МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ им. В. Г. ШУХОВА» (БГТУ им. В.Г. Шухова)

Кафедра программного обеспечения вычислительной техники и автоматизированных систем

Лабораторная работа №3

по дисциплине: Теория информации тема: «Исследование возможности применения методов энтропийного кодирования для обработки двоичных последовательностей»

Выполнил: ст. группы ПВ-233 Мороз Роман Алексеевич

Проверил: Твердохлеб Виталий Викторович

Белгород 2025 г.

Задания лабораторной работы

- 1) Открыть файл Лабораторная работа 3 (задание).txt. Рассмотреть возможность построения кода по методам Хаффмана и Шеннона-Фано для бинарной последовательности. Сделать выводы.
 - 2) Рассмотреть варианты обработки цепочек символов, а именно:
 - 2 символа;
 - 4 символа;
 - 8 символов.

Для этого разработать консольное приложение, разбивающее сплошной массив символов на цепочки заданной длины.

- 3) Рассматривая каждую цепочку (2, 4 и 8 символов длиной) как отдельный символ, построить коды по методу Хаффмана и Шеннона-Фано.
- 4) Составить последовательности из полученных кодов символов для каждого случая.
- 5) По результатам работы в п.3 сделать выводы по поводу полученных результатов для каждого из методов (простота, скорость, полученные результаты (рассчитать коэффициенты сжатия)).
- 6) Написать программу, восстанавливающую последовательности, полученные в п.3 в исходный вид согласно вариантам, приведенным в п.2.
- 7) Восстановить исходный текст из полученных последовательностей, пользуясь сервисом https://onlineutf8tools.com/convert-binary-to-utf8

2. Рассмотреть варианты обработки цепочек символов, а именно: 2 символа; 4 символа; 8 символов. Для этого разработать консольное приложение, разбивающее сплошной массив символов на цепочки заданной длины.

```
std::string codeParsing(const std::string &str, const size_t n) {
    std::string out;

    size_t l = str.length() / n;
    for (size_t i = 0; i < 1; i++) {
        int id = 0;
        for (size_t j = 0; j < n; j++)
            id = id * 2 + str[i * n + j] - '0';

        out.push_back(id);
    }

    return out;
}</pre>
```

3. Рассматривая каждую цепочку (2, 4 и 8 символов длиной) как отдельный символ, построить коды по методу Хаффмана и Шеннона-Фано.

Результаты работы:

Для 2-х символьного алфавита

FANO Table: <0> = 00 <0> = 01 <▼> = 10 <> = 11 HUFFMAN Table: <> = 11 <∇> = 10 <0> = 01 <0> = 00

Для 4-х символьного алфавита

FANO Table:

```
<
        > = 00000000
<>> = 00000001
<4> = 00000010
> = 11
< > = 10
<ð> = 011
<0> = 010
<#>> = 00011
> = 0011
<0> = 0010
 = 0000011
<+> = 00010
<9> = 000011
<♥> = 000010
> = 0000010
<> = 00000011
HUFFMAN
Table:
<♂> = 111
> = 1101
<0> = 1100
<0> = 000
<#>> = 00101
<+> = 00100
<♥> = 001100
       > = 00110100
<+> = 00110101
> = 0011011
 = 0011100
<4> = 00111010
<> = 00111011
< = 001111
< > = 01
> = 10
```

Для 8-и символьного алфавита

FANO		
Table:		
<й>>	=	000000000
<t></t>	=	00000001
< 4>>	=	000000100
<Φ>	=	000000101
	=	01001
<r>></r>	=	0000101
<m>></m>	=	01011
>	=	01101
<,>	=	00110
<c></c>	=	0111
<u>></u>	=	10000
<л>	=	10001
<t></t>	=	01100
<.>	=	0000100
<e></e>	=	1010
<ы>	=	01010
<0>	=	110
<h>></h>	=	1001
<a>>	=	1011
<ц>	=	00000011
< >	=	111
<д>	=	000011
<y></y>	=	01000
< K >	=	00111
<6>	=	00101
<ю>	=	00000101
<n></n>	=	001001
<r>></r>	=	001000
<	=	000111
<>>	=	000000001
	=	00000100
<3>	=	000110
<ж>	=	000101
<x></x>	=	000100
<ш>	=	0000011

HUFFMAN

Table	

Table.		
<n></n>	=	111111
<6>	=	111110
<c></c>	=	11110
<0>	=	1110
>	=	11011
<t></t>	=	11010
<y></y>	=	01100
	=	01011
< R>	=	000100
< K >	=	01010
<h>></h>	=	0100
<u>></u>	=	0011
<,>	=	00011
<л>	=	0010
<ш>	=	0001011
<д>	=	000011
<ц>	=	0001010
< >	=	101
<.>	=	000010
<r>></r>	=	00000
<3>	=	011010
<x></x>	=	011011
<ж>	=	011100
<ю>	=	0111010
< 4>>	=	01110110
<Φ>	=	01110111
<m>></m>	=	01111
<e></e>	=	1000
<ы>	=	10010
<t></t>	=	10011000
<	=	100111
	=	10011001
<∋>	=	10011010
<a>>	=	1100
<й>>	=	10011011

