



ЗВІТ

до лабораторної роботи №1

з дисципліни

“Операційні системи для мобільних платформ”

Тема: “Аналіз розмірів файлів”

Виконав

Студент 3 курсу

Групи ТТП-31

Факультету комп’ютерних наук та кібернетики

Олександр БАЖИН

Зміст

Постановка задачі.....	3
Теоретичні відомості та обчислення.....	4
Висновки.....	6
Додаток 1.....	7
Джерела.....	9

Постановка задачі

Необхідно було проаналізувати частотну характеристику частотний розподіл) розміру файлів у файловій системі на власному комп'ютері (залежність кількості файлів від їх розміру) та надати відповідний звіт.

Теоретичні відомості та обчислення

Я вирішив зібрати розмір кожного файлу на моєму ПК за допомогою `bash` команди та за допомогою програми візуалізувати дані у вигляді гістограми.

Для виконання подібного завдання я використав дистрибутив Linux Ubuntu 24.02 LTS, адже в ньому є всі необхідні команди на відміну від Windows PowerShell.

Для збору даних я використав наступну команду з терміналу:
`sudo find / -type f -ls 2>/dev/null | awk '{print $7}' | sort -n > file_sizes.txt`

Проте в подальшому аналізі я виявив, що він нез'ясованої причини не враховує каталог `/media`, де є всі підключені диски. Оскільки моя основна операційна система це Windows, а не Linux, я вважаю це критичним для відображення дійсної інформації стосовно ситуації в моїй файловій системі.

Отже, я покращив команду:
`sudo find / /mnt /media -type f -printf "%s\n" 2>/dev/null | sort -n > gathered_file_sizes.txt`

Оскільки `find /` не охоплює ті ж місця пошуку, що й друга, дублювання даних відсутнє.

Для візуалізації даних, я написав програму мовою програмування Python. Програма зазначена в Додаток 1 разом з посиланням на репозиторій.

Я отримав наступні результати розбиті по діапазоном. Там де було найбільше файлів, я розбив на окремі діапазони.

0 В - 0.5 KB: 36.22% файлів (1933457 шт.)

0.5 KB - 1 KB: 7.72% файлів (411897 шт.)

1 KB - 5 KB: 19.84% файлів (1059349 шт.)

5 KB - 10 KB: 7.06% файлів (376946 шт.)

10 KB - 100 KB: 20.14% файлів (1075065 шт.)

100 KB - 1 MB: 6.98% файлів (372768 шт.)

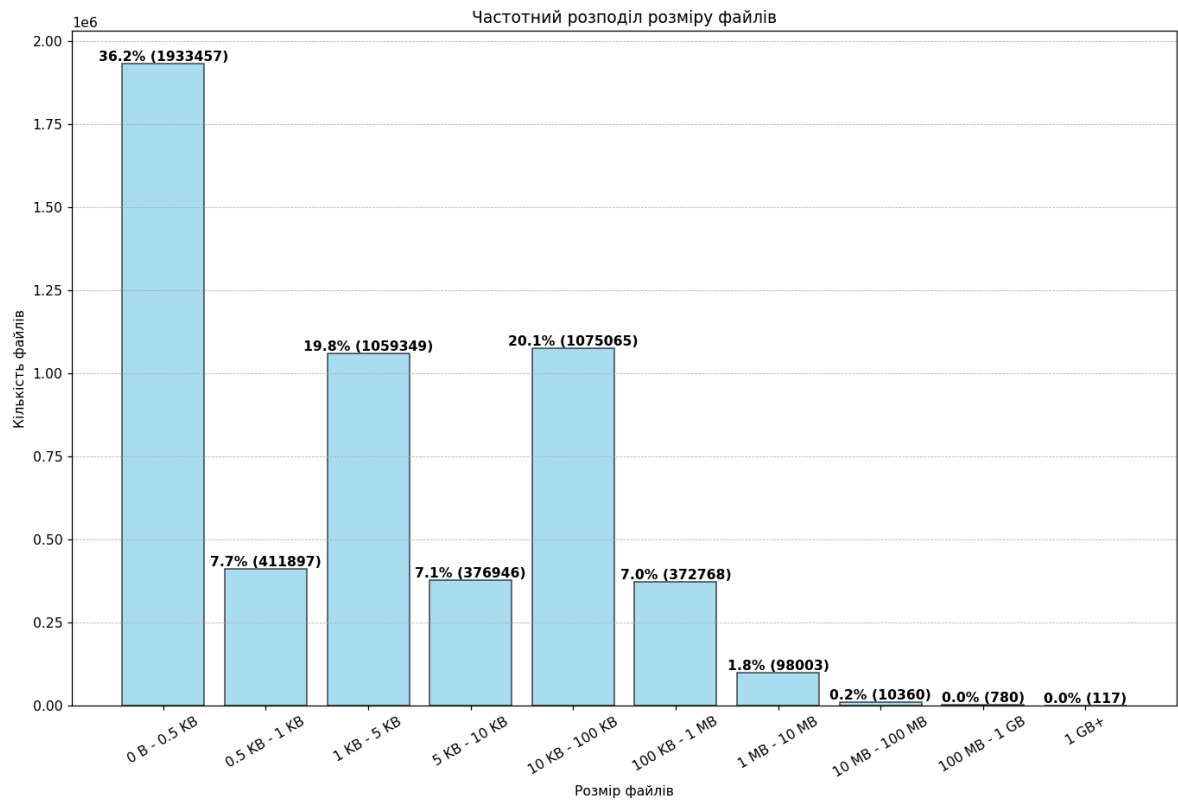
1 MB - 10 MB: 1.84% файлів (98003 шт.)

10 MB - 100 MB: 0.19% файлів (10360 шт.)

100 MB - 1 GB: 0.01% файлів (780 шт.)

1 GB+: 0.00% файлів (117 шт.)

Графічно зображено:



Гістограма побудована на основі зібраних результатів

Висновки

Переважає більшість файлів (90,9%) має розміри у діапазоні від 0 байтів до 100 КБ.

Зібрані та візуалізовані дані дозволили здійснити більш глибокий аналіз файлової системи мого комп'ютера, враховуючи також зовнішні пристрої зберігання, що важливо для адекватної оцінки розподілу розмірів файлів на всіх носіях.

Додаток 1

1. <https://github.com/OleksandrBazhyn/Analysis-of-file-sizes.git>

```
import numpy as np
import matplotlib.pyplot as plt

try:
    with open("gathered_file_sizes.txt") as f:
        sizes = np.array([int(line.strip()) for line in f if
line.strip().isdigit()])
except FileNotFoundError:
    print("Файл не знайдено. Переконайтесь, що `gathered_file_sizes.txt`
існує.")
    exit()

if sizes.size == 0:
    print("Файл `gathered_file_sizes.txt` порожній або містить
некоректні дані.")
    exit()

bins = [0, 512, 1024, 5_120, 10_240, 100_240, 1_024_000, 10_240_000,
100_240_000,
        1_024_000_000, float("inf")]
labels = ["0 В - 0.5 KB", "0.5 KB - 1 KB", "1 KB - 5 KB", "5 KB - 10
KB",
          "10 KB - 100 KB", "100 KB - 1 MB", "1 MB - 10 MB", "10 MB -
100 MB",
          "100 MB - 1 GB", "1 GB+"]

hist, _ = np.histogram(sizes, bins=bins)

total_files = sizes.size
percentages = (hist / total_files) * 100

plt.figure(figsize=(12, 6))
bars = plt.bar(labels, hist, color="skyblue", edgecolor="black",
alpha=0.7)

for bar, perc, count in zip(bars, percentages, hist):
    plt.text(bar.get_x() + bar.get_width()/2, bar.get_height(),
```

```
        f"{perc:.1f}% ({count})",
        ha='center', va='bottom', fontsize=10, fontweight='bold')

plt.xlabel("Розмір файлів")
plt.ylabel("Кількість файлів")
plt.title("Частотний розподіл розміру файлів")
plt.xticks(rotation=30)
plt.grid(axis="y", linestyle="--", linewidth=0.5)

plt.show()

print()
for label, perc, count in zip(labels, percentages, hist):
    print(f"{label}: {perc:.2f}% файлів ({count} шт.)")
print()
```


Джерела

1. <https://linuxcommand.org/> LinuxCommand
2. <https://matplotlib.org/> Matplotlib