**Исследование кодирования по методу Хаффмана. Оценка эффективности кода.**

**Краткие теоретические сведения**

Физический смысл кодирования метода Хаффмана заключается в замене наиболее часто встречающихся символов сообщения короткими кодовыми комбинациями, а символов, встречающихся редко – более длинными.

Один из способов получения кода по Хаффману – использование **метода деревьев Хаффмана**, а именно:

1. В таблицах символов и их вероятностей каждая пара «символ-вероятность» рассматривается как один узел-лист дерева Хаффмана.
2. На первом шаге построения дерева выбирается пара узлов-листов, которым соответствуют минимальные величины вероятностей. Для них строится новый узел, вес которого будет равно сумме вероятностей входящих в него узлов-листов. Если на этом шаге, и на любом последующем, присутствует возможность выбора нескольких вариантов действий (например, есть более 2 узлов с минимальными и равными между собой вероятностями), можно выбирать любой из вариантов.
3. Новый узел, полученный на шаге 2, будет являться родительским по отношению к формирующим его узлам. С ними он соединяется ребрами, каждому из которых присваивается вес – 0 или 1 (можем назначать произвольно, но лучше в рамках текущего дерева придерживаться какого-то единого правила. Например – первым идет 1, или наоборот).
4. Пункты 2-3 повторяются до тех пор, пока не будет получена единая вершина (корень дерева), вес которой будет равен 1.
5. Двигаясь от корня по направлению к каждому узлу-листу, соответствующему тому или иному символу, считываются веса ребер. Из них формируется цепочка двоичных элементов, которая и будет кодом символа.

Для каждого сообщения строится кодовое представление. Далее полученный код оценивается в соответствии с формулами 1-4.

**Для вычисления коэффициента сжатия считаем, что каждый из символов исходного сообщения представлен 8 битами.**

Если для какого-либо сообщения существует возможность построения ряда деревьев, т.е., соответственно, и кодов, оптимальный из них выбирается исходя из величин дисперсии (формула 4) и коэффициента сжатия (формула 2). Здесь коэффициент сжатия должен быть максимально возможным, а дисперсия – наоборот, должна минимизироваться.

**В работе необходимо:**

- построить коды для 2 сообщений (задание 1 и задание 2).

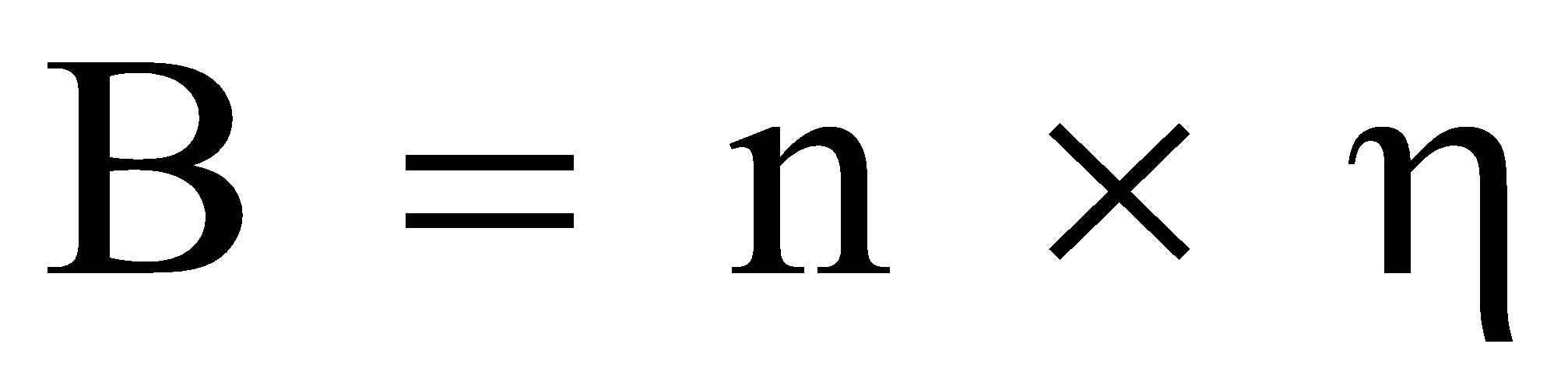
- вычислить величины дисперсии и коэффициента сжатия.

