МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное автономное образовательное учреждение высшего образования

«САНКТ-ПЕТЕРБУГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»

КАФЕДРА № 43

ОТЧЕТ   
ЗАЩИЩЕН С ОЦЕНКОЙ

ПРЕПОДАВАТЕЛЬ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Старший преподаватель |  |  |  | Фоменкова А.А. |
| Ассистент |  |  |  | Величко М.В. |
| должность, уч. степень, звание |  | подпись, дата |  | инициалы, фамилия |

|  |
| --- |
| ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ № 7 |
| «Оценка количества информации в сообщении и эффективное кодирование» |
| по курсу: ИНФОРМАТИКА |
|  |
|  |

РАБОТУ ВЫПОЛНИЛ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| СТУДЕНТ ГР. № | 4134К |  |  |  | Опарин С.Н. |
|  |  |  | подпись, дата |  | инициалы, фамилия |

Санкт-Петербург 2021

***Цель работы:***

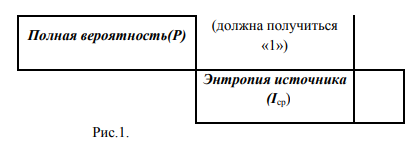
Получение практических навыков численного определения количества информации, содержащегося в сообщении. Освоение методов построения кодов дискретного источника информации используя конструктивный метод, предложенный К.Шенноном и Н.Фано, и метод Хаффмана. На примере показать однозначность раскодирования имеющегося сообщения.

***ЧАСТЬ 1.***

**«*Определение количества информации, содержащейся в сообщении»***

***Порядок выполнения части 1 лабораторной работы***

1. Создать таблицу (50 рабочих строк) в Excel аналогичную рис.1.

2Заполнить столбец Символ следующими значениями:

¬ 33 буквы русского алфавита;

¬ 10 цифр (0 — 9);

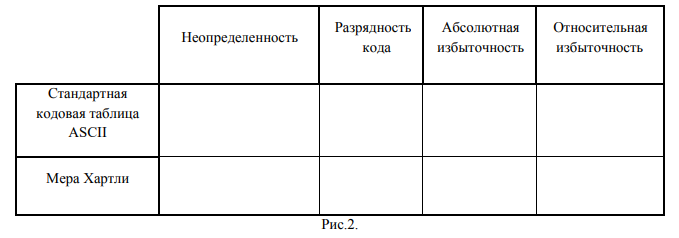
¬ Знаки препинания – «.», «,», «:», «;», «-», « », «(».

3. Заполнить столбец Код символа используя функцию «КОДСИМВ(…)», находящуюся в категории «Текстовые».

4. Открыв каскадом текст по варианту и таблицу и используя в Word «Правка ⇒ Заменить» заполнить столбец Число вхождений символа в текст. (Предполагается, что других символов в тексте НЕТ.) Сосчитать общее число символов.

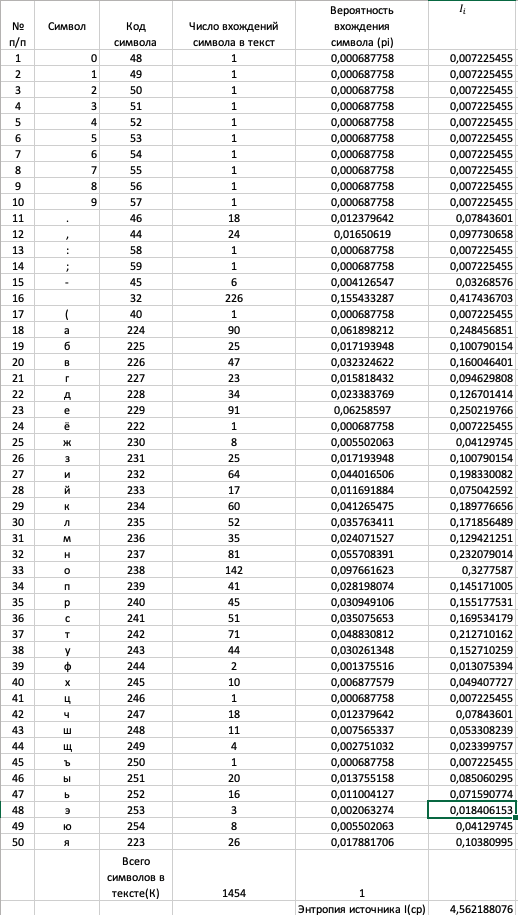
5. По формулам заполнить столбцы «рi» и «Ii». Сосчитать полную вероятность и энтропию источника.

