|  |  |
| --- | --- |
| Череповецкий государственный университет  Кафедра «Математического и программного обеспечения ЭВМ» | |
| ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ  по дисциплине «Теория информации»  КОДЫ ФАНО И ХАФФМАНА | |
|  | Принял:  преподаватель Е.Н. Руденко    подпись, дата  Выполнил:  студент гр. 1ПИб-02-3оп-22  Маркелов С.А.    подпись, дата |
| Череповец, 2023 | |

Реферат

Предметом исследования являются коды Фано и Хаффмана, которые позволяют закодировать информацию самым оптимальным способом (т.е. избыточность кода будет минимальна).

Цель работы – научиться кодировать и декодировать по методу Фано и Хаффмана

В ходе работы проводились теоретические исследования различных методов кодирования.

В результате аналитических исследований были найдены наиболее оптимальные методы кодирования. Численно были получены значения оптимальности кодирования. Проведено кодирование различных сообщений методами Фано и Хаффмана с применением как традиционного табличного способв кодирования, так и "кодовых деревьев".

Коды Фано и Хаффмана применяются при сжатии информации и кодировании сообщений.

1 Метод Фано

При кодировании по Фано все сообщения записываются в таблицу по степени убывания вероятности и разбиваются на две группы примерно (насколько это возможно) равной вероятности. Соответственно этой процедуре из корня кодового дерева исходят два ребра, которым в качестве весов присваиваются полученные вероятности. Двум образовавшимся вершинам приписывают кодовые символы 0 и 1. Затем каждая из групп вероятностей вновь делится на две подгруппы примерно равной вероятности. В соответствии с этим из каждой вершины 0 и 1 исходят по два ребра с весами, равными вероятностям подгрупп, а вновь образованным вершинам приписывают символы 00 и 01, 10 и 11. В результате многократного повторения процедуры разделения вероятностей и образования вершин приходим к ситуации, когда в качестве веса, приписанного ребру бинарного дерева, выступает вероятность одного из данных сообщений. В этом случае вновь образованная вершина оказывается листом дерева, т.к. процесс деления вероятностей для нее завершен. Задача кодирования считается решенной, когда на всех ветвях кодового бинарного дерева образуются листья.

**Задание 1.**Проведите кодирование по методу Фано алфавита из четырех букв, вероятности которых равны 0,4; 0,3; 0,2 и 0,1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Сообщение | 1 | 2 | 3 | 4 |
| Вероятность | 0,4 | 0,3 | 0,2 | 0,1 |

Решение с использованием кодового дерева:



