



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ _____ «Информатика и системы управления»

КАФЕДРА _____ «Теоретическая информатика и компьютерные технологии»

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

Студент группы ИУ9-21Б Горбунов А. Д.

Преподаватель Посевин Д. П.

Москва 2023

1 Задание

Найти все файлы с расширением «html» в заданном каталоге, для каждого файла очистить его текст от скриптов и сохранить в файл с тем же именем в текущий каталог. Скрипты в HTML-файле записываются внутри тега script.

Например:

```
<script>
```

```
document.getElementById("demo").innerHTML = "Hello JavaScript!";
```

```
</script>
```

Работоспособность программы нужно проверить на наборе HTML-файлов, загруженных из интернета.

2 Результаты

Исходный код программы представлен в листинге 1, 2, 3

Листинг 1 — main.cpp

```
1 #include <iostream>
2 #include <cstring>
3 #include <sys/types.h>
4 #include <dirent.h>
5 #include <fstream>
6 using namespace std;
7 int main(int argc, char *argv[])
8 {
9     if (argc != 2) {
10         cout << "Usage: " << argv[0] << " directoryPath" << endl;
11         return 1;
12     }
13     DIR *dir = opendir(argv[1]);
14     if (dir == NULL)
15     {
16         cout << "Failed to open directory!" << endl;
17         return 1;
18     }
19     struct dirent *ent;
20     char path[1000], outfilename[1000], buffer[1024];
21     ifstream infile;
22     ofstream outfile;
```

Листинг 2 — main.cpp

```

1  while ((ent = readdir(dir)) != NULL)
2      if (ent->d_type == DT_REG && strstr(ent->d_name, ".html"))
3      {
4          cout << "Processing file " << ent->d_name << "... " << endl;
5          strcpy(path, argv[1]);
6          strcat(path, "/");
7          strcat(path, ent->d_name);
8          infile.open(path);
9          if (!infile.is_open())
10         {
11             cout << "Failed to open input file!" << endl;
12             return 1;
13         }
14         strcpy(outfilename, ent->d_name);
15         outfile.open(outfilename);
16         if (!outfile.is_open())
17         {
18             cout << "Failed to open output file!" << endl;
19             return 1;
20         }
21         bool in_script = false;
22         while(infile.getline(buffer, 1024))
23         {
24             char *pos = buffer;
25             while (*pos)
26             {
27                 if (in_script)
28                     if (strncmp(pos, "</script>", 9) == 0)
29                     {
30                         pos += 9;
31                         in_script = false;
32                     }
33                     else
34                         ++pos;
35                 else
36                 {
37                     char *start = strstr(pos, "<script>");
38                     if (start)
39                     {
40                         outfile.write(pos, start - pos);
41                         pos = start;
42                         in_script = true;
43                     }
44                     else
45                     {
46                         outfile << *pos;
47                         if(*pos++ == '>'){
48                             outfile << endl;
49                         }
50                     }
51                 }
52             }
53             if (in_script)
54             {
55                 outfile << endl;
56             }
57         }
58         infile.close();
59         outfile.close();
60     }
61     closedir(dir);
62     return 0;
63 }
```

Листинг 3 — Один из html файлов

```
1 <!DOCTYPE html>
2 <html>
3   <head>
4     <meta charset="UTF-8">
5     <title>Script example JavaScript </title>
6   </head>
7   <body>
8     <script>
9       alert("Hello , world!");
10    </script>
11  </body>
12</html>
```

