

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)



ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И УПРАВЛЯЮЩИХ СИСТЕМ

Лабораторная работа №2
по дисциплине: Теория информации
тема: «Исследование кодов Шеннона-Фано.»

Выполнил: ст. группы ПВ-223
Пахомов Владислав Андреевич

Проверили:
пр. Твердохлеб Виталий Викторович

Белгород 2024 г.

Лабораторная работа №2

Исследование кодов Шеннона-Фано

Цель работы: исследовать кодирование по методу Шеннона-Фано. Научиться оценивать эффективности кода.

Задание 1. Построить код для сообщения, содержащего строку панграммы «в чашах юга жил бы цитрус? Да, но фальшивый экземпляр!». Для полученного кода рассчитать показатели коэффициента сжатия и дисперсии.

| Символ | Вероятность | Этапы | | | | | | Код |
|--------|-------------|-------|----|-----|----|---|----|--------|
| | | I | II | III | IV | V | VI | |
| пробел | 0,166666667 | 1 | 1 | 1 | | | | 111 |
| а | 0,092592593 | | | 0 | | | | 110 |
| л | 0,055555556 | | 0 | 1 | 1 | | | 1011 |
| и | 0,055555556 | | | | 0 | | | 1010 |
| в | 0,037037037 | | | 0 | 1 | 1 | | 10011 |
| ы | 0,037037037 | | | | | 0 | | 10010 |
| р | 0,037037037 | | | | 0 | 1 | | 10001 |
| п | 0,018518519 | | | | | 0 | | 10000 |
| ? | 0,018518519 | 1 | 1 | 1 | 1 | 1 | | 01111 |
| о | 0,018518519 | | | | | 0 | 1 | 011101 |
| н | 0,018518519 | | | | | | 0 | 011100 |
| м | 0,018518519 | | | 0 | 1 | 1 | | 01101 |
| к | 0,018518519 | | | | | 0 | 1 | 011001 |
| й | 0,018518519 | | | | 0 | | 0 | 011000 |
| з | 0,018518519 | | 0 | 1 | 1 | 1 | | 01011 |
| ж | 0,018518519 | | | | | 0 | 1 | 010101 |
| е | 0,018518519 | | | | | | 0 | 010100 |
| г | 0,018518519 | | | 0 | 1 | 1 | | 010011 |
| б | 0,018518519 | | | | | 0 | 1 | 010010 |
| , | 0,018518519 | | | | 0 | | 1 | 010001 |
| ! | 0,018518519 | | | | | | 0 | 010000 |
| д | 0,018518519 | | 0 | 1 | 1 | 1 | | 00111 |
| я | 0,018518519 | | | | | 0 | 1 | 001101 |
| ю | 0,018518519 | | | | | | 0 | 001100 |
| э | 0,018518519 | | | 0 | 1 | 1 | | 001011 |
| ь | 0,018518519 | | | | | | 0 | 00101 |
| щ | 0,018518519 | | | | 0 | 1 | | 001001 |
| ш | 0,018518519 | | | | | | 0 | 001000 |
| ч | 0,018518519 | | 0 | 1 | 1 | 1 | | 00011 |
| ц | 0,018518519 | | | | | 0 | 1 | 000101 |
| х | 0,018518519 | | | | | | 0 | 000100 |
| ф | 0,018518519 | | | 0 | 1 | 1 | | 000011 |
| у | 0,018518519 | | | | | | 0 | 000010 |
| т | 0,018518519 | | | | 0 | 1 | | 000001 |
| с | 0,018518519 | | | | | | 0 | 000000 |

Код сообщения: 10011 111 00011 110 001001 110 000100 111 001100 010011 110 111 010101 1010 1011 111 010010 10010 111 000101 1010 000001 10001 000010 000000 01111 111 00111 110 010001 111 011100 011101 111 000011 110 1011 001010 001000 1010 10011 10010 011000 111 001011 011001 01011 010100 01101 10000 1011 001101 10001 010000

$$B = 54 \cdot 8 = 432$$

$$B' = 258$$

$$K_{comp} = \frac{B}{B'} = \frac{432}{258} = 1 \frac{29}{43}$$

$$l = 4.7778$$

$$\delta = 1.5062$$

Задание 2. Построить код для сообщения, содержащего строку «Victoria nulla est, Quam quae confessos animo quoque subjugat hostes» Для полученного кода рассчитать показатели коэффициента сжатия и дисперсии.