6. Создать таблицу, аналогичную рис.2 и заполнить ее по формулам.

7. Выписать применяемые формулы с расшифровкой используемых символов

**Ход выполнения:**

**Таблица №1**

****

**Таблица №2**



**Текст**

**(сгенерированный)**

Любопытно, что прошепчет Тутошкин кому-нибудь обо мне; я будто бы сколочен без особого изъяна и лицо в норме, но очкарик, без очков двух шагов не ступлю. Я долго форсил, и в десятом классе ходил еще без очков, у карты безнадежно плавал, тыча указкой 1 наугад, но не подавал виду, что уже ничего не различаю. Видимо, этим себе основательно повредил. Паренек на кровати, что стоит к моей спинка в спинку, укладывается спать: уже давно 0 свечерело, легла на парусину роса. В палатке прохладно. Он недавний школьник, этот круглощекий Володя Гришечкин. У него голубые глаза. И физика, не столько дух, сколько букву которой он усердно усваивал с шестого по десятый, как видно, пошла ему впрок. На ночь он "заземляет" себя какой-то проволокой3, в прямом смысле привязывает за ногу к спинке кровати - по его понятиям, все то электричество, которое вобрано им за день, уйдет в землю и облегчит ему сон. Гришечкин смолоду бережется от перегрузок. Он далеко пойдет. Долго не могу4 заснуть. В соседней палатке знакомые5 мне латыши с устрашающим рыком исполняют на губах и на всяких подсобных инструментах импортные джазово-негритянские мотивы. Неунывающий6 народ! Нам бы в палатку такой накал, а то и впрямь дрожь пробирает. Утром заглядывает плотный симпатяга в кожаной куртке на "молниях" - такие парни7 успешно занимаются борьбой. 8Оказывается, он тоже занимался борьбой, но потянуло к 9"интеллектуальному" спорту. Хочет записаться к нам в отделение. Звать - Ким Попов.ё

**Формулы, примененные в работе:**

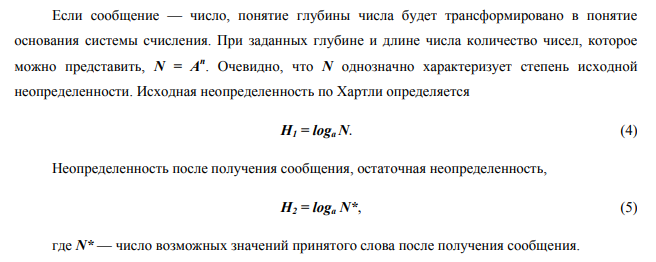
p(i) = k/n

Где: 1) k – количество вхождения символа в текст

2) n – Количество символов в тексте

I(i) = p(i)\*log2(1/p(i)) – Значение энтропии отдельного элемента

Где: 1) p(i) – вероятность вхождение символа

 2) i – символ

**Количество информации по Хартли:**



средняя информация, доставляемая одним опытом,



K - кол-во символов алфавита.

А - количество исходов опыта.

ki - количество появления символа в тексте.

Ii - количество вносимой этим символом информации.

**Вывод:**

В ходе выполнения первой части лабораторной работы были изучены метода определения информации по Шеннону и Хартли, метод Хаффмана

Мера Хартли и неопределенность по Шеннону совпадают, так как все символы в таблице ascii имеют одинаковую вероятность нахождения. В таком случае результаты будут аналогичными. Абсолютная избыточность равна нулю так как предполагается передача без потерь.