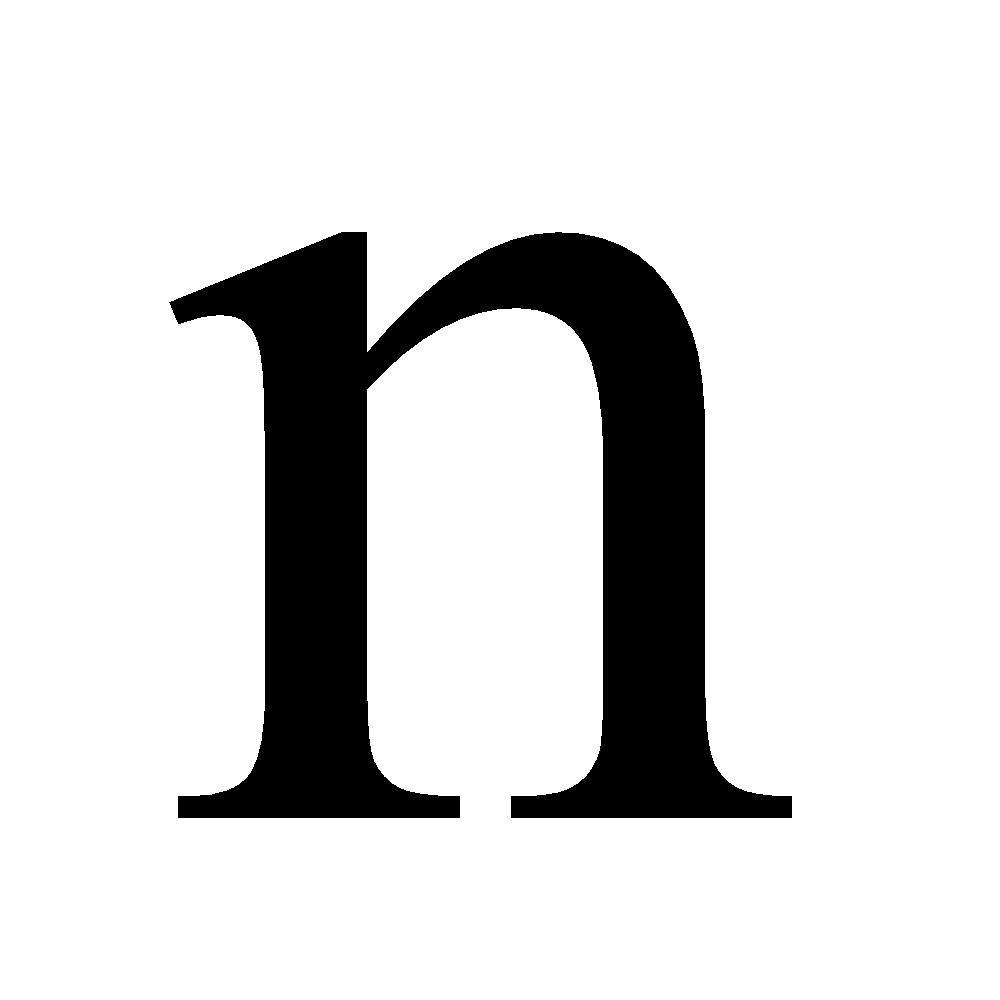
- там, где будут различные варианты построения кодов, на основе величин дисперсии и коэффициента сжатия выбрать оптимальный (задание 3).

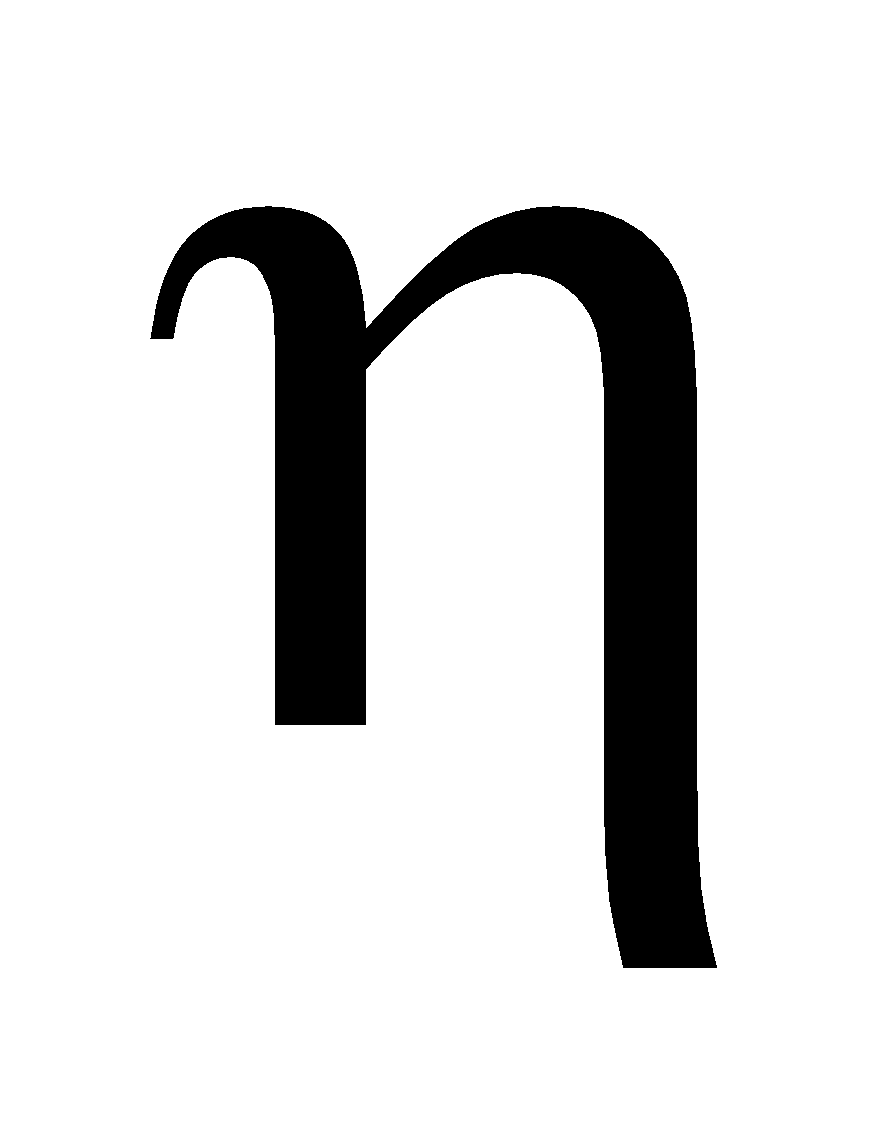
- в консольном приложении Хаффман для разных сообщений – не более 10 разных (вписываются в отдельный файл, см. архив), получить их коды и параметры, которые рассчитывает программа. Сделать выводы о том, как связаны между собой особенности сообщения (и какие) и результаты кодирования. Процесс заскринить (не более 3 сообщений).

