

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ

ЗВІТ

до лабораторної роботи № 3 на тему
«Аналіз розміру файлів»

Студентки 3 курсу
факультету комп'ютерних
наук та кібернетики
групи ТТП-32
Сушко Діани

Київ-2024

ЗМІСТ

Вступ	3
Код програми	4
Виконання програми	7
Висновки	9

ВСТУП

Проаналізувати частотну характеристику (частотний розподіл) розміру файлів у файловій системі на комп'ютері (залежність кількості файлів від їх розміру).

ПРОГРАММНЫЙ КОД

Main.py

```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.widgets import Slider

file_sizes = pd.read_csv('file_sizes.txt', header=None,
names=['size'])

intervals = [0, 1024, 1024*1024, 10*1024*1024,
100*1024*1024, 1024*1024*1024, float('inf')]
labels = ['<1KB', '1KB-1MB', '1MB-10MB', '10MB-100MB',
'100MB-1GB', '>1GB']

file_sizes['size_interval'] = pd.cut(file_sizes['size'],
bins=intervals, labels=labels, right=False)

file_size_counts =
file_sizes['size_interval'].value_counts().sort_index()

total_sizes_mb =
file_sizes.groupby('size_interval')['size'].sum() / (1024
* 1024)
total_sizes_gb = total_sizes_mb / 1024
```

```

fig, ax = plt.subplots(figsize=(10, 9))
bar_plot = ax.bar(file_size_counts.index,
file_size_counts.values, color='skyblue')
ax.set_title(f'Гістограма кількості файлів за їх розміром
(усього {len(file_sizes)} файлів)')
ax.set_xlabel('Розмір файлу (байт)')
ax.set_ylabel('Кількість файлів')
ax.grid(axis='y')
ax.set_xticklabels(labels=file_size_counts.index,
rotation=0, ha='center')

for i, v in enumerate(file_size_counts.values):
    ax.text(i, v, str(v), ha='center', va='bottom',
rotation=0)

ax_slider = plt.axes([0.15, 0.01, 0.7, 0.03])
slider = Slider(ax_slider, 'Кількість файлів', 25,
max(file_size_counts), valinit=max(file_size_counts),
valstep=25)

def update(val):
    max_val = int(slider.val)
    ax.set_ylim(0, max_val)
    plt.draw()

slider.on_changed(update)

conclusions = []

```

```

for interval, count in zip(labels,
file_size_counts.values):
    percentage = count / len(file_sizes) * 100
    conclusion = f"Переважає більшість файлів
({percentage:.2f}%) має розміри у діапазоні {interval}.
Загальний обсяг: {total_sizes_mb[interval]:.2f} MB
({total_sizes_gb[interval]:.2f} GB)"
    conclusions.append((interval, percentage, conclusion))

conclusions.sort(key=lambda x: x[1], reverse=True)

with open('conclusions.txt', 'w') as f:
    for _, _, conclusion in conclusions:
        f.write(conclusion + '\n')

plt.show()

```

get_file_sizes.sh

```

#!/bin/bash

read -p "Enter the directory path: " dir_path

find "$dir_path" -type f -exec ls -l {} + | grep -v
'^total' | awk '{print $5}' > file_sizes.txt

echo "File sizes have been saved to file_sizes.txt"

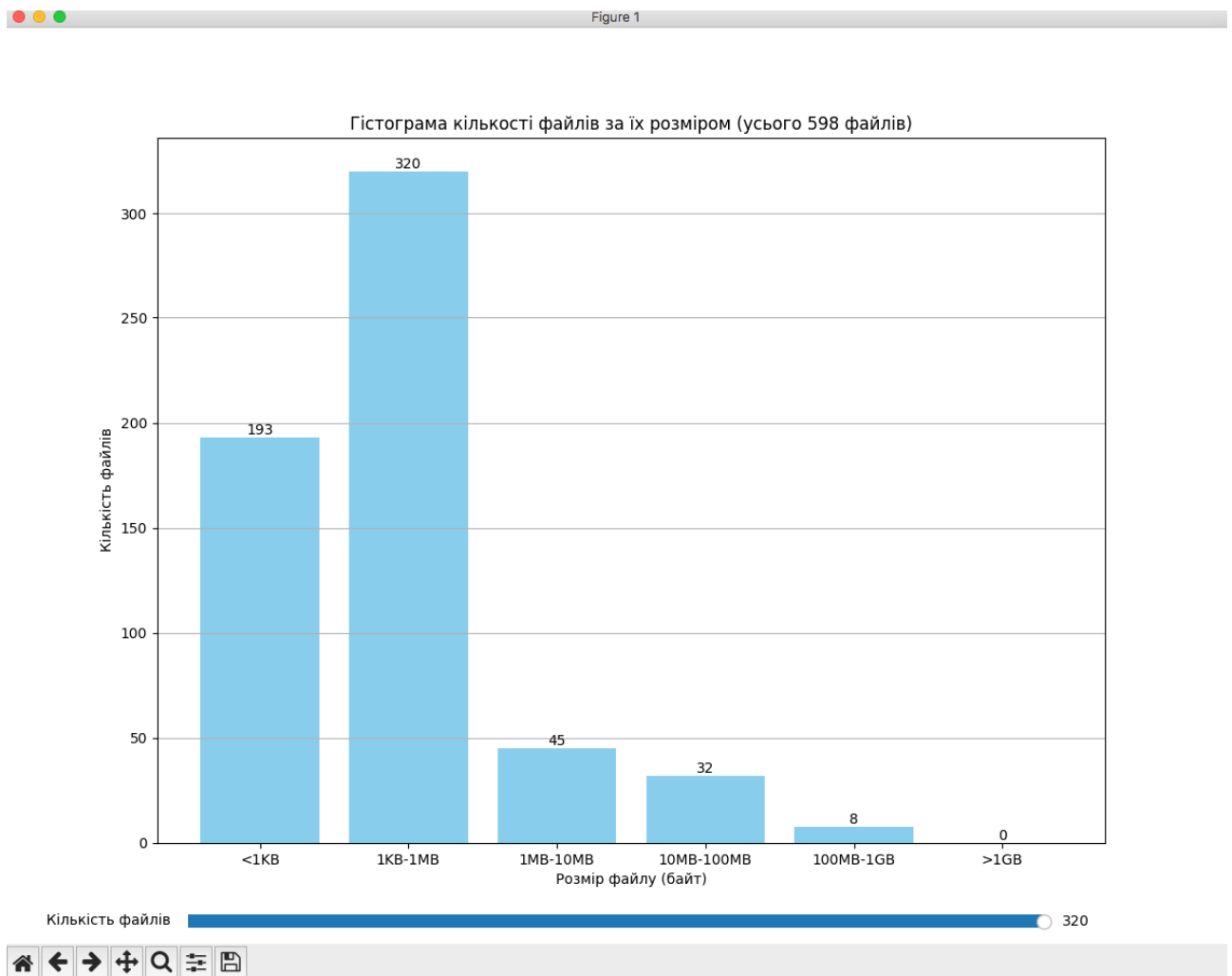
python3 main.py

