

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана»
(МГТУ им. Н.Э. Баумана)

Факультет: Информатика и системы управления

Кафедра: Теоретическая информатика и компьютерные технологии

Лабораторная работа №12-1
по курсу: «Языки и методы программирования»
«Обработка текстовых файлов»

Выполнил:
Студент группы ИУ9-21Б

Проверил:
Посевин Д. П.

Москва, 2024

1. Цель

Целью лабораторной работы является приобретение навыка разработки на языке C++ программ, осуществляющих анализ и преобразование текстовых файлов, записанных в различных форматах.

2. Персональный вариант

Найти все файлы с расширением «txt» в указанном каталоге, разбить текст из каждого файла на слова и сформировать в текущем каталоге два файла: all.txt и shared.txt, содержащие объединение и пересечение множеств слов из найденных файлов. Каждое слово в сформированных файлах должно располагаться в отдельной строке. Слова должны быть отсортированы лексикографически. Работоспособность программы нужно проверить на наборе текстовых файлов, содержащих текст на английском языке.

3. Решение

3.1. Код

```
#include <bits/stdc++.h>
#include <sys/dir.h>

using namespace std;

using stringSet = unordered_multiset<string>;

string getFileExtension(string path) { return
    → path.substr(path.find_last_of('.') + 1); }

stringSet getWordsFromFile(string path) {
    stringSet result;
    ifstream input(path);

    assert(input.good());

    for (string w; input >> w;)
        result.insert(w);

    return result;
}

vector<string> getAllFilesInDirectory(string path, string
    → extension) {
    vector<string> result;
    DIR *wd = opendir(path.c_str());
```

```

if (wd == NULL) {
    perror("opendir");
    return result;
}

for (struct dirent *ent; (ent = readdir(wd)) != NULL;) {
    if (ent->d_type == DT_REG) {
        string filename = path + "/" + ent->d_name;
        if (getFileExtension(filename) == extension) {
            result.push_back(filename);
        }
    }
}

return result;
}

vector<string> getAllWords(string path, string extension) {
    stringSet all;

    for (const auto &x : getAllFilesInDirectory(path,
        extension)) {
        auto words = getWordsFromFile(x);
        all.merge(words);
    }

    vector<string> all_v(all.begin(), all.end());
    sort(all_v.begin(), all_v.end());
    return all_v;
}

vector<string> getSharedWords(string path, string extension) {
    stringSet shared;
    bool share = true;

    for (const auto &x : getAllFilesInDirectory(path,
        extension)) {
        auto words = getWordsFromFile(x);
        if (share) {
            for (const auto &x : words) {
                shared.insert(x);
            }
            share = false;
        } else {
            stringSet temp;

```

```

        for (const auto &x : words) {
            if (shared.find(x) != shared.end()) {
                shared.erase(x);
                temp.insert(x);
            }
        }
        shared = temp;
    }
}

vector<string> shared_v(shared.begin(), shared.end());
sort(shared_v.begin(), shared_v.end());
return shared_v;
}

void saveWordsInFile(vector<string> v, string s) {
    ofstream output(s);
    assert(output.good());

    for (const auto &x : v)
        output << x << '\n';
}

int main() {
    string path = "test directory";

    bool share = true;

    auto all = getAllWords(path, "txt");
    auto shared = getSharedWords(path, "txt");

    saveWordsInFile(all, "all.txt");
    saveWordsInFile(shared, "shared.txt");
}

```

Код 1: main.cc

3.2. Скриншоты

```
[trolo@trolo lab12]$ cat all.txt | tail -n 3
то
хотя
ядро,
[trolo@trolo lab12]$ cat shared.txt | tail -n 3
int
{
}
[trolo@trolo lab12]$ |
```

Рис. 1: Пример работы программы