

Міністерство освіти і науки України  
Національний технічний університет України «КПІ ім. Ігоря Сікорського»  
Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії  
Програмування інтелектуальних інформаційних систем

**ЗВІТ**  
до лабораторної роботи №1

**Виконав**  
**студент**

ІТ-03 Кобилинський В.О.  
(№ групи, прізвище, ім'я, по батькові )

**Прийняв**

ас. Очеретяний О. К.  
(посада, прізвище, ім'я, по батькові )

Київ 2021

## 1. Завдання лабораторної роботи

Обчислювальна задача тут тривіальна: для текстового файлу ми хочемо відобразити N (наприклад, 25) найчастіших слів і відповідну частоту їх повторення, упорядковано за зменшенням. Слід обов'язково нормалізувати використання великих літер і ігнорувати стоп-слова, як «the», «for» тощо. Щоб все було просто, ми не піклуємося про порядок слів з однаковою частотою повторень. Ця обчислювальна задача відома як **term frequency**.

## 2. Опис використаних технологій

Для написання програми згідно до завдання лабораторної було використано мову програмування Pascal. Використовував Free Pascal для запуску і налагодження коду, VS Code для приємного написання коду :)

## 3. Опис програмного коду

Код знаходиться за посиланням: <https://github.com/TheVitik/MP>

Дуже короткий опис логіки програми

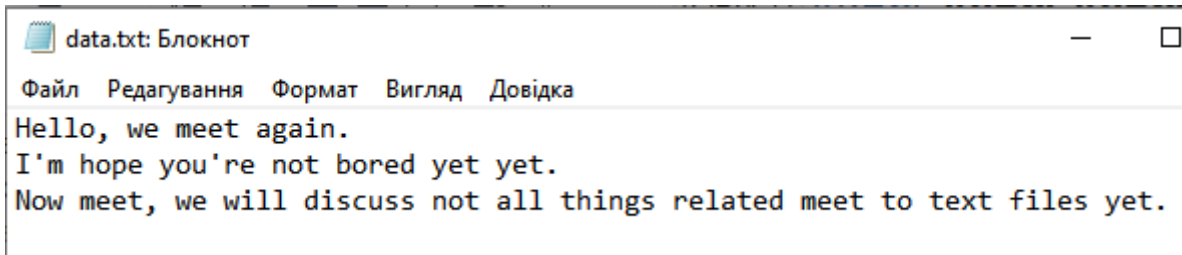
Відкриваємо файл з текстом, читаємо по рядку, а потім по символу. Одразу перевіряємо, чи є символ в діапазоні a-Z (за виключенням деяких символів в таблиці ASCII). Якщо літера велика, то її перетворюємо в маленьку. Якщо підряд йдуть букви, вони додаються в слово. Якщо є інший символ, то відбувається перевірка на стоп слово (якщо це стоп слово, тоді пропускаємо наступні кроки) та цикл по масиву зі словами, перевірка на наявність слова. Якщо слово знайдено, то до лічильника додається 1. Якщо слова немає, воно додається в масив слів, кількість 1 додається в масив лічильника. Потім слово обнуляється, і процес повторюється, доки не закінчиться файл.

Наступний етап це сортування бульбашкою. Спочатку сортуються масиви за словами від a до z, а потім відбувається сортування за кількістю слів у тексті, за спаданням.

Третій етап - виведення. Після закінчення сортування, виводиться масив з лімітом слів MAX (можна вказати своє число, у коді це 8). Виводиться слово та його кількість у тексті.

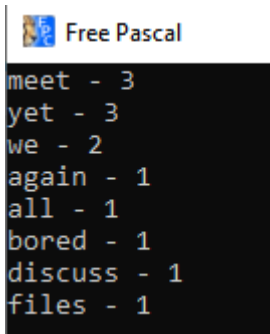
#### 4. Скріншоти роботи програмного застосунку

Вхідний файл



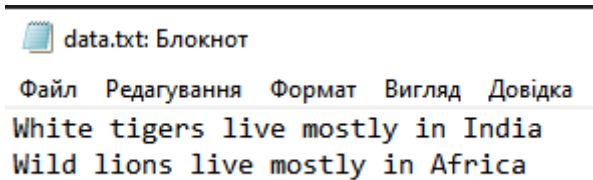
```
data.txt: Блокнот
Файл  Редагування  Формат  Вигляд  Довідка
Hello, we meet again.
I'm hope you're not bored yet yet.
Now meet, we will discuss not all things related meet to text files yet.
```

Результат програми




```
Free Pascal
meet - 3
yet - 3
we - 2
again - 1
all - 1
bored - 1
discuss - 1
files - 1
```

Вхідний файл



```
data.txt: Блокнот
Файл  Редагування  Формат  Вигляд  Довідка
White tigers live mostly in India
Wild lions live mostly in Africa
```

Результат програми



```
Free Pascal
live - 2
mostly - 2
africa - 1
india - 1
lions - 1
tigers - 1
white - 1
wild - 1
```