# Міністерство освіти і науки України Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського" Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

## Звіт

з лабораторної роботи № 1 з дисципліни «Мультипарадигменне програмування»

«Імперативне програмування»

Виконав: ІП-02 Ілющенко В. Т.

#### Лабораторна робота 1

Практична робота складається із трьох завдань, які самі по собі є досить простими. Але, оскільки задача - зрозуміти, як писали код наші славні пращури у 1950-х, ми введемо кілька обмежень:

- Заборонено використовувати функції
- Заборонено використовувати цикли
- Для виконання потрібно взяти мову, що підтримує конструкцію GOTO

# Завдання 1

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.

# Завдання 2

Тепер, нам потрібно виконати задачу, що називається словниковим індексуванням. Для текстового файлу виведіть усі слова в алфавітному порядку разом із номерами сторінок, на яких Ці слова знаходяться. Ігноруйте всі слова, які зустрічаються більше 100 разів. Припустимо, що сторінка являє собою послідовність із 45 рядків.

#### Завдання 1

#### Алгоритм

Для реалізації даної задачі необхідно виконати наступні дії:

- 1. Створити змінні для збереження інформації про слова: wordsArraySize (розмір масиву, що зберігає зчитані слова), wordsNum (кількість слів, унікальних зчитаних ЩО знаходяться В масиві), wordsArray[wordsArraySize] (масив, ЩО зберігає зчитані слова), wordsFrequency[wordsArraySize] (масив, кожен елемент якого зберігає частоту слова, що знаходиться в масиві wordsArray на відповідному індексі).
- 2. Обробити кожен рядок вхідних даних:
  - 2.1. Знайти довжину поточного рядка та присвоїти її змінній *lineSize*.
  - 2.2. Знайти індекс початку слова (заголовна або мала літера латинського алфавіту) та присвоїти його змінній *wordStartIdx*.
  - 2.3. Знайти індекс кінця слова (**HE** заголовна або мала літера латинського алфавіту, цифра, або символи " та '-' **AБO** кінець поточного рядка) та присвоїти його змінній wordEndIdx.
  - 2.4. Знайти довжину слова: wordSize = wordEndIdx wordStartIdx.
  - 2.5. **ЯКЩО** wordSize > 1 (не є шумовим словом) **ТО** обробити поточне слово, ІНАКШЕ присвоїти wordStartIdx = wordEndIdx і пропустити обробку слова (перейти до пункту 2.7)
  - 2.6. Обробити поточне слово:
    - 2.6.1. Присвоїти змінній *currentWord* поточне слово.
    - 2.6.2. Для кожного слова в *STOP\_WORDS*:
      - 2.6.2.1. **ЯКЩО** *currentWord* == **поточне стоп слово** ( $\epsilon$  *шумовим* (*cmon*) *словом*) **ТО** пропустити обробку слова (перейти до пункту 2.2).
    - 2.6.3. Для кожного слова в wordsArray:

- 2.6.3.1. Присвоїти stringsExEqual значення виразу currentWord == nomoчне записане слово
- 2.6.3.2. **ЯКЩО НЕ** *stringsExEqual* **ТО** обробити наступне записане слово **ІНАКШЕ** перейти до пункту 2.6.4
- 2.6.4. ЯКЩОstringsExEqualTOwordsFrequency[comparedExWordIdx]++IHAKШЕзаписати нове слово у відповідні масиви та збільшитикількість слів на 1
- 2.6.5. **ЯКЩО** wordsNum >= 0.8 \* wordsArraySize **TO** збільшити wordsArraySize вдвічі та створити нові масиви з цим розміром та перенести до них всі дані з старих масивів. Старі масиви видалити
- 2.7. **ЯКЩО** wordEndIdx < lineSize **TO** продовжити обробку поточного рядка вхідних даних (перейти до пункту 2.2).
- 2.8. **ЯКЩО** не кінець вхідних даних **ТО** почати обробку наступного рядка вхідних даних (перейти до пункту 2).
- 3. Сортування масивів wordsArray та wordsFrequency бульбашкою за спаданням кількості повторень слів (при перестановках в wordsFrequency проводимо аналогічні перестановки в wordsArray).
- 4. Форматовано вивести перші N слів та частоту їх повторень у файл.