Для 2-х символьного алфавита

FANO Code:

HUFFMAN

Code:

FANO Code:

HUFFMAN

Code:

Для 8-и символьного алфавита

FANO Code:

10011011101001100000001100001011011100010011011100111011000 01110000100110000101100000001000000100

HUFFMAN

Code:

5. По результатам работы в п.3 сделать выводы по поводу полученных результатов для каждого из методов (простота, скорость, полученные результаты (рассчитать коэффициенты сжатия)).

Для 2-х символьного алфавита

FANO

Начальная длина: 4120 Длина кода: 4120 Коэффициент: 1

Средняя длина кода: 2

Дисперсия: 0

HUFFMAN

Начальная длина: 4120 Длина кода: 4120 Коэффициент: 1

Средняя длина кода: 2

Дисперсия: 0

Для 4-х символьного алфавита

FANC

Начальная длина: 4120 Длина кода: 3121

Коэффициент: 1.3209

Средняя длина кода: 3.0301

Дисперсия: 1.7923

HUFFMAN

Начальная длина: 4120 Длина кода: 3121

Коэффициент: 1.3209

Средняя длина кода: 3.0301

Дисперсия: 1.7923

Для 8-и символьного алфавита

FANO

Начальная длина: 4120

Длина кода: 1298 Ко∋ффициент: 3.17411

Средняя длина кода: 4.58657

Дисперсия: 1.59233

HUFFMAN

Начальная длина: 4120

Длина кода: 1298 Коэффициент: 3.18639

Средняя длина кода: 4.5689

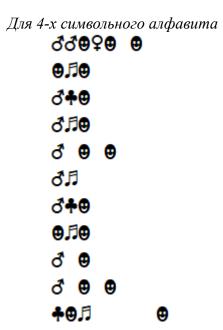
Дисперсия: 1.16398

6. Написать программу, восстанавливающую последовательности, полученные в п.3 в исходный вид согласно вариантам, приведенным в п.2.

```
std::unordered map<std::vector<bool>, wchar t>
                  GetDecodingTable(const std::unordered map<wchar t,</pre>
                  std::vector<bool>> &code) {
    std::unordered map<std::vector<bool>, wchar t> out;
    for (auto &el: code)
       out[el.second] = el.first;
   return out;
}
std::wstring getDecodedMessage(const std::unordered map<std::vector<bool>,
                               wchar t> &code, const std::wstring &str) {
    std::wstring out;
    std::vector<bool> c;
    int i = 0;
    while (i < str.length()) {</pre>
        c.clear();
        while (!code.contains(c)) {
            c.push back(str[i] - '0');
            i++;
        out.push back(code.find(c)->second);
   return out;
int main() {
    setlocale(LC ALL, "RU");
    std::string bit_code = "52";
    std::vector<int> k = \{2, 4, 8\};
    for (auto &i: k) {
        auto msg = changeCode(codeParsing(bit code, i));
        auto table = GetTable(GetFanoCode(ParseString(msg)));
        auto coded = CodeMessage(table, msg);
        std::wcout << "FANO\n"<< "\nНачальная длина: " <<bit code.length()
                   << " Коэффициент: " << float(bit code.length()) /
                                                coded.length()
                   << " Средняя длина кода: "<< GetAvg(table, msg)
                   << " Дисперсия: " << GetDispersion(table, msg) << "\n";
        std::wcout << " Декодирование: \n" <<
        getDecodedMessage(GetDecodingTable(table),coded) << "\n";</pre>
    for (auto &i: k) {
        auto msg = changeCode(codeParsing(bit code, i));
        auto table = GetTable(GetHuffmanCode(ParseString(msq)));
        auto coded = CodeMessage(table, msg);
        std::wcout <<"HUFFMAN\n"<<"\nНачальная длина:" <<bit code.length()
                   << " Коэффициент: " << float(bit code.length()) /
                                                coded.length()
                   << " Средняя длина кода: "
                   << GetAvg(table, msg) << " Дисперсия : " <<
                   GetDispersion(table, msg) << "\n";</pre>
        std::wcout << " Декодирование :\n" <<
       getDecodedMessage(GetDecodingTable(table),
                                                          coded) << "\n";</pre>
    }
   return 0;
}
```