- закрепить методику построения кода Хаффама, для чего просмотреть процесс генерирования кода по методу деревьев с приложением greedy. Данным пункт в отчете не обязателен.

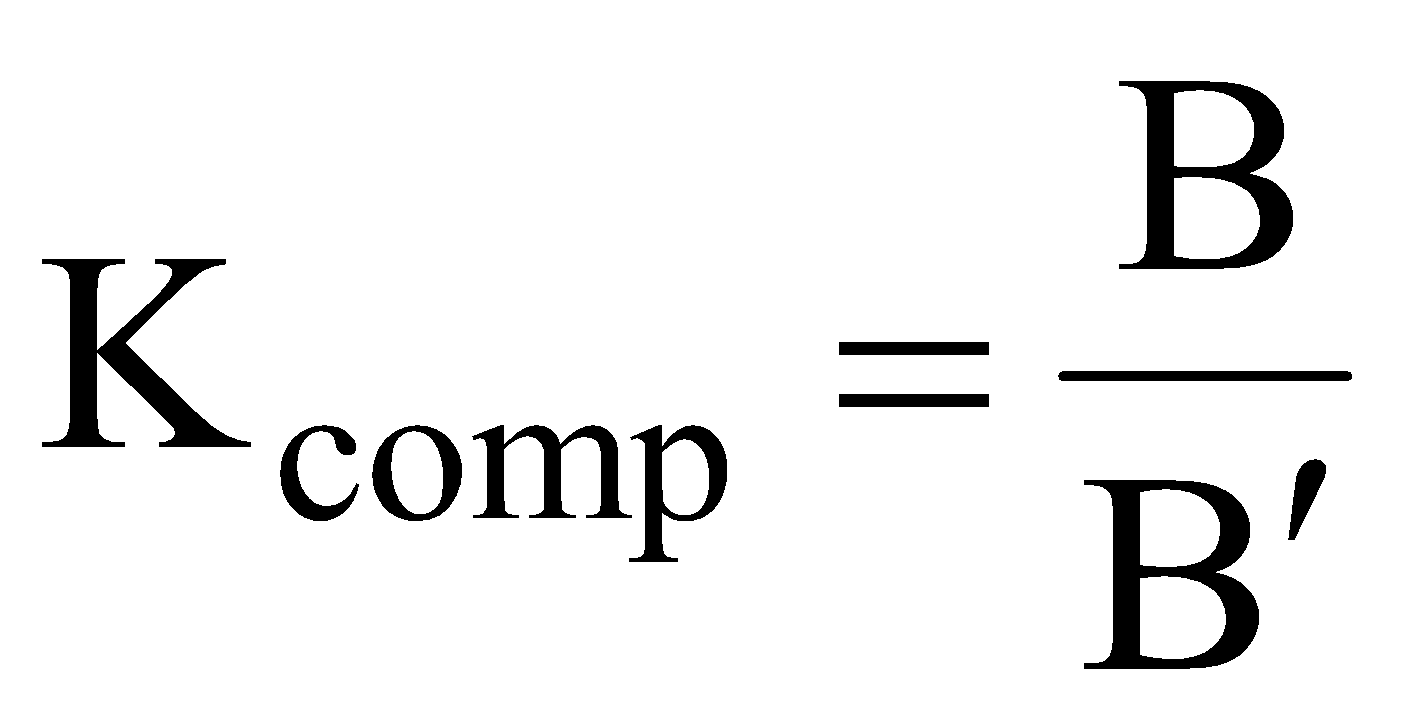
Общий объем бит, который необходимо использовать для описания исходного сообщения произвольного содержания:

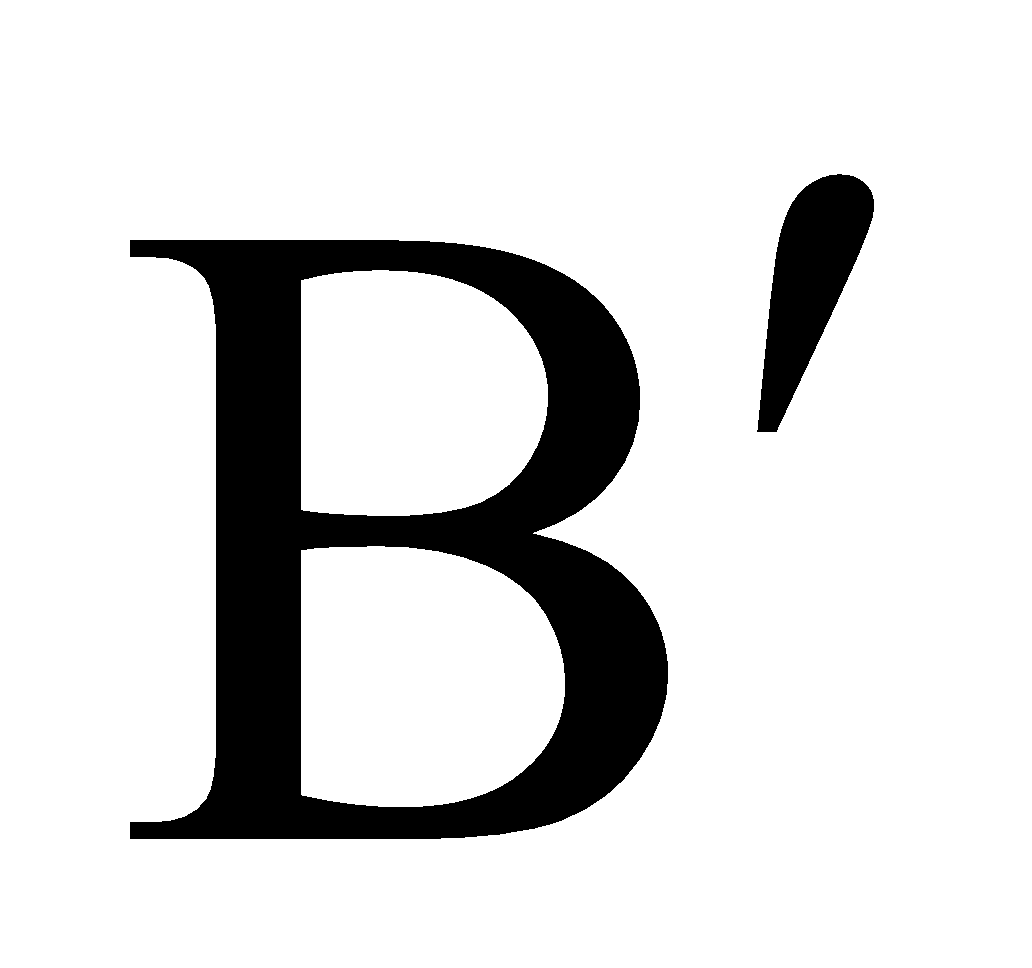
, (1)

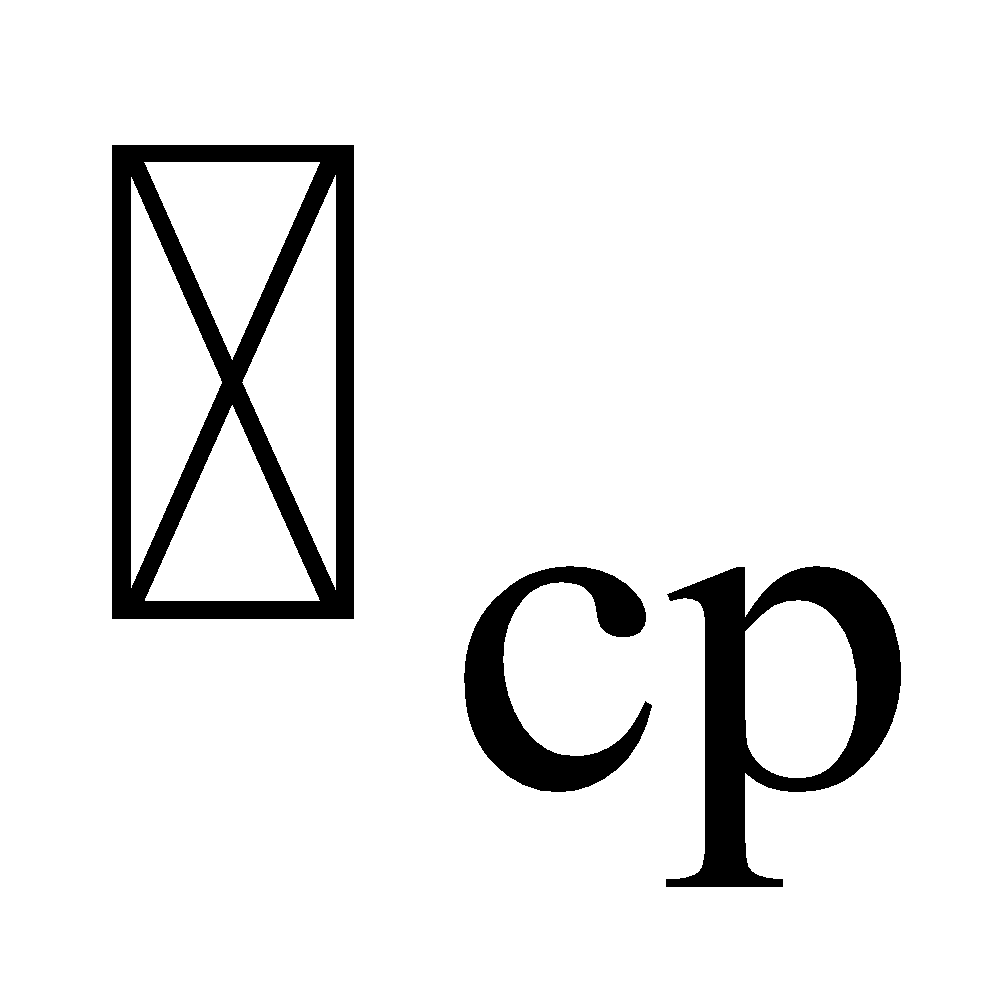
где – количество символов в сообщении;

