

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 № 4
по дисциплине: Теория информации
тема: «Исследование кода Гилберта-Мура»

Выполнил: ст. группы ПВ-223
Игнатъев Артур Олегович

Проверил:
Твердохлеб Виталий Викторович

Белгород 2024г.

Лабораторная работа №4

«Исследование кода Гилберта-Мура»

Цель работы: исследовать кодирование по методу Гилберта-Мура.

1. Построить обработчик, выполняющий компрессию по алгоритму Гилберта-Мура.

```
public static class HilbertMurielTableElement {
    public byte symbol = 0;
    public double p = 0;
    public double d = 0;
    public double delta = 0;
    public int logp = 0;
    public int code = 0;
    public int amount = 0;
    public HilbertMurielTableElement(byte symbol) {this.symbol = symbol;}
}

public static List<HilbertMurielTableElement> getHilbertMurielTableElement(List<Byte> input) {
    List<HilbertMurielTableElement> result = new ArrayList<>();
    List<TableElement> segTable = getSegmentisedTable(input);
    if (segTable.isEmpty()) return new ArrayList<>();
    HilbertMurielTableElement hmElement = new HilbertMurielTableElement(segTable.get(0).symbol);
    hmElement.p = (1.0 * segTable.get(0).amount) / input.size();
    hmElement.delta = hmElement.p / 2;
    hmElement.amount = segTable.get(0).amount;
    int i = 0;
    while (true) {
        hmElement.logp = (int) Math.ceil(-(Math.log(hmElement.p) / Math.log(2))) + 1;
        double tmp = hmElement.delta;
        for (int j = 0; j < hmElement.logp; j++) {
            hmElement.code = hmElement.code * 2 + (((int) tmp) & 1);
            tmp *= 2;
        }
        result.add(hmElement);
        i++;
        if (i >= segTable.size()) break;
        TableElement element = segTable.get(i);
        hmElement = new HilbertMurielTableElement(segTable.get(i).symbol);
        hmElement.amount = segTable.get(i).amount;
        hmElement.p = (1.0 * element.amount) / input.size();
        for (int j = 0; j < i; j++) {
            hmElement.d += result.get(j).p;
        }
        hmElement.delta = hmElement.d + hmElement.p / 2;
    }
    return result;
}
```

2. Создать генераторы данных, работающих как источники Хартли и Бернулли (в двоичном алфавите).

```
#include <codecvt>
#include <iostream>
#include <map>
#include <random>
#include <string>
#include <locale>
#include <fstream>

std::string HartliGenerator(int n) {
    std::string t;
    std::random_device rd;
    std::mt19937 gen(rd());
    std::uniform_int_distribution<unsigned char> d(0, 127);
    for (int i = 0; i < n; i++) {
        t.push_back(d(gen));
    }
    return t;
}

std::string BernoulliGenerator(int n, float pivot) {
    std::string t;
    std::random_device rd;
    std::mt19937 gen(rd());
    std::bernoulli_distribution d(pivot);
    for (int i = 0; i < n; i++) {
        unsigned char res;
        for (int j = 0; j < 8; j++) {
            res += res * 2 + d(gen);
        }
        t.push_back(res);
    }
    return t;
}

int main() {
    std::ofstream out("out.txt");
    out << BernoulliGenerator(100, 0.5);
    out.flush();
    out.close();
}
```

3. Построить коды Гилберта-Мура для последовательностей длиной 100 символов, сгенерированных источником Хартли, Бернулли и любой русский текст, для чего предварительно сегментировать каждую цепочку по 8 символов. Вычислить полученные коэффициенты сжатия и величину дисперсии для каждой последовательности. Результаты кодирования занести в сводную таблицу.