***Часть 2.***

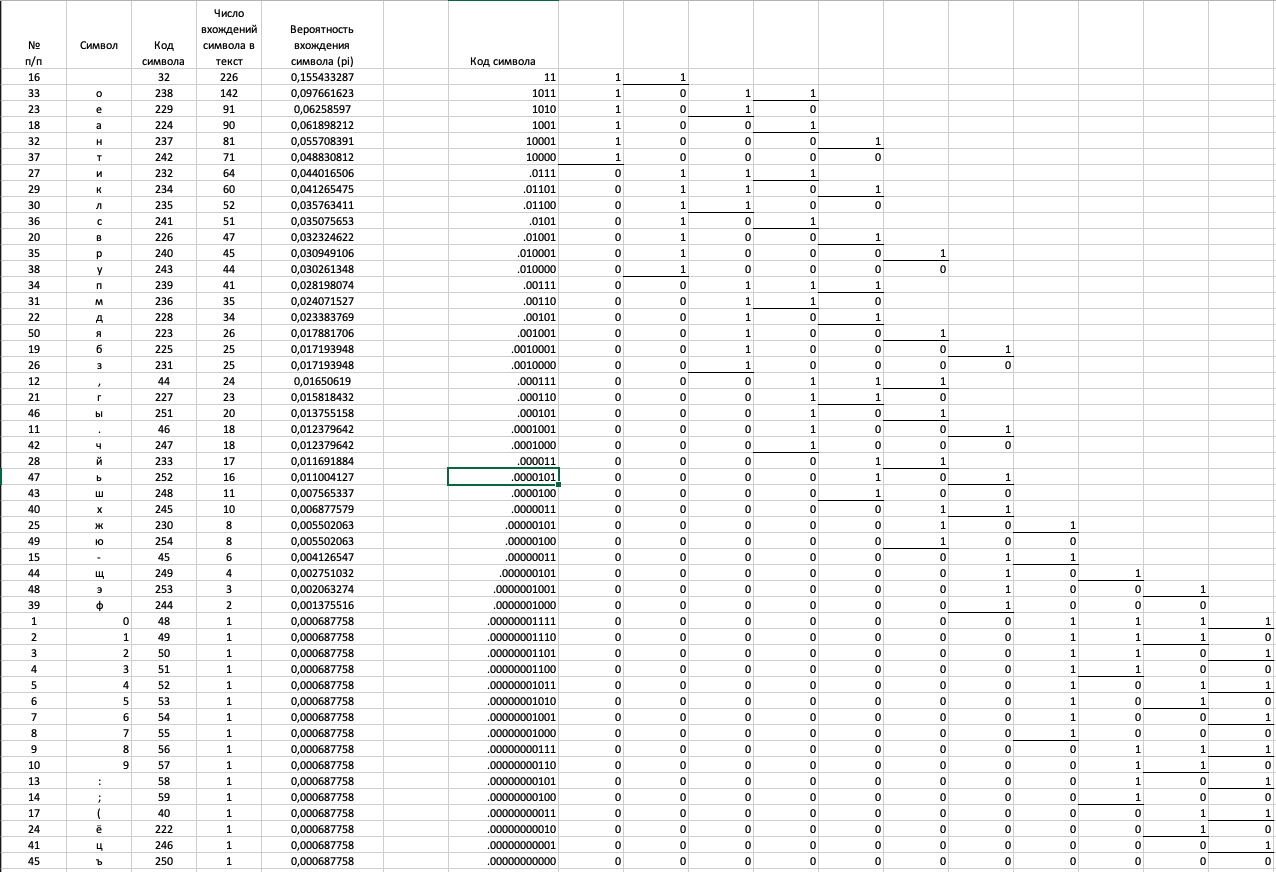
***«Кодирование дискретных источников информации методом Шеннона-Фано»***

***Порядок выполнения части 2 лабораторной работы***

Исходными данными для данной лабораторной работы являются результаты статистической обработки текста, выполненной в предыдущей лабораторной работе. Из лабораторной работы «Определение количества информации, содержащегося в сообщении» для данной работы необходимо взять: 1) список символов данного текста; 2) оценку вероятностей появления символов в тексте; 3) значение энтропии источника. Расчеты рекомендуется выполнять в табличной форме, используя MSExcel. 1. Отсортировать символы в порядке убывания их вероятности появления в тексте. 2. Построить один из возможных вариантов по правилу Шеннона-Фано для посимвольного кодирования заданного текста. 3. Определить энтропию и среднее количество двоичных разрядов, необходимых для передачи текста при использовании эффективных кодов. 4. Проверить возможность однозначного декодирования полученных кодов, рассмотрев пример передачи слова, состоящего из не менее 10 символов.

***Ход выполнения:***

**Таблица для 50-ти символов, содержащая список символов, значения вероятностей, кодовые комбинации, ступени.**



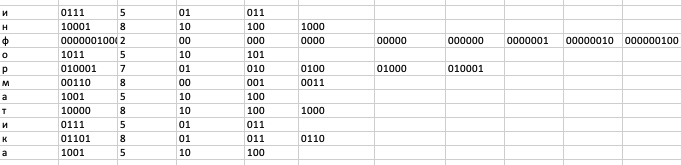
**Значение средней информации в битах =** 7,66

**Применялись формулы подсчета вероятности и энтропии каждого элемента.**

**Закодированное сообщение:**

****

Декодированное сообщение:



**Вывод:**

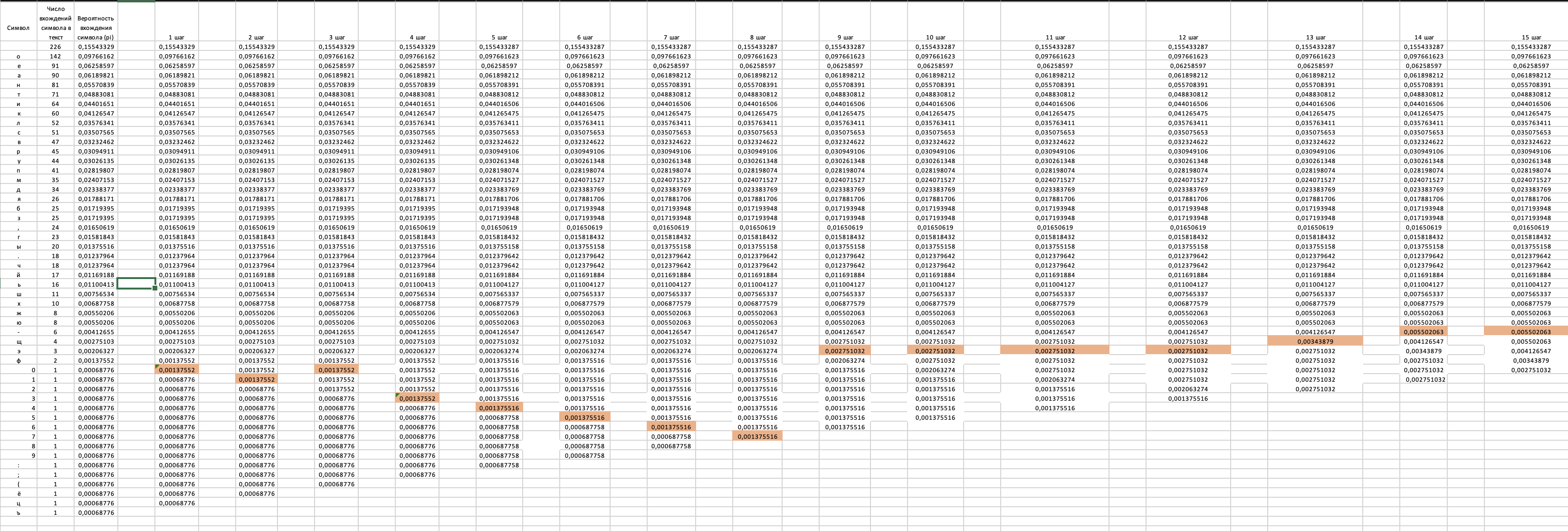
За вторую часть работы я воспользовался методом Шеннона-Фано, собрав все двоичные коды для каждого символа. Проверил на корректность и составил таблицу ступеней.

***Часть 3.***

***«Кодирование дискретных источников информации по методике Хаффмана»***

***Порядок выполнения части 3 лабораторной работы***

Из лабораторной работы «Кодирование дискретных источников информации методом Шеннона-Фано» необходимо взять вычисленное значение средней информации. Расчеты рекомендуется выполнять в табличной форме, используя MSExcel. 1. Отсортировать символы в порядке убывания их вероятности появления в тексте. 2. Построить таблицу по правилу Д. Хаффмана для посимвольного кодирования заданного текста (См. Табл.3.1). 3. Определить энтропию и среднее количество двоичных разрядов, необходимых для передачи текста при использовании эффективных кодов. 4. Построить кодовое дерево (См.рис.3.1). 5. Создать таблицу кодов. 6. Проверить возможность однозначного декодирования полученных кодов, рассмотрев пример передачи слова, состоящего из не менее 10 символов.



**Значение средней информации в битах =** 7,16

**Были применены формулы:**

p(i) = k/n

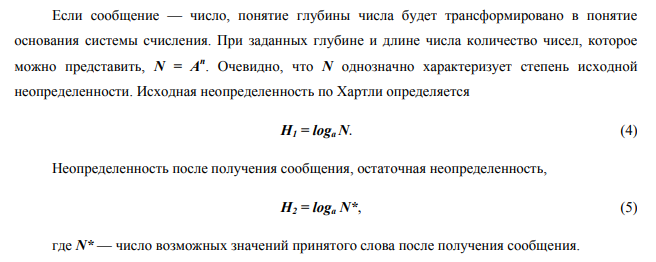
Где: 1) k – количество вхождения символа в текст