Декодированные сообщения: Для 2-х символьного алфавита

Y0 **0**0 **0Y**0 **0Y**0**0Y**0 **00 0Y**0 **0Y**0 **0Y**0 **9 Y**9 9**9** 9**Y**9 **9Y** 9**Y**9 **9¥9 ¥9 99 9Y9 99 9Y9 9Y99Y9 9 Y9 9Y 9Y9** 8Y8 Y9 8Y9YY9 8Y98Y9 8Y Y9 8Y8Y 8Y 8 Y9 8Y88Y9 ♥⊕ 9Y YY0 09 YY0 9Y0 Y0 9Y00Y0 9Y9Y 9 Y0 9YY0Y0 9Y 9 Y0 **Y**0 **0Y Y**0 **9Y9YY9 9YY9Y9 99 9YY9** Θ **9**¥99 **9** ¥9 **9**¥9¥ **Y**0 **0Y Y**0 **0Y**0 **Y**0 **00 0Y 0Y0 0 Y0 00 0Y0 0YY0 0 Y0 0Y00Y**0 **9Y9Y9 9Y 9Y9 977979 99 9779 977 9 84 48 848448** YO OB BYO BYYBYO BYYBYO BYOOYO BYYOYO OB Y YO BYBBYO **848 A0 844 8 A0 84 A40 844840 848440 844840 88 840** 84840 844840 844 84 8 40 08 840 8448 8 40 84 440 08 **YO OO YYO OY OYO OO OYYO OYY O YO OY OYO OY YO 9Y9 8Y88Y9 8Y Y9 98 8Y9 8YY8Y9 8YY 8 Y9 98 ♥**⊕ 87 70 08 070 877770 870070 87 870 87 70 8787 8 70 977979 9779 9 79 99 979 97 979 977979 99 79 9 79 97 9 9YY9Y9 8Y88Y9 98 Y 8Y8 8 Y9 889 Y9 **⊕♥ ⊕♥⊕ ♥♥⊕♥⊕** 977979 977979 97 79 98 **Y**9 **9Y**9 **9 Y**9 **99 9YY**9 **99** 9 40 640440 **64 40 644 40** 9♥9 99 ♥ **⊕**♥⊕⊕♥⊕ ⊕**♥**⊕ 9YY@Y@ 9YY9 9 Y@ 9YY Y@ 9Y9 Y@ 9Y YY@ 9Y Y@ 9Y9YY@ 946 6 46 66 949 848846 84 846 844846 848446 88 **YO OYYOYO OYY YO OYYOYO OO OYYO** YO 8Y8YYO 8YY8YO 8YYYYO OB Y YO OB YY 8 YO OB OYO 8YY9YO **848848 84 448 84 844848 84 848 848 48 ⊕♥ ♥ ♥ ⊎** 9Y9YY9 98 Y Y9 8YY9Y9 8YY8 8 Y9 98 9Y9 98 8YY9 8YYYY9 8Y YO BY BYO OB B YO BYB YO BYOOYO OB OYO OB YY B YO 977979 97 **9 Y**9 9**9 9Y**9 9**9 Y**9 9**8 8Y 8Y 9 Y**9 **8Y**9**Y**9 **8Y ♥**⊕ **♥**♥⊕ 9Y9YY9 9Y9 Y9 98 98Y9 98 Y 8Y 8 Y9 8YY9Y9 8Y **⊕** ♥⊕ ⊕**⊕ YOYO 9Y88YO 9Y8 YO 9YYYYO 9Y YO 9Y08YO 9Y8 8Y 8 \@@\@\\@@\\@\@** @**@ @ \@ @** **@ @\@** \@ **@\@@\@ @\@@ @** \@ **84444**® **88 ♥**⊕ 9₩99 **9¥¥999¥9 84464**8 **99 Y00Y0 Y0 00 00 0Y0**



Для 8-и символьного алфавита

Ветер свистел, визжал, кряхтел и гудел на разные лады. То жалобным тоненьким голоском, то грубым басовым раскатом распевал он свою боевую песенку. Фонари чуть заметно мигали сквозь огромные белые хлопья снега, обильно сыпавшиеся на тротуары, на улицу, на экипажи, лошадей и прохожих.

Вывод: в ходе выполнения лабораторной работы выяснено, что эффективнее всего кодируются последовательности длиной в 8 бит. 2-битные символы сжать практически невозможно, поскольку при таком маленьком алфавите и относительно большом кол-ве символов будет наблюдаться низкая энтропия, при которой сжатие методами Хаффмана и Шеннона-Фано не имеет смысла.

Оба алгоритма имеют одинаковую вычислительную сложность декодирования (если записать коды с помощью хеш-таблицы). Алгоритм Хаффмана при кодировании требует дополнительной памяти, в отличие от алгоритма Шеннона-Фано, однако выполняется итеративно с линейной сложностью, в то время, как алгоритм Шеннона Фано имеет квадратичную вычислительную сложность и выполняется рекурсивно