Хартли:

Таблица:

Код символа	Вероятность	α	Дельта	Количество бит	Код
73	0,039	0,000	0,020	6	000000
55	0,029	0,039	0,054	7	0000011
56	0,029	0,069	0,083	7	0000101
18	0,029	0,098	0,113	7	0000111
77	0,029	0,127	0,142	7	0001001
41	0,029	0,157	0,172	7	0001010
14	0,029	0,186	0,201	7	0001100
118	0,020	0,216	0,225	7	0001110
72	0,020	0,235	0,245	7	0001111
48	0,020	0,255	0,265	7	0010000
29	0,020	0,275	0,284	7	0010010
78	0,020	0,294	0,304	7	0010011
123	0,020	0,314	0,324	7	0010100
102	0,020	0,333	0,343	7	0010101
58	0,020	0,353	0,363	7	0010111
105	0,020	0,373	0,382	7	0011000
10	0,020	0,392	0,402	7	0011001
13	0,020	0,412	0,422	7	0011010
11	0,020	0,431	0,441	7	0011100
86	0,020	0,451	0,461	7	0011101
106	0,020	0,471	0,480	7	0011110
24	0,020	0,490	0,500	7	0011111
96	0,020	0,510	0,520	7	0100001
101	0,020	0,529	0,539	7	0100010
63	0,020	0,549	0,559	7	0100011
91	0,020	0,569	0,578	7	0100101
114	0,010	0,588	0,593	8	01001011
79	0,010	0,598	0,603	8	01001101
22	0,010	0,608	0,613	8	01001110
7	0,010	0,618	0,623	8	01001111
68	0,010	0,627	0,632	8	01010000
27	0,010	0,637	0,642	8	01010010
111	0,010	0,647	0,652	8	01010011
23	0,010	0,657	0,662	8	01010100
54	0,010	0,667	0,672	8	01010101
28	0,010	0,676	0,681	8	01010111
64	0,010	0,686	0,691	8	01011000
112	0,010	0,696	0,701	8	01011001
127	0,010	0,706	0,711	8	01011010
1	0,010	0,716	0,721	8	01011100
35	0,010	0,725	0,730	8	01011101
108	0,010	0,735	0,740	8	01011110
31	0,010	0,745	0,750	8	01011111

65	0,010	0,755	0,760	8	01100001
84	0,010	0,765	0,770	8	01100010
117	0,010	0,775	0,779	8	01100011
12	0,010	0,784	0,789	8	01100101
116	0,010	0,794	0,799	8	01100110
124	0,010	0,804	0,809	8	01100111
61	0,010	0,814	0,819	8	01101000
34	0,010	0,824	0,828	8	01101010
9	0,010	0,833	0,838	8	01101011
5	0,010	0,843	0,848	8	01101100
83	0,010	0,853	0,858	8	01101101
30	0,010	0,863	0,868	8	01101111
47	0,010	0,873	0,877	8	01110000
98	0,010	0,882	0,887	8	01110001
103	0,010	0,892	0,897	8	01110010
3	0,010	0,902	0,907	8	01110100
59	0,010	0,912	0,917	8	01110101
2	0,010	0,922	0,926	8	01110110
42	0,010	0,931	0,936	8	01110111
92	0,010	0,941	0,946	8	01111001
82	0,010	0,951	0,956	8	01111010
36	0,010	0,961	0,966	8	01111011
50	0,010	0,971	0,975	8	01111100
66	0,010	0,980	0,985	8	01111110
19	0,010	0,990	0,995	8	01111111

Закодированное сообщение:

```
01111111010010101000110100010011111100001100010000100111110111110001
11101101111010011110010011110001110101110111011101100011100000101001
11010100010010111010001110010001101000110010001001011100010011000011
10000000011101101111011011010010111001010101101100000100101000100000
00001010000011000110101101101010000011100111000001010000010100000001
10100000100110010010000001101100111001111000100000110011000110000000
00001000001100101001011101100011011000100001100011000010011101010001
10011111010111110101111000100100010100010111010000101010000100001010
10111000010101010110100101100101011000010101110010011000101000011110
1010101000001100011100101010001010011000111101010010000111001010000
00000000011101001111010011100100110100110100011001000001101001010100
1011
```

Коэффициент сжатия: 1.0851063829787233

Средняя длина: 7.37254901960783

Дисперсия: 0.312187620146097

Бернулли:

Таблица:

Код символа	Вероятность	d	Дельта	Количество бит	Код	
=====						
35	0,030	0,000	0,015	7	0000000	
112	0,020	0,030	0,040	7	0000010	
-84	0,020	0,050	0,060	7	0000011	
80	0,020	0,070	0,080	7	0000101	
18	0,020	0,090	0,100	7	0000110	
-102	0,020	0,110	0,120	7	0000111	
95	0,020	0,130	0,140	7	0001000	
92	0,020	0,150	0,160	7	0001010	
-37	0,020	0,170	0,180	7	0001011	
64	0,020	0,190	0,200	7	0001100	

30	0,020	0,210	0,220	7	0001110
-75	0,020	0,230	0,240	7	0001111
96	0,020	0,250	0,260	7	0010000
16	0,020	0,270	0,280	7	0010001
-3	0,020	0,290	0,300	7	0010011
-70	0,020	0,310	0,320	7	0010100
3	0,010	0,330	0,335	8	00101010
-2	0,010	0,340	0,345	8	00101100
12	0,010	0,350	0,355	8	00101101
-89	0,010	0,360	0,365	8	00101110
-26	0,010	0,370	0,375	8	00110000
62	0,010	0,380	0,385	8	00110001
-114	0,010	0,390	0,395	8	00110010
41	0,010	0,400	0,405	8	00110011
45	0,010	0,410	0,415	8	00110101
-29	0,010	0,420	0,425	8	00110110
-77	0,010	0,430	0,435	8	00110111
94	0,010	0,440	0,445	8	00111000
26	0,010	0,450	0,455	8	00111010
0	0,010	0,460	0,465	8	00111011
31	0,010	0,470	0,475	8	00111100
53	0,010	0,480	0,485	8	00111110
88	0,010	0,490	0,495	8	00111111
-85	0,010	0,500	0,505	8	01000000
73	0,010	0,510	0,515	8	01000001
75	0,010	0,520	0,525	8	01000011
-9	0,010	0,530	0,535	8	01000100
15	0,010	0,540	0,545	8	01000101
13	0,010	0,550	0,555	8	01000111
-111	0,010	0,560	0,565	8	01001000
-105	0,010	0,570	0,575	8	01001001
-118	0,010	0,580	0,585	8	01001010
-115	0,010	0,590	0,595	8	01001100
78	0,010	0,600	0,605	8	01001101
74	0,010	0,610	0,615	8	01001110
-20	0,010	0,620	0,625	8	01010000
-7	0,010	0,630	0,635	8	01010001
-27	0,010	0,640	0,645	8	01010010
-112	0,010	0,650	0,655	8	01010011
-36	0,010	0,660	0,665	8	01010101
-126	0,010	0,670	0,675	8	01010110
-19	0,010	0,680	0,685	8	01010111
69	0,010	0,690	0,695	8	01011000
4	0,010	0,700	0,705	8	01011010
-48	0,010	0,710	0,715	8	01011011
11	0,010	0,720	0,725	8	01011100
68	0,010	0,730	0,735	8	01011110
102	0,010	0,740	0,745	8	01011111
-128	0,010	0,750	0,755	8	01100000
-21	0,010	0,760	0,765	8	01100001
84	0,010	0,770	0,775	8	01100011
1	0,010	0,780	0,785	8	01100100
36	0,010	0,790	0,795	8	01100101
-32	0,010	0,800	0,805	8	01100111
-23	0,010	0,810	0,815	8	01101000
19	0,010	0,820	0,825	8	01101001
6	0,010	0,830	0,835	8	01101010
70	0,010	0,840	0,845	8	01101100
-5	0,010	0,850	0,855	8	01101101
-124	0,010	0,860	0,865	8	01101110
107	0,010	0,870	0,875	8	01110000

