



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 ИУ «Информатика и системы управления»

КАФЕДРА ИУ-7 «Программное обеспечение ЭВМ и информационные технологии»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №5
по дисциплине «Защита информации»
«Сжатие файлов. Алгоритм сжатия
Хаффмана.»

Студент группы ИУ7-76Б

(Подпись, дата) В. М. Мансуров
(И.О. Фамилия)

Руководитель

(Подпись, дата) И. С. Чиж
(И.О. Фамилия)

2023 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
1 Аналитическая часть	5
1.1 Алгоритм Хаффмана	5
2 Конструкторская часть	7
2.1 Разработка алгоритмов	7
3 Технологическая часть	8
3.1 Средства реализации	8
3.2 Реализация алгоритма	8
Заключение	12
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	13

ВВЕДЕНИЕ

В XXI веке количество информации, с которым встречается человек в течении дня, намного больше чем в другие периоды существования человечества. Информация хранится на физических носителях, передаётся по проводам и по беспроводной связи. Часто возникают ситуации, когда объём передаваемых или хранимых данных слишком велик, и возникает необходимость его уменьшить.

Сжатие данных — обратимое преобразование данных, производимое с целью уменьшения занимаемого ими объёма с целью уменьшения объёма данных, который требуется для хранения или передачи информации.

Целью данной работы является реализация в виде программы алгоритма сжатия данных Хаффмана, обеспечить сжатие и разжатие произвольного файла с использованием разработанной программы, рассчитывать коэффициент сжатия. Предусмотреть работу с пустым, однобайтовым файлами.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- 1) изучить алгоритм сжатия Хаффмана;
- 2) реализовать алгоритм сжатия Хаффмана в виде программы, обеспечив возможности сжатия и разжатия произвольного файла и расчёт коэффициента сжатия;
- 3) протестировать разработанную программу, показать, что удаётся сжимать и разжимать файлы разных форматов;
- 4) описать и обосновать полученные результаты в отчёте о выполненной лабораторной работе.

1 Аналитическая часть

В этом разделе будет рассмотрен алгоритм сжатия Хаффмана, шаги для сжатия и разжатия, а также его преимущества, недостатки и возможные способы решения возникающих проблем.

1.1 Алгоритм Хаффмана

Алгоритм Хаффмана — алгоритм сжатия данных, который формирует основную идею сжатия файлов. Кодирование Хаффмана — это тип кодирования с переменной длиной слова. Был разработан в 1952 году аспирантом Массачусетского технологического института Дэвидом Хаффманом при написании им курсовой работы. В настоящее время используется во многих программах сжатия данных.

Основные этапы алгоритма сжатия с помощью кодов Хаффмана:

Сбор статистической информации для последующего построения таблиц кодов переменной длины

Построение кодов переменной длины на основании собранной статистической информации

Кодирование (сжатие) данных с использованием построенных кодов

Алгоритм состоит из следующих шагов:

- сортировка выходных символов, не меняя местоположения символа, по вероятности их встречаемости в убывающем порядке;
- объединение двух символов с наименьшими вероятностями в композицию символов с вероятностью, равной сумме исходных вероятностей;
- повторения предыдущего шага до тех пор, пока не получится композиция с вероятностью 1, которая называется корнем. Полученная структура называется деревом Хаффмана;
- проход по дереву от корня до соответствующего символа и присвоение 0

(1) — левой и 1 (0) — правой ветви.

Описанный выше алгоритм сжатия требует хранения и передачи вместе с кодированными данными дополнительной информации, которая позволяет однозначно восстановить таблицу соответствия кодируемых символов и кодирующих битовых цепочек.

Следует отметить, что в некоторых случаях можно использовать постоянную таблицу (или набор таблиц), которые заранее известны как при кодировании, так и при декодировании. Или же строить таблицу адаптивно в процессе сжатия и восстановления. В этих случаях хранение и передача дополнительной информации не требуется, а также отпадает необходимость в предварительном сборе статистической информации (этап 1).

2 Конструкторская часть

В этом разделе будут представлены описания модулей программы, а также схема алгоритма шифрования RSA, SHA1.