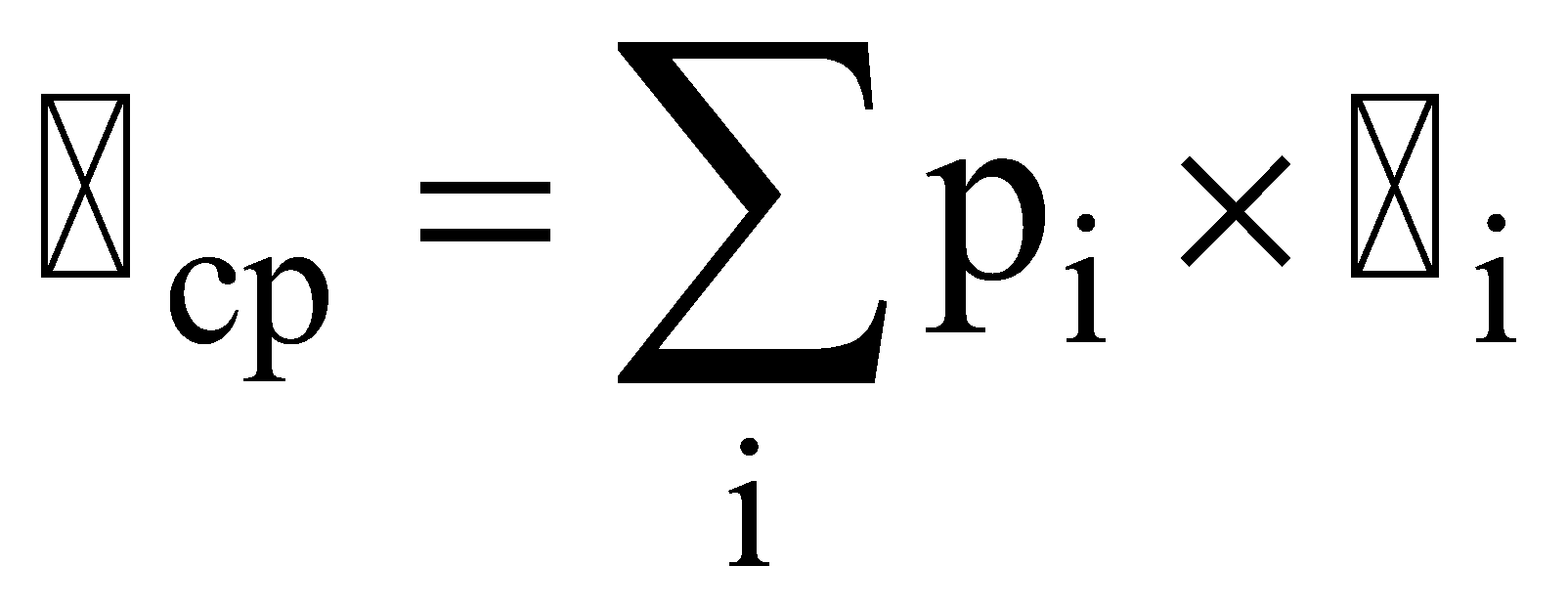
- необходим объем бит для представления одного символа.