-12	0,010	0,880 0,885 8	01110001
21	0,010	0,890 0,895 8	01110010
87	0,010	0,900 0,905 8	01110011
-42	0,010	0,910 0,915 8	01110101
127	0,010	0,920 0,925 8	01110110
-1	0,010	0,930 0,935 8	01110111
81	0,010	0,940 0,945 8	01111000
85	0,010	0,950 0,955 8	01111010
83	0,010	0,960 0,965 8	01111011
-62	0,010	0,970 0,975 8	01111100
56	0,010	0,980 0,985 8	01111110
-24	0,010	0,990 0,995 8	01111111

Закодированное сообщение:

```
00100110010011001000100100110111111000100000010011011110110010011011
11010011110000010011011101100010011000111001110011011100100010011011
10000000110000100110010011001001101101100011010100110100100100110001
01000100110110010101100100011000110010011010111110010001010111100001
11001011100001001101011010010110000010011001001100010000010011001001
10010011001001100001100010011001001100001100100111001001101001001100
10011001001100100110100011100100000100010100100110100001101000001001
00110010011001111110011111000001010011110000100110000000001110110010
01100111010000110000111000001001100100110001010001101010011001100001
01000100000000000010011000001000100110000010001001100110001001001100
00000001001100101101001001100101010
```

Коэффициент сжатия: 1.0853146853146853

Средняя длина: 7.6700000000000005

Дисперсия: 0.22110000000000009

Обычный текст:

Таблица:

Код символа	Вероятность	d	Дельта	Количество бит	Код
=====	=====	=====	=====	=====	=====
-48	0,277	0,000	0,138	3	000
-47	0,177	0,277	0,365	4	0010
32	0,071	0,454	0,489	5	00111
-66	0,057	0,525	0,553	6	010001
-80	0,035	0,582	0,599	6	010011
-126	0,035	0,617	0,635	6	010100
-78	0,028	0,652	0,667	7	0101010
-117	0,028	0,681	0,695	7	0101100
-128	0,028	0,709	0,723	7	0101110
-69	0,021	0,738	0,748	7	0101111
-67	0,021	0,759	0,770	7	0110001
-127	0,021	0,780	0,791	7	0110010
46	0,014	0,801	0,809	8	01100111
-75	0,014	0,816	0,823	8	01101001
-120	0,014	0,830	0,837	8	01101011
-125	0,014	0,844	0,851	8	01101100
-76	0,014	0,858	0,865	8	01101110
-65	0,014	0,872	0,879	8	01110000
-79	0,014	0,887	0,894	8	01110010
-68	0,014	0,901	0,908	8	01110100
-71	0,007	0,915	0,918	9	011101011
-113	0,007	0,922	0,926	9	011101100
-94	0,007	0,929	0,933	9	011101110
-123	0,007	0,936	0,940	9	011110000

-72	0,007	0,943	0,947	9	011110010
-116	0,007	0,950	0,954	9	011110100
-119	0,007	0,957	0,961	9	011110110
-77	0,007	0,965	0,968	9	011110111
-121	0,007	0,972	0,975	9	011111001
44	0,007	0,979	0,982	9	011111011
-70	0,007	0,986	0,989	9	011111101
-92	0,007	0,993	0,996	9	011111111

Закодированное сообщение:

```
0000111111100100101110000010001001001100100010010100000011101000000100010010
01011100000110001001110000111001000100101100000010111100111000010101000100101
1000000111110100001000100001010100000100110000110001011111011001110010011111
00100100101000000100010000111001000100101100001110000111000000001000100001111
01110000101111000010001001001111011000001001100100101000010011110100001110000
11011100010011011000010011010110000111100100011100100110010000011101000000110
10010010010111000100101000000110001001001011000010011110000011001110011100001
11011100000101010000010001001001110110000111000011011100010011011000010011010
11000010011001110010011001000100101000000100110000101111000010011001110000111
00000000110100100100101110000010101000001000100001110101101100111
```

Коэффициент сжатия: 1.4881266490765173

Средняя длина: 5.375886524822696

Дисперсия: 4.248780242442529

Вывод: в ходе лабораторной работы исследовали кодирование по методу ГилбертаМура.