#### Реалізація

```
// Used for input and output data
#include <fstream>
// Used only as a container, without methods
#include <string>
using namespace std;

const string INPUT_FILE_ADDRESS = "task1_input.txt";
const string OUTPUT FILE ADDRESS = "task1 output.txt";

const int NUM_OF_DISPLAYED_WORDS = 25;

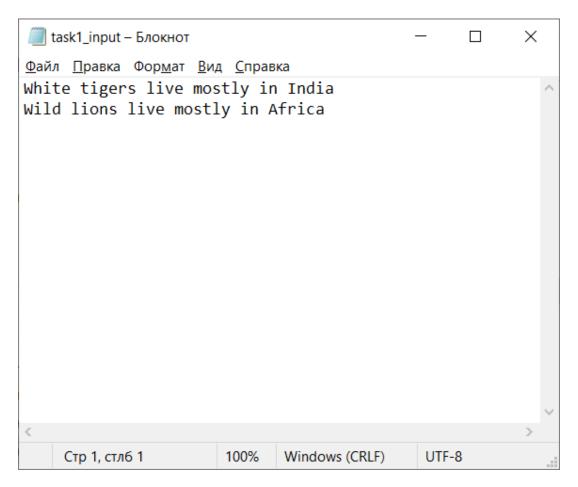
// Just as example of stop words handling
const int NUM_OF_STOP_WORDS = 8;
const char *STOP_WORDS[] = {"in", "on", "out", "of", "the", "an", "and",
"for"};
int main() {
```

```
ifstream input file;
PROCESS FILE:
PROCESS LINE:
FIND WORD START:
if (!(((currentLine[wordStartIdx] >= 'A' && currentLine[wordStartIdx] <=</pre>
      ((currentLine[wordStartIdx] >= 'a' && currentLine[wordStartIdx] <=</pre>
FIND WORD END:
    currentWord[wordSize] = '\0';
    COPY SUBSTR:
            lastCharIdx++;
```

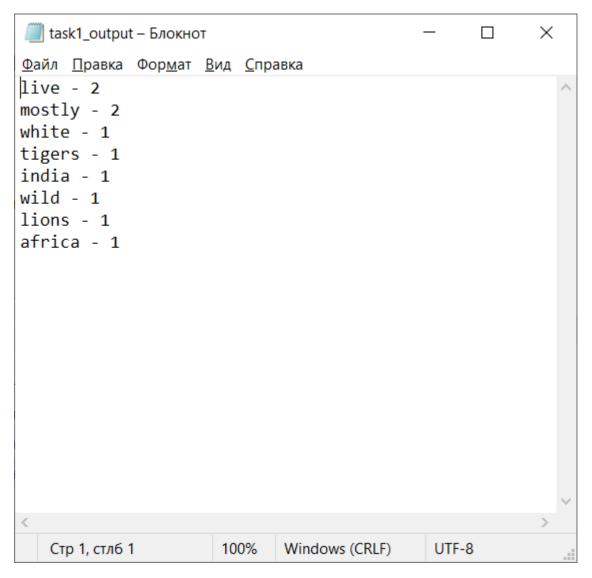
```
CHECK EXISTING WORDS:
    wordsFrequency[wordsNum] = 1;
```

```
bool swapped = false;
    if (wordsFrequency[j] < wordsFrequency[j + 1]) {</pre>
        string wordTmp = wordsArray[j];
        wordsFrequency[j] = wordsFrequency[j + 1];
        wordsArray[j] = wordsArray[j + 1];
        swapped = true;
    if (swapped) {
    output_file.open(OUTPUT_FILE_ADDRESS);
wordsFrequency[outputWordIdx] << '\n';</pre>
```

## Приклад роботи:



Зміст файлу «task1\_input.txt»



Зміст файлу «task1\_output.txt»

#### Завдання 2

#### Алгоритм

Для реалізації даної задачі необхідно виконати наступні дії:

- 1. Створити змінні для збереження інформації про слова: wordsArraySize (розмір масиву, що зберігає зчитані слова), wordsNum (кількість слів, унікальних зчитаних ЩО знаходяться В масиві), (масив, wordsArray[wordsArraySize] зберігає ЩО зчитані слова), wordsFrequency[wordsArraySize] (масив, кожен елемент якого зберігає частоту слова, що знаходиться в масиві wordsArray на відповідному індексі), wordsPages (масив масивів для збереження інформації про сторінки), pagesNums (масив для збереження інформації про кількості сторінок), pagesCapacities (масив для збереження інформації про розмір масивів у wordsPages), currentPageNum та currentLineNum (для інформації про поточну сторінку)
- 2. Обробити кожен рядок вхідних даних:
  - 2.1. **ЯКЩО** currentLineNum % LINES\_PER\_PAGE == 0 **TO** currentPageNum++
  - 2.2. Знайти довжину поточного рядка та присвоїти її змінній *lineSize*.
  - 2.3. Знайти індекс початку слова (заголовна або мала літера латинського алфавіту) та присвоїти його змінній *wordStartIdx*.
  - 2.4. Знайти індекс кінця слова (**HE** заголовна або мала літера латинського алфавіту, цифра, або символи "" та "-" **AБO** кінець поточного рядка) та присвоїти його змінній *wordEndIdx*.
  - 2.5. Знайти довжину слова: wordSize = wordEndIdx wordStartIdx.
  - 2.6. **ЯКЩО** wordSize > 0 **ТО** обробити поточне слово, IHAКШЕ присвоїти *wordStartIdx* = *wordEndIdx* і пропустити обробку слова (перейти до пункту 2.8)
  - 2.7. Обробити поточне слово:
    - 2.7.1. Присвоїти змінній *currentWord* поточне слово.

- 2.7.2. Для кожного слова в wordsArray:
  - 2.7.2.1. Присвоїти *stringsExEqual* значення виразу currentWord == *nomoчне записане слово*
  - 2.7.2.2. **ЯКЩО НЕ** *stringsExEqual* **ТО** обробити наступне записане слово **ІНАКШЕ** перейти до пункту 2.7.3
- 2.7.3. **ЯКЩО** stringsExEqual **TO** збільшити кількість повторів відповідного слова на 1 та записати інформацію про його повторення на поточній сторінці, якщо вона не була вказана раніше, при необхідності збільшити розмір масиву сторінок вдвічі **ІНАКШЕ** записати інформацію про нове слово (слово, сторінка) у відповідні масиви та збільшити кількість слів на 1
- 2.7.4. **ЯКЩО** wordsNum >= 0.8 \* wordsArraySize **TO** збільшити wordsArraySize вдвічі та створити нові масиви з цим розміром та перенести до них всі дані з старих масивів. Старі масиви видалити
- 2.8. **ЯКЩО** wordEndIdx < lineSize **TO** продовжити обробку поточного рядка вхідних даних (перейти до пункту 2.3).
- 2.9. **ЯКЩО** не кінець вхідних даних **ТО** почати обробку наступного рядка вхідних даних (перейти до пункту 2).
- 3. Перезаписати у масиви інформацію лише про ті слова, кількість повторень яких не перевищує 100.
- 4. Сортування масивів wordsArray, wordsPages, pagesNums бульбашкою за алфавітним порядком слів в wordsArray (при перестановках в wordsArray проводимо аналогічні перестановки в wordsPages ma pagesNums).
- 5. Форматовано вивести всі слова та сторінки, на яких вони зустрічаються, у файл.

#### Реалізація

```
#include <string>
   int wordStartIdx;
   int lineSize;
   int currentPageNum = 0;
   PROCESS FILE:
   lineSize = 0;
   CALCULATE LINE LENGTH:
       goto CALCULATE LINE LENGTH;
   wordEndIdx = wordStartIdx;
   FIND WORD END:
```

```
goto FIND WORD END;
    int wordSize = wordEndIdx - wordStartIdx;
        if (wordStartIdx < wordEndIdx) {</pre>
                lastCharIdx++;
            wordStartIdx++;
        bool stringsExEqual = false;
        CHECK EXISTING WORDS:
wordsArray[comparedExWordIdx][comparedExCharIdx]);
currentPageNum) {
wordsPages[comparedExWordIdx][pagesNums[comparedExWordIdx]] = currentPageNum;
pagesCapacities[comparedExWordIdx]) {
```