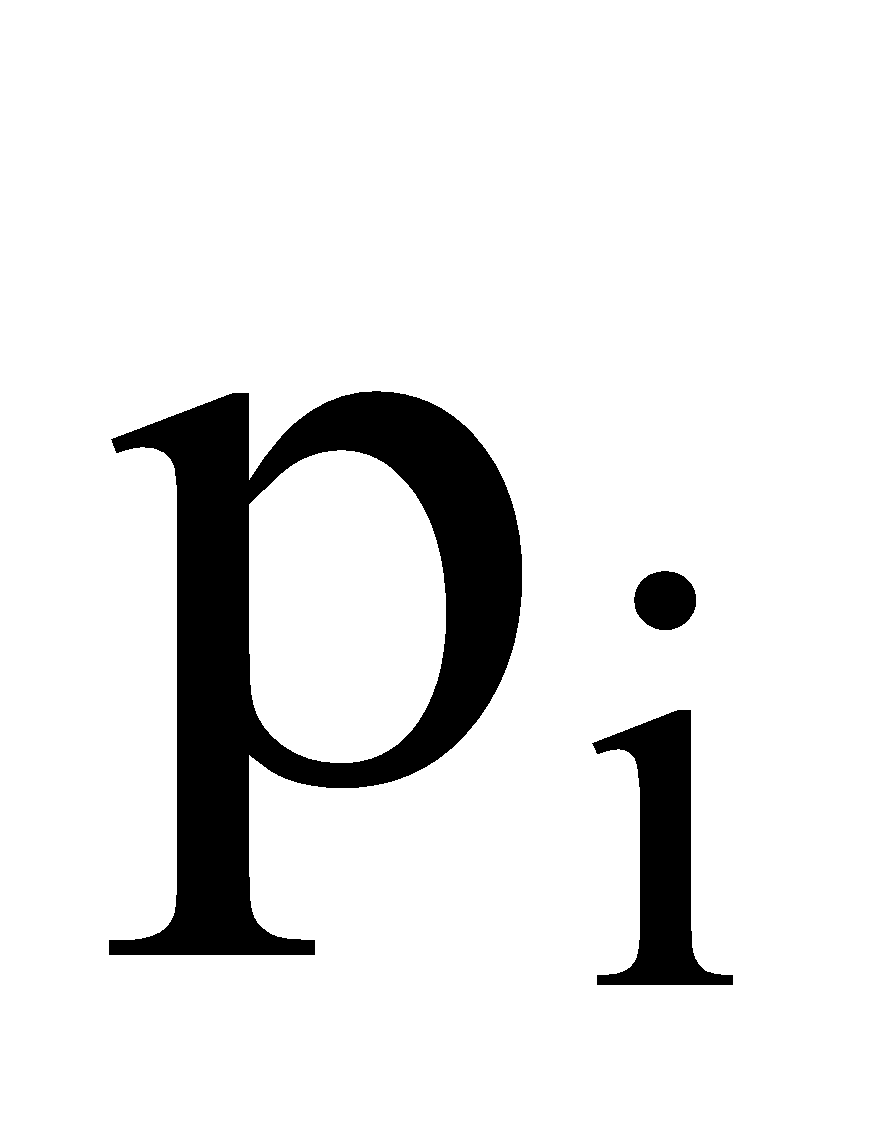
Коэффициент сжатия:

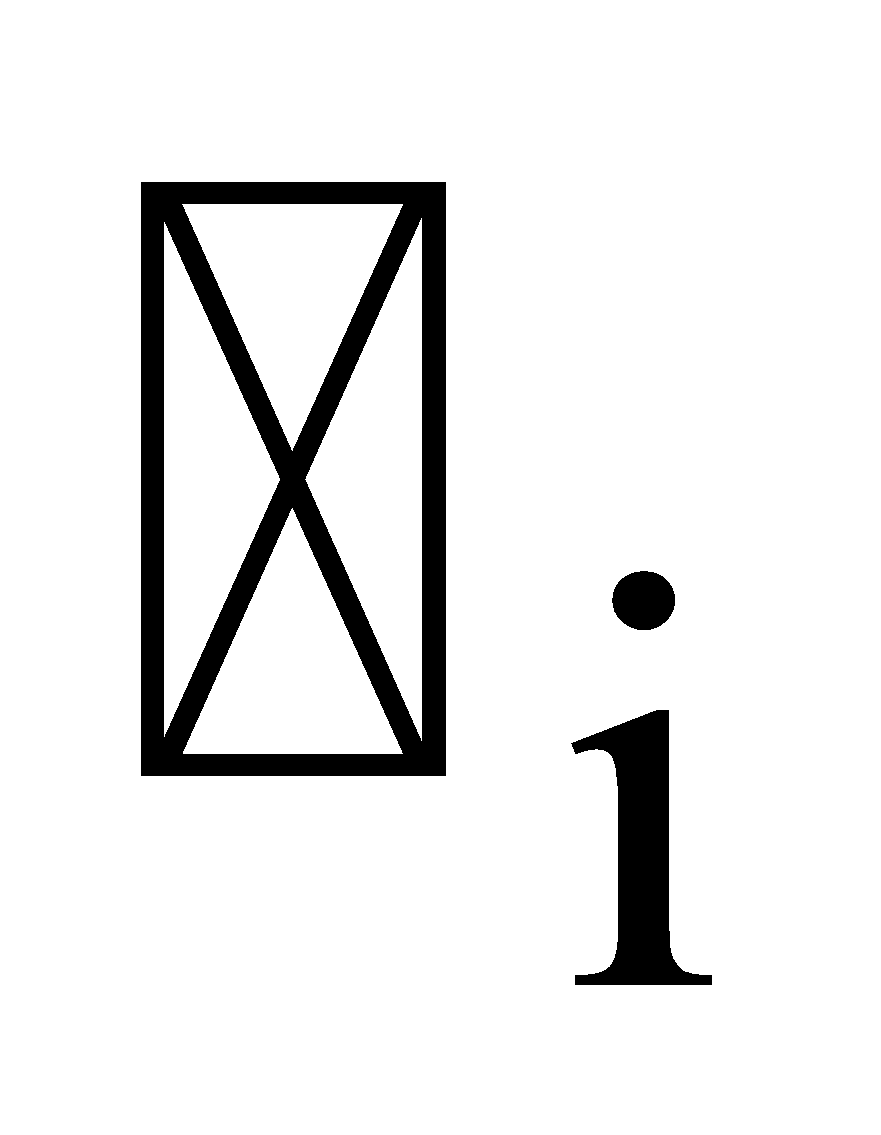
, (2)

где – количество бит после кодирования.