2) n – Количество символов в тексте

I(i) = p(i)\*log2(1/p(i)) – Значение энтропии отдельного элемента

Где: 1) p(i) – вероятность вхождение символа

2) i – символ



**Количество информации по Хартли:**



средняя информация, доставляемая одним опытом,

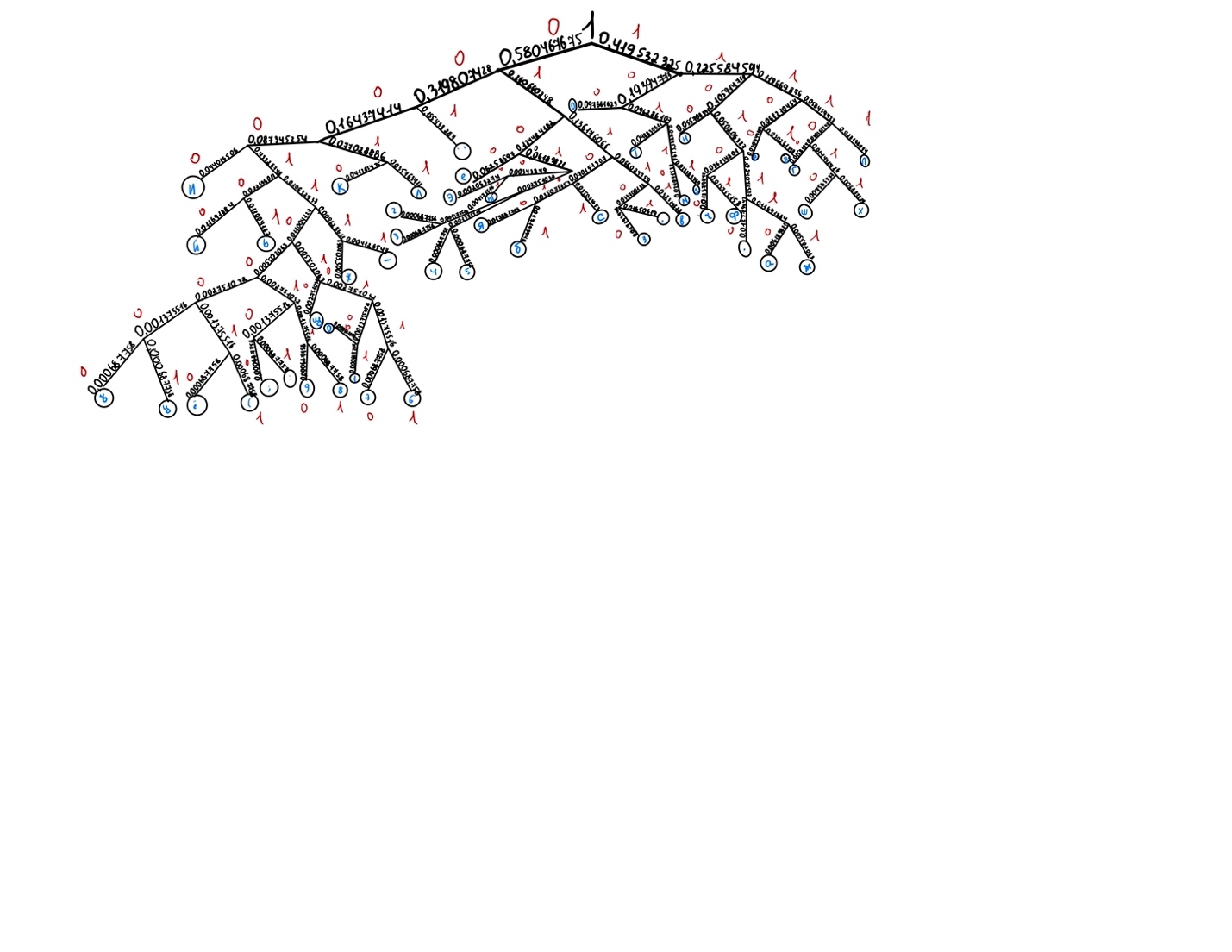


K - кол-во символов алфавита.

А - количество исходов опыта.

ki - количество появления символа в тексте.

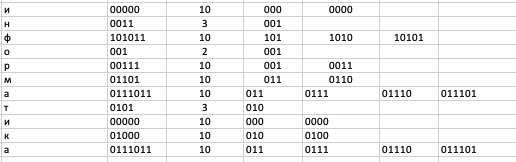
Ii - количество вносимой этим символом информации.



**Закодированное сообщение:**

****

**Декодированное сообщение:**

****

**Вывод:**

Кодировка методом Хаффмана оказалась труднее, но более эффективной. Средняя длинна бит стремиться к энтропии с большим успехом. В ходе третьей части я научился строить дерево Хаффмана, проверил наличие условия единственной раскодировки сообщения.