Рис. 1

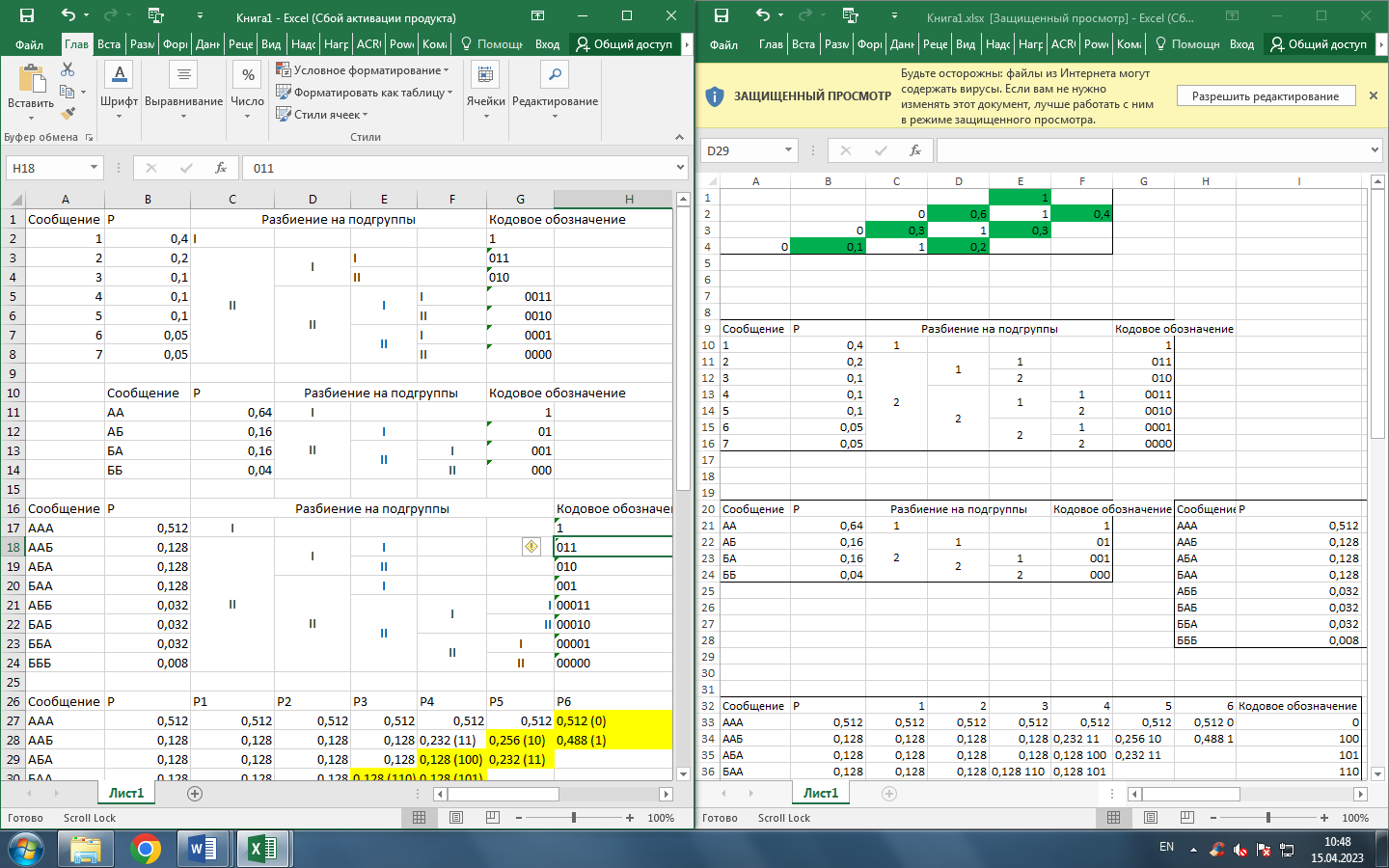
Листья кодового дерева представляют собой кодируемые сообщения с присвоенными им кодовыми словами.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Сообщение | 1 | 2 | 3 | 4 |
| Код | 1 | 01 | 000 | 001 |

**Задание 2.** Алфавит содержит 7 букв, которые встречаются с вероятностями 0,4; 0,2; 0,1; 0,1; 0,1; 0,05; 0,05. Осуществите кодирование по методу Фано.

Решение табличным способом:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сообщение | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Вероятность | 0,4 | 0,2 | 0,1 | 0,1 | 0,1 | 0,05 | 0,05 |