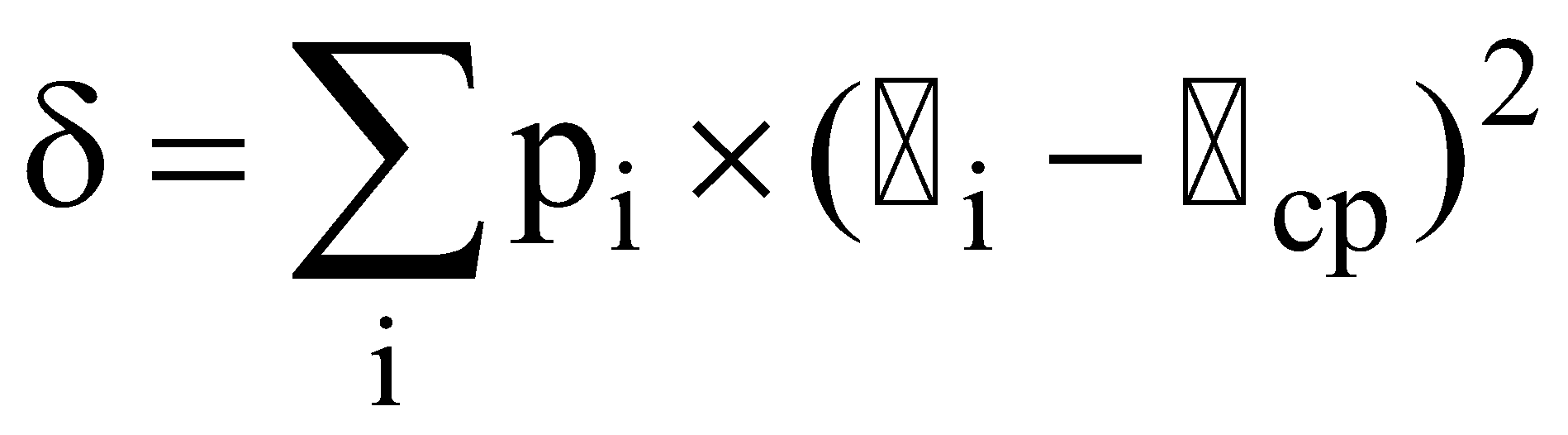
Значение  средней длины образовавшейся кодовой конструкции:

, (3)

где - вероятность появления символа;

- длина одного символа сообщения.

Величина дисперсии:

 (4)

**Исходные данные для работы**

Задание 1. Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.1.