| Символ | Вероятность | Этапы | | | | | | | Код |
|--------|-------------|-------|----|-----|----|---|----|-----|---------|
| | | I | II | III | IV | V | VI | VII | |
| пробел | 9 | 1 | 1 | 1 | | | | | 111 |
| u | 7 | | | 0 | | | | | 110 |
| s | 7 | | 0 | 1 | | | | | 101 |
| o | 6 | | | 0 | 1 | | | | 1001 |
| a | 6 | | | | 0 | | | | 1000 |
| e | 5 | 0 | 1 | 1 | 1 | | | | 0111 |
| t | 4 | | | 0 | 0 | | | | 0110 |
| q | 3 | | | 0 | 1 | | | | 0101 |
| n | 3 | | | | 0 | | | | 0100 |
| i | 3 | | 1 | 1 | 1 | | | | 00111 |
| m | 2 | | | 0 | 0 | | | | 00110 |
| l | 2 | | | | 1 | | | | 00101 |
| c | 2 | | | | 0 | | | | 00100 |
| V | 1 | | 0 | 1 | 1 | 1 | | | 000111 |
| r | 1 | | | | 0 | 0 | | | 000110 |
| Q | 1 | | | | 0 | 1 | | | 000101 |
| , | 1 | | | | | 0 | | | 000100 |
| j | 1 | | | 0 | 1 | 1 | | | 000011 |
| h | 1 | | | | | 0 | | | 000010 |
| g | 1 | | | | 0 | 1 | | | 000001 |
| f | 1 | | | | | 0 | 1 | | 0000001 |
| b | 1 | | | | | | 0 | | 0000000 |

Код сообщения: 000111 00111 00100 0110 1001 000110 00111 1000 111 0100 110 00101 00101 1000 111 0111 101 0110 000100 111 000101 110 1000 00110 111 0101 110 1000 0111 111 00100 1001 0100 0000001 0111 101 101 1001 101 111 1000 0100 00111 00110 1001 111 0101 110 1001 0101 110 0111 111 101 110 0000000 000011 110 000001 1000 0110 111 000010 1001 101 0110 0111 101

$$B = 68 \cdot 8 = 544$$

$$B' = 278$$

$$K_{comp} = \frac{B}{B'} = \frac{544}{278} = 1\frac{133}{139}$$

$$l = 4.0882$$

$$\delta = 0.5515$$

Задание 3. Построить консольное приложение, реализующее процесс кодирования по методу Шеннона-Фано (с возможностью расчета коэффициента сжатия и дисперсии).

```
import java.io.BufferedReader;
import java.io.IOException;
import java.io.InputStreamReader;
import java.util.*;
import java.util.stream.Collectors;

public class Main {
    public static class TableElement {
        char symbol;
        int amount;
        List<Boolean> code;

        public TableElement(char symbol) {
            this.symbol = symbol;
            this.amount = 1;
            this.code = new ArrayList<>();
        }
    }

    public static List<TableElement> schennonFano(String input) {
        List<TableElement> table = new ArrayList<>();
        for (char symbol : input.toCharArray()) {
            Optional<TableElement> result = table.stream().filter((el) -> el.symbol == symbol).findAny();
            if (result.isPresent()) {
                result.get().amount++;
            } else {
                table.add(new TableElement(symbol));
            }
        }

        table.sort(Comparator.comparingInt(o -> o.amount));
        Collections.reverse(table);

        schennonFano(table, 0, table.size());

        return table;
    }

    public static void schennonFano(List<TableElement> table, int beginIndex, int endIndex) {
        if (endIndex - beginIndex <= 1) return;
        if (endIndex - beginIndex == 2) {
            table.get(beginIndex).code.add(true);
            table.get(beginIndex + 1).code.add(false);
            return;
        }
    }
}
```

```

    int separateIndex = getSeparateIndex(table, beginIndex, endIndex);

    for (int i = beginIndex; i < endIndex; i++) {
        if (i < beginIndex + separateIndex) {
            table.get(i).code.add(true);
        } else {
            table.get(i).code.add(false);
        }
    }
}

schennonFano(table, beginIndex, separateIndex);
schennonFano(table, separateIndex, endIndex);
}

private static int getSeparateIndex(List<TableElement> table, int beginIndex, int endIndex) {
    int sum = 0;
    for (int i = beginIndex; i < endIndex; i++) {
        sum += table.get(i).amount;
    }
    int sumBefore = table.get(beginIndex).amount;
    int sumAfter = sum - sumBefore;
    int separateIndex = beginIndex + 1;
    while (separateIndex < endIndex - 1 && sumAfter - table.get(separateIndex).amount -
        (sumBefore + table.get(separateIndex).amount) > 0 ) {
        sumAfter -= table.get(separateIndex).amount;
        sumBefore += table.get(separateIndex).amount;
        separateIndex++;
    }

    if (Math.abs(sumBefore - sumAfter) > Math.abs(sumAfter - table.get(separateIndex).amount -
        (sumBefore + table.get(separateIndex).amount))) {
        separateIndex++;
    }
    return separateIndex;
}

public static void main(String[] args) throws IOException {
    System.out.println("Введите сообщение: ");
    String input;
    BufferedReader r = new BufferedReader(new InputStreamReader(System.in));
    input = r.readLine();
    System.out.println();

    List<TableElement> table = schennonFano(input);

    System.out.println("Таблица: ");
    for (TableElement element : table) {
        System.out.print(element.symbol);
        System.out.print(" " + element.amount + " ");
        for (int i = 0; i < element.code.size(); i++) {
            System.out.print(element.code.get(i) ? "1" : "0");
        }
        System.out.println();
    }
}