```

ВИКОНАННЯ КОДУ

```
FileAnalizator — -bash — 100x29
Last login: Sun Apr 14 23:10:40 on ttys000
[MacBook-Pro-Diana:~ dianasushko$ cd desktop
[MacBook-Pro-Diana:desktop dianasushko$ cd FileAnalizator
[MacBook-Pro-Diana:FileAnalizator dianasushko$ ls -l
total 24
-rw-r--r--  1 dianasushko  staff   960 Apr 14 23:10 conclusions.txt
-rw-r--r--  1 dianasushko  staff    0 Apr 14 23:10 file_sizes.txt
-rwxr-xr-x  1 dianasushko  staff   226 Apr 10 22:12 get_file_sizes.sh
-rw-r--r--@ 1 dianasushko  staff  2120 Apr 10 22:47 main.py
MacBook-Pro-Diana:FileAnalizator dianasushko$
```

```
FileAnalizator — -bash — 100x29
[MacBook-Pro-Diana:FileAnalizator dianasushko$ pwd
/Users/dianasushko/desktop/FileAnalizator
[MacBook-Pro-Diana:FileAnalizator dianasushko$ ./get_file_sizes.sh
Enter the directory path: /Users/dianasushko/desktop
File sizes have been saved to file_sizes.txt
/Users/dianasushko/Desktop/FileAnalizator/main.py:14: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain current behavior or observed=True to adopt the future default and silence this warning.
  total_sizes_mb = file_sizes.groupby('size_interval')['size'].sum() / (1024 * 1024)
/Users/dianasushko/Desktop/FileAnalizator/main.py:23: UserWarning: set_ticklabels() should only be used with a fixed number of ticks, i.e. after set_ticks() or using a FixedLocator.
  ax.set_xticklabels(labels=file_size_counts.index, rotation=0, ha='center')
MacBook-Pro-Diana:FileAnalizator dianasushko$
```



Результати виконання програмного коду `main.py` було збережено у файл `conclusions.txt`.

Усього було знайдено 598 файлів.

Переважна більшість файлів 53.51% має розміри у діапазоні 1KB-1MB, загальний обсяг: 52.24 MB (0.05 GB); 32.27% файлів має розміри у діапазоні <1KB, загальний обсяг: 0.04 MB (0.00 GB); 7.53% файлів має розміри у діапазоні 1MB-10MB, загальний обсяг: 103.31 MB (0.10 GB); 5.35% файлів має розміри у діапазоні 10MB-100MB, загальний обсяг: 1366.27 MB (1.33 GB); 1.34% файлів має розміри у діапазоні 100MB-1GB, загальний обсяг: 2405.01 MB (2.35 GB); 0.00% файлів має розміри у діапазоні >1GB, загальний обсяг: 0.00 MB (0.00 GB).

ВИСНОВКИ

На основі виконаного аналізу розмірів файлів директорії /Users/dianasushko/desktop можна зробити такі висновки:

Усього було знайдено і проаналізовано 598 файлів.

Найбільш поширеними є файли з розмірами у діапазоні від 1KB до 1MB, що становлять 53.51% усіх файлів, тобто більшість. Загальний обсяг цих файлів становить 52.24 MB (0.05 GB).

Наступними йдуть файли розміром <1KB, що становлять 32.27% усіх файлів. Загальний обсяг таких файлів складає 0.04 MB (0.00 GB).

Файли у діапазоні від 1MB до 10MB зустрічаються в 7.53%, їх значно менше, і їх загальний обсяг становить 103.31 MB (0.10 GB).

Ще меншу частку становлять файли з розмірами в діапазоні від 10MB до 100MB, що складає 5.35%. Їх загальний обсяг становить 1366.27 MB (1.33 GB).

Частка файлів з розмірами від 100MB до 1GB складає всього 1.34%, є найбільш ваговими (важкими) і займають 2405.01 MB (2.35 GB).

А вот файлів з розмірами понад 1GB в досліджуваній директорії взагалі немає, тобто їхній загальний обсяг становить 0.00 MB (0.00 GB).