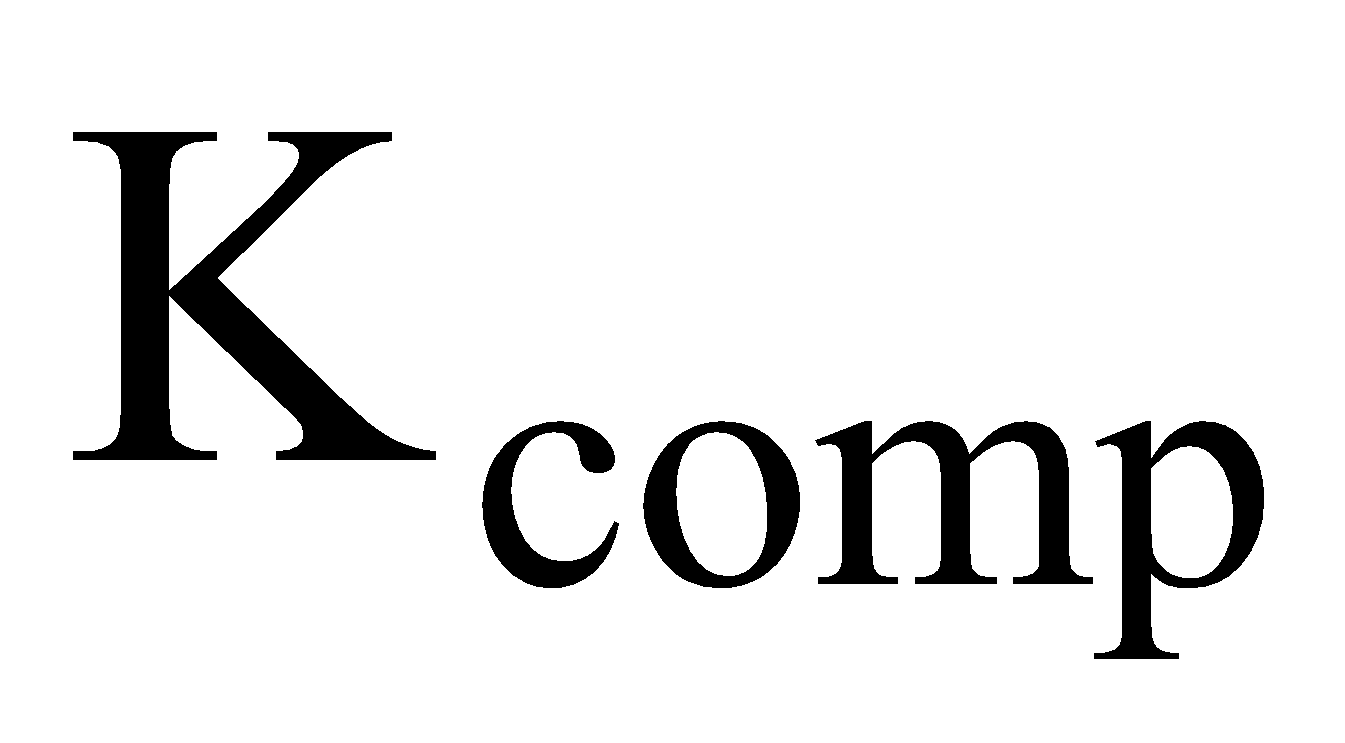
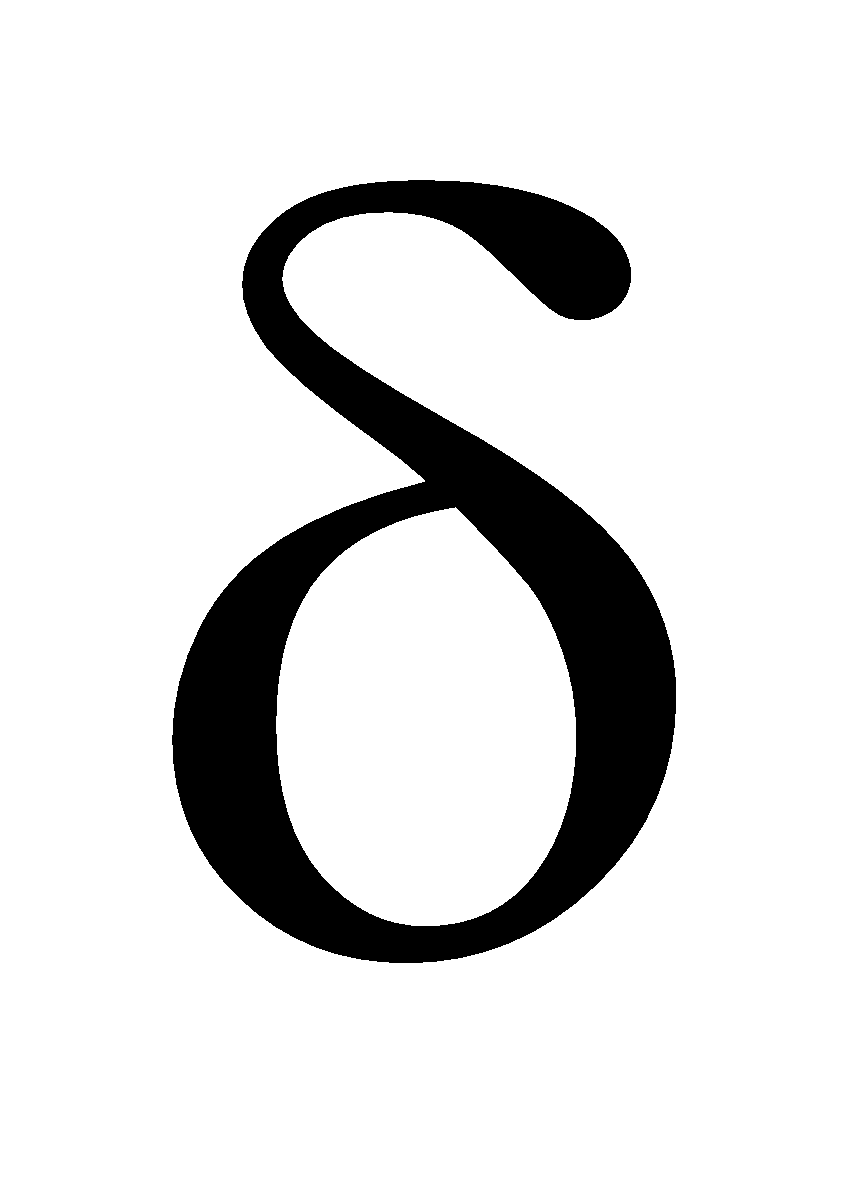
Таблица 1 – Вероятности появления символов в пределах алфавита исходного сообщения

| Символ | s1 | s2 | s3 | s4 | s5 | s6 | s7 | s8 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Вероятность | 0.23 | 0.19 | 0.16 | 0.16 | 0.10 | 0.10 | 0.05 | 0.01 |

Задание 2. Построить кодовое представление сообщения, вероятности появления символов в пределах алфавита которого приведены в табл.2.

Таблица 2 – Вероятности появления символов в пределах алфавита исходного сообщения

| Символ | s1 | s2 | s3 | s4 | s5 | s6 | s7 | s8 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Вероятность | 0.25 | 0.22 | 0.13 | 0.11 | 0.1 | 0.09 | 0.07 | 0.03 |

Задание 3. Для условий, приведенных в заданиях 1 и 2, выявить возможность построения альтернативных кодовых моделей сообщения. В случае обнаружения таковых, выявить наиболее эффективные из них по критериям  и  .

**Содержание отчета**

- решения заданий 1-3 (процесс и результаты построения кодов Хаффмана, можно сжато, оценка полученных кодов и выводы касательно оптимального кода, если где-то будут возможны его разные варианты);

- скриншоты и краткое описание (пояснение процесса) исследования зависимости особенностей сообщений и результирующих кодов (приложение Хаффман);

- размещение сведений о работе с greedy – по желанию;

- выводы