```

```

System.out.println();
System.out.println("Закодированное сообщение: ");
String code = input;
int sum = 0;
for (TableElement element : table) {
    String elementCode = element.code.stream().map((el) -> el ? "1" : "0").collect(Collectors.joining(""));
    code = code.replace("" + element.symbol, elementCode);
    sum += element.amount;
}
System.out.println(code);
System.out.println();

int codedLength = code.length();
int uncodedLength = input.length() * 8;
System.out.println("Коэффициент сжатия: " + 1.0 * uncodedLength / codedLength);
System.out.println();
double midLen = 0;
for (TableElement element : table) {
    String elementCode = element.code.stream().map((el) -> el ? "1" : "0").collect(Collectors.joining(""));
    midLen += elementCode.length() * (1.0 * element.amount / sum);
}
System.out.println("Средняя длина: " + midLen);
System.out.println();

double delta = 0;
for (TableElement element : table) {
    String elementCode = element.code.stream().map((el) -> el ? "1" : "0").collect(Collectors.joining(""));
    delta += (1.0 * element.amount / sum) * (elementCode.length() - midLen) * (elementCode.length() - midLen);
}
System.out.println("Дисперсия: " + delta);
}
}

```

Задание 4. Получить кодовые представления сообщений из пунктов 1 и 2 задания по методу Хаффмана. Сравнить полученные результаты с методом Шеннона-Фано по показателям сжатия и дисперсии. Сделать соответствующие выводы.

Результаты выполнения программы для задания 1:

Введите сообщение:

в чащах юга жил бы цитрус? Да, но фальшивый экземпляр!

Таблица:

```

9 111
а 5 110
л 3 1011
и 3 1010
р 2 10111
ы 2 10110
в 2 10011
! 1 10010
я 1 01111
п 1 011111
м 1 011110

```

е 1 01111
э 1 011111
к 1 011110
э 1 01111
й 1 011111
ш 1 011110
ь 1 011111
ф 1 011110
о 1 011111
н 1 011110
, 1 01111
д 1 011111
? 1 011110
с 1 011111
у 1 011110
т 1 011111
ц 1 011110
б 1 01111
ж 1 001111
г 1 001110
ю 1 001111
х 1 001110
щ 1 001111
ч 1 001110

Закодированное сообщение:

100111110011101100011111100011101110011
11001110110111001111101010111110111101
1011101111010100111111011110111100111110
1111011101111111001111110111100111111
1011110110101111110111101010100111011
00111111101111011110011111011110111100
111111011011111011110010

Коэффициент сжатия: 1.6744186046511629

Средняя длина: 4.777777777777776

Дисперсия: 1.5061728395061729

Результаты выполнения программы для задания 2:

Введите сообщение:

Victoria nulla est, Quam quae confessos animo quoque subjugat hostes

Таблица:

9 111
s 7 110
u 7 101
a 6 1011
o 6 1010
e 5 0111
t 4 0110

```
q 3 0111
n 3 0110
i 3 01111
m 2 01110
l 2 01111
c 2 01110
h 1 011111
g 1 001110
j 1 001111
b 1 001110
f 1 001111
Q 1 001110
, 1 001111
r 1 0011111
V 1 0011110
```

Закодированное сообщение:

```
001111001111011100110101000111110111110
111110110101011110111110111110111110011
00011111110011101011011011101111011
011011111101110101001100011110111110110
1010110111101101100111101110101110111
101101001111010111111110101001110001111
10100111010110111011111111010110011001
11110
```

Коэффициент сжатия: 1.9568345323741008

Средняя длина: 4.088235294117648

Дисперсия: 1.1392733564013837

Вывод: в ходе лабораторной работы исследовали кодирование по методу Шеннона-Фано. Научиться оценивать эффективности кода. Выполнили реализацию метода кодирования по методу Шеннона-Фано в виде консольного приложения.