2.1 Разработка алгоритмов



Рисунок 2.1 – Схема кодирования Хаффмана

3 Технологическая часть

3.1 Средства реализации

Для программной реализации шифровальной машины был выбран язык C++ [1]. В данном языке есть все требующиеся инструменты для данной лабораторной работы. В качестве среды разработки была выбрана среда CLion [2].

3.2 Реализация алгоритма

Листинг 3.1 – Алгоритм сжатия файла

```
1 void decompress(const char* filename, const ll Filesize)
2 {
3     const std::string fl = filename;
4     FILE* iptr = fopen(std::string("../data/" + fl).c_str(),
5         "rb");
6     FILE* optr = fopen(std::string("../data/d_" +
7         fl).c_str(), "wb");
8
9     if (iptr == NULL)
10    {
11        perror("Error: File not found");
12        exit(-1);
13    }
14
15    auto [padding, headersize] = decode_header(iptr);
16    store_huffman_value();
17    print_tree();
18
19    byte ch;
20    char counter = 7;
```

```

19     ll size = 0;
20     const ll filesize = Filesize - headersize;
21     Node* traverse = root;
22     ch = fgetc(iptr);
23     while (size != filesize)
24     {
25         while (counter >= 0)
26         {
27             traverse = ch & (1 << counter) ? traverse->right
28                 : traverse->left;
29             --counter;
30             if (!traverse->left && !traverse->right) {
31                 fputc(traverse->character, optr);
32                 if (size == filesize - 1 && padding ==
33                     counter + 1) {
34                     break;
35                 }
36                 traverse = root;
37             }
38             ++size;
39             counter = 7;
40             ch = fgetc(iptr);
41         }
42         fclose(iptr);
43         fclose(optr);
44     }

```

Листинг 3.2 – Алгоритм разжатия файла

```

1 void decompress(const char* filename, const ll Filesize)
2 {
3     const std::string fl = filename;
4     FILE* iptr = fopen(std::string("../data/" + fl).c_str(),
5         "rb");
6     FILE* optr = fopen(std::string("../data/d_" + fl).c_str(),

```



```

        "wb");
6
7     if (iptr == NULL)
8     {
9         perror("Error: File not found");
10        exit(-1);
11    }
12
13    auto [padding, headersize] = decode_header(iptr);
14    store_huffman_value();
15    print_tree();
16
17    byte ch;
18    char counter = 7;
19    ll size = 0;
20    const ll filesize = Filesize - headersize;
21    Node* traverse = root;
22    ch = fgetc(iptr);
23    while (size != filesize)
24    {
25        while (counter >= 0)
26        {
27            traverse = ch & (1 << counter) ? traverse->right :
                traverse->left;
28            --counter;
29            if (!traverse->left && !traverse->right) {
30                fputc(traverse->character, optr);
31                if (size == filesize - 1 && padding == counter +
                    1) {
32                    break;
33                }
34                traverse = root;
35            }
36        }
37        ++size;

```

```
38         counter = 7;
39         ch = fgetc(iptr);
40     }
41     fclose(iptr);
42     fclose(optr);
43 }
```

Заключение

Целью данной работы является реализация в виде программы алгоритма сжатия данных Хаффмана, обеспечить сжатие и разжатие произвольного файла с использованием разработанной программы, рассчитывать коэффициент сжатия. Предусмотреть работу с пустым, однобайтовым файлами.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- 1) изучить алгоритм сжатия Хаффмана;
- 2) реализовать алгоритм сжатия Хаффмана в виде программы, обеспечив возможности сжатия и разжатия произвольного файла и расчёт коэффициента сжатия;
- 3) протестировать разработанную программу, показать, что удаётся сжимать и разжимать файлы разных форматов;
- 4) описать и обосновать полученные результаты в отчёте о выполненной лабораторной работе.

Список использованных источников

1. Язык программирования C++. <https://learn.microsoft.com/en-us/cpp/cpp/cpp-language-reference?view=msvc-170>. дата обращения: 15.10.2023.
2. CLion. [jetbrains.com](https://www.jetbrains.com/idea/). дата обращения: 15.10.2023.