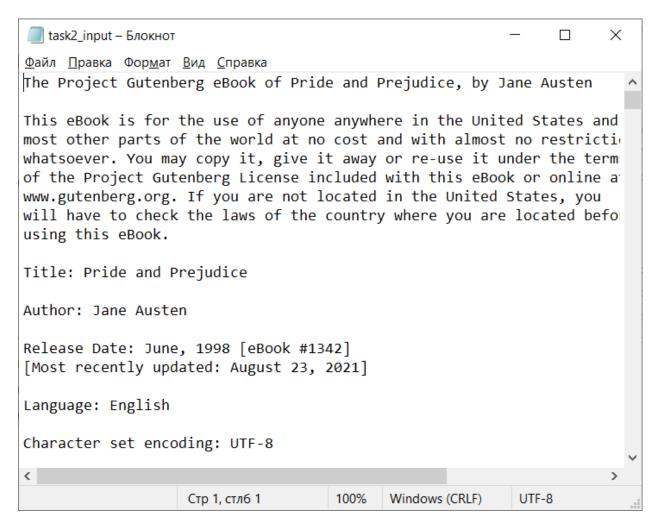
```
oldPages[wordIndex];
                           delete[] oldPages;
                COPY TO NEW:
                wordsArray[wordIndex] = oldWords[wordIndex];
                delete[] oldFrequency;
delete[] oldPages;
delete[] oldPagesCapacities;
delete[] oldPagesNums;
     if (!file.eof()) {
     int **oldPages = wordsPages;
```

```
int *oldPagesCapacities = pagesCapacities;
        wordsArray[newWordsNum] = oldWords[wordIndex];
delete[] oldPagesCapacities;
SORT ALPHABETICALLY:
bool swapped = false;
if (wordsArray[j][leftLength] != '\0') {
    if (wordsArray[j][charIdx] == wordsArray[j + 1][charIdx]) {
        leftGreater = (wordsArray[j][charIdx] > wordsArray[j +
```

```
pagesNums[j + 1] = pageNumTmp;
    swapped = true;
}
j++;
if (j < wordsNum - 1) {
    goto SORT_INNER;
}
if (swapped) {
    goto SORT_ALPHABETICALLY;
}

int wordOutputIdx = 0;
ofstream output_file;
output_file.open(OUTPUT_FILE_ADDRESS);
OUTPUT_RESULT:
if (wordOutputIdx < wordsNum) {
    output_file << wordsArray[wordOutputIdx] << " - ";
    int outputWordPageIdx = 0;
OUTPUT_PAGES:
    if (outputWordPageIdx < pagesNums[wordOutputIdx]) {
        if (outputWordPageIdx != 0) {
            output_file << ", ";
        }
        output_file << wordsPages[wordOutputIdx][outputWordPageIdx];
        output_file << \", ";
        }
        output_file << \"\n';
        wordOutputIdx++;
        goto OUTPUT_PAGES;
    }
    output_file << '\n';
        wordOutputIdx++;
    goto OUTPUT_RESULT;
}
output_file.close();
}</pre>
```

Приклад роботи:



Зміст файлу «task2\_input.txt»

```
X
 ____ task2_output – Блокнот
<u>Ф</u>айл <u>Правка Формат Вид Справка</u>
a-shooting - 305
abatement - 99
abhorrence - 111, 160, 167, 263, 299, 306
abhorrent - 276
abide - 174, 318
abiding - 177
abilities - 72, 74, 107, 155, 171, 194
able - 19, 37, 58, 78, 84, 86, 88, 91, 98, 101, 107, 109, 110, 120, 13
ablution - 119
abode - 59, 60, 66, 110, 122, 130, 176, 260
abominable - 32, 51, 71, 122, 161
abominably - 48, 133, 269, 299
abominate - 263, 296
abound - 101
above - 11, 32, 153, 179, 195, 202, 210, 212, 213, 214, 218, 220, 232
abroad - 194, 196, 233, 288
abrupt - 203
abruptly - 41, 155
abruptness - 198
absence - 54, 56, 64, 77, 78, 90, 99, 100, 106, 110, 111, 127, 150, 1
absent - 31, 199, 225, 229
                    Стр 1, стлб 1
                                      100%
                                             Windows (CRLF)
                                                             UTF-8
```

Зміст файлу «task2\_output.txt»

# Використані функції

Було використано лише функції введення-виведення з файлу, оскільки умова лабораторної роботи передбачала заборону інших функцій. Весь алгоритм знаходиться у функції main.

#### Висновки

Під час виконання даної лабораторної роботи було реалізовано задачі term frequency та словникового індексування на мові C++ без використання функцій (окрім зчитування/запису в файл), без використання циклів та з використанням конструкції GOTO. На власному досвіді ми переконалися чому GOTO є небезпечним інструментом і у більшості випадків його використання є недоцільним. Переконалися у перевагах, які надають нам сучасні парадигми програмування, а також зрозуміли наскільки процес написання програм у 1950-х відрізнявся від сучасного.