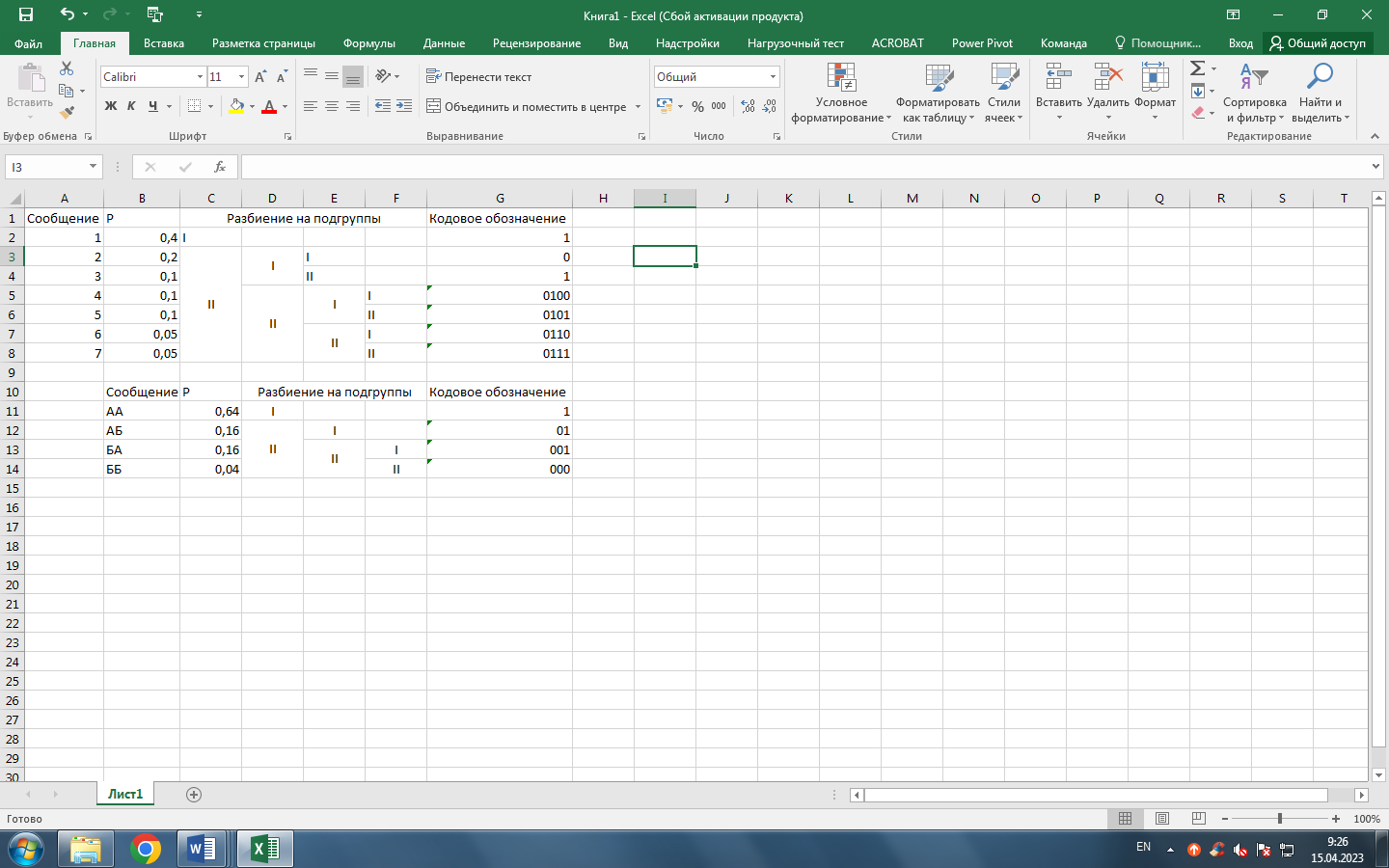


|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сообщение | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Код | 1 | 000 | 001 | 0100 | 0101 | 0110 | 0111 |

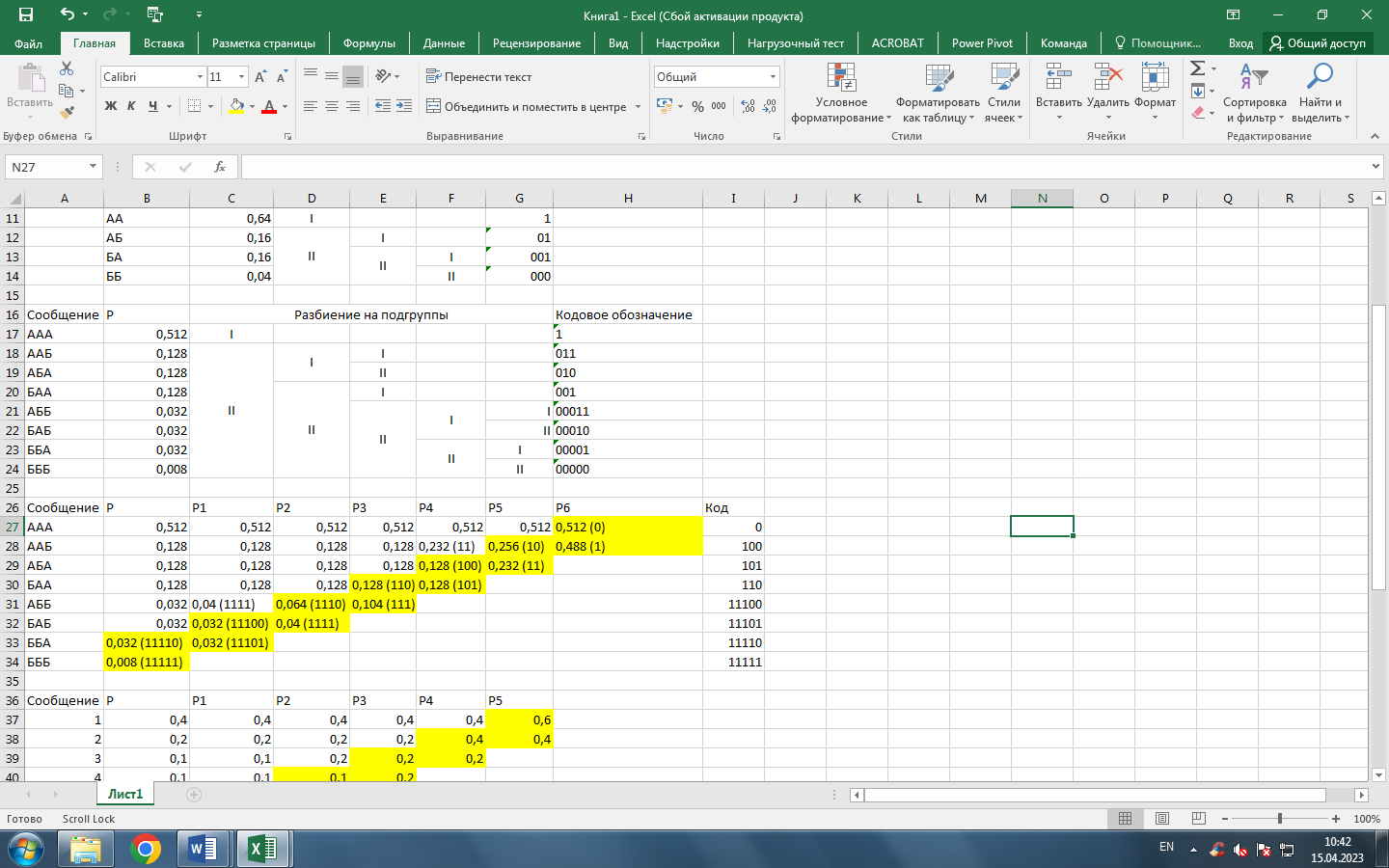
Цена кодирования (средняя длина кодового слова *l*) является критерием степени оптимальности кодирования. Вычислим ее в нашем случае.

**Задание 3.** Алфавит состоит из двух букв, А и Б, встречающихся с вероятностями *P(А)* = 0,8 и *P(Б)* = 0,2. Примените метод Фано к кодированию всевозможных двухбуквенных и трехбуквенных комбинаций.

Решение.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Сообщение | АА | АБ | БА | ББ |
| Код | 1 | 01 | 001 | 000 |

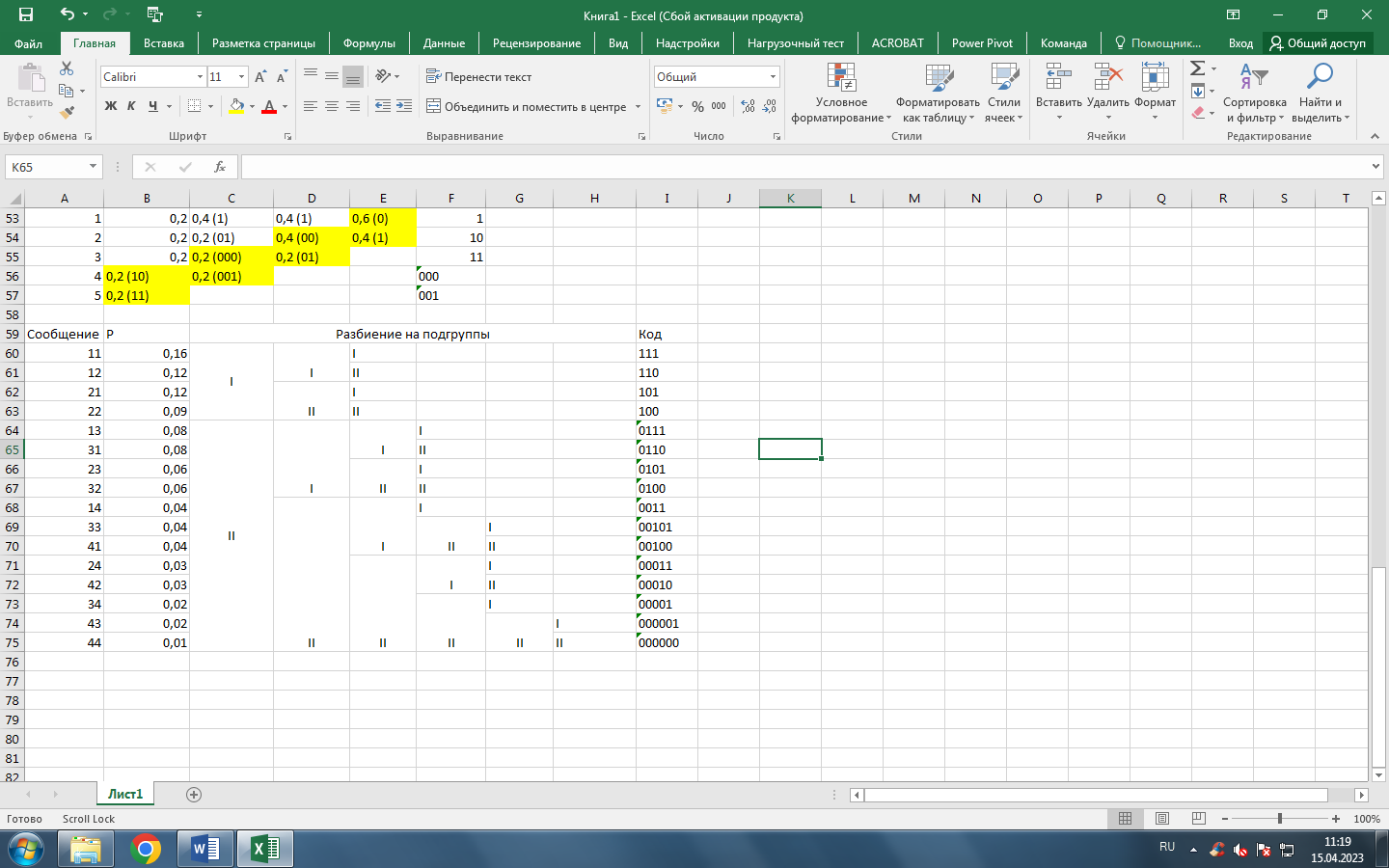


|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Сообщение | ААА | ААБ | АБА | АББ | БАА | БАБ | ББА | БББ |
| Код | 1 | 011 | 010 | 0011 | 0010 | 0001 | 00001 | 00000 |

Цена кода при алфавите из 2 букв: , и на одну букву алфавита приходится 0,78 бита информации. При побуквенном кодировании на каждую букву приходится следующее количество информации: (бит).

Цена кода при алфавите из 3 букв: , и на одну букву алфавита приходится 0,728 бита информации. При побуквенном кодировании на каждую букву также приходится 0,72 бита информации.

**Задание 7.** Осуществите кодирование двухбуквенных комбинаций четырех букв из задачи 1.

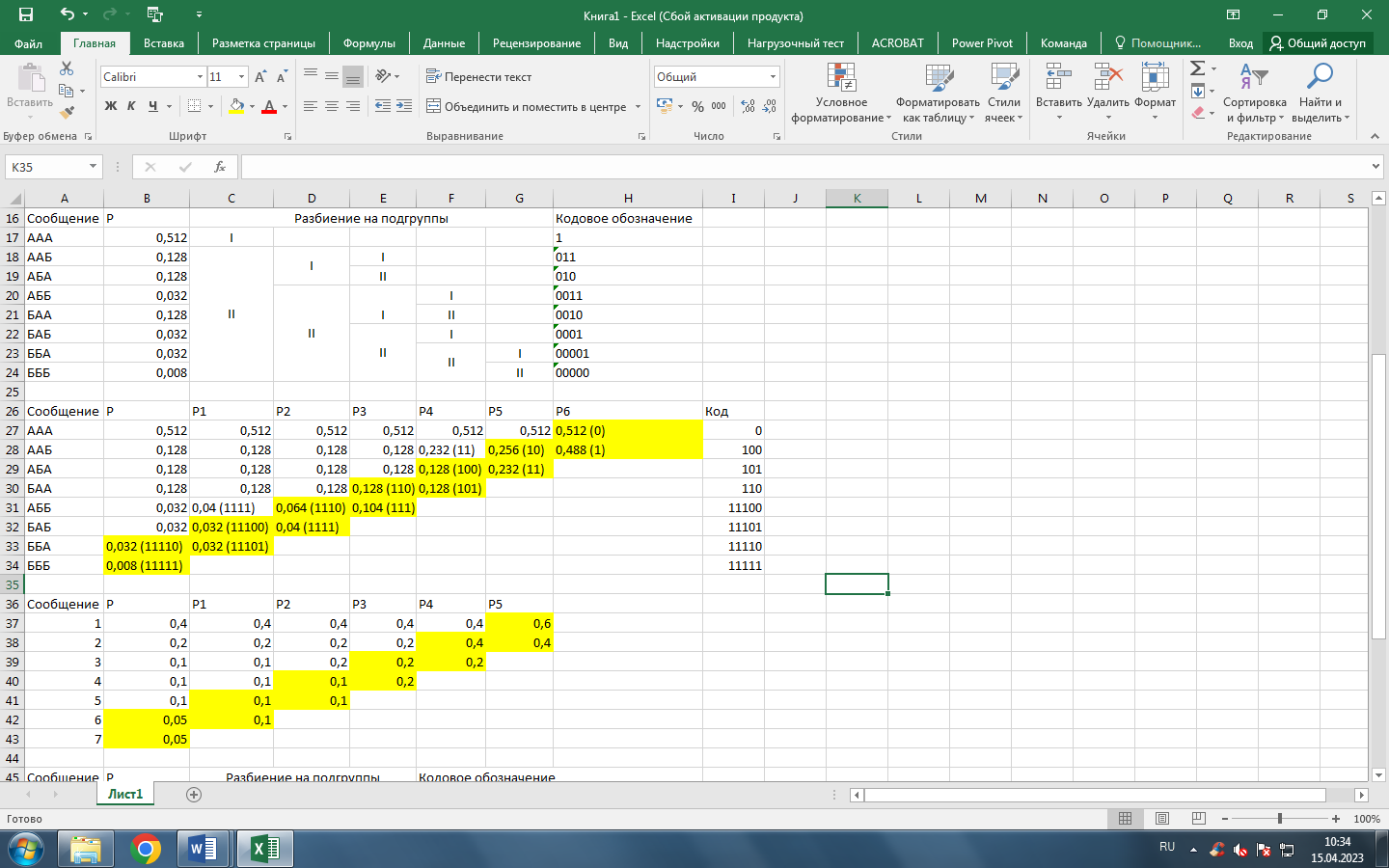


2 Метод Хаффмана

Для того, чтобы закодировать сообщения по Хаффману, предварительно преобразуется таблица, задающая вероятности сообщений. Исходные данные записываются в столбец, две последние (наименьшие) вероятности в котором складываются, а полученная сумма становится новым элементом таблицы, занимающим соответствующее место в списке убывающих по величине вероятностей. Эта процедура продолжается до тех пор, пока в столбце не останутся всего два элемента.

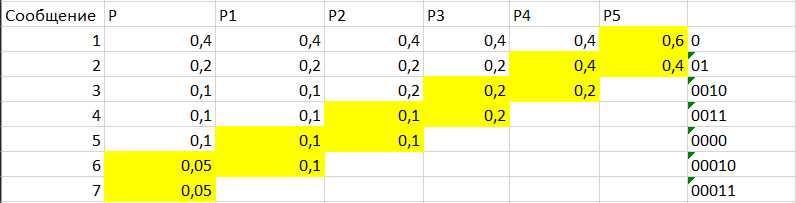
**Задание 4.** Проведите кодирование по методу Хаффмана трехбуквенных слов из задания 3.

Решение табличным способом:



Первым шагом складываются вероятности. Вторым шагом производим кодирование, "проходя" по таблице справа налево

**Задание 5.** Проведите кодирование 7 букв из задачи 2 по методу Хаффмана.



Построение кодового дерева начинается с корня. Двум исходящим из него ребрам приписывается в качестве весов вероятности 0,6 и 0,4, стоящие в последнем столбце. Образовавшимся при этом вершинам дерева приписываются кодовые символы 0 и 1. Далее "идем" по таблице справа налево. Поскольку вероятность 0,6 является результатом сложения двух вероятностей 0,4 и 0,2, из вершины 0 исходят два ребра с весами 0,4 и 0,2 соответственно, что приводит к образованию двух новых вершин с кодовыми символами 00 и 01. Процедура продолжается до тех пор, пока в таблице остаются вероятности, получившиеся в результате суммирования. Построение кодового дерева заканчивается образованием семи листьев, соответствующих данным сообщениям с присвоенными им кодами. Дерево, полученное в результате кодирования по Хаффману, имеет следующий вид:

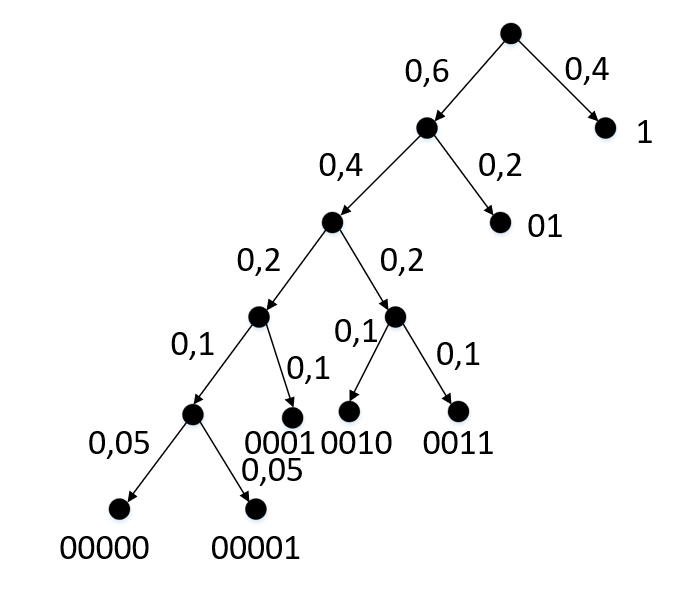


Рис. 2

Листья кодового дерева представляют собой кодируемые сообщения с присвоенными им кодовыми словами. Таблица кодов имеет вид:

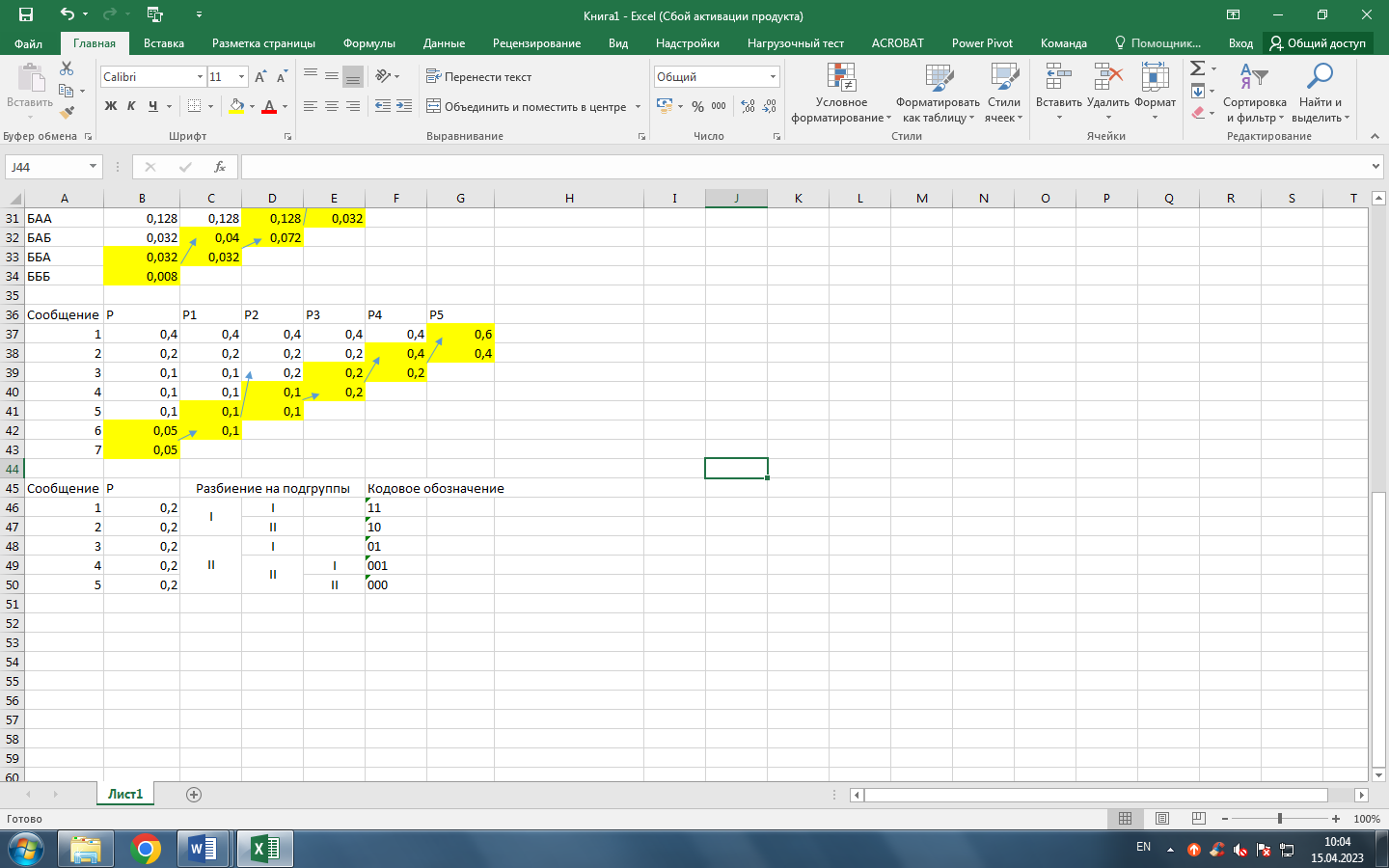
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сообщение | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Код | 1 | 01 | 0010 | 0011 | 0001 | 00000 | 00001 |

Цена кодирования здесь будет равна

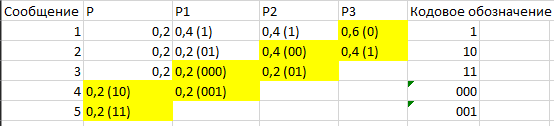
Для передачи данных сообщений можно перейти от побуквенного (поцифрового) кодирования к кодированию "блоков", состоящих из фиксированного числа последовательных "букв".

**Задание 6.** Проведите кодирование по методам Фано и Хаффмана пяти букв, равновероятно встречающихся.

Метод Фано:



Метод Хаффмана:



Заключение

В ходе лабораторной работы повторен теоретический материал по темам «Кодирование информации», «Метод Фано», «Метод Хаффмана». С помощью методов Фано и Хаффмана составлены оптимальные коды для различных сообщений. С помощью Excel построены коды табличным способом. С помощью Visio построены коды методом «кодовых деревьев». Вычислены значения цены кодирования (среднего значения кодового слова).

Список использованных источников

1 Советов, Б. Я. Информационная технология [Текст] : Учеб. для студ. вузов по спец. «Автоматизир. системы обраб. информ. и управления» / Б. Я. Советов. – М. : Высш. шк., 1994. – 366 c.

2 Дмитриев, В. И. Прикладная теория информации [Текст] : Учеб. для студ. вузов по спец. «Автоматизированные системы обработки информации и управления» / В. И. Дмитриев. – М. : Высш. шк., 1989. – 320 с. : ил.

3 ГОСТ 7.32–2001. Отчёт о научно-исследовательской работе. Структура и правила оформления [Текст]. – Взамен ГОСТ 7.32–91 ; введ. 2001–07–01. – Минск : Межгос. совет по стандартизации, метрологии и сертификации ; М. : Изд-во стандартов, 2001. – 16 с. – (Система стандартов по информации, библиотечному и издательскому делу).

4 Теоретические основы информационных процессов [Текст] : Учеб. пособие для вузов по спец. «Автоматизация и механизация процессов обработки и выдачи информации» / Л. Ф. Куликовский, В. В. Мотов. – М. : Высш. шк., 1987. – 248 с.

Скрытый служебный текст:

рисунков 2,

всего таблиц 2,

источников 4.