 2023, Kyiv, Ukraine / Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Publisher Individual entrepreneur Picha Y.V., 2023.

6. Dmytro Prokhorchuk, Andrii Khlevnyi. Heart rate variability analysis software. // Information Technology and Implementation (Satellite): Conference Proceedings, November 20-21, 2023, Kyiv, Ukraine / Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Publisher Individual entrepreneur Picha Y.V., 2023.

7. Yevhenii Rudkivskyi, Andrii Khlevnyi. Software for recognizing ships on digital images. // Information Technology and Implementation (Satellite): Conference Proceedings, November 20-21, 2023, Kyiv, Ukraine / Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Publisher Individual entrepreneur Picha Y.V., 2023.

8. Bohdan Oryshchak, Andrii Khlevnyi. A WEB-based platform to improve the coordination of volunteerism. // Information Technology and Implementation (Satellite): Conference Proceedings, November 20-21, 2023, Kyiv, Ukraine / Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Publisher Individual entrepreneur Picha Y.V., 2023.

Електронні джерела:

1. <https://www.kaggle.com/>

2. <https://data.gov.ua/>

3. <https://uteka.ua/ua/publication/news-14-delovye-novosti-36-spisok-otkrytyx-baz-dannyx>

ДОДАТКИ

Додаток А

Теми для виконання практичних:

1. Адміністрація Президента України.
2. Секретаріат Кабінету Міністрів України.
3. Рада Національної Безпеки України.
4. Міністерство оборони України.
5. Міністерство закордонних справ України.
6. Міністерство фінансів України.
7. Секретаріат Верховної Ради.
8. Міністерство внутрішніх справ.
9. Київський національний університет імені Тараса Шевченка.
10. Міністерство інформаційної політики України.
11. Міністерство податків і зборів.
12. Міністерство освіти і науки України.
13. Прокуратура.
14. Національний банк України.
15. Футбольний клуб «Динамо Київ».
16. Київська міська державна адміністрація.
17. НАК «Нафтогаз України»
18. Державний концерн «Укроборонпром».
19. Державне агентство автомобільних доріг України.
20. Прогноз розвитку футболу в Україні та міжнародних успіхів клубних команд та збірної.
21. Організація інформаційного впливу на абітурієнтів для вступу на ФІТ.
22. Діагностика причин впливу ігрової індустрії на соціальну адаптацію людей.
23. Інформаційне забезпечення пошуку та аналітики запозичень контенту в електронних документах.
24. Інформаційне забезпечення формування наукових груп в Інтернет ресурсах для виконання спільних проєктів
25. Аналітика та прогнозування забруднення навколишнього середовища.
26. Інформаційне забезпечення визначення емоційного тону електронного тексту.
27. Прогнозування розвитку систем голосового управління сучасними технологіями (моб. телефони, комп'ютери, GPS навігатори і т.п).
28. Інформаційне забезпечення проєкту створення додатку голосового управління мобільним телефоном.

29. Інформаційний аналіз та прогнозування даних онлайн-сервісів ринку нерухомості.
30. Інформаційне забезпечення та автоматизація календарного планування реалізації ІТ проекту.
31. Автоматизація аналізу та візуалізація даних у сфері онлайн-маркетингу.
32. Інформаційне забезпечення проекту створення поточних аналітичних звітів фінансової діяльності в комерційній сфері.
33. Прогнозування розвитку проектного менеджменту в Україні.
34. Діагностика причин низького рівня вакцинації дітей в Україні.
35. Інформаційне забезпечення проекту створення інформаційної системи електронного проектного менеджменту.
36. Прогнозування розвитку ринку електромобілів, та інформаційне забезпечення процесу прийняття рішень по виробництву електромобілів в Україні.
37. Діагностика причин занепаду тваринництва в Україні, та пропозиція стратегії її розвитку.
38. Аналіз інформаційних потоків для прийняття управлінських рішень в системі технологічної підготовки виробництва машинобудівних підприємств.
39. Організація інформаційного впливу на працівників підприємства для успішного впровадження методології управління проектами.
40. Ваша пропозиція.