Результат запуска представлен на рисунке 1, 2, 3, 4

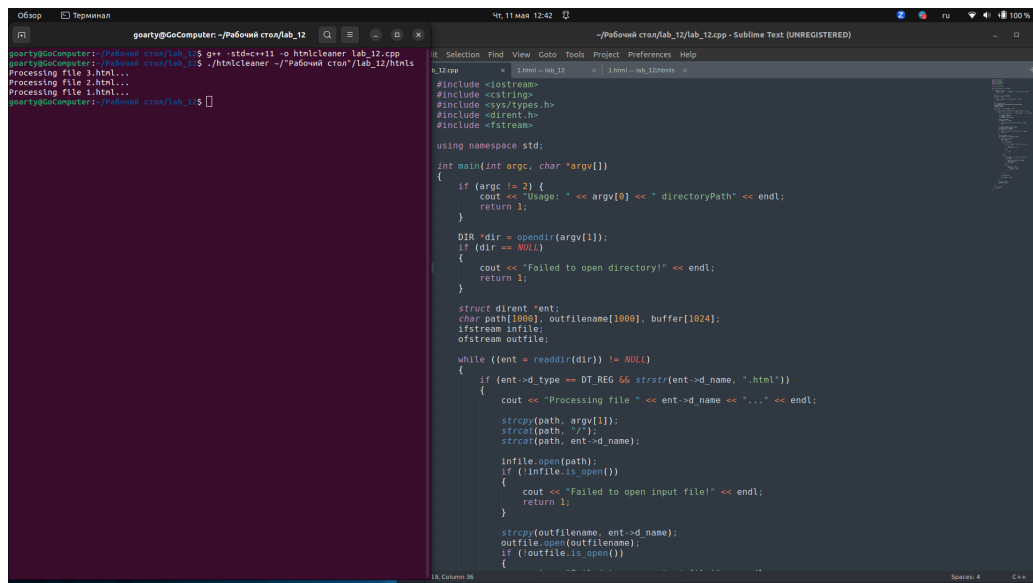


Рис. 1 — Реализация main.cpp

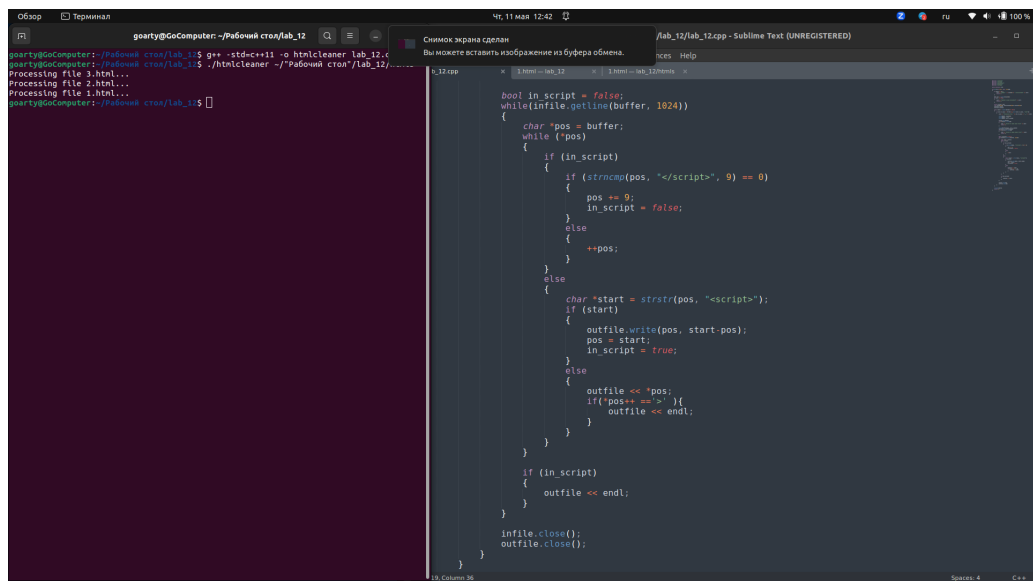


Рис. 2 — Реализация main.cpp(продолжение)

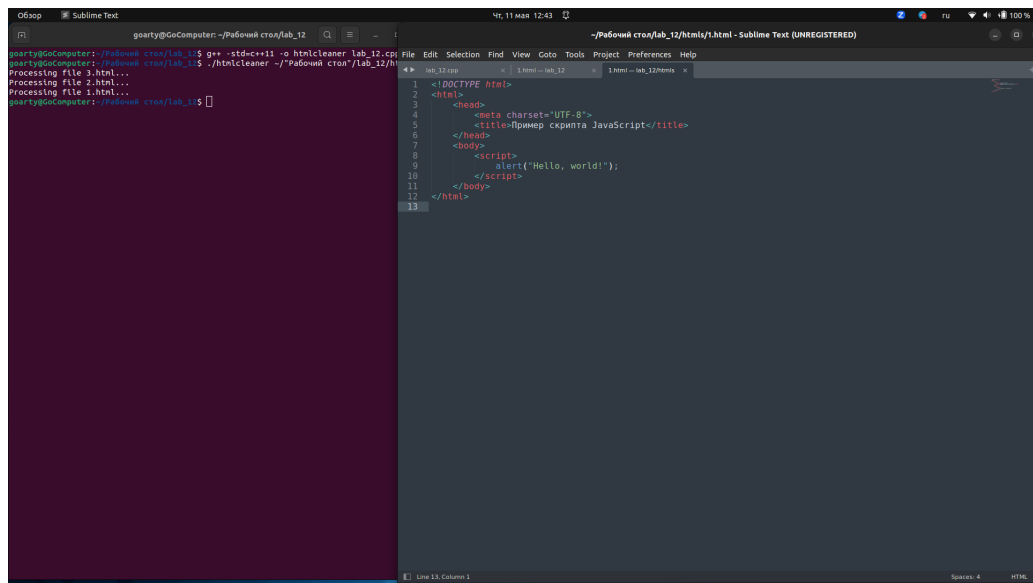


Рис. 3 — Реализация 1.html

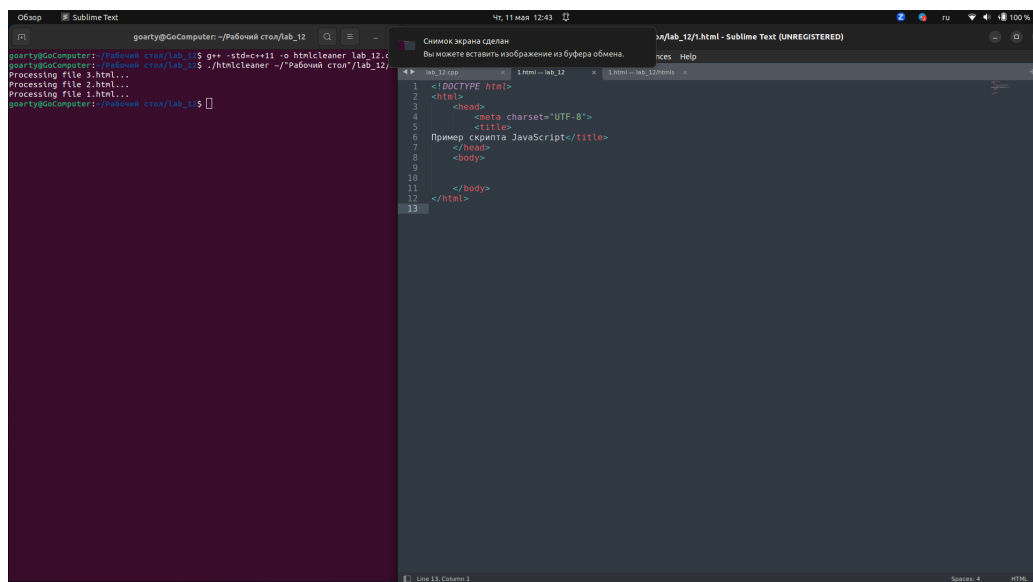


Рис. 4 — Реализация 1.html(после работы программы).java