# Звіт до лабораторної 3 (Аналіз розмірів файлів.)

Студента групи ТТП-32

Остренка Олександра

## 1. Умова лабораторної:

## Изображение выглядит как текст, снимок экрана, Шрифт, документ Автоматически созданное описание

## 2. Опис виконання:

## 2.1. Підготовка даних.

Щоб витягти дані про файли на диску ‘D:’ буде використано наступну команду для Windows PowerShell – “Get-ChildItem -Path D:\ -File -Recurse | Select-Object FullName, Length, CreationTime, Extension | Export-Csv -Path file\_info\_d.csv -NoTypeInformation”:

* «Get-ChildItem»: Ця команда отримує список файлів та папок на вказаному шляху.
* «-Path D:»: Це параметр, який вказує, що ми хочемо отримати список файлів та папок з диска D:.
* «-File»: Цей параметр вказує, що потрібно повертати тільки файли, а не папки або інші типи об'єктів.
* «-Recurse»: Цей параметр вказує, що команда повинна рекурсивно переглядати всі підпапки та їх вміст.
* «Select-Object FullName, Length, CreationTime, Extension»: Ця частина команди вибирає певні властивості для кожного файлу, включаючи повне ім'я (FullName), розмір в байтах (Length), дату створення (CreationTime) та розширення файлу (Extension).
* «Export-Csv -Path file\_info\_d.csv -NoTypeInformation»: Ця частина команди експортує вибрані властивості у CSV файл з ім'ям file\_info\_d.csv. Параметр -NoTypeInformation вказує, що необхідно не додавати типи даних до CSV файлу.

Отже, в результаті ця команда виконує дії:

* Отримує список файлів з диска D:\
* Вибирає властивості (FullName, Length, CreationTime, Extension) для кожного файлу
* Записує ці властивості у файл CSV з ім'ям file\_info\_d.csv.

## 2.2. Аналіз та візуалізація залежностей.

1. Зіставити тип файлу з середнім розміром для даного типу файлу. Даним аналізом визначимо, які типи файлів (в середньому) займають найбільше місця.Изображение выглядит как текст, снимок экрана, График, линия

   Автоматически созданное описание  
   Можна побачити, що в загальному на диску є до біса файлів різних форматів, але значним розміром виділяються лише декілька:  
   Изображение выглядит как текст, снимок экрана, График, линия

   Автоматически созданное описание  
     
   Изображение выглядит как текст, снимок экрана, Шрифт

   Автоматически созданное описание
2. Зіставити тип файлу з сумарним розміром всіх файлів. Даним аналізом можна дослідити якого типу файли реально займають найбільше місця і зіставити результати з попереднім аналізом, перевіривши гіпотезу «багато маленьких переможуть кількох великих».  
   Изображение выглядит как текст, линия, График, диаграмма

   Автоматически созданное описание  
   Изображение выглядит как текст, снимок экрана, линия, График

   Автоматически созданное описание  
   Изображение выглядит как текст, Шрифт, снимок экрана

   Автоматически созданное описание  
    Бачимо, що хоча .mp4 досі фаворит, але на друге та третє місце приходять нові переможці - .dll та .js, які беруть саме кількістю.
3. Останній аналіз – визначення файлів якого розміру найбільше на диску D:. Визначимо шість областей – [1e1; 1e2), [1e2; 1e3), [1e3; 1e4),   
   [1e4; 1e5), [1e5; 1e6), [1e6; inf) bytes. Розподілимо всі файли на ці області по обсягу пам’яті і порахуємо кількості.  
   Изображение выглядит как текст, снимок экрана, диаграмма, График

   Автоматически созданное описание  
   Изображение выглядит как текст, Шрифт, снимок экрана, черный

   Автоматически созданное описание  
   Бачимо, що залежність кількості файлів від розміру має вид нормального розподілу – спочатку збільшується, але після проміжку [100; 10000) байтів починає значно зменшуватись.

## 3. Висновки.

Аналізуючи описані дослідження, можна дійти наступних висновків:

1. Більшість файлів на диску, а саме 76.8% мають розмір від 100 байт 10 кілобайт.
2. Топ три типи файлів, які в середньому займають найбільше місця (тобто загальна пам’ять, виділена під ці файли / на кількість даних файлів) – .mp4, .zip, .pack.
3. Топ три типи файлів, які в загальному займають найбільше місця (тобто загальна кількість пам’яті, виділена під всі файли даного типу) – .mp